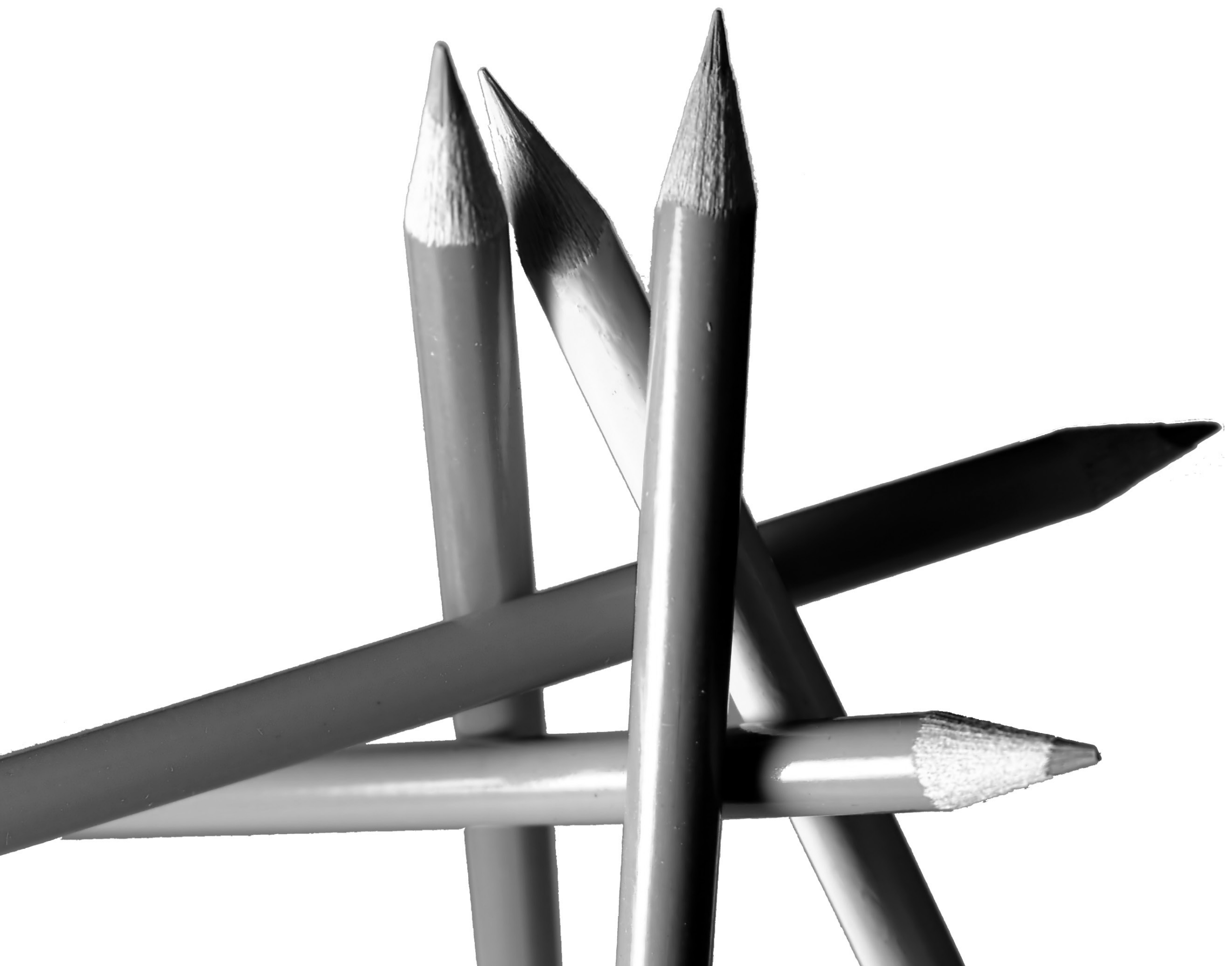
Lineamientos para la Evaluación y Enseñanza en Educación Estadística, Reporte (GAISE)

UN MARCO PARA EL CURRICULO DE PRE-K-12



CHRISTINE FRANKLIN
University of Georgia

GARY KADER
Appalachian State University

DENISE MEWBORN
University of Georgia

JERRY MORENO
John Carroll University

ROXY PECK
California Polytechnic State University, San Luis Obispo

MIKE PERRY
Appalachian State University

RICHARD SCHEAFFER
University of Florida

Reconocido por la Asociación Americana de Estadística (ASA)
Agosto 2005

Consejeros

Susan Friel
The University of North Carolina

Landy Godbold
Westminster Schools

Brad Hartlaub
Kenyon College

Peter Holmes
Nottingham Trent University

Cliff Konold
University of Massachusetts/Amherst

Equipo de Producción

Lucia Zapata Cardona
Universidad de Antioquia, Colombia

Christine Franklin
University of Georgia

Nicholas Horton
Smith College

Gary Kader
Appalachian State University

Jerry Moreno
John Carroll University

Megan Murphy
American Statistical Association

Valerie (Snider) Nirala
American Statistical Association

M. Alejandra Sorto
Texas State University

Daren Starnes
Fountain Valley School of Colorado

Agradecimientos Especiales

Los autores extienden un especial agradecimiento al Consejo de Directores de la Asociación Americana de Estadística por financiar el proceso de elaboración de GAISE como una iniciativa estratégica, así como al Comité Conjunto ASA/NCTM por financiar la producción del Marco de GAISE.

Traducción: Lucia Zapata Cardona, Universidad de Antioquia, Colombia

Trabajo editorial de la traducción en español: M. Alejandra Sorto, Texas State University

Fotografia de portada por: Andres Rodriguez

Diseño del libro por: Valerie (Snider) Nirala

Contenidos

Índice de Tablas y Figuras

Introducción

El objetivo final: La alfabetización estadística. Cada mañana, el periódico y otros medios nos confrontan con información en tópicos que van desde la economía a la educación, desde las películas hasta los deportes, desde los alimentos hasta las medicinas, y desde la opinión pública hasta el comportamiento social. Tal información guía las decisiones en nuestra vida personal y nos permite satisfacer nuestras responsabilidades como ciudadanos. En el trabajo, podemos enfrentarnos con información cuantitativa en presupuestos, suministros, especificaciones de manufactura, demandas de mercado, predicciones de ventas o volumen de trabajo. Los profesores pueden ser enfrentados con estadísticas educativas relacionadas con el desempeño de los estudiantes o de su propia rendición de cuentas. Los científicos médicos tienen que entender los resultados estadísticos de experimentos usados para probar la efectividad y seguridad de los medicamentos. La aplicación de la ley depende de las estadísticas del crimen. Si consideramos cambiar de trabajo y mudarnos a otra comunidad, nuestra decisión puede ser afectada por las estadísticas sobre el costo de vida, índice de delincuencia y calidad educativa.

Nuestras vidas son gobernadas por números. Cada graduado de la escuela preparatoria[1] debe estar en condiciones de usar un sólido razonamiento estadístico para afrontar inteligentemente los requerimientos de la ciudadanía, el empleo, la familia y estar preparado para una vida sana, feliz y productiva.

1 Escuela preparatoria se refiere a los años escolares de noveno a décimo segundo.

Ciudadanía

Las encuestas de opinión pública son los ejemplos más visibles de la aplicación de la estadística que tienen impacto en nuestras vidas. Además de informar directamente a los ciudadanos, los sondeos electorales son usados por otros en formas que nos afectan. Los procesos políticos emplean encuestas de opinión de diferentes formas. Los candidatos a ciertos cargos usan sondeos electorales para guiar sus estrategias de campaña. Un sondeo electoral puede determinar las fortalezas de un candidato con los votantes, lo cual puede a su vez, ser destacado en la campaña. Los ciudadanos también pueden sospechar que los resultados electorales pueden influenciar a los candidatos a tomar posiciones sólo porque son populares.

Un ciudadano informado por sondeos electorales necesita entender que los resultados fueron determinados de una muestra de la población bajo estudio, que la confiabilidad de los resultados depende de cómo fue seleccionada la muestra, y que los resultados están sujetos al error de muestreo. Los ciudadanos alfabetizados estadísticamente deben entender el comportamiento de muestras "aleatorias" y estar en condiciones de interpretar un "margen de error muestral".

El Gobierno Federal de los Estados Unidos de América ha utilizado la estadística desde su creación. El censo de los Estados Unidos fue establecido en 1790 para proveer una cuenta oficial de la población con el propósito de asignar los representantes al congreso. El rol del Departamento de Censo de los Estados Unidos no sólo se ha expandido considerablemente para incluir la recolección de un amplio espectro de datos socioeconómicos, sino que también otros

"Cada graduado de la escuela preparatoria debe estar en condiciones de usar un sólido razonamiento estadístico para afrontar inteligentemente los requerimientos de la ciudadanía, el empleo, la familia y estar preparado para una vida sana, feliz y productiva."

departamentos federales producen extensas estadísticas "oficiales" relacionadas con agricultura, salud, educación, medio ambiente y comercio. La información recogida por estas agencias influencia la creación de políticas y ayuda a determinar las prioridades para el gasto gubernamental. También está disponible para el uso general de individuos o grupos privados. Así, las estadísticas recogidas por las agencias de gobierno tienen un tremendo impacto en la vida del ciudadano común.

Elecciones Personales

La alfabetización estadística es requerida para las elecciones personales diarias. La estadística provee información sobre la calidad nutricional de los alimentos y de este modo nos ayuda en nuestras decisiones en el supermercado. La estadística ayuda a establecer la seguridad y la efectividad de los medicamentos, ayudando a los médicos en la prescripción de un tratamiento. La estadística también ayuda a establecer la seguridad de los juguetes para asegurar que nuestros niños no estén en riesgo. Nuestras decisiones con respecto a inversiones están guiadas por abundante información estadística sobre acciones y bonos financieros. Los índices de audiencia Nielsen ayudan a determinar cuál programa sobrevivirá en la televisión, afectando así lo que está disponible. Muchos productos tienen una historia estadística, y nuestra elección de productos puede ser afectada por el conocimiento de esa historia. El diseño de un automóvil es ayudado por la antropometría—la estadística del cuerpo humano—para mejorar la comodidad del pasajero. Las valoraciones estadísticas de eficiencia en combustible, seguridad y confiabilidad están disponibles para ayudarnos a seleccionar un vehículo.

El Lugar de Trabajo y las Profesiones

Los individuos que están preparados para usar el razonamiento estadístico en sus carreras tendrán la oportunidad de avanzar a posiciones más gratificantes y desafiantes. Una fuerza de trabajo competente estadísticamente permitirá a los Estados Unidos competir más efectivamente en el mercado global y mejorar su posición en la economía mundial. Una inversión en alfabetización estadística es una inversión en el futuro económico de nuestra nación como también en el bienestar de los individuos.

El mercado competitivo demanda calidad. Los esfuerzos para mejorar la calidad y la rendición de cuentas son prominentes entre las muchas formas que el pensamiento estadístico y las herramientas estadísticas se pueden utilizar para mejorar la productividad. Las prácticas de control de calidad como el monitoreo estadístico de procesos de diseño y manufactura, identifican donde pueden tener lugar mejoras para refinar la calidad del producto. Los sistemas de rendición de cuentas pueden ayudar a promover empleados y organizaciones más efectivos, sin embargo muchos sistemas de rendición de cuentas ahora no se basan en principios estadísticos profundos y pueden, de hecho, tener el efecto contrario. Los buenos sistemas de rendición de cuentas requieren un uso apropiado de las herramientas estadísticas para determinar y aplicar los criterios apropiados.

Ciencia

La esperanza de vida en los Estados Unidos es casi el doble de la del siglo XX; este rápido crecimiento en la duración de la vida es la consecuencia de la ciencia. La ciencia nos ha permitido mejorar los cuidados y los procedimientos médicos, la producción de alimentos y la detección y prevención de epidemias. La estadística juega un rol destacado en el progreso científico.

La Administración de Alimentos y Medicamentos de los Estados Unidos requiere de extensas pruebas en los medicamentos para determinar la efectividad y los efectos colaterales antes que ellos puedan ser comercializados. Una publicidad reciente de una droga diseñada para reducir coágulos de sangre estableció, "PLAVIX, complementada con la aspirina y con sus medicamentos actuales, ayuda a aumentar su protección contra ataque al corazón o infarto". Pero la publicidad también advirtió, "el riesgo de hemorragia puede incrementarse con PLAVIX…"

La alfabetización estadística involucra una buena dosis de escepticismo con respecto a hallazgos "científicos". ¿Es confiable la información relacionada con efectos colaterales de los tratamientos con PLAVIX? Una persona alfabetizada estadísticamente debe hacer esas preguntas y estar en condiciones de responderlas inteligentemente. Un graduado de la preparatoria alfabetizado estadísticamente estaría en condiciones de entender las conclusiones de investigaciones científicas y ofrecer una opinión informada acerca de la legitimidad de los resultados reportados. De acuerdo a *Matemática y Democracia: El caso para la Alfabetización Cuantitativa* (Steen, 2001), tal conocimiento "empodera a la gente dándole herramientas para pensar por ellos mismos, para hacer preguntas inteligentes como los expertos, y para confrontar la autoridad con confianza. Estas son habilidades requeridas para sobrevivir en el mundo moderno".

La alfabetización estadística es esencial en nuestra vida personal como consumidores, ciudadanos y profesionales. La estadística juega un rol en nuestra salud y felicidad. Las habilidades de razonamiento estadístico sólido toman tiempo para desarrollarse. Ellas no pueden perfeccionarse al nivel que se requiere en el mundo moderno mediante un curso de preparatoria. El camino más seguro para ayudar a los estudiantes a obtener el nivel de habilidades necesarias es empezar el proceso de educación estadística en los grados elementales y continuar fortaleciendo las habilidades de pensamiento estadístico de los estudiantes a través de los años de secundaria[2] y preparatoria. Un graduado de la preparatoria alfabetizado estadísticamente sabrá cómo interpretar la información del periódico de la mañana y hará las preguntas apropiadas sobre las afirmaciones estadísticas. Él estará cómodo manejando decisiones cuantitativas que surgen en el trabajo y estará en condiciones de tomar decisiones informadas en asuntos de calidad de vida.

El resto de este documento presenta un marco curricular para los programas de educación de preescolar a grado 12 el cual está diseñado para ayudar a los estudiantes a alcanzar la alfabetización estadística.

El Caso de la Educación Estadística

Durante el último cuarto de siglo, la estadística (frecuentemente llamada análisis de datos y probabilidad) se ha convertido en un componente clave del currículo de matemáticas de preescolar a grado 12. Los avances en tecnología y los métodos modernos de análisis de datos de

2 Secundaria se refiere a los años escolares de sexto a octavo.

> “La educación estadística como se propone en este Marco puede promover las habilidades que los graduados ‘tienen que tener’ para ‘prosperar en el mundo moderno’.”

los años 1980, junto con la riqueza de datos de la sociedad en la era de la información, condujeron al desarrollo de materiales curriculares orientados hacia la introducción de conceptos estadísticos en el currículo escolar tan temprano como en los grados elementales. Este esfuerzo de base fue aprobado por el Concilio Nacional de Profesores de Matemáticas (NCTM por sus siglas en inglés) cuando su influyente documento, *Estándares de Currículo y Evaluación para la Matemática Escolar* (NCTM, 1989), incluyó “análisis de datos y probabilidad” como una de las cinco áreas de contenido. Como este documento y su reemplazo del 2000, *Principios y Estándares para la Matemática Escolar* (NCTM, 2000), se convirtió en la base de la reforma del currículo escolar en muchos estados, la aceptación y el interés de la estadística como parte de la educación matemática ganó fuerza. En años recientes, muchos educadores matemáticos y estadísticos han dedicado grandes segmentos de sus carreras a mejorar los materiales para la educación estadística y las técnicas pedagógicas.

El NCTM no es el único grupo que llama a mejorar la educación estadística empezando en el nivel escolar. La Evaluación Nacional de Progreso Educativo (NAEP, 2005) fue desarrollada alrededor de las mismas cinco áreas de contenido como los estándares del NCTM, con las preguntas de análisis de datos y probabilidad jugando un rol cada vez más destacado en el examen NAEP. En 2006, el College Board [Consejo de Educación Superior] publicó sus *Estándares para el Éxito Universitario: Matemáticas y Estadística*, los cuales incluyen “Datos y Variación”, y “Azar, Justicia y Riesgo” entre su lista de ocho áreas temáticas que son “centrales al conocimiento y a las habilidades desarrollas en los años de escuela secundaria y preparatoria”. Un examen de los estándares recomendados por este documento revela un énfasis consistente con el análisis de datos, la probabilidad y la estadística en cada curso.

La emergencia del movimiento de alfabetización cuantitativa llama por un énfasis en habilidades cuantitativas prácticas que ayudarán a asegurar el éxito de los graduados de la preparatoria en la vida y el trabajo; muchas de estas habilidades son de naturaleza estadística. Citando a *Matemáticas y Democracia: El caso para la Alfabetización Estadística* (Steen, 2001):

> La alfabetización cuantitativa, también llamada numeramiento, es la herramienta natural para comprender información en la era de los computadores. La esperanza que los ciudadanos comunes sean alfabetizados cuantitativamente es primariamente un fenómeno del siglo XX… Infortunadamente, a pesar de años de estudio y experiencia de vida en un ambiente inmerso en los datos, muchos adultos educados permanecen iletrados funcionalmente…. La alfabetización estadística empodera la gente dándole herramientas para pensar por ellos mismos, para hacer preguntas inteligentes de expertos, y para confrontar la autoridad con confianza. Estas son las habilidades requeridas para prosperar en el mundo moderno.

Un estudio reciente de American Diploma Project (Proyecto de Diploma Americano), titulado en inglés Ready or Not: Creating a High School Diploma That Counts (*www.achieve.org/publications/ready-or-not-creating-high-school-diploma-counts*), recomienda las competencias “que se deben tener” para que los graduados de la preparatoria "tengan éxito en la educación superior o en el trabajo de alto desempeño o de alto crecimiento." Estos incluyen, además del algebra y la geometría, aspectos del análisis de datos, la estadística y otras aplicaciones que son de vital importancia para otras áreas, como también para el empleo en la economía de hoy, rica en datos.

La educación estadística como se propone en este *Marco* puede promover las habilidades que los graduados "deben tener" para "prosperar en el mundo moderno."

Los Estándares del NCTM y el Marco

El objetivo principal de este documento es proveer un *Marco* para la educación estadística de preescolar a grado 12. El fundamento de este *Marco* descansa en los *Principios y Estándares para la Matemática Escolar* del NCTM (2000).

El *Marco* está intencionado a complementar las recomendaciones de los *Principios y Estándares para la Matemática Escolar* del NCTM.

Los Principios y Estándares del NCTM describen los contenidos del área de estadística de la siguiente manera:

Análisis de Datos y Probabilidad

Los programas de enseñanza desde el preescolar hasta el grado 12 deben permitir a los estudiantes:

- formular preguntas que puedan ser direccionadas con datos y recolectar, organizar, y presentar datos relevantes para responderlas;
- seleccionar y usar métodos estadísticos apropiados para analizar los datos;
- desarrollar y evaluar inferencias y predicciones que estén basadas en los datos; y
- entender y aplicar conceptos básicos de probabilidad.

El estándar de "Análisis de Datos y Probabilidad" recomienda que los estudiantes formulen preguntas que puedan ser respondidas usando datos y traten sabiamente la recolección y uso de esos datos. Los estudiantes deben aprender como recolectar datos, organizar sus propios datos o los de otros, y presentar los datos en gráficos y tablas que les sean útiles para responder a sus preguntas. Este estándar también incluye aprender métodos para analizar datos y formas de hacer inferencias y sacar conclusiones de los datos. También se tratan los conceptos básicos y las aplicaciones de la probabilidad con énfasis en cómo la probabilidad y la estadística están relacionadas.

Los Principios y Estándares del NCTM profundizan en estos temas y proveen ejemplos de los tipos de lecciones y actividades que podrían ser usadas en un salón de clases. Ejemplos más completos se pueden encontrar en *Navigation Series on Data Analysis and Probability* [*Serie Navegación en Análisis de Datos y Probabilidad*] del NCTM (2002–2004). La estadística, sin embargo, es un área relativamente nueva para muchos profesores, quienes no han tenido la oportunidad de desarrollar un conocimiento sólido de los principios y conceptos subyacentes a las prácticas de análisis de datos que ahora ellos son llamados a enseñar. Estos profesores no comprenden claramente la diferencia entre estadística y matemática. Ellos no ven el currículo de estadística de preescolar a grado 12 como un área cohesiva y coherente del currículo. Estos profesores pueden no ver el currículo de estadística como "un todo" que provee una secuencia evolutiva de experiencias de aprendizaje.

Este *Marco* provee una estructura conceptual para la educación estadística que ofrece una imagen coherente del currículo global.

La Diferencia entre Estadística y Matemáticas

"La estadística es una disciplina metodológica. No existe por sí misma, sino que ofrece a otros campos de estudio un conjunto coherente de ideas y herramientas para el tratamiento de datos. La necesidad para tal disciplina surge de *la omnipresencia de la variabilidad*." (Moore y Cobb, 1997)

Un objetivo mayor de la educación estadística es ayudar a los estudiantes a desarrollar el pensamiento estadístico. El pensamiento estadístico, en gran parte, debe tratar con esta omnipresencia de la variabilidad; la resolución de problemas estadísticos y la toma de decisiones dependen de la comprensión, explicación y cuantificación de la variabilidad.

Es este enfoque en *la variabilidad de los datos* que diferencia la estadística de la matemática.

La Naturaleza de la Variabilidad

Hay muchas fuentes de variación en los datos. Algunas de las fuentes importantes se describen a continuación.

Medidas de Variabilidad—Las medidas repetidas en el mismo individuo varían. Algunas veces dos medidas varían porque los dispositivos de medida producen resultados no confiables, como cuando tratamos de medir una gran distancia con una regla pequeña. En otras ocasiones la variación resulta de cambios en el sistema a ser medido. Por ejemplo, aún con un dispositivo de medida preciso, tu registro de presión arterial podría ser diferente de un momento a otro.

Variabilidad Natural—La variabilidad es inherente a la naturaleza. Los individuos son diferentes. Cuando medimos la misma cantidad en diferentes individuos, estamos obligados a obtener diferencias en las medidas. Aunque algo de esto podría ser debido a nuestros instrumentos de medida, la mayor parte es simplemente debido al hecho que los individuos difieren. La gente naturalmente tiene diferentes estaturas, diferentes aptitudes y habilidades, y diferentes opiniones y respuestas emocionales. Cuando medimos alguno de estos rasgos, estamos obligados a obtener variabilidad en las medidas. Diferentes semillas de la misma variedad de frijol crecerán a diferente tamaño aun estando sujetas al mismo ambiente porque dos semillas no son exactamente iguales; hay variación de semilla a semilla en las medidas de crecimiento.

Variabilidad Inducida—Si plantamos un paquete de semillas de frijol en un campo y otro paquete de semillas en otro terreno con diferente clima, entonces la diferencia observada en crecimiento entre las semillas de un terreno con las del otro terreno podría ser debida a la diferencia inherente a las semillas (variabilidad natural), o la diferencia observada podría ser debida al hecho que los terrenos no son iguales. Si un tipo de fertilizante es usado en un terreno y otro tipo en otro, entonces las diferencias observadas podrían ser debidas a las diferencias en el fertilizante. Para el caso, la diferencia observada podría ser debida a un factor que no hemos siquiera pensado. Un experimento diseñado más cuidadosamente puede ayudarnos a determinar los efectos de diferentes factores.

Esta idea básica, comparar variabilidad natural con la variabilidad inducida por otros factores, constituye la esencia de la estadística moderna. Lo anterior ha permitido a la ciencia médica concluir que algunas drogas son efectivas y seguras, mientras otras son inefectivas o tienen efectos colaterales dañinos. Ello ha sido empleado por científicos de la agricultura para demostrar que una variedad de maíz crece mejor en un clima que en otro, que un fertilizante es más efectivo que otro, o que un tipo de alimento es mejor para el ganado vacuno que otro.

Variabilidad Muestral—En las encuestas políticas, parece razonable usar la proporción de votantes encuestados (una muestra estadística) como una cantidad estimada de la proporción desconocida de todos los votantes quienes apoyan a un candidato particular. Pero si una segunda muestra del mismo tamaño es usada, es casi seguro que no habrá exactamente la misma proporción de votantes en la muestra que apoya al candidato. El valor de la proporción muestral variará de muestra a muestra. Esto es llamado variabilidad muestral. Así que, ¿qué es lo que permite a una muestra estimar que la verdadera proporción es .60 y a otra decir que es .40? Esto es posible, pero improbable, si han sido usadas técnicas apropiadas de muestreo. Los resultados de encuestas de sondeo son útiles porque el uso de estas técnicas y de un adecuado tamaño de muestra puede asegurar que las diferencias inaceptables entre muestras sean improbables.

Una discusión excelente sobre la naturaleza de la variabilidad se da en *Seeing Through Statistics* [*Mirando a través de la Estadística*] (Utts, 1999).

El Rol del Contexto

"El enfoque en la variabilidad naturalmente da a la estadística un contenido particular que la diferencia de las matemáticas, y de otras ciencias matemáticas, pero es más que sólo el contenido lo que distingue el pensamiento estadístico de las matemáticas. La estadística requiere una forma diferente de pensar, porque *los datos no son sólo números, ellos son números en un contexto*. En matemáticas, el contexto oscurece la estructura. En análisis de datos, el contexto proporciona significado". (Moore y Cobb, 1997)

Muchos problemas matemáticos surgen de contextos aplicados, pero el contexto se remueve para revelar patrones matemáticos. Los estadísticos, como los matemáticos, buscan patrones, pero el significado de los patrones depende del contexto.

Una gráfica que ocasionalmente aparece en la sección de negocios del periódico muestra un diagrama del Promedio Industrial Dow Jones (DJIA por sus siglas en inglés—Dow Jones Industrial Average) en un periodo de 10 años. La variabilidad de los precios de las acciones roba la atención de un inversionista. El índice de acciones puede subir o bajar en intervalos de tiempo, y puede caer o subir abruptamente en periodos cortos de tiempo. En contexto,

"En matemáticas, el contexto oscurece la estructura. En análisis de datos, el contexto proporciona significado."

el diagrama plantea preguntas. Un inversionista serio no sólo está interesado en cuándo o qué tan rápido el índice sube o baja, sino también por qué. ¿Qué estaba pasando en el mundo cuando el mercado subió, qué estaba pasando en el mundo cuando él bajó? Ahora quite el contexto. Remueva el tiempo (años) del eje horizontal y llámelo "X", remueva el valor de la acción (DJIA) del eje vertical y llámelo "Y", y ¡allí hay una gráfica de muy poco interés o contenido matemático!

Probabilidad

La probabilidad es una herramienta para la estadística.

La probabilidad es un componente importante de cualquier formación matemática. Es una parte de la matemática que enriquece el área como un todo por su interacción con otros usos de las matemáticas. La probabilidad es una herramienta esencial en matemáticas aplicada y en modelación matemática. Es también una herramienta esencial en estadística.

El uso de la probabilidad como un modelo matemático y el uso de la probabilidad como una herramienta en estadística no sólo emplea diferentes aproximaciones, sino también diferentes tipos de razonamiento. Los dos problemas siguientes y la naturaleza de su solución ilustrarán la diferencia.

Problema 1:

Asuma que una moneda es "justa."[3]

Pregunta: ¿Si lanzamos la moneda cinco veces, cuántas caras obtendríamos?

Problema 2:

Tú eliges una moneda.

Pregunta: ¿Es esta una moneda justa?

El Problema 1 es un problema de probabilidad matemática.

El Problema 2 es un problema estadístico que puede usar el modelo de probabilidad matemática determinado en el problema 1 como una herramienta para buscar la solución.

Ninguna de las respuestas es determinista. El lanzamiento de una moneda produce resultados aleatorios, los cuales sugieren que el resultado es probabilístico. La solución al Problema 1 empieza con la suposición que la moneda es justa y procede a *deducir* lógicamente las probabilidades numéricas para cada posible número de caras: 0, 1, …, 5.

La solución al Problema 2 inicia con una moneda extraña, no sabemos si es justa o sesgada. La búsqueda por una respuesta es experimental—lance la moneda y mire que pasa. Examine los resultados para ver si lucen como si vinieran de una moneda justa o de una moneda sesgada. Hay varias aproximaciones posibles, lance la moneda cinco veces y registre el número de caras. Luego, hágalo de nuevo: Lance la moneda cinco veces y registre el número de caras. Repítalo 100 veces. Compile las frecuencias de los resultados para cada número posible de caras. Compare estos resultados con las frecuencias predichas en el modelo matemático para una moneda justa en el Problema 1. Si las frecuencias

3 "Justa" se refiere en este contexto a una moneda equilibrada. Es decir, que la probabilidad de que al lanzarla salga cara es ½ y no está cargada hacia uno u otro lado.

empíricas del experimento son diferentes de las predichas por el modelo matemático para una moneda justa y no es probable que sean causadas por la variación aleatoria en el lanzamiento de la moneda, entonces concluimos que la moneda no es justa. En este caso, *inducimos* una respuesta sacando una conclusión general de observaciones de resultados experimentales.

Probabilidad y Variabilidad del Azar

Dos usos importantes de la "aleatorización" en el trabajo estadístico ocurren en el muestreo y en el diseño de experimentos. En muestreo, nosotros "seleccionamos aleatoriamente", y en experimentos, aleatoriamente asignamos individuos a diferentes tratamientos. La aleatorización hace mucho más que remover el sesgo en la selección y asignación. La aleatorización conduce a *variabilidad del azar* en resultados que pueden ser descritos con modelos probabilísticos.

La probabilidad de algo nos dice del porcentaje de las veces que se espera que suceda cuando el proceso básico es repetido una y otra vez. La teoría probabilística no dice mucho acerca de un lanzamiento de una moneda; hace predicciones acerca del comportamiento a largo plazo de muchos lanzamientos de la moneda.

La probabilidad nos dice poco acerca de las consecuencias de la selección aleatoria para una muestra, pero describe la variación que esperamos ver en muestras cuando el proceso de muestreo es repetido un gran número de veces. La probabilidad nos dice poco acerca de las consecuencias de la asignación aleatoria para un experimento, pero describe la variación que esperamos ver en los resultados cuando el experimento es repetido muchas veces.

Cuando la aleatorización está presente, los estadísticos quieren saber si el resultado observado es debido al azar o a algo más. Esta es la idea de *significancia estadística*.

El Rol de las Matemáticas en la Educación Estadística

La evidencia que la estadística es diferente de las matemáticas no se presenta para argumentar que las matemáticas no son importantes en la educación estadística o que la educación estadística no debe ser parte de la educación matemática. Por el contrario, la educación estadística llega a ser altamente matemática cuando el nivel de comprensión aumenta. Sin embargo, el diseño de recolección de información, la exploración de datos y la interpretación de resultados deben estar acentuados en la educación estadística para alcanzar la alfabetización estadística. Estos dependen en gran medida del contexto y en un nivel introductorio, involucra matemática formal limitada.

La probabilidad juega un importante rol en el análisis estadístico, pero la probabilidad matemática formal debe tener su propio lugar en el currículo. La educación estadística preuniversitaria debe enfatizar las formas en las que la probabilidad es usada en el pensamiento estadístico; una comprensión intuitiva de la probabilidad será suficiente a esos niveles.

En Esta Sección

El Marco

La resolución de problemas estadísticos es un proceso investigativo que involucra cuatro componentes:

I. Formulación de Preguntas

→ Clarificar el problema en cuestión

→ Formular una o más preguntas que puedan ser respondidas con datos

II. Recolección de Datos

→ Diseñar un plan para recolección de los datos apropiados

→ Emplear el plan para recolección de los datos

III. Análisis de Datos

→ Seleccionar métodos gráficos y numéricos apropiados

→ Usar estos métodos para analizar los datos

IV. Interpretación de Resultados

→ Interpretar el análisis

→ Relacionar la interpretación con la pregunta original

El Rol de la Variabilidad en el Proceso de Resolución de Problemas

I. Formulación de Preguntas

Anticipar la Variabilidad—Hacer la Distinción en las Preguntas Estadísticas

La formulación de una pregunta estadística requiere comprensión de la diferencia entre una pregunta que anticipa una respuesta determinista y una pregunta que anticipa una respuesta basada en datos y variabilidad.

La pregunta "¿qué tan alto soy?" será respondida con una simple estatura. Esta no es una pregunta estadística. La pregunta "¿Qué tan altos son los hombres adultos en USA?" ¡No sería una pregunta estadística si todos estos hombres tuvieran exactamente la misma estatura! El hecho que haya diferentes estaturas, sin embargo, implica que nos anticipamos a una respuesta basada en medidas de estaturas que varían. Esto es una pregunta estadística.

El autor de la pregunta, "¿Cómo la luz del sol afecta el crecimiento de una planta?" debe anticipar que el crecimiento de dos plantas del mismo tipo expuestas a la misma luz del sol probablemente diferirá. Esta es una pregunta estadística.

II. Recolección de Datos

Reconocer la Variabilidad—Diseñar para las Diferencias

Los diseños de recolección de datos deben reconocer la variabilidad en los datos, y frecuentemente están destinados a reducir la variabilidad. El muestreo aleatorio está destinado a reducir las diferencias entre la muestra y la población. El tamaño de la muestra influencia el efecto de variabilidad de la muestra (error). Los diseños experimentales son elegidos para reconocer las diferencias entre grupos sometidos a diferentes tratamientos. La asignación

"La formulación de una pregunta estadística requiere comprensión de la diferencia entre una pregunta que anticipa una respuesta determinista y una pregunta que anticipa una respuesta basada en datos y variabilidad."

aleatoria a los grupos es intencionada para reducir las diferencias entre los grupos debido a los factores que no son manipulados en el experimento.

Algunos diseños experimentales emparejan sujetos de manera que sean similares. Los gemelos generalmente son emparejados en experimentos médicos para que las diferencias observadas puedan ser más probablemente atribuidas a la diferencia en tratamiento, más que a la diferencia en los sujetos.

La comprensión de los diseños de recolección de datos que reconoce las diferencias se requiere para la efectiva recolección de datos.

III. Análisis de Datos

Dar cuenta de la variabilidad—Usar Distribuciones

El propósito principal del análisis estadístico es dar cuenta de la variabilidad en los datos. Cuando los resultados de una encuesta establecen: "42% de los encuestados apoyan un candidato particular con un margen de error +/- 3% con un nivel de confianza del 95%", el foco es en la variabilidad de la muestra. Las encuestas de sondeo dan una estimación del apoyo entre los votantes. El margen de error indica qué tan lejos los resultados de la muestra (42% +/-3%) pueden diferir del porcentaje real de todos los votantes que apoyan el candidato. El nivel de confianza nos dice qué tan frecuente las estimaciones producidas por el método empleado producirán resultados correctos. Este análisis está basado en la distribución de estimaciones de muestreo aleatorio repetido.

Cuando los resultados de una prueba están descritos como "normalmente distribuidos con media 450 y desviación estándar de 100", el enfoque está en cómo las diferentes puntuaciones difieren desde la media. La distribución normal describe un patrón en forma de campana de los puntajes, y la desviación estándar indica el nivel de variación de los puntajes desde la media.

Dar cuenta de la variabilidad con el uso de distribuciones es la idea clave en el análisis de datos.

IV. Interpretación de Resultados

Permitir la Variabilidad—Mirar Más Allá de los Datos

Las interpretaciones estadísticas son hechas en presencia de la variabilidad y debe permitirla.

Los resultados de una encuesta de elección deben ser interpretados como una estimación que puede variar de muestra a muestra. La generalización de resultados de encuestas a la población de votantes mira más allá de la muestra de votantes encuestados y debe permitir la posibilidad de variabilidad de los resultados entre muestras diferentes. Los resultados de un experimento médico comparativo aleatorizado debe ser interpretado en presencia de la variabilidad debido al hecho que diferentes individuos responden diferente al mismo tratamiento y a la variabilidad debido a la aleatorización. La generalización de los resultados miran más allá de los datos recogidos de sujetos participantes en el experimento y debe permitir esas fuentes de variabilidad.

Mirar más allá de los datos para hacer generalizaciones debe permitir variabilidad en los datos.

Maduración de los Niveles

El profesional estadístico maduro entiende el rol de la variabilidad en el proceso de resolución de problemas estadísticos. Al momento de formular preguntas, el estadístico anticipa la recolección de datos, la naturaleza del análisis, y las posibles interpretaciones—los cuales involucran posibles fuentes de variabilidad. Al final, el profesional reflexiona sobre todos los aspectos de la recolección de datos y análisis como también sobre la pregunta, cuando interpreta los resultados. Igualmente, él o ella vinculan la recolección de datos y el análisis a cada uno de los otros dos componentes.

No puede esperarse que los estudiantes principiantes hagan todas estas conexiones. Ellos requieren años de experiencia y entrenamiento. La educación estadística debe ser mirada como un proceso de desarrollo. Para conseguir los objetivos propuestos, este reporte ofrece un marco teórico para la educación estadística en tres niveles. Si el objetivo fuera generar un profesional estadístico maduro, ciertamente habría varios niveles más allá de estos. No se intenta ligar estos niveles a grados educativos específicos.

El Marco usa tres niveles de desarrollo: A, B y C. Aunque estos tres niveles podrían ser análogos a grados educativos, están basados en el desarrollo de la alfabetización estadística, no en la edad. Así, un estudiante de secundaria quien no ha tenido experiencia previa con la estadística necesitará empezar con conceptos y actividades del Nivel A antes de pasar al Nivel B. Esto es cierto también para estudiantes de preparatoria. Si un estudiante no ha tenido experiencias de Nivel A y B antes de la escuela preparatoria, entonces no es apropiado para ese estudiante saltar a las expectativas del nivel C. El aprendizaje es guiado por el profesor en el Nivel A pero se hace más guiado por el estudiante en los Niveles B y C.

El Modelo *del Marco*

La estructura conceptual para la educación estadística se imparte en un modelo bi-dimensional mostrado en la Tabla 1. Una dimensión es definida por los componentes del proceso de resolución de problemas más la naturaleza de la variabilidad considerada y cómo nosotros nos enfocamos en variabilidad. La segunda dimensión está compuesta de los tres niveles de desarrollo.

Cada una de las cuatro primeras filas describe un componente de proceso y cómo este se desarrolla a lo largo de los niveles. La quinta fila indica la naturaleza de la variabilidad considerada en un nivel dado. Se entiende que el trabajo al Nivel B asume y desarrolla más a fondo los conceptos del Nivel A; similarmente, el Nivel C asume y usa conceptos de niveles inferiores.

Leyendo hacia abajo una columna describirá un problema de investigación completo para un nivel particular junto con la naturaleza de la variabilidad considerada.

Tabla 1: El Marco

Componente de Proceso	Nivel A	Nivel B	Nivel C
I. Formulación de Pregunta	**Comienzo de toma de conciencia de *la distinción de la pregunta estadística*** Los profesores plantean preguntas de interés Las preguntas se restringen al salón de clase	**Incrementa la toma de conciencia de *la distinción de la pregunta estadística*** Los estudiantes empiezan a plantear sus propias preguntas de interés Las preguntas no se restringen al salón de clase	**Los estudiantes *pueden distinguir las preguntas estadísticas*** Los estudiantes plantean sus propias preguntas de interés Las preguntas buscan generalización
II. Recolección de Datos	Todavía no hay *diseño para las diferencias* Censo en el salón de clase Experimento simple	Inicia la toma de conciencia de *diseño para las diferencias* Encuestas por muestreo; inicia el uso de selección aleatoria Experimentos comparativos; inicia el uso de asignación aleatoria	Los estudiantes hacen *diseños para las diferencias* Diseños muestrales con selección aleatoria Diseño experimental con aleatorización
III. Análisis de Datos	Uso de propiedades particulares de distribuciones en el contexto de un ejemplo específico Despliega variabilidad dentro de un grupo Compara individuo con individuo Compara individuo con grupo Inicia la conciencia de grupo con grupo Observa asociación entre dos variables	Aprende a usar propiedades particulares de distribuciones como herramientas de análisis Cuantifica la variabilidad dentro de un grupo Compara grupo con grupo en representaciones gráficas Reconoce el error muestral Alguna cuantificación de la asociación; modelos simples para la asociación	Entiende y usa distribuciones en el análisis como un concepto global Mide la variabilidad dentro de un grupo; mide la variabilidad entre grupos Compara grupo con grupo usando representaciones y medidas de variabilidad Describe y cuantifica el error muestral Cuantificación de la asociación; ajuste de modelos para la asociación

Componente de Proceso	Nivel A	Nivel B	Nivel C
IV. Interpretación de Resultados	Los estudiantes no miran *mas allá de los datos* No generalizan más allá del salón de clase Notan la diferencia entre dos individuos con diferentes condiciones Observan asociación en las representaciones gráficas	Los estudiantes reconocen que es posible mirar *más allá de los datos* Reconocen que una muestra puede o no ser representativa de la población mayor Notan la diferencia entre dos grupos con diferentes condiciones Son conscientes de la diferencia entre estudios observacionales y experimentos Notan diferencias en la fuerza de la asociación Dan interpretaciones básicas de modelos de asociación Son conscientes de la distinción entre asociación y causa y efecto	Los estudiantes pueden mirar *más allá de los datos* en algunos contextos Generalizan de muestra a población Son conscientes del efecto de la aleatorización en los resultados de experimentos Entienden la diferencia entre estudios observacionales y experimentos Interpretan medidas de fuerza de la asociación Interpretan modelos de asociación Distinguen entre conclusiones de estudios de asociación y experimentos
Naturaleza de la Variabilidad	Medidas de variabilidad Variabilidad natural Variabilidad inducida	Variabilidad muestral	Variabilidad del azar
Centrarse en la Variabilidad	Variabilidad dentro de un grupo	Variabilidad dentro de un grupo y variabilidad entre grupos Covariación	Variabilidad en el ajuste de modelos

> Las ilustraciones de actividades de aprendizaje dadas aquí son intencionadas para clarificar las diferencias entre los niveles de desarrollo para cada componente de proceso de resolución de problemas.

Ilustraciones

Los cuatro pasos del proceso de resolución de problemas son usados en los tres niveles, pero la comprensión profunda y la sofisticación de métodos usados incrementan a través de los niveles A, B, y C. Esta maduración en comprensión en el proceso de resolución de problemas y sus conceptos subyacentes es paralela al aumento de la complejidad en el rol de la variabilidad. Las ilustraciones de actividades de aprendizaje dadas aquí son intencionadas para clarificar las diferencias entre los niveles de desarrollo para cada componente de proceso de resolución de problemas. Las últimas secciones darán ilustraciones del proceso completo de resolución de problemas para actividades de aprendizaje en cada nivel.

I. Formulación de Preguntas

Ejemplo sobre la Longitud de una Palabra

Nivel A: ¿Qué tan largas son las palabras en esta página?

Nivel B: ¿Son las palabras en un capítulo de un libro de quinto grado más largas que las palabras en un capítulo de un libro de tercer grado?

Nivel C: ¿Los libros de quinto grado usan palabras más largas que los libros de tercer grado?

Ejemplo sobre la Música de Moda

Nivel A: ¿Qué tipo de música es más popular entre los estudiantes de nuestra clase?

Nivel B: ¿Cómo los tipos de música favoritos se comparan entre diferentes clases?

Nivel C: ¿Qué tipo de música es más popular entre los estudiantes de nuestra escuela?

Ejemplo sobre la Estatura y la Extensión de los Brazos

Nivel A: En nuestra clase, ¿es la estatura y la extension de los brazos de los estudiantes aproximadamente las mismas?

Nivel B: ¿Es la relación entre la extensión de los brazos y la estatura para los estudiantes en nuestra clase igual que la relación entre la extensión de los brazos y la estatura para los estudiantes en otra clase?

Nivel C: ¿Es la estatura un predictor útil de la extensión de los brazos para los estudiantes de nuestra escuela?

Ejemplo sobre el Crecimiento de una Planta

Nivel A: ¿Una planta ubicada cerca a la ventana crecerá más alta que una planta ubicada lejos de la ventana?

Nivel B: ¿Cinco plantas ubicadas cerca de la ventana crecerán más altas que cinco plantas ubicadas lejos de la ventana?

Nivel C: ¿Cómo el nivel de luz solar afecta el crecimiento de las plantas?

II. Recolección de Datos

Ejemplo sobre la Longitud de una Palabra

Nivel A: ¿Qué tan largas son las palabras en esta página?

Se determina la longitud de cada palabra en la página y se registra.

Nivel B: ¿Son las palabras en un capítulo de un libro de quinto grado más largas que las palabras en un capítulo de un libro de tercer grado?

Se usa una muestra aleatoria de palabras de cada capítulo.

Nivel C: ¿Usan los libros de quinto grado palabras más largas que los libros de tercer grado?

Se consideran y comparan diferentes diseños muestrales, y se usan algunos. Por ejemplo, en vez de seleccionar una muestra aleatoria simple de páginas del libro y se usan todas las palabras en las páginas elegidas como muestra.

Nota: En cada nivel, se deben direccionar los asuntos relacionados con la medición. La longitud de una palabra depende de la definición de "palabra." Por ejemplo, ¿Es un número una palabra? La consistencia de la definición ayuda a reducir variabilidad de la medición.

Ejemplo sobre el Crecimiento de una Planta

Nivel A: ¿Una planta ubicada cerca a la ventana crecerá más alto que una planta ubicada lejos de la ventana?

Un vástago se siembra en una maceta que es ubicado en el umbral de la ventana. Un segundo vástago del mismo tipo y tamaño se planta en una maceta que se ubica lejos del umbral de la ventana. Después de seis semanas, el cambio en la altura de cada una se mide y se registra.

Nivel B: ¿Cinco plantas ubicadas cerca de la ventana crecerán más altas que cinco plantas del mismo tipo ubicadas lejos de la ventana?

Cinco vástagos del mismo tipo y tamaño se siembran en una maceta que se ubica en el umbral de la ventana. Cinco vástagos del mismo tipo y tamaño se siembran en una maceta que se ubica lejos del umbral de la ventana. Se usan números aleatorios para decidir cuales plantas van en la ventana. Después de seis semanas, se mide y registra el cambio en la altura de cada vástago.

Nivel C: ¿Cómo el nivel de luz solar afecta el crecimiento de las plantas?

Se seleccionan quince vástagos del mismo tipo y tamaño. Se usan tres macetas, se siembran cinco de estos vástagos en cada una. Se seleccionan quince vástagos de otra variedad para determinar el efecto de la luz solar es el mismo en diferentes tipos de plantas. Cinco de estos se siembran en cada una de las tres macetas. Las tres macetas son ubicadas en sitios con tres niveles de luz diferentes. Se usan números aleatorios para decidir cuáles plantas van en cuál maceta. Después de seis semanas, se mide y registra el cambio en la altura de cada vástago.

Nota: En cada nivel, se deben direccionar los asuntos relacionados con la medición. El método para medir el cambio en la altura debe ser claramente entendido y aplicado para reducir la variabilidad de la medición.

III. Análisis de Datos

Ejemplo sobre la Música de Moda

Nivel A: ¿Qué tipo de música es más popular entre los estudiantes de nuestra clase?

Se usa un diagrama de barras para representar gráficamente el número de estudiantes que eligen cada categoría musical.

Nivel B: ¿Cómo los tipos de música favoritos se comparan entre diferentes clases?

Para cada clase, se usa un diagrama de barras para representar gráficamente el porcentaje de estudiantes que eligen cada categoría musical. Las mismas escalas se usan para ambos gráficos de manera que puedan comparar fácilmente.

Nivel C: ¿Qué tipo de música es más popular entre los estudiantes de nuestra escuela?

Se usa un diagrama de barras para representar gráficamente el porcentaje de estudiantes que eligen cada categoría musical. Como se usa una muestra aleatoria, se da un estimado del margen de error.

Nota: En cada nivel, se deben direccionar los asuntos relacionados con la medición. Se usará un cuestionario para recoger las preferencias musicales de los estudiantes. El diseño y la redacción del cuestionario deben ser cuidadosamente considerados para evitar posibles sesgos en las respuestas. La elección de las categorías musicales también podrían afectar los resultados.

Ejemplo sobre la Estatura y la Extensión de los Brazos

Nivel A: En nuestra clase, ¿es la estatura y la extension de los brazos de los estudiantes aproximadamente las mismas?

Se determina la diferencia entre la estatura y la extensión de los brazos por cada individuo. Se construye un gráfico X-Y (diagrama de dispersión) con X = estatura, Y = extensión de los brazos. Se dibuja la línea Y = X en este gráfico.

Nivel B: ¿Es la relación entre la extensión de los brazos y la estatura para los estudiantes en nuestra clase igual que la relación entre la extensión de los brazos y la estatura para los estudiantes en otra clase?

Para cada clase, se construye un gráfico X-Y con X = estatura, Y = extensión de brazos. Se dibuja una línea “a ojo” en cada gráfica para describir la relación entre estatura y extensión del brazo. Se determina la ecuación de esta línea. Se calcula una medida elemental de asociación.

Nivel C: ¿Es la estatura un predictor útil de la extensión de los brazos para los estudiantes de nuestra escuela?

Se determina y evalúa la línea de regresión por mínimos cuadrados para usar como un modelo predictor.

Nota: En cada nivel, se deben direccionar los asuntos de medición. Los métodos usados para medir la estatura y la extensión de los brazos deben ser claramente entendidos y aplicados para reducir la variabilidad de la medición. Por ejemplo, ¿medimos la estatura con los zapatos puestos?

IV. Interpretación de Resultados

Ejemplo sobre la Longitud de una Palabra

Nivel A: ¿Qué tan largas son las palabras en esta página?

Se examina y se resume el diagrama de puntos de la longitud de todas las palabras. En particular, los estudiantes notarán la longitud de la palabra más larga y más corta, las longitudes más comunes y menos comunes, y la longitud en el medio.

Nivel B: ¿Son las palabras en un capítulo de un libro de quinto grado más largas que las palabras en un capítulo de un libro de tercer grado?

Los estudiantes interpretan una comparación de la distribución de una muestra de longitud de palabras de un libro de quinto grado con la distribución de la longitud de palabras de un libro de tercer grado usando un diagrama de cajas y bigotes para representar cada uno de ellos. Los estudiantes también reconocen que las muestras que están siendo usadas podrían o no podrían ser representativas de los capítulos completos.

El diagrama de cajas y bigotes para una muestra de la longitud de la palabra de un libro de quinto grado se ubica al lado de un diagrama de cajas y bigotes de una muestra de un libro de tercer grado.

Nivel C: ¿Usan los libros de quinto grado palabras más largas que los libros de tercer grado?

La interpretación al Nivel C incluye la interpretación al Nivel B, pero también debe considerar generalización de los libros incluidos en el estudio a una población mayor de libros.

Ejemplo sobre el Crecimiento de una Planta

Nivel A: ¿Una planta ubicada cerca a la ventana crecerá más alta que una planta ubicada lejos de la ventana?

En este experimento simple, la interpretación es sólo una cuestión de comparar una medida de cambio en tamaño con otra.

Nivel B: ¿Cinco plantas ubicadas cerca de la ventana crecerán más altas que cinco plantas ubicadas lejos de la ventana?

En este experimento, los estudiantes deben interpretar una comparación de un grupo de cinco medidas con otro grupo. Si se nota una diferencia, entonces el estudiante reconoce que es causada probablemente por la diferencia en las condiciones de iluminación.

Nivel C: ¿Cómo el nivel de luz solar afecta el crecimiento de las plantas?

Hay varias comparaciones de grupos posibles con este diseño. Si se percibe una diferencia, entonces el estudiante reconoce que es causada probablemente por la diferencia en las condiciones de iluminación o la diferencia en los tipos de plantas. También se reconoce que la aleatorización usada en el experimento puede resultar en algunas de las diferencias observadas.

Naturaleza de la Variabilidad

El enfoque en la variabilidad se vuelve cada vez más sofisticado a medida que los estudiantes progresan a través de los niveles de desarrollo.

Variabilidad dentro de un Grupo

Este es el único tipo que se considera en el Nivel A. En el ejemplo de la longitud de la palabra, se consideran las

“El enfoque en la variabilidad se vuelve cada vez más sofisticado a medida que los estudiantes progresan a través de los niveles de desarrollo.”

diferencias entre las longitudes de palabras en una página individual; esta es la variabilidad dentro de un grupo de longitudes de palabras. En el ejemplo de la música de moda, se consideran las diferencias en cómo los estudiantes eligen cada categoría de música; esta es la variabilidad dentro de un grupo de frecuencias.

Variabilidad dentro de un Grupo y Variabilidad entre Grupos

Al Nivel B, los estudiantes empiezan a hacer comparaciones de grupos de medidas. En el ejemplo de la longitud de palabras, se compara un grupo de palabras de un libro de quinto grado con un grupo de palabras de un libro de tercer grado. Tal comparación no sólo nota cuánto difieren la longitud de las palabras dentro de cada grupo, sino que también toma en cuenta las diferencias entre los dos grupos, tales como la diferencia entre las medianas o las medias de las longitudes de las palabras.

Covariación

Al Nivel B, los estudiantes también empiezan a investigar la relación "estadística" entre dos variables. La naturaleza de esta relación estadística se describe en términos de cómo las dos variables "co-varian". En el ejemplo sobre la estatura y la extensión de los brazos, por ejemplo, si las estaturas de dos estudiantes difieren por dos centímetros, entonces nos gustaría que nuestro modelo de la relación nos dijera cuanto podríamos esperar que sus extensiones de brazos difieran.

Variabilidad en el Ajuste del Modelo

Al Nivel C, los estudiantes evalúan qué tan bien una línea de regresión predecirá los valores de una variable desde valores de otra variable usando gráficos residuales. En la situación sobre la estatura y la extensión de los brazos, por ejemplo, esta evaluación está fundamentada en examinar si las diferencias entre la extensión actual de los brazos y la extensión predicha de los brazos por el modelo varia aleatoriamente cerca de la línea horizontal de "no diferencia" en el gráfico residual. La inferencia sobre un valor predicho de y para un valor dado de x es válida sólo si los valores de y varían aleatoriamente de acuerdo a una distribución normal centrada en la línea de regresión. Los estudiantes al Nivel C aprenden a estimar esta variabilidad alrededor de la regresión lineal usando la desviación estándar estimada de los residuos.

Variabilidad Inducida

En el ejemplo del crecimiento de la planta al Nivel B, el experimento se diseña para determinar si habrá una diferencia entre el crecimiento de plantas en luz solar y plantas alejadas de la luz solar. Queremos determinar si una diferencia impuesta en los ambientes inducirá una diferencia en el crecimiento.

Variabilidad Muestral

En el ejemplo de la longitud de palabras al Nivel B, se usan muestras de palabras de un capítulo. Los estudiantes observan que dos muestras producirán diferentes grupos de longitudes de palabras. Esta es la variabilidad muestral.

Variabilidad Aleatoria del Muestreo

Cuando se usa selección aleatoria, las diferencias entre las muestras serán debido al azar. Entender esta variación del azar es lo que conduce a la predictibilidad de los resultados. En el ejemplo de la música de moda, al Nivel C, no sólo se considera esta variación de azar, sino que es la base para comprender el concepto de margen de error.

Variabilidad de Azar Resultando de la Asignación a Grupos en Experimentos

En el ejemplo del crecimiento de la planta al Nivel C, las plantas son aleatoriamente asignadas a los grupos. Los estudiantes consideran cómo esta variación del azar en la asignación aleatoria puede producir diferencias en resultados, aunque un análisis formal no se lleve a cabo.

Descripción Detallada de Cada Nivel

A medida que este documento evoluciona en descripción detallada de cada nivel, es importante notar que los ejemplos seleccionados para ilustrar los conceptos claves y el proceso de resolución de problemas de razonamiento estadístico están fundamentados en datos reales y en contextos del mundo real. *Ustedes que leen este documento, necesitarán ser flexibles adaptando estos ejemplos de modo que se ajusten a sus circunstancias de enseñanza.*

En Esta Sección

Nivel A

Los niños están rodeados de datos. Ellos pueden pensar sobre los datos como un registro de las preferencias de los estudiantes, tal como el tipo favorito de música, o como medidas, tales como la extensión de los brazos de los estudiantes y el número de libros en sus bolsos escolares.

Es en el Nivel A que los niños necesitan desarrollar el sentido de los datos—una comprensión que los datos son más que números. La estadística cambia los números en información.

Los estudiantes deben aprender que los datos son generados con respecto a contextos o situaciones particulares y pueden ser usados para responder preguntas acerca del contexto o situación.

Se deben suministrar oportunidades a los estudiantes para generar preguntas acerca de un contexto particular (tal como su salón de clase) y determinar qué datos pueden ser recolectados para responder estas preguntas.

Los estudiantes también deben aprender cómo usar herramientas estadísticas básicas para analizar los datos y hacer inferencias informales al responder las preguntas propuestas.

Finalmente, los estudiantes deben desarrollar ideas básicas de probabilidad para apoyar su uso posterior de probabilidad al extraer inferencias en los Niveles B y C.

Es preferible que los estudiantes recolecten datos, pero no necesariamente en cada caso. Los profesores deben tomar ventaja de situaciones que ocurren naturalmente en las cuales los estudiantes notan un patrón acerca de algunos datos y empiezan a formular preguntas. Por ejemplo, cuando se toma asistencia en la mañana, los estudiantes pueden notar que muchos estudiantes están ausentes. El profesor podría capitalizar esta oportunidad para que los estudiantes formulen preguntas que podrían ser respondidas con datos de la asistencia.

Específicamente, las recomendaciones en el proceso investigativo del Nivel A incluyen:

I. Formular la Pregunta

→ Los profesores ayudan a proponer preguntas (preguntas en contextos de interés para los estudiantes).

→ Los estudiantes distinguen entre solución estadística y respuesta fija.

II. Recolección de Datos para Responder la Pregunta

→ Los estudiantes realizan un censo de la clase.

→ Los estudiantes entienden la variabilidad natural de individuo a individuo.

→ Los estudiantes ejecutan experimentos simples con asignación no aleatoria de tratamientos.

→ Los estudiantes entienden la variabilidad inducida atribuible a una condición experimental.

III. Análisis de Datos

→ Los estudiantes comparan individuo con individuo.

→ Los estudiantes comparan individuos con un grupo.

→ Los estudiantes empiezan a ser conscientes de la comparación de grupo con grupo.

→ Los estudiantes entienden la idea de una distribución.

→ Los estudiantes describen una distribución.

→ Los estudiantes observan asociación entre dos variables.

→ Los estudiantes usan herramientas para explorar distribuciones y asociación, incluyendo:

- Gráfico de barras
- Gráfico de puntos
- Diagrama de tallos y hojas
- Diagrama de dispersión
- Tablas (usando conteos)
- Media, Mediana, Moda, Rango
- Categoría Modal

IV. Interpretación de Resultados

→ Los estudiantes hacen inferencias al salón de clase.

→ Los estudiantes reconocen que los resultados pueden ser diferentes en en otro salón de clase o grupo.

→ Los estudiantes reconocen la limitación de alcance de inferencia a los salones de clase.

Los niños en el Nivel A pueden estar interesados en el tipo de música favorito entre los estudiantes en cierto nivel. Una fiesta de fin de año se está planeando y sólo hay dinero suficiente para contratar un grupo musical. La clase puede investigar la pregunta: *¿Qué tipo de música está de moda entre los estudiantes?*

Ejemplo 1: Eligiendo la Banda para la Fiesta de Fin de Año—Realizando una Encuesta

Esta pregunta intenta medir una característica en la población de niños en el grado en el cual se tendrá la fiesta. La característica, tipo de música favorito, es una variable categórica—cada niño en ese grado será ubicado en una categoría particular no-numérica basado en su tipo de música favorito. Los datos resultantes frecuentemente son llamados *datos categóricos*.

La clase de Nivel A muy probablemente llevaría a cabo un censo de los estudiantes en un salón particular para estimar cuál puede ser el tipo de música favorito para todo el grado. Al Nivel A, queremos que los estudiantes reconozcan que habrá variación de individuo a individuo.

Por ejemplo, se toma una encuesta de 24 estudiantes en uno de los salones en un grado particular. Los datos se resumen en la tabla de frecuencia siguiente. Esta *tabla de*

Tabla 2: Tabla de Conteo de Frecuencia

Favorito	Conteo de Frecuencia
Country	8
Rap	12
Rock	4

frecuencia es una *representación tabular* que toma los estudiantes de Nivel A a un nivel sumativo para datos categóricos. Los estudiantes pueden primero usar *marcas de conteo* para registrar las medidas de los datos categóricos antes de encontrar frecuencias (conteo) para cada categoría.

Al Nivel A, el estudiante puede primero usar una *imagen gráfica* para representar los recuentos para cada categoría. Una imagen gráfica usa una foto de algún tipo (tal como una banda musical) para representar cada individuo. De este modo, cada niño que prefiere un tipo particular de música pondría un recorte de ese tipo de banda directamente en la gráfica que el profesor ha creado en el pizarrón. En vez de una foto de una banda, otra representación—como una foto de una guitarra, una X, o un cuadrado de color—puede ser usado para representar la preferencia de cada individuo. Un niño que prefiere "country" iría al pizarrón y ubicaría una guitarra, un punto, una X, o un cuadrado de color sobre la columna nombrada "country". En ambos casos, hay un registro deliberado de cada valor, uno por uno.

Note que una imagen gráfica se refiere a una gráfica donde un objeto, tal como una construcción de papel recortado, se usa para representar un individuo en la gráfica. (Un recorte de un diente podría ser usado para

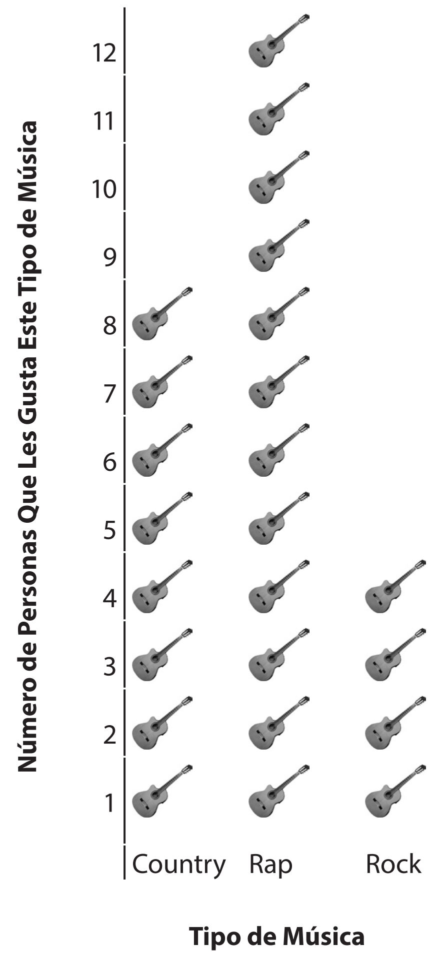


Figura 1. Imagen gráfica de las preferencias musicales

registrar cuántos dientes se le cayeron a los niños de una clase de preescolar cada mes.) El término *pictograma* frecuentemente se usa para referirse a una gráfica en la cual

> “Los estudiantes deben comprender que la moda es la categoría que contiene más datos, frecuentemente se llama la categoría modal.”

una foto o un símbolo se usan para representar varios ítems que pertenecen a la misma categoría. Por ejemplo, en una gráfica que muestra la distribución de conductores de carro, caminantes, y de quienes montan en bus en una clase, un recorte de un bus escolar podría ser usado para representar cinco que montan en bus. Así, si la clase tiene 13 que montan en bus, habrán aproximadamente 2.5 buses en la gráfica.

Este tipo de gráfica requiere una comprensión básica de razonamiento proporcional o multiplicativo, y por esta razón no recomendamos su uso en el Nivel A. Similarmente, las gráficas circulares requieren una comprensión de razonamiento proporcional, así que no recomendamos su uso en el Nivel A.

Un gráfico de barras lleva al estudiante al nivel de síntesis con los datos resumidos de otra representación, tal como una imagen gráfica o una tabla de conteo de frecuencias. La barra en un gráfico de barras se dibuja como un rectángulo, elevándolo al número deseado en el eje *y*.

Un gráfico de barras de las preferencias musicales de los estudiantes se despliega en la Figura 2 para el censo llevado a cabo por los estudiantes y que es representado en la tabla de conteo de frecuencias y en la imagen gráfica previa.

Los estudiantes al Nivel A deben reconocer la *moda* como una forma de describir un valor “representativo” o “típico” de la distribución.

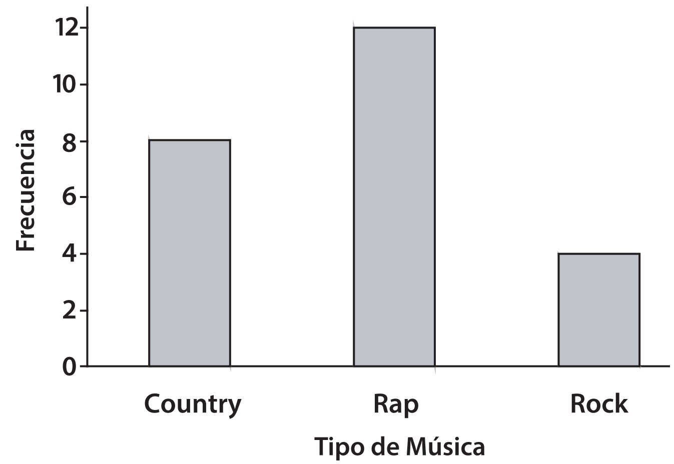


Figura 2: Gráfico de barras de preferencias musicales

La moda es más útil para datos categóricos. Los estudiantes deben comprender que la moda es la categoría que contiene más datos, frecuentemente se llama la *categoría modal*. En nuestro ejemplo de la música favorita, la música rap fue preferida por más niños, así la moda o categoría modal del conjunto de datos es la música rap. Los estudiantes podrían usar esta información para ayudar a los profesores a buscar un grupo musical para la fiesta de fin de año que sea especialista en música rap.

El eje vertical de la gráfica de barras en la Figura 2 se podría escalar en términos de la proporción de porcentaje de la muestra para cada categoría. Como esto involucra razonamiento proporcional, convertir frecuencias a proporciones (o porcentajes) será desarrollado en el Nivel B.

Como la mayoría de los datos recogidos en el Nivel A involucrarán un censo de los estudiantes de la clase, la primera etapa es que los estudiantes aprendan a leer y a interpretar a un nivel básico lo que los datos muestran acerca de su propia clase. Leer e interpretar viene antes de la inferencia. Es importante considerar la pregunta:

¿Qué podría haber causado que los datos luzcan así?

Es también importante que los niños piensen si y cómo los hallazgos podrían *"ampliarse" a un grupo mayor*, tal como todo el grado, toda la escuela, todos los niños en el sistema escolar, o toda la gente en la nación. Ellos deben notar variables (tales como edad o ubicación geográfica) que pueda afectar los datos en un conjunto más grande. En el ejemplo previo de la música, los estudiantes podrían especular que si ellos recolectan datos de preferencias musicales de sus profesores, los profesores pueden preferir un tipo de música diferente. O ¿qué pasaría si ellos recolectan las preferencias musicales de los estudiantes de la escuela secundaria en su sistema escolar? Los estudiantes al Nivel A deben empezar reconociendo las limitaciones del alcance de la inferencia a un salón específico.

Comparando Grupos

Los estudiantes al Nivel A pueden estar interesados en comparar dos grupos distintos con respecto a algunas características de esos grupos. Por ejemplo, ¿hay una diferencia entre dos grupos—niños y niñas—con respecto a la participación de los estudiantes en deportes? La característica "participación en deportes" es categórica (sí o no). Los datos categóricos resultantes para cada género pueden ser analizados usando una tabla de conteos de frecuencia o un gráfico de barras. Otra pregunta que los estudiantes al Nivel A podrían preguntar es si hay diferencia entre niños y niñas con respecto a la distancia que ellos pueden saltar, es un ejemplo de toma de medidas en una variable numérica. Los datos en variables numéricas se obtienen de situaciones que involucran la toma de medidas, tales como la estatura o temperaturas, o situaciones en las cuales se cuentan objetos (por ejemplo, determinar el número de letras en tu primer nombre, el número de bolsillos en la ropa que visten los niños de la clase, o el número de hermanos que cada uno tiene). Tales datos frecuentemente se llaman datos numéricos.

Retornando a la pregunta de comparar niños con niñas con respecto a saltar distancias, los estudiantes pueden medir la distancia saltada por todos sus compañeros de clase. Una vez los datos numéricos han sido recogidos, los niños pueden comparar la longitud de los saltos de

Pulgadas Saltadas en el Salto de Longitud en Parada

Niñas		Niños
	8	
	7	
	6	1
	5	2 6 9
9 7 2	4	1 3 5 5 5
5 5 3 3 3 2 1	3	1 1 2 5 6 7
9 8 7 7 6 4 4 3 2	2	2 3 4 6
	1	

Figura 3: Diagrama de tallos y hojas de las distancias saltadas

las niñas con la de los niños usando un diagrama de tallos y hojas doble, como el que se presenta en la Figura 3.

Del diagrama de tallos y hojas, los estudiantes pueden obtener un sentido sobre la forma de la distribución de los datos—más simétrica para los niños que para las niñas—y el hecho que los niños tienden a tener saltos más largos. Más adelante en el Nivel C, se discutirán más formalmente los ejemplos previos de diseño de recolección de datos como ejemplos de estudios observacionales. El investigador no tiene control sobre cuales estudiantes van en el grupo de niños y cuales en el grupo de niñas (la condición pre—existente de género define los grupos). El investigador sólo observa y recolecta medidas de características dentro de cada grupo.

El Experimento Simple

Otro tipo de diseño para recolectar datos apropiados al Nivel A es el *experimento simple*, el cual consiste en tomar medidas en una condición particular de un grupo. Los estudiantes en el Nivel A pueden estar interesados en tomar el tiempo de oscilación de un péndulo o mirar que tan lejos un carro de juguete corre hasta el final de la pendiente de un punto de inicio fijo (¿futuros participantes en Pinewood Derby?) También, medir la misma cosa varias veces y encontrar una media ayuda a establecer los fundamentos para el hecho que la media tenga menos variabilidad como una estimación del valor de la verdadera media que una simple lectura. Esta idea será desarrollada más plenamente en el Nivel C.

Ejemplo 2: Cultivando Frijoles—Un Experimento Comparativo Simple

Un *experimento comparativo simple* es como un experimento científico en el cual los niños comparan los resultados de dos o más condiciones. Por ejemplo, los niños pueden plantar frijoles secos en tierra y dejarlos germinar, y luego comparar cuál crece más rápido—aquel en la luz o aquel en la oscuridad. Los niños deciden cuáles frijoles serán expuestos a un tipo particular de iluminación. Las condiciones a ser comparadas aquí son los dos tipos de ambientes de iluminación—luz y oscuridad. El tipo de ambiente de iluminación es un ejemplo de variable categórica. Las medidas de la altura de las plantas se pueden tomar al final de un periodo de tiempo especificado para responder la pregunta de si un ambiente de iluminación es mejor para el crecimiento de los frijoles. Las alturas recolectadas son un ejemplo de datos numéricos. En el Nivel C, será desarrollado a fondo el concepto de un experimento (donde las condiciones son impuestas por el investigador).

Otra *representación gráfica apropiada para datos numéricos en una variable* (además del diagrama de tallos y hojas) para el Nivel A es *el diagrama de puntos*. Ambos, el diagrama de puntos y el diagrama de tallos y hojas, pueden ser usados para comparar fácilmente dos o más conjuntos de datos numéricos similares. Al crear un diagrama de puntos, el eje x debe ser rotulado con un rango de los valores que la variable numérica pueda asumir. El eje x, para cualquier gráfico de una variable, convencionalmente es el eje que representa los valores de la variable bajo estudio. Por ejemplo, en el experimento del crecimiento del frijol, los niños pueden registrar en un gráfico de puntos la altura de las plantas de frijoles (en centímetros) que fueron cultivados en la oscuridad (etiquetados O) y en la luz (etiquetados L) usando un diagrama de puntos.

Es obvio al observar el diagrama de puntos que las plantas en el ambiente de luz tienden a tener alturas mayores que las plantas en el ambiente de oscuridad.

Buscar conglomerados o brechas en la distribución ayuda a los estudiantes a identificar la *forma* de la distribución. Los estudiantes deben desarrollar un sentido de por qué una distribución toma una forma particular para el contexto de la variable considerada.

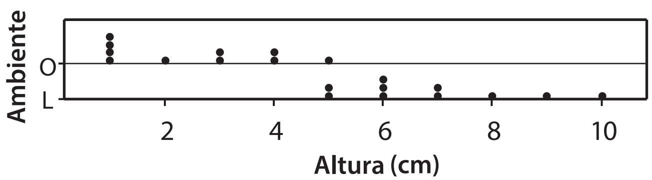


Figura 4: Diagrama de puntos de ambiente vs. altura

→ ¿Tiene la distribución un conglomerado (o montículo) principal con grupos más pequeños de tamaño similar en cada lado del conglomerado? Si es así, la distribución podría ser descrita como *simétrica*.

→ ¿Tiene la distribución un conglomerado principal con grupos más pequeños en cada lado que no son del mismo tamaño? Los estudiantes pueden clasificar esta como "desequilibrada", o pueden usar el término asimétrica.

→ ¿Por qué la distribución toma esta forma? Usando el diagrama de puntos anterior, los estudiantes reconocerán que ambos grupos tienen distribuciones que son "desequilibradas", con el conglomerado principal en el extremo inferior de la distribución y unos pocos valores a la derecha del montículo principal.

Usando Datos Disponibles

A la mayoría de los niños les encanta comer perros calientes (o hot dogs), pero son conscientes que mucho sodio no es necesariamente saludable. ¿Hay diferencia en el contenido de sodio de perros calientes de carne de res (etiquetado B en la Figura 5) y perros calientes de aves (etiquetado P en la Figura 5)? Para investigar esta pregunta, los estudiantes pueden hacer uso de datos disponibles. Usando datos de la revista *Consumers Reports* [*Reporte de Consumidores*] de la edición de Junio de 1993, se pueden construir gráficos de puntos paralelos.

Los estudiantes notarán que la distribución de los perros calientes de aves tiene dos conglomerados diferentes. ¿Qué

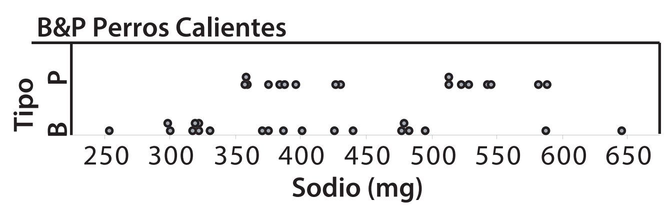


Figura 5: Diagrama paralelo de puntos del contenido de sodio

“A medida que los estudiantes avanzan al Nivel B, considerar la forma de la distribución conducirá a una comprensión de que medidas son apropiadas para describir el centro y la dispersión.”

puede explicar la brecha y los dos conglomerados? Podría ser otra variable, tal como el precio de los perros calientes de aves, siendo más costosos aquellos que tienen menos sodio. También se puede observar que las cantidades de sodio en los perros calientes de carne de res son más dispersas (o varían más) que en los de aves. Adicionalmente, parece que el centro de la distribución para los perros calientes de aves es mayor que el centro de la distribusion para los perros calientes de carne de res.

A medida que los estudiantes avanzan al Nivel B, considerar la forma de la distribución conducirá a una comprensión de que medidas son apropiadas para describir el centro y la dispersión.

Describiendo Centro y Dispersión

Los estudiantes deben entender que *la mediana* describe el centro de un conjunto de datos numéricos en términos de cuántos datos están por encima y por debajo de ella. El mismo número de datos (aproximadamente la mitad) yacen a la izquierda de la mediana y a la derecha de la mediana. Los niños pueden crear una gráfica humana para mostrar cuantas letras hay en sus nombres. Todos los niños con nombres de dos letras pueden estar en una fila, con todos los niños con nombres de tres letras parados en una fila paralela. Una vez los niños estén organizados, el profesor puede pedir a un niño de cada uno de los extremos de la gráfica que se siente, repitiendo este proceso hasta que un solo niño permanezca parado, representando la mediana. Con estudiantes de Nivel A, recomendamos usar un número impar de datos para que así la mediana sea clara hasta que los estudiantes hayan dominado la idea de punto medio.

Los estudiantes deben entender *la media como una medida de centro de reparto justo* al Nivel A. En el ejemplo de la longitud del nombre, la media sería interpretada como “¿Qué tan largos deben ser nuestros nombres si ellos son todos de la misma longitud?” Esto puede ser ilustrado en pequeños grupos haciendo que los niños tomen un cubo por cada letra en sus nombres. En pequeños grupos, haga que los estudiantes pongan todos los cubos en el centro de la mesa y redistribúyanlos uno a la vez de manera que cada niño tenga igual número de cubos. Dependiendo de la experiencia de los niños con el uso de fracciones, ellos pueden decir que la longitud media del nombre es 4 R 2 o 4 ½ o 4.5. Otro ejemplo sería que el profesor recogiera ocho lápices de los niños, de longitudes variadas y los pusiera extremo con extremo en el carril de la tiza. Encontrar la media será responder la pregunta “¿Qué tan largo sería cada lápiz si ellos fueran todos de la misma longitud?” Esto es, si pudiéramos pegar todos los lápices y cortarlos en ocho secciones iguales, ¿qué tan larga sería cada sección? Esto puede ser modelado usando cinta de papel de maquinas registradoras (o una cuerda), cortando un pedazo de cinta de la misma longitud de los ocho lápices y unir extremo con extremo. Luego, doblar la cinta a la mitad tres veces para obtener octavos, mostrar la longitud de un lápiz de los ocho de igual longitud. Ambas demostraciones pueden ser alineadas directamente en el algoritmo para encontrar la media: combinar todos los valores (poner todos los cubos en el medio, poner los lápices extremo con extremo y medir, sumar todos los valores) y repartir equitativamente (distribuya los cubos, doble la cinta, y divida por el número de datos). Los estudiantes del Nivel A deben dominar el cálculo (manual o usando la tecnología apropiada) de la media de manera que se puedan desarrollar interpretaciones más sofisticadas de la media en el Nivel B y el Nivel C.

La media y la mediana son *medidas de localización* para describir el centro de un conjunto de datos numéricos. Determinar los valores máximos y mínimos de un conjunto de datos ayuda a los niños a describir la posición de los valores más pequeños y más grandes en un conjunto de datos. Además de describir el centro de un conjunto de datos, es útil saber cómo los datos varían o qué tan dispersos están.

Una *medida de dispersión* para la distribución es el rango, el cual es la diferencia entre el valor máximo y mínimo. Las medidas de dispersión sólo tienen sentido con datos numéricos.

Mirando al diagrama de tallos y hojas formado por las distancias saltadas (Figura 3), el rango difiere para los niños (rango = 39 pulgadas) y las niñas (rango = 27 pulgadas). Las niñas son más consistentes en sus distancias de salto que los niños.

Buscando Asociación

Los estudiantes deben estar en condiciones de buscar una posible *asociación de una variable numérica y una variable categórica* mediante la comparación de diagramas de puntos de una variable numérica desagregada por una varia-ble categórica. Por ejemplo, usando los diagramas paralelos de puntos mostrando el crecimiento de las plantas de frijoles en la luz y en la oscuridad, los estudiantes deben buscar similitudes dentro de cada categoría y diferencias entre las categorías. Como se mencionó antes, los estudiantes deben reconocer fácilmente en los diagramas de puntos que, en general, las plantas de los frijoles en el ambiente de luz han crecido más altos, y por tal razón es mejor para estas plantas tener un ambiente de luz. También se pueden comparar las medidas de centro y dispersión. Por ejemplo, los estudiantes podrían calcular o hacer estimados visuales de la altura media de los frijoles cultivados en ambientes de luz y los frijoles cultivados en ambientes de oscuridad para fundamentar sus afirmaciones que la condiciones de luz es mejor para los frijoles. Ellos también pueden notar que el rango para las plantas cultivadas en la oscuridad es 4 cm, y 5 cm para plantas cultivadas en la luz. Juntar esa información con la media debe habilitar a los estudiantes a consolidar aún más sus conclusiones acerca de las ventajas de cultivar los frijoles en la luz.

Considerando los datos de los perros calientes, una impresión general del diagrama de puntos es que hay más variación en el contenido de sodio para los perros calientes de carne de res. Para estos, el contenido de sodio está entre 250 mg y 650 mg, mientras para los perros calientes de aves, el contenido de sodio está entre 350 mg 600 mg. Ni los centros ni las formas de la distribución son obvios desde los diagramas de puntos. Es interesante notar los dos conglomerados de datos aparentes para los perros calientes de aves. Nueve de los 17 perros calientes de aves tienen contenido de sodio entre 350 mg y 450 mg, mientras ocho de los 17 perros calientes de aves tienen contenido de sodio entre 500 mg y 600 mg. Una posible explicación para esta división es que algunos perros calientes de aves son hechos de pollo y otros de pavo.

Ejemplo 3: Comprando Sudaderas—El Rol de la Estatura y la Extensión de los Brazos

¿Qué pasa con la asociación entre dos variables numéricas? Las organizaciones de profesores y padres de familia de las escuelas de educación primaria tienen una recaudación de fondos popular "vestir el espíritu", tales como camisetas o pantalones de ejercicio con el nombre de la escuela y la mascota. Los organizadores necesitan tener algunas orientaciones acerca de cuántas prendas de cada talla ordenar. ¿Deben ellos ofrecer la camiseta y el pantalón separadamente u ofrecerlos como un conjunto? ¿Están cercanamente relacionadas las estaturas y las extensiones de los brazos de los estudiantes de la escuela o ellos difieren considerablemente debido al patrón de crecimiento individual de los niños? Por consiguiente, algunas preguntas útiles para responder son:

¿Existe asociación entre la estatura y la extensión del brazo?

¿Qué tan fuerte es la asociación entre la estatura y la extensión del brazo?

Un *diagrama de dispersión* se puede usar para representar datos gráficamente cuando los valores de dos variables numéricas son obtenidos del mismo individuo u objeto. ¿Podemos usar la estatura de una persona para obtener su extensión de brazos? Los estudiantes pueden medirse

"Con el uso de un diagrama de dispersión, los estudiantes de Nivel A pueden visualmente buscar por patrones o tendencias."

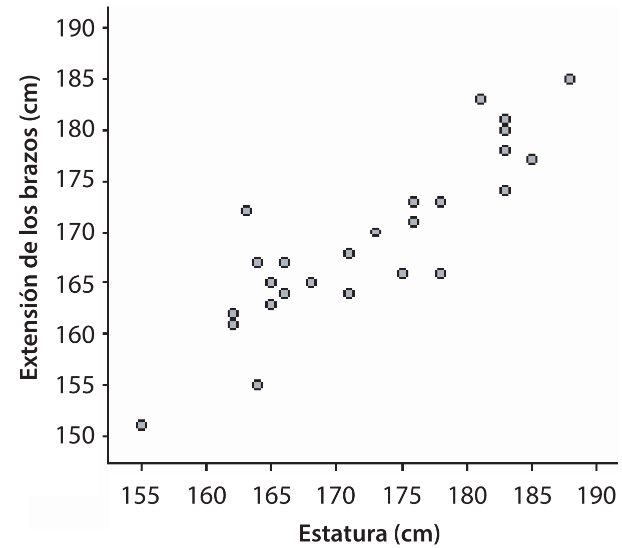


Figura 6: Diagrama de dispersión de la extensión de los brazos vs. estatura

unos a otros la estatura y la extensión de los brazos, y luego construir un diagrama de dispersión para observar la relación entre estas dos variables numéricas. Se mide la estatura y la extensión de los brazos (en centímetros) de 26 estudiantes. Los datos que se presentan a continuación son para estudiantes universitarios y se incluyen con propósitos ilustrativos.

Con el uso de un diagrama de dispersión, los estudiantes de Nivel A pueden visualmente *buscar por patrones o tendencias.*

Por ejemplo, en el anterior diagrama de dispersión de la extensión de los brazos versus la estatura, los estudiantes deben estar en condiciones de identificar la relación consistente entre las dos variables: generalmente mientras la una crece, la otra también. Basado en estos datos, los organizadores se podrían sentir cómodos encargando algunos conjuntos completos de camiseta y pantalones basados en las tallas. Sin embargo, algunos estudiantes podrían necesitar encargar la camiseta y el pantalón separadamente basados en las tallas. Otra pregunta importante que los organizadores necesitarán hacerse es si esta muestra es representativa de todos los estudiantes en la escuela. ¿Cómo se seleccionó la muestra?

Los estudiantes en el Nivel A también pueden usar diagramas de dispersión para observar gráficamente el cambio en los valores de una variable numérica en el tiempo, éste se refiere como un *diagrama de tiempo*. Por ejemplo, los niños pueden graficar la temperatura exterior en varios momentos del día registrando ellos mismos los valores o usando datos de un periódico o de la Internet.

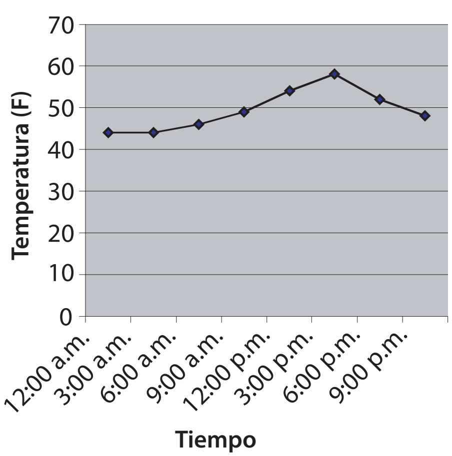


Figura 7: Diagrama de tiempo de temperatura vs. tiempo

Cuando los estudiantes avancen al Nivel B, cuantificarán estas tendencias y patrones con medidas de asociación.

Comprendiendo la Variabilidad

Los estudiantes deben explorar las posibles razones por las cuales los datos lucen de la forma que lucen y *diferenciar entre variación y error*. Por ejemplo, al graficar los colores de dulces en un paquete, los niños podrían esperar que los colores estén uniformemente distribuidos (o que puedan saber por la experiencia previa que no lo están). Los niños podrían especular sobre por qué ciertos colores aparecen más o menos frecuente debido a la variación (por ejemplo, los costos de colorantes, investigación de mercado en las preferencias de las personas, etc.). Los niños también podrían identificar posibles espacios donde los errores puedan haber ocurrido en el manejo de datos/dulces (por ejemplo, dulces caídos, dulces atorados en la bolsa, dulces que se han comido, dulces regalados a otros, colores no contados porque no se ajustan a las preferencias personales, mal conteo). Los profesores deben sacar provecho de la *ocurrencia natural del "error"* que sucede cuando se recogen datos en el salón de clase y ayudar a los estudiantes a especular sobre el *impacto de estos errores* en los resultados finales. Por ejemplo, cuando se pide a los estudiantes votar por su alimento favorito, es común que los estudiantes voten dos veces, olviden votar, registren su voto en el lugar equivocado, no comprendan lo que se les pregunta, cambien su opinión, o quieran votar por una opción que no está enlistada. Los errores de conteo también son comunes entre los niños pequeños, lo cual puede conducir a recuentos incorrectos de datos en las categorías. Los profesores pueden ayudar a los estudiantes a pensar sobre cómo estos eventos pueden afectar los resultados finales si sólo una persona hiciera esto, si varias personas lo hicieran, o si muchas personas lo hicieran. Los estudiantes pueden generar ejemplos adicionales de las formas en que los errores pueden ocurrir en una situación particular de recolección de datos.

Las nociones de error y variabilidad deben ser usadas para explicar los datos extremos , conglomerados y brechas que los estudiantes observan en las representaciones gráficas de los datos. Una comprensión del error frente a la variabilidad natural ayudará a los estudiantes a interpretar si un dato extremo es un valor legítimo que es inusual o si el dato extremo es debido a la forma de registrar de los datos.

Al Nivel A, es imperativo que los estudiantes empiecen a comprender el concepto de variabilidad. A medida que los estudiantes progresan del Nivel A, al Nivel B, al Nivel C, es importante mantener en primer plano que la *comprensión de la variabilidad es la esencia del desarrollo del sentido de los datos*.

El Rol de la Probabilidad

Los estudiantes en el Nivel A necesitan desarrollar ideas básicas de probabilidad para apoyar su uso posterior de probabilidad para inferir en los Niveles B y C.

En el Nivel A, los estudiantes deben entender que la *probabilidad es una medida de la posibilidad de que algo ocurrirá. Es una medida de certidumbre o incertidumbre*. Los eventos deben ser vistos como yaciendo en un continuo de imposible a seguro, con las opciones menos probable, igualmente probable, y más probable dentro de este continuo. Los estudiantes informalmente aprenden a asignar números a la probabilidad de que algo ocurrirá.

Un ejemplo de asignación de números en una línea numérica se da a continuación:

0	1/4	1/2	3/4	1
Imposible	Improbable o menos probable	Igualmente probable de ocurrir y no ocurrir	Probable o más probable	Seguro

Los estudiantes deben tener experiencias *estimando probabilidades usando datos empíricos*. Mediante la experimentación (o simulación), los estudiantes deben desarrollar una comprensión explícita de la noción que mientras más veces se repita un fenómeno aleatorio, más cercanos serán los resultados al modelo matemático esperado. En el Nivel A, nosotros consideramos sólo modelos simples basados en resultados igualmente probables o, a lo sumo, basado en ello, como ser la suma de las caras de dos dados. Por ejemplo, los niños muy pequeños pueden establecer que al lanzar una moneda, esta puede caer en cara la mitad de las veces y en cruz la mitad de las veces. El estudiante ha dado el modelo esperado y la probabilidad de obtener una cara o una cruz, asumiendo que la moneda es "justa."

Si un niño lanza una moneda 10 veces para obtener datos empíricos, es muy posible que no obtenga cinco caras y cinco cruces. Sin embargo, si el niño lanza la moneda cien veces, podríamos esperar que los resultados empezarán a *estabilizarse* a las probabilidades esperadas de 0.5 para caras y 0.5 para cruces. Esto se conoce como *la Ley de los Números Grandes*. Por lo tanto, en el Nivel A, los experimentos probabilísticos deben enfocarse en obtener datos empíricos para desarrollar interpretaciones de frecuencias relativas que los niños puedan fácilmente trasladar a modelos con probabilidades "matemáticas" conocidas y comprensibles. Los clásicos lanzamientos de monedas, giros de ruletas y lanzamiento de dados son herramientas confiables para ayudar a estudiantes del Nivel A en la comprensión de probabilidad. El concepto de interpretación de frecuencias relativas será importante al Nivel B cuando los estudiantes trabajen con razonamiento proporcional—ir de conteos o frecuencias a proporciones o porcentajes.

Mientras los estudiantes trabajan con resultados de fenómenos aleatorios repetidos, pueden desarrollar una comprensión del concepto de *aleatoriedad*. Ellos verán que cuando lanzan una moneda 10 veces, aunque esperáramos cinco caras y cinco cruces, los resultados variarán de un estudiante al otro. También verán que si una cara resulta en un lanzamiento, eso no significa que el siguiente resultado será una cruz. Porque el lanzamiento de una moneda es un experimento aleatorio, hay siempre incertidumbre en cómo la moneda caerá de un lanzamiento al otro. Sin embargo, en el Nivel A, los estudiantes pueden empezar a desarrollar la noción que aunque tenemos incertidumbre y variabilidad en nuestros resultados, examinando qué pasa con el proceso aleatorio a *largo plazo*, nosotros podemos cuantificar la incertidumbre y variabilidad con probabilidades—dando un número predictivo para la probabilidad de un resultado en el largo plazo. En el Nivel B, los estudiantes verán el rol que la probabilidad juega en el desarrollo del concepto de muestra aleatoria simple y el rol que la probabilidad juega en la aleatoriedad.

Usos Indebidos de la Estadística

Los estudiantes en el Nivel A deben aprender que el uso apropiado de terminología estadística es tan importante como el uso apropiado de herramientas estadísticas. En particular, se debe enfatizar en el uso apropiado de la media o mediana. Estos resúmenes numéricos son apropiados para describir variables numéricas, no variables categóricas. Por ejemplo, cuando se recolectan datos categóricos como el tipo de música favorito, *el número* de niños en la muestra que prefieren cada tipo de música se resume como una frecuencia. Es fácil confundir datos categóricos y datos numéricos en este caso y tratar de encontrar la media o mediana de las frecuencias del tipo de música favorito. Sin embargo, uno no puede usar los conteos de frecuencia para calcular la media o mediana de una variable categórica. Los conteos de frecuencia *son* los resúmenes numéricos para variables categóricas.

Otro error común en estudiantes del Nivel A es el uso inapropiado del gráfico de barras con datos numéricos. Un gráfico de barras se usa para resumir variables categóricas. Si una variable es numérica, la visualización gráfica apropiada con barras se llama *histograma*, el cual es introducido en el Nivel B. Al Nivel A, la visualización gráfica apropiada para datos numéricos son los diagramas de puntos y el diagrama de tallos y hojas.

Resumen del Nivel A

Si los estudiantes se sienten cómodos con las ideas y conceptos descritos previamente, estarán preparados para desarrollar y mejorar su comprensión de conceptos clave para el sentido de datos en el Nivel B.

Es también importante reconocer que ayudar a los estudiantes a desarrollar el sentido de datos en el Nivel A permite que la instrucción matemática se imparta utilizando datos. Las ramas tradicionales de las matemáticas como algebra, funciones, geometría, y medida pueden todas ser desarrolladas utilizando datos. Dar sentido a los datos debe ser una parte integrada del currículo de matemáticas, empezando en el preescolar.

En Esta Sección

Nivel B

La enseñanza en el Nivel B debe construirse en la base estadística desarrollada en el Nivel A y sentar las bases estadísticas para el Nivel C. Las actividades de enseñanza en el Nivel B deben continuar enfatizando los cuatro componentes principales en el proceso investigativo y tener el espíritu de la práctica estadística genuina. Los estudiantes que completen el Nivel B deben ver el razonamiento estadístico como un proceso de resolución de problemas mediante datos y razonamiento cuantitativo.

En el Nivel B, los estudiantes están más conscientes de la distinción de la pregunta estadística (una pregunta con una respuesta basada en datos que varían, versus una pregunta con una respuesta determinista). Ellos también deben tomar decisiones acerca de qué variables medir y cómo medirlas para dar respuesta a la pregunta propuesta.

Los estudiantes deben usar y ampliar los resúmenes gráficos, tabulares, y numéricos introducidos en el Nivel A para investigar problemas más sofisticados. También al seleccionar una muestra, los estudiantes deben desarrollar una comprensión básica del rol que la probabilidad juega en la selección aleatoria—y en la asignación aleatoria cuando se lleva a cabo un experimento.

En el Nivel B, los estudiantes investigan problemas poniendo mayor énfasis en la posible asociación entre dos o más variables y entienden como una colección más sofisticada de resúmenes gráficos, tabulares y numéricos se usa para dar respuesta a esas preguntas. Finalmente, los estudiantes reconocen formas en las cuales la estadística se usa bien o mal en el mundo.

Específicamente, las recomendaciones en *el proceso investigativo* incluyen:

I. Formular Preguntas

→ Los estudiantes empiezan a proponer sus propias preguntas.

→ Los estudiantes dan respuesta a preguntas que involucran grupos mucho más grandes que su salón de clases y empiezan a reconocer la diferencia entre una población, un censo y una muestra.

II. Recolección de Datos

→ Los estudiantes realizan censos de dos o más salones de clase.

→ Los estudiantes diseñan y ejecutan encuestas por muestreo no aleatorio y empiezan a usar selección aleatoria.

→ Los estudiantes diseñan y ejecutan experimentos comparativos y empiezan a usar asignación aleatoria.

III. Análisis de Datos

→ Los estudiantes amplían su comprensión de distribuciones de datos.

→ Los estudiantes cuantifican la variabilidad dentro de un grupo.

→ Los estudiantes comparan dos o más distribuciones utilizando visualizaciones gráficas y resúmenes numéricos.

→ Los estudiantes utilizan herramientas mucho más sofisticadas para resumir y comparar distribuciones, incluyendo:

- Histogramas
- El RIC (Rango Inter-Cuartílico) y DMA (Desviación Media Absoluta)
- Las cinco medidas de resumen y diagramas de cajas

→ Los estudiantes reconocen el error muestral.

→ Los estudiantes cuantifican la fuerza de la asociación entre dos variables, desarrollan modelos simples para la asociación entre dos variables numéricas, y usan herramientas ampliadas para explorar asociación, incluyendo:

- Tablas de contingencia para dos variables categóricas
- Gráficos de series de tiempo
- La CCC (Cociente de Conteo de Cuadrante) como una medida de fuerza de la asociación
- Líneas simples para modelar la asociación entre dos variables numéricas

IV. Interpretación de Resultados

→ Los estudiantes describen diferencias entres dos o más grupos con respecto al centro, la dispersión y la forma de las distribuciones.

→ Los estudiantes reconocen que una muestra puede o no ser representativa de una población mayor.

→ Los estudiantes comprenden interpretaciones básicas de medidas de asociación.

→ Los estudiantes empiezan a distinguir entre un estudio observacional y un experimento diseñado.

→ Los estudiantes empiezan a distinguir entre "asociación" y "causa y efecto".

→ Los estudiantes reconocen variabilidad muestral en resúmenes estadísticos, tales como media muestral y proporción muestral.

Ejemplo 1: Nivel A Revisitado—Eligiendo una Banda para el Baile de la Escuela

Muchos de los resúmenes gráficos, tabulares y numéricos introducidos en el Nivel A pueden ser mejorados y usados para investigar problemas más sofisticados en el Nivel B. Revisitemos el problema de planificar el baile de la escuela introducido en el Nivel A, que mediante un censo de la clase, una clase de Nivel A investigó la siguiente pregunta:

¿Qué tipo de música es más popular entre los estudiantes?

Recordemos que la clase fue considerada como la totalidad de la población, y los datos fueron recolectados de cada miembro de la población. Una investigación similar en el Nivel B incluiría el reconocimiento de que un salón de clase podría no ser representativo de las opiniones de todos los estudiantes de la escuela. Los estudiantes del Nivel B podrían comparar las opiniones de sus salones de clase con las opiniones de otros salones de su escuela. Un salón de clase de Nivel B podría investigar las siguientes preguntas:

¿Qué tipo de música es más popular entre los estudiantes en nuestra escuela?

¿Cómo los tipos de música favoritos difieren entre los salones de clase?

Como los tamaños de los salones de clase pueden ser diferentes, los resultados deben ser resumidos con frecuencias relativas o porcentajes para ser comparados. Los porcentajes son útiles porque nos permiten pensar en resultados comparables para grupos de tamaño 100. Los estudiantes de Nivel B estarán expuestos a más énfasis en el razonamiento proporcional a lo largo del currículo de matemáticas, y ellos deben sentirse cómodos resumiendo e interpretando información en términos de porcentajes o fracciones.

Tabla 3: Frecuencias y Frecuencias Relativas

	Salón de Clase 1	
Favorita	Frecuencia	Frecuencia Relativa Porcentaje
Country	8	33%
Rap	12	50%
Rock	4	17%
Total	**24**	**100%**
	Salón de Clase 2	
Favorita	Frecuencia	Frecuencia Relativa Porcentaje
Country	5	17%
Rap	11	37%
Rock	14	47%
Total	**30**	**101%**

Los resultados de dos salones de clase se resumen en la Tabla 3 usando ambas frecuencias y frecuencias relativas (porcentajes).

El gráfico de barra en la Figura 8 compara los porcentajes de cada categoría de música favorita para los dos salones de clase.

Los estudiantes en el Nivel B deben empezar a reconocer que no sólo hay variabilidad de un individuo a otro dentro de un grupo, sino también en resultados de un grupo a otro. Este segundo tipo de variabilidad se ilustra por el hecho que la música más popular es rap para el Salón de Clase 1, mientras que es la música rock para el Salón de Clase 2.

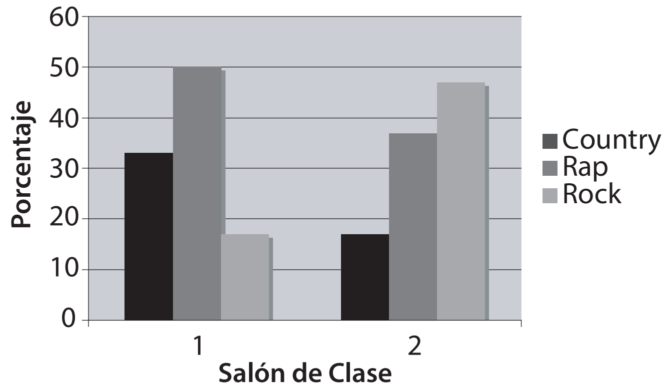


Figura 8: Gráfico de barras comparativo para las preferencias musicales

Esto es, la moda para la Clase 1 es la música rap, mientras que la moda para la Clase 2 es música rock.

Los resultados de las dos muestras pueden ser combinadas para tener una muestra mucho más grande de la escuela. Los resultados combinados indican que la música rap es el tipo de música favorito para 43% de los estudiantes, la música rock es preferida por 33%, mientras que sólo un 24% de los estudiantes seleccionaron música country como su favorita. Los estudiantes de Nivel B deben reconocer que aunque esta es una muestra más grande, aun podría no ser representativa de toda la población (todos los estudiantes en la escuela). En estadística, la aleatoriedad y la probabilidad son incorporadas en el procedimiento de selección de la muestra para proveer un método que es "justo" y para mejorar las posibilidades de seleccionar una muestra representativa.

Por ejemplo, si la clase decide seleccionar lo que se llama una muestra aleatoria simple de 54 estudiantes, entonces cada posible muestra de 54 estudiantes tiene la misma probabilidad de ser seleccionada. Esta aplicación ilustra uno de los roles de la probabilidad en la estadística. Aunque los estudiantes de Nivel B puede no utilizar un procedimiento de selección aleatorio cuando se recolectan datos, las cuestiones relacionadas con obtener muestras representativas deben ser discutidas en este nivel.

Conectando Dos Variables Categóricas

Como el rap fue el tipo de música más popular para los dos salones de clases combinadas, los estudiantes pueden argumentar a favor de un grupo de rap para el baile. Sin embargo, más de la mitad de los encuestados prefirieron música rock o country. ¿Estarían tristes estos estudiantes si una banda de rap es elegida? No necesariamente, ya que a muchos estudiantes que les gusta la música rock también les podría gustar la música rap. Para investigar este problema, los estudiantes podrían explorar dos preguntas adicionales:

¿A los estudiantes que les gusta la música rock tiende a gustarles o disgustarles la música rap?

¿A los estudiantes que les gusta la música country tienden a gustarles o disgustarles la música rap?

Para abordar estas preguntas, la encuesta debe preguntar a los estudiantes no sólo su tipo de música favorito, sino también si a ellos les gusta la música rap, rock y country.

La siguiente *tabla de frecuencias de dos entradas* (o *tabla de contingencia*) provee una forma de investigar posibles conexiones entre las dos variables categóricas.

Tabla 4: Tabla de Frecuencia de Dos Entradas

		¿Le Gusta la Música Rap?		
		Sí	**No**	**Total Filas**
¿Le Gusta la Música Rock?	**Sí**	27	6	33
	No	4	17	21
Total Columnas		31	23	54

De acuerdo a estos resultados, de los 33 estudiantes que les gusta la música rock, 27 también les gusta la música rap. Esto es, 82% (27/33) de los estudiantes que les gusta la música rock también les gusta la música rap. Esto indica que los estudiantes que les gusta la música rock también tiende a gustarles la música rap. Una vez más, note el uso del razonamiento proporcional en la interpretación de estos resultados. Un análisis similar se puede hacer para determinar si a los estudiantes que les gusta la música country tiende a gustarles o disgustarles la música rap. Una discusión más detallada de este ejemplo y una medida de asociación entre dos variables categóricas se da en el Apéndice para el Nivel B.

Cuestionarios y Sus Dificultades

En el Nivel B, los estudiantes deben comenzar a aprender acerca de las encuestas y los muchos escollos a evitar en el diseño y en la realización de ellas. Un problema es la redacción de preguntas. Las preguntas tienen que ser sin ambigüedades y fáciles de entender. Por ejemplo, la pregunta:

¿Estás en contra que la escuela implemente una política de no puertas en los cuartos de baño?

está redactada en una forma confusa. Una forma alternativa de formular esta pregunta es:

La escuela está considerando implementar una política de no puertas en los cuartos de baño. ¿Cuál es tu opinión con respecto a esta política?

Fuertemente Opuesto Opuesto No opinión Apoyo Apoyo Firmemente

Las preguntas deben evitar llevar al encuestado a una respuesta. Por ejemplo la pregunta:

Como nuestro equipo de futbol americano no ha ganado en 20 años y está costando dinero a la escuela en vez de generar recursos, ¿crees que deberíamos concentrarnos más en otro deporte, tal como el futbol o baloncesto?

está redactada en una forma que está sesgada en contra del equipo de futbol americano.

Las respuestas a preguntas con respuestas codificadas deben incluir todas las posibles opciones, y las respuestas no se deben traslapar. Por ejemplo, para la pregunta:

¿Cuánto tiempo invierte estudiando en casa en una noche típica?

las respuestas:

ninguna 1 hora o menos 1 hora o más

confundiría a un estudiante que invierte una hora por noche estudiando.

Hay muchas otras consideraciones acerca de la formulación de preguntas y gestión de encuestas que pueden ser introducidas al Nivel B. Dos de estos asuntos son cómo el entrevistador hace las preguntas y la precisión al registrar las respuestas. Es importante que los estudiantes se den cuenta que las conclusiones de sus estudios dependen de la precisión de sus datos.

“En el Nivel A, la media es interpretada como el valor de ”reparto justo” para los datos.”

Medidas de Localización—Media Como un Punto de Equilibrio

Otra idea desarrollada en el Nivel A que puede ser extendida al Nivel B es la media como un resumen numérico de centro en un conjunto de datos numéricos. En el Nivel A, la media es interpretada como el valor de “reparto justo” para los datos. Esto es, la media es el valor que obtendrías si todos los datos de los sujetos son combinados y luego redistribuidos equitativamente de manera que el valor de cada sujeto sea el mismo. Otra interpretación de la media es como punto de balance de la correspondiente distribución de datos. A continuación se presenta un bosquejo de una actividad que ilustra la noción de la media como punto de balance. A nueve estudiantes se les preguntó:

¿Cuántas mascotas tienes?

Los datos resultantes fueron 1, 3, 4, 4, 4, 5, 7, 8, 9. Estos datos se resumen en el diagrama de puntos mostrado en la Figura 9. Note que en la actividad real realizada en el aula, notas adhesivas se usaron como “puntos” en vez de Xs.

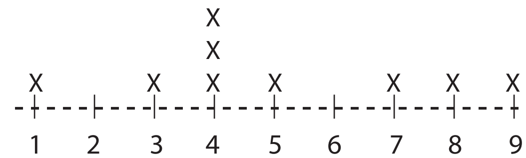


Figura 9: Diagrama de puntos para el conteo de mascotas

Si las mascotas se combinan en un grupo, hay un total de 45 mascotas. Si las mascotas se redistribuyen uniformemente entre los nueve estudiantes, entonces cada estudiante tendría cinco mascotas. Es decir, el número medio de mascotas es cinco. El diagrama de puntos representando los resultados que los nueve estudiantes tienen exactamente cinco mascotas se muestra a continuación:

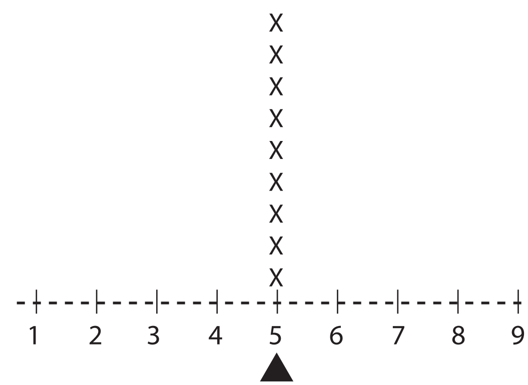


Figura 10: Diagrama de puntos mostrando las mascotas uniformemente distribuidas

Es obvio que si un pivote se ubica en el valor 5, entonces el eje horizontal se "balanceará" en este punto pivote. Es decir, el "punto de equilibrio" del eje horizontal para este diagrama de puntos es 5. ¿Cuál es punto de equilibrio para el diagrama de puntos que muestra los datos originales?

Empezamos por señalar lo que sucede si uno de los puntos sobre 5 se elimina y se ubica sobre el valor 7, como se muestra a continuación.

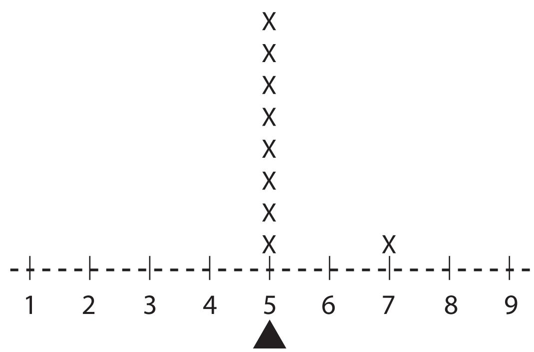


Figura 11: Diagrama de puntos con un dato movido

Claramente, si el pivote permanece en 5, el eje horizontal se inclinará hacia la derecha. ¿Qué se puede hacer con los puntos restantes sobre 5 para "reequilibrar" el eje horizontal en el punto pivote? Como 7 es dos unidades por encima de 5, una solución es mover un punto dos unidades por debajo de 5 a 3, como se muestra a continuación:

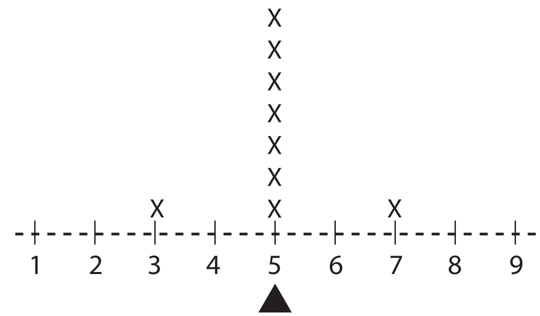


Figura 12: Diagrama de puntos con dos datos movidos

El eje horizontal está reequilibrado ahora en el punto pivote. ¿Es esta la única forma de reequilibrar el eje en 5? No. Otra forma de reequilibrar el eje en el punto pivote sería mover dos puntos sobre la columna de 5 a la columna de 4, como se muestra a continuación:

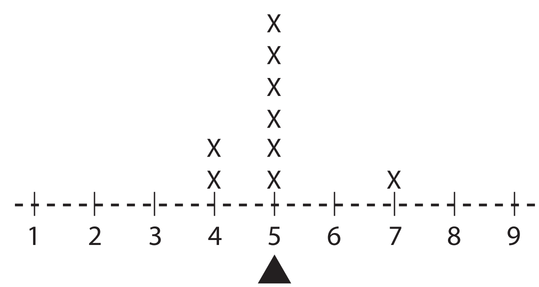


Figura 13: Diagrama de puntos con diferentes datos movidos

El eje horizontal está reequilibrado ahora en el punto pivote. Es decir, el "punto de equilibrio" del eje horizontal en este diagrama de puntos es 5. Reemplazando cada "X" (punto) en este diagrama con la distancia entre el valor y 5, tenemos:

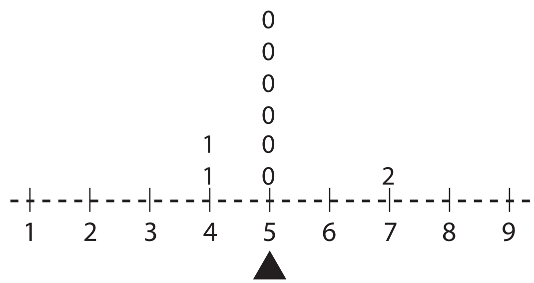


Figura 14: Diagrama de puntos mostrando la distancia desde 5

Note que la distancia total para los dos valores por debajo de 5 (los dos 4s) es la misma que la distancia total para un valor por encima de 5 (el 7). Por esta razón, el punto de equilibrio del eje horizontal es 5. Reemplazando cada valor en el diagrama de puntos de los datos originales por su distancia desde 5 produce el siguiente diagrama:

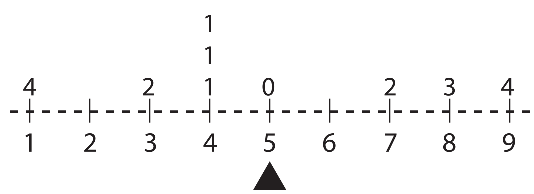


Figura 15: Diagrama de puntos mostrando los datos originales y la distancia desde 5

La distancia total para los valores por debajo de 5 es 9, la misma que la distancia total para los valores por encima de 5. Por esta razón, la media (5) es el punto de equilibrio del eje horizontal.

Frecuentemente se refiere a ambos, la media y la mediana, como las medidas de tendencia central. En el Nivel A, también se introdujo la mediana como la cantidad que tiene el mismo número de datos en cada lado de ella en los datos ordenados. Esta "similitud de cada lado" es la razón que la mediana sea una medida de tendencia central. La actividad anterior demuestra que la distancia total para los valores por debajo de la media es la misma que la distancia total para los valores por encima de la media, e ilustra por qué la media también se considera una medida de tendencia central.

Una Medida de Dispersión—La Desviación Media Absoluta

La estadística se ocupa de la variabilidad en los datos. Una idea importante es cuantificar cuánta variabilidad existe en un conjunto de datos numéricos. Las cantidades que miden el grado de variabilidad en los datos se llaman *medidas de dispersión*. A los estudiantes de Nivel A, se les introduce el *rango* como una medida de dispersión en datos numéricos. En el Nivel B, los estudiantes deben ser introducidos a la idea de comparar los valores de los datos a un valor central, tal como la media o mediana, y cuantificar qué tan diferentes son los datos de este valor central.

En el ejemplo del número de mascotas, ¿qué tan diferentes a la media son los valores originales? Una forma de medir el grado de variabilidad desde la media es determinar la distancia total de todos los valores desde la media. Usando el diagrama de puntos final del ejemplo anterior, la distancia total de todos los valores desde la media de 5 mascotas es 18 mascotas. La magnitud de esta cantidad depende de varios factores, incluyendo el número de medidas. Para ajustar por el número de medidas, la distancia total desde la media se divide por el número de medidas. La cantidad resultante se llama la *Desviación Media Absoluta*, o DMA. La DMA es la distancia promedio de cada dato desde la media. Es decir:

$$\text{DMA} = \frac{(\text{Distancia Total desde la Media para Todos los Valores})}{(\text{Número de Datos })}$$

La DMA para los datos del número de mascotas de la actividad anterior es:

$$\text{DMA} = 18/9 = 2$$

La DMA indica que el número real de mascotas para los nueve estudiantes difiere de la media de cinco mascotas, en promedio en dos mascotas. Kader (1999) provee una explicación mas detallada de esta actividad y de la DMA.

La DMA es un indicador de la dispersión basada en todos los datos y ofrece una medida de variación promedio en los datos desde la media. La DMA también sirve como un precursor de la desviación estándar, la cual será desarrollada en el Nivel C.

Representando Distribuciones de Datos— La Tabla de Frecuencias y el Histograma

En el Nivel B, los estudiantes deben desarrollar dispositivos tabulares y gráficos adicionales para representar distribuciones de datos de variables numéricas. Varios de estos se construyen sobre las representaciones desarrolladas en el Nivel A. Por ejemplo, los estudiantes al Nivel B pueden explorar el problema de elaborar una orden de compra de sombreros. Para preparar la orden de compra, uno necesita saber cuáles tamaños de sombrero son más comunes y cuáles ocurren en menor frecuencia. Para obtener información de los tamaños de los sombreros, es necesario medir el perímetro cefálico. Los tamaños de sombreros europeos se basan en el sistema métrico. Por ejemplo, un tamaño de sombrero europeo de 55 está diseñado para ajustarse a una persona con un perímetro cefálico entre 550 mm y 559 mm. En la planificación

```
51 | 3
52 | 5
53 | 133455
54 | 2334699
55 | 12222345
56 | 0133355588
57 | 113477
58 | 02334458
59 | 1558
60 | 13
61 | 28

51 | 3 significa 513 mm
```

Figura 16: Diagrama de tallos y hojas de las circunferencias de cabezas

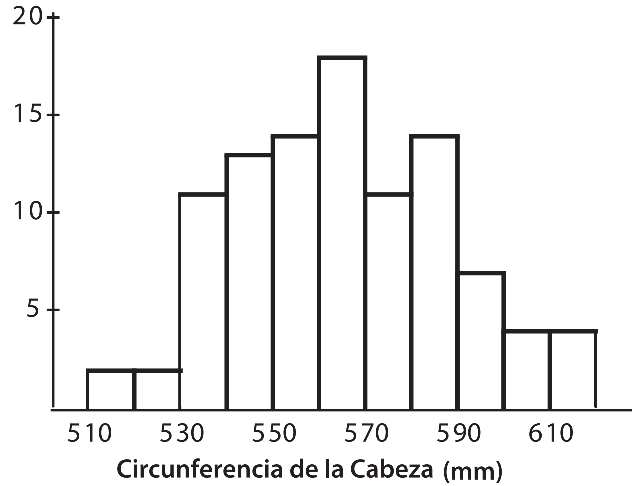


Figura 17: Histograma de Frecuencia Relativa

de una orden de compra para adultos, los estudiantes pueden recolectar datos preliminares con los perímetros cefálicos de sus padres, encargados u otros adultos. Tales datos serían el resultado de una muestra no aleatoria. Los datos resumidos en el siguiente diagrama de tallos y hojas son perímetros cefálicos medidos en milímetros para una muestra de 55 adultos.

Basados en el diagrama de tallos y hojas, algunos tamaños de cabeza parecen ser más comunes que otros. Los perímetros cefálicos en los 560s son más comunes. Los perímetros cefálicos caen de una manera simétrica en ambos lados de los 560s, con muy pocos más pequeños que 530 mm o más grandes que 600 mm.

En la práctica, una decisión de cuántos sombreros ordenar estaría basada en una muestra mucho más grande, posiblemente cientos o incluso miles de adultos. Si una muestra más grande estuviera disponible, un diagrama de tallos y hojas no sería un dispositivo práctico para resumir la distribución de datos. Una alternativa al diagrama de tallos y hojas es formar una distribución basada en dividir los datos en grupos o intervalos. Este método puede ser ilustrado mediante un conjunto de datos más pequeño, tal como los 55 perímetros cefálicos, pero también es aplicable a un conjunto de datos más grande. Las *distribuciones de frecuencia agrupada y frecuencia relativa agrupada* y el *histograma de frecuencia relativa* que corresponde al diagrama de tallos y hojas anterior son en la Figura 17.

Si el fabricante de sombreros requiere que las ordenes sean en múltiplos de 250 sombreros, entonces basado en los resultados previos, ¿Cuántos sombreros de cada tamaño deberían ser ordenados? Usando la distribución de frecuencias relativas, el número de sombreros de cada tamaño para una orden de 250 sombreros se muestra en la Tabla 6.

Una vez más, note cómo los estudiantes al Nivel B deben utilizar razonamiento proporcional para determinar el número de cada tamaño a ordenar. Kader y Perry (1994) dan una descripción detallada del problema de "La Tienda de Sombreros."

Comparando Distribuciones—El Diagrama de Caja

Los problemas que requieren comparar distribuciones para dos o más grupos son comunes en estadística. Por ejemplo, al Nivel A los estudiantes compararon

Tabla 5: Distribuciones de Frecuencia Agrupada y Frecuencia Relativa Agrupada

Tallo	**Limites en las Medidas Registradas del Perímetro Cefálico**	**Intervalo de los Perímetros Cefálicos Reales**	**Frecuencia**	**Frecuencia Relativa (%)**
51	510-519	510-<520	1	1.8
52	520-529	520-<530	1	1.8
53	530-539	530-<540	6	10.9
54	540-549	540-<550	7	12.7
55	550-559	550-<560	8	14.5
56	560-569	560-<570	10	18.2
57	570-579	570-<580	6	10.9
58	580-589	580-<590	8	14.5
59	590-599	590-<600	4	7.3
60	600-609	600-<610	2	3.6
61	610-619	610-<620	2	3.6
		Total	**55**	**99.8**

la cantidad de sodio en perros calientes (o hot dogs) de carne de res y de aves examinando diagramas de puntos paralelos. Al Nivel B, se deben desarrollar representaciones más sofisticadas para comparar distribuciones. Uno de los dispositivos gráficos más útiles para comparar distribuciones de datos numéricos es el *diagrama de cajas*. El diagrama de caja (también llamado diagrama de caja y bigotes) es un gráfico basado en una división de los datos ordenados en cuatro grupos, con el mismo número de datos en cada grupo (aproximadamente un cuarto). Los cuatro grupos se determinan de *Las Cinco Medidas de Resumen* (el valor mínimo, el primer cuartil, la mediana, el tercer cuartil y el valor máximo). Las cinco medidas de resumen y los diagramas de caja comparativos para los datos de contenido de sodio en los perros calientes de carne de res (etiquetados B) y de aves (etiquetados P) introducidos en el Nivel A se presentan en la Tabla 7 y en la Figura 18.

Interpretar los resultados apoyados en tal análisis requiere comparaciones basadas en características globales de cada distribución (centro, dispersión, y forma). Por ejemplo, el contenido medio de sodio para perros calientes de aves es 430 mg, casi 50 mg más que el contenido medio de sodio para perros calientes de carne de res. Las medianas indican que un valor típico para el contenido de

Tabla 6: Datos del Tamaño del Sombrero

Tamaño del Sombrero	Número de Órdenes
51	5
52	5
53	27
54	32
55	36
56	46
57	27
58	36
59	18
60	9
61	9

sodio de los perros calientes de aves es mayor que un valor típico para perros calientes de carne de res. El rango para los perros calientes de carne de res es 392 mg, versus 231 mg para los perros calientes de aves. Los rangos indican que, en general, hay más dispersión (variación) en el contenido de sodio de los perros calientes de carne de res que en los de aves. Otra medida de dispersión que debe ser introducida al Nivel B es el *rango intercuartílico*, o RIQ. El RIQ es la diferencia entre el tercer y el primer cuartil e indica el rango de la mitad 50% de los datos. Los RIQs para el contenido de sodio es 157.5 mg para perros calientes de carne de res y 156 mg para perros calientes de aves. Los RIQs sugieren que la dispersión dentro de la mitad central de los datos para los perros calientes de carne de res es similar a la dispersión de la mitad central de los datos para los perros calientes de

Tabla 7: Las Cinco Medidas de Resumen para el Contenido de Sodio

	Perros Calientes de Carne de Res n=20	Perros Calientes de Aves n=17
Mínimo	253	357
Primer Cuartil	320.5	379
Mediana	380.5	430
Tercer Cuartil	478	535
Máximo	645	588

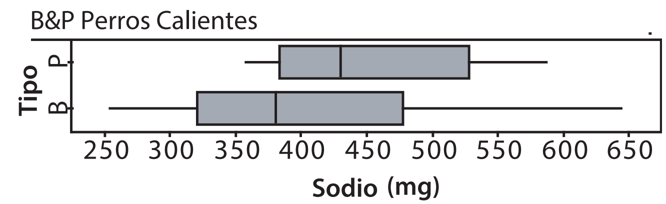


Figura 18: Diagrama de caja para el contenido de sodio

aves. Los diagramas de caja también sugieren que cada distribución es de alguna manera sesgada a la derecha. Es decir, cada distribución parece tener un poco más de variación en la mitad superior. Considerando el grado de variación en los datos y la cantidad de solapamientos en los diagramas de caja, una diferencia de 50 mg entre las medianas realmente no es tan grande. Finalmente, es interesante notar que más del 25% de los perros calientes de carne de res tienen menos sodio que todos los perros calientes de aves. En contraste, los niveles de sodio más altos están en los perros calientes de carne de res.

Note que hay diversas variaciones en los diagramas de caja. Al Nivel C, ejecutar un análisis usando diagramas de caja puede incluir una prueba para *datos extremos* (valores que son extremadamente grandes o pequeños en comparación con la variación en la mayoría de los datos). Si se identifican datos extremos, frecuentemente están separados de los "bigotes" del diagrama. El análisis de datos extremos no se recomienda al Nivel B, puesto que los bigotes se extienden a los valores mínimos y máximo. Sin embargo, los estudiantes del Nivel B podrían encontrar datos extremos cuando usen software estadístico o calculadoras gráficas.

Midiendo la Fuerza de la Asociación entre Dos Variables Cuantitativas

Al Nivel B, se deben desarrollar representaciones más sofisticadas para la investigación de problemas que involucren el estudio de la relación entre dos variables numéricas. Al Nivel A, el problema de empacar sudaderas (camisetas y pantalones juntos o separados) se examinó mediante un estudio de la relación entre estatura y extensión de los brazos. Hay varias preguntas estadísticas relacionadas con este problema que pueden ser orientadas al Nivel B con un análisis más profundo de los datos de estatura/extensión de los brazos. Por ejemplo:

¿Qué tan fuerte es la asociación entre la estatura y la extensión de los brazos?

¿Es la estatura un predictor útil de la extensión de brazos?

Tabla 8: Datos de Estatura y Extensión de Brazos

Estatura	Extensión de Brazos	Estatura	Extensión de Brazos
155	151	173	170
162	162	175	166
162	161	176	171
163	172	176	173
164	167	178	173
164	155	178	166
165	163	181	183
165	165	183	181
166	167	183	178
166	164	183	174
168	165	183	180
171	164	185	177
171	168	188	185

La Tabla 8 aporta los datos de estatura y extensión de brazos (medida en centímetros) para 26 estudiantes. Por conveniencia, los datos de la estatura se han ordenado.

Los datos de la estatura y la extensión de los brazos se presentan en la Figura 19. El diagrama de dispersión sugiere una relación bastante fuerte y creciente entre la estatura y la extensión de los brazos. Además, la relación parece ser lineal.

La medición de la fuerza de la asociación entre dos variables es un concepto estadístico importante que puede introducirse en el Nivel B. El diagrama de dispersión en la Figura 20 para los datos de estatura/extensión de los

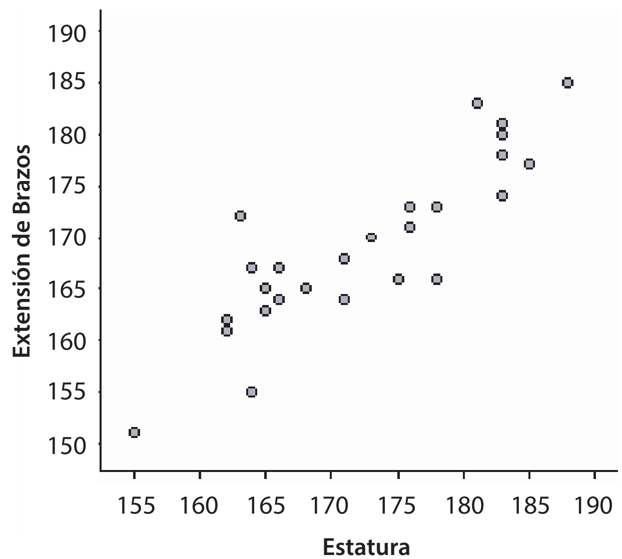


Figura 19: Diagrama de dispersión de la extensión de brazos versus estatura

brazos incluye una línea vertical trazada a través de la estatura media ($x = 172.5$) y una línea horizontal trazada a través de la extensión media de los brazos ($y = 169.3$).

Las dos líneas dividen el diagrama de dispersión en cuatro regiones (o cuadrantes). La región superior derecha (Cuadrante 1) contiene puntos que corresponden a individuos con estaturas y extensión de brazos por encima del promedio. La región superior izquierda (Cuadrante 2) contiene puntos que corresponden a individuos con estatura por debajo del promedio y extensión de los brazos por encima del promedio. La región inferior izquierda (Cuadrante 3) contiene puntos que corresponden a individuos con

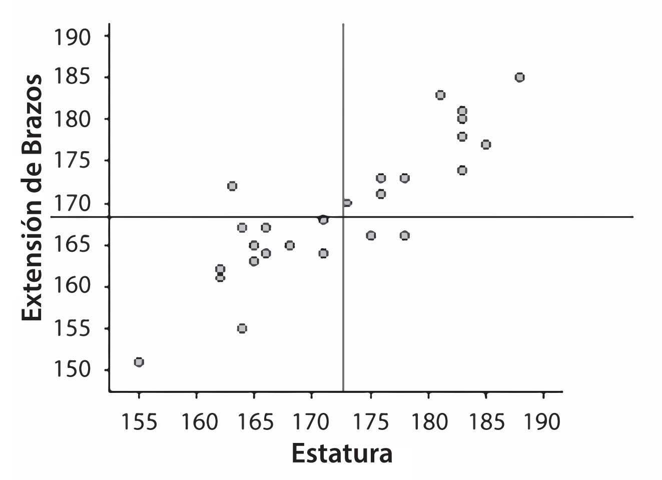


Figura 20: Diagrama de dispersión mostrando las medias

estatura y extensión de brazos por debajo del promedio. La región inferior derecha (Cuadrante 4) contiene puntos que corresponden a individuos con estatura por encima del promedio y extensión de brazos por debajo del promedio.

Note que la mayoría de los puntos en el diagrama de dispersión están en el Cuadrante 1 o Cuadrante 3. Es decir, la mayoría de las personas con estatura por encima del promedio también tienen extensión de brazos por encima del promedio (Cuadrante 1) y la mayoría de las personas con estatura por debajo del promedio también tienen extensión de los brazos por debajo del promedio (Cuadrante 3). Una persona tiene estatura por debajo del promedio y extensión de brazos por encima del promedio (Cuadrante 2) y dos personas tienen estatura por encima del promedio y extensión de brazos por debajo del promedio (Cuadrante 4). Estos resultados indican que hay una *asociación positiva* entre las variables *estatura* y *extensión de brazos*. Generalmente se afirma, dos variables numéricas están *asociadas positivamente* cuando los valores por encima del promedio de una variable tienden a ocurrir con los valores por encima del promedio de la otra variable. La *asociación negativa* entre dos variables numéricas ocurre cuando los valores por debajo del promedio de una variable tienden a ocurrir con los valores por encima del promedio de la otra variable y cuando los valores por encima del promedio de una variable tienden a ocurrir con los valores por debajo del promedio de la otra variable.

Un *coeficiente de correlación* es una cantidad que mide la dirección y fuerza de una asociación entre dos variables. Note que en el ejemplo previo, los puntos en los Cuadrantes 1 y 3 contribuyen a la asociación positiva entre la estatura y la extensión de los brazos, y hay un total de 23 puntos en estos dos cuadrantes. Los puntos en los Cuadrantes 2 y 4 no contribuyen a la asociación positiva entre la estatura y la extensión de los brazos y hay un total de tres puntos en estos dos cuadrantes. Un coeficiente de correlación entre la estatura y la extensión de los brazos está dado por CCC (*Cociente de Conteo de Cuadrante*):

$$CCC = (23 - 3) / 26 = .77$$

Una CCC de .77 indica que hay una relación positiva bastante fuerte entre las dos variables estatura y extensión de los brazos. Esto indica que la estatura de una persona es un predictor útil de la extensión de sus brazos.

En general, la CCC se define como:

((El Número de Datos en los Cuadrantes 1 y 3) - (El Número de Datos en el Cuadrantes 2 y 4)) / (Número de Datos en los Cuatro Cuadrantes)

El CCC tiene las siguientes propiedades:

→ El CCC no tiene unidades

→ El CCC siempre está entre -1 y +1 inclusive

Holmes (2001) da una discusión detallada del CCC. Un coeficiente de correlación similar para tablas de contingencia 2x2 se describe en Conover (1999) y se discute en el Apéndice para el Nivel B. El CCC es una medida de la fuerza de la asociación que se basa únicamente en el número de datos de cada cuadrante, y como muchas medidas resumen tienen sus deficiencias. Al Nivel C, las deficiencias del CCC pueden ser abordadas y usadas como la base para desarrollar el coeficiente de correlación de Pearson.

Modelando Asociación Lineal

Los datos de estatura/extensión de brazos se recogieron en el Nivel A para estudiar el problema de empacar las sudaderas. ¿Deberían una camiseta y una sudadera ser empacadas separadamente o juntas? Un CCC de 0.77 sugiere una relación positiva bastante fuerte entre la estatura y la extensión de los brazos, lo cual indica que la estatura es un predictor útil para la extensión de los brazos y que una camiseta y un pantalón pueden ser empacados juntos. Si se empacan juntos ¿cómo podría una persona decidir cuál talla de sudadera comprar? Ciertamente, la talla del

“Un coeficiente de correlación es una cantidad que mide la dirección y fuerza de una asociación entre dos variables.”

pantalón de una sudadera depende de la estatura de la persona y la talla de la camiseta depende de la extensión de brazos de la persona. Como mucha gente conoce su estatura, pero no su extensión de brazos, ¿puede la estatura ser usada para ayudar a la gente a decidir cual talla de sudadera usar? Específicamente:

¿Puede la relación entre la estatura y la extensión de brazos ser descrita usando una función lineal?

Los estudiantes en el Nivel B estudiarán relaciones lineales en otras áreas de sus currículos de matemáticas. El grado en el cual estas ideas han sido desarrolladas determinará cómo podemos proceder en este punto. Por ejemplo, si los estudiantes aún no han sido introducidos a la ecuación de la recta, entonces simplemente pueden trazar una línea a través del "centro de los datos" como se muestra en la Figura 21.

Esta línea puede ser usada para predecir la extensión de los brazos de una persona si su estatura se conoce. Por ejemplo, para predecir la extensión de los brazos de una persona que mide 170 cm de estatura, se dibuja un segmento vertical desde el eje X a la estatura = 170. En el punto que este segmento vertical intercepta el segmento, una línea horizontal se dibuja en el eje Y. El valor donde este segmento horizontal intercepta el eje Y es el valor predicho para la extensión de los brazos. Basados en la gráfica anterior, parece que predeciríamos una extensión de los brazos de aproximadamente 167 cm para una persona que tiene 170 cm de estatura.

Si los estudiantes se familiarizan con la ecuación de la recta y saben cómo encontrar la ecuación a partir de dos

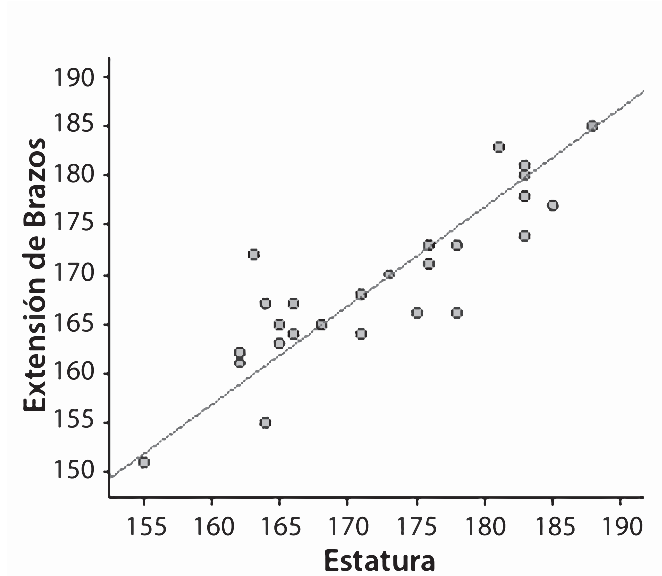


Figura 21: Línea trazada a ojo

puntos, entonces pueden usar la recta Media – Media, la cual se determina como sigue. Ordene los datos de acuerdo a las coordenadas X y divida los datos en dos "mitades" basadas en este orden. Si hay un número impar de medidas, remueva el punto medio del análisis. Determine las medias para las coordenadas-X y para las coordenadas-Y en cada mitad y encuentre la ecuación de la recta que pasa por estos dos puntos. Usando los datos previos:

Mitad Inferior (13 puntos)	Mitad Superior (13 puntos)
Estatura Media = 164.8	Estatura Media = 180.2
Extensión Promedio de los Brazos = 163.4	Extensión Promedio de los Brazos =175.2

La ecuación de la recta que pasa por los puntos (164.8, 163.4) y (180.2, 175.2) es Extensión de Brazos Predicha ≈ 37.1 + 0.766 (Estatura). Esta ecuación puede ser usada para predecir la estatura de una persona más precisamente que una recta a ojo. Por ejemplo, si una persona mide 170 cm de estatura, entonces su estatura predicha podría ser aproximadamente 37.1 + 0.766 (170)= 167.3 cm. Un enfoque mucho más sofisticado (mínimos cuadrados) para determinar la recta de "mejor ajuste" a través de los datos se introducirá en el Nivel C.

La Importancia de la Selección Aleatoria

En la estadística, frecuentemente queremos extender los resultados más allá de un grupo particular estudiado a un grupo mayor, la *población*. Intentamos obtener información de la población examinando una porción de la población, llamada *muestra*. Tales generalizaciones son válidas sólo si los datos son representativos del grupo mayor. Una muestra representativa es aquella en la cual las características relevantes de los miembros de la muestra son generalmente las mismas de la población. La selección de muestras inapropiadas o sesgadas tiende a favorecer sistemáticamente ciertos resultados, y puede producir resultados engañosos y conclusiones erróneas.

El muestreo aleatorio es una forma de remover el sesgo en la selección de la muestra, y tiende a producir muestras representativas. En el Nivel B, los estudiantes deben experimentar las consecuencias de selección no aleatoria y desarrollar una comprensión básica de los principios involucrados en los procesos de selección aleatoria. A continuación se describe una actividad que permite a los estudiantes comparar resultados de muestras basados en selección personal (no aleatoria) versus resultados de muestras basados en selección aleatoria.

Considere los 80 círculos en la página siguiente. ¿Cuál es el diámetro promedio para esos 80 círculos? Cada estudiante deberá tomar aproximadamente 15 segundos y seleccionar cinco círculos que crean que mejor representan los tamaños de los 80 círculos. Después de seleccionar la muestra, cada estudiante debe encontrar el diámetro promedio para los círculos en su muestra personal. Note que el diámetro es 1 cm para los círculos pequeños, 2 cm para los círculos medianos, y 3 cm para los círculos grandes.

Luego, cada estudiante debe enumerar los círculos de 1 a 80 y usar un generador de dígitos aleatorios para seleccionar una muestra aleatoria de tamaño cinco. Cada estudiante debe encontrar el diámetro promedio para los círculos de su muestra aleatoria. La media muestral de los diámetros de la clase pueden ser resumidos para los dos procesos de selección con diagramas de tallos y hojas doble.

¿Cómo se comparan las medias de los dos procesos de selección de muestras con el verdadero diámetro medio de 1.25 cm? La selección personal usualmente tenderá a producir medias muestrales que son mayores que 1.25. Es decir, la selección personal tiende a ser sesgada favoreciendo sistemáticamente los círculos grandes y sobrestimando la media de la población. La selección aleatoria tiende a producir algunas medias de la muestra que subestiman la media de la población y algunas que

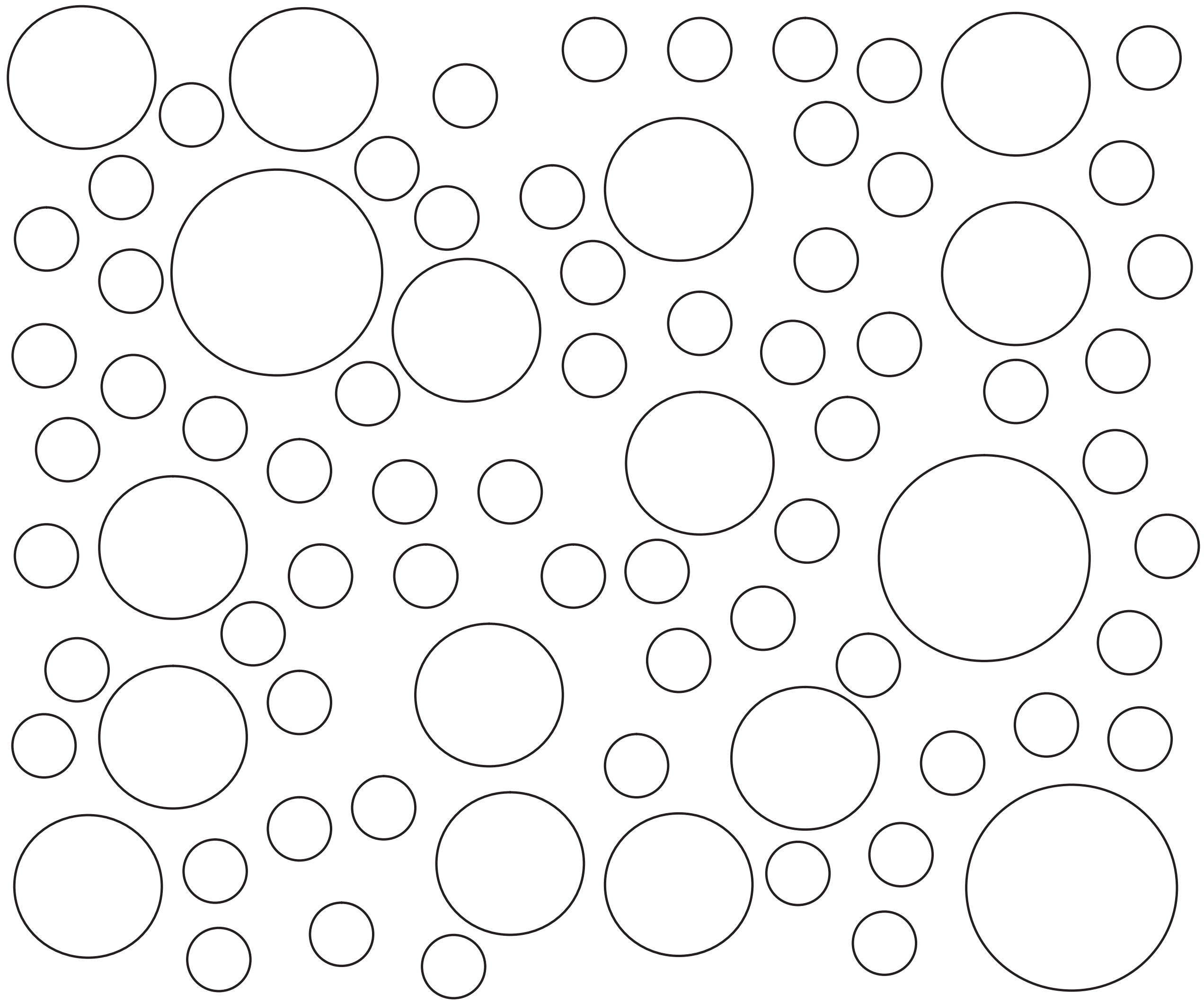


Figura 22: Ochenta círculos

sobrestiman la media de la población, de tal manera que la media de la muestra se agrupa de manera uniforme alrededor del valor de la media poblacional (es decir, la selección aleatoria tiende a ser *insesgada*).

En el ejemplo anterior, el hecho que las medias de las muestras varíen de una muestra a otra ilustra una idea que se introdujo previamente en la encuesta del tipo de música favorito. Esta es la noción de variabilidad en el muestreo. La imposición de aleatoriedad en el proceso de muestreo nos permite *usar la probabilidad* para describir el comportamiento a largo plazo en la variabilidad de la media de la muestra resultante de un muestreo aleatorio. La variación en resultados de muestreo repetido se describe mediante lo que se llama *distribución muestral*. La distribución muestral se explorará en más detalle en el Nivel C.

Experimentos Comparativos

Otro método estadístico importante que se debe introducir en el Nivel B es el de los *estudios comparativos experimentales*. Los estudios comparativos experimentales involucran la comparación de los efectos de dos o más *tratamientos* (condiciones experimentales) en algunas variables de respuesta. Al Nivel B, los estudios que comparan dos tratamientos son adecuados. Por ejemplo, los estudiantes quisieran estudiar los efectos de escuchar música rock en la habilidad de memorizar. Antes de emprender un estudio como este, es importante para los estudiantes tener la oportunidad de identificar y, en lo posible, controlar alguna fuente potencial externa que pueda interferir con nuestra habilidad de interpretar los resultados. Para abordar este problema, la clase necesita desarrollar una estrategia de diseño para recolectar los datos experimentales apropiados.

Un experimento simple *sería* dividir la clase en dos grupos de igual tamaño (o cerca de igual tamaño). La asignación aleatoria provee una manera justa de asignar los estudiantes a los dos grupos porque tiende a promediar las diferencias en la habilidad de los estudiantes y otras características que puedan afectar la respuesta. Por ejemplo, suponga que una clase tiene 28 estudiantes. Los 28 estudiantes se asignan aleatoriamente a dos grupos de 14. Una forma de lograr esto es ubicar 28 papeles en una caja—14 etiquetados con "M" y 14 etiquetados con "S". Mezcle el contenido de la caja y haga que cada estudiante aleatoriamente elija una pieza de papel. Los 14 M escucharán música y los 14 S estarán en silencio.

A cada estudiante se le mostrará una lista de palabras. Se determinarán reglas sobre cuánto tiempo los estudiantes tienen que estudiar las palabras y cuánto tiempo tienen para reproducirlas. Por ejemplo, los estudiantes pueden tener dos minutos para estudiar las palabras, una pausa de

Tabla 9: Las Cinco Medidas de Resumen

	Música	**Silencio**
Mínimo	3	6
Primer Cuartil	6	8
Mediana	7	10
Tercer Cuartil	9	12
Máximo	15	14

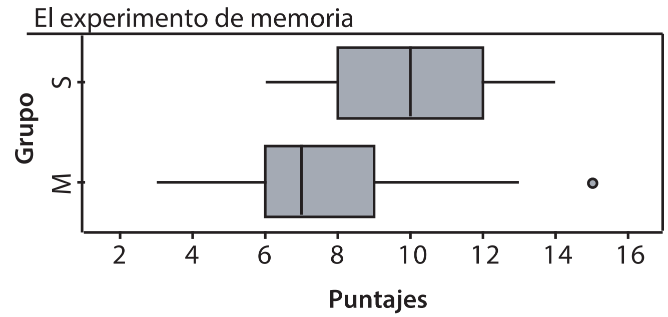


Figura 23: Diagrama de caja para los datos de memoria

un minuto, y luego dos minutos para reproducir (escribirlas) tantas palabras como sea posible. El número de palabras recordadas bajo cada condición (escuchando música o en silencio) es la respuesta a la variable de interés.

Las Cinco Medidas de Resumen y los diagramas de caja comparativos para un conjunto hipotético de datos se muestran en la Tabla 9 y la Figura 23. Estos resultados sugieren que los estudiantes generalmente memorizan menos palabras cuando están escuchando música que cuando hay silencio. Con excepción del valor máximo en el grupo que escuchó música (el cual se clasifica como un dato extremo), todas las medidas de resumen para el grupo que escuchó música (etiquetado M en la Figura 23) son mucho más bajas que las medidas de resumen correspondientes para el grupo que estuvo en silencio (etiquetado S en la Figura 23). Sin el dato extremo, el grado de variación en los puntajes parece ser similar para ambos grupos. La distribución S parecer ser razonablemente simétrica, mientras la distribución M es ligeramente sesgada a la derecha. Considerando el grado

Tabla 10: Datos de Nacimientos Vivos

Año	**Nacimientos (x 1,000)**	**Año**	**Nacimientos (x 1,000)**
1970	3,731	1985	3,761
1971	3,556	1986	3,757
1972	3,258	1987	3,809
1973	3,137	1988	3,910
1974	3,160	1989	4,041
1975	3,144	1990	4,158
1976	3,168	1991	4,111
1977	3,327	1992	4,065
1978	3,333	1993	4,000
1979	3,494	1994	3,979
1980	3,612	1995	3,900
1981	3,629	1996	3,891
1982	3,681	1997	3,881
1983	3,639	1998	3,942
1984	3,669	1999	3,959

de variación en los puntajes y la separación en los diagramas de caja, una diferencia de 3 entre las medianas es bastante grande.

Series de Tiempo

Otra herramienta estadística importante que se debe introducir al Nivel B es un gráfico de serie de tiempo. Los problemas que exploran tendencias en los datos sobre un periodo de tiempo son bastante comunes. Por ejemplo, la población de los Estados Unidos y del mundo continúa creciendo, y hay varios factores que afectan el tamaño de

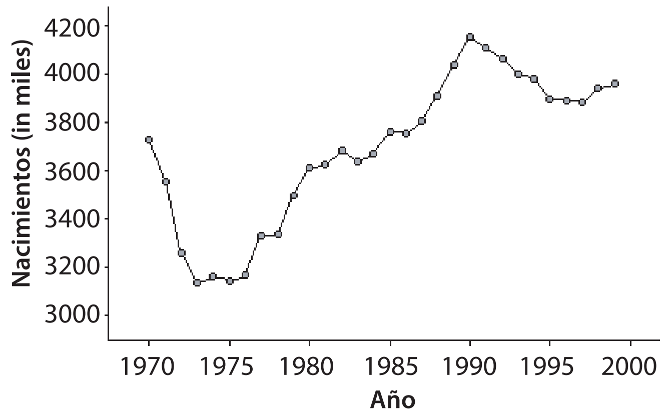


Figura 24: Gráfica de serie de tiempo de nacimientos vivos

una población, tales como el número de nacimientos y el número de muertes por año. Una pregunta que hacemos es:

¿Cómo el número de nacimientos vivos ha cambiado en los últimos 30 años?

La U.S. Census Bureau [Oficina de Censo de los Estados Unidos] publica estadísticas vitales en su serie anual *Resúmenes Estadísticos de los Estados Unido*s. Los datos siguientes son de *Los Resúmenes Estadísticos de los Estados Unidos* (2004–2005) y representan el número de nacimientos vivos por año (en miles) por residentes de los Estados Unidos desde 1970. Note que en 1970, el valor 3,731 representa 3,731,000 nacimientos vivos.

La gráfica de serie de tiempo en la Figura 24 muestra el número de nacimientos vivos en el tiempo. La gráfica indica que:

→ de 1970 a 1975, el número de nacimientos vivos disminuyó

→ de 1976 a 1990, el número de nacimientos vivos incrementó

→ de 1991 a 1997, el número de nacimientos vivos disminuyó

Y parece que el número de nacimientos vivos puede haber empezado a incrementar desde 1997.

Usos Indebidos de la Estadística

La introducción de este documento apunta que los datos gobiernan nuestras vidas. Por esto, cada graduado de la preparatoria merece tener un fundamento sólido en el razonamiento estadístico. Junto con la identificación de los usos apropiados de la estadística en cuestionarios y gráficas, los estudiantes del Nivel B deben empezar a ser conscientes de los usos indebidos comunes de la estadística.

El razonamiento proporcional permite que los estudiantes del Nivel B interpreten información resumida en una variedad de formas. Un tipo de gráfico que usualmente es usado incorrectamente para representar información es el pictograma. Por ejemplo, suponga que el poder adquisitivo de un dólar hoy es 50% de lo que fue hace 20 años. ¿Cómo uno representaría eso en un pictograma? Sea el poder adquisitivo de un dólar 20 años atrás representado por el siguiente billete de dólar :

Si el poder adquisitivo hoy es la mitad de lo que fue hace 20 años, uno podría pensar en reducir el ancho y

> “Las gráficas estadísticas pobremente diseñadas se encuentran comúnmente en periódicos y otros medios populares.”

el largo de este billete a la mitad como se ilustra en el pictograma siguiente:

El dólar de hoy debería lucir la mitad del tamaño del dólar de hace 20 años. ¿Luce así? Como el largo y el ancho del dólar se cortaron a la mitad, el área del dólar de hoy mostrado previamente es un cuarto del área original, no la mitad.

El dólar de hoy a la “mitad” del tamaño, representa que se compra sólo la mitad de lo que se compraba hace 20 años.

Los dos pictogramas siguientes muestran la reducción correcta en el área. El que está encima cambia sólo una dimensión, mientras que el otro cambia ambas dimensiones, pero en proporción correcta para que el área sea la mitad del área de la representación original. Este ejemplo proporciona a los estudiantes del Nivel B un excelente ejercicio en razonamiento proporcional.

Las gráficas estadísticas pobremente diseñadas se encuentran comúnmente en periódicos y otros medios populares. Varios ejemplos de gráficas mal elaboradas, incluyendo el uso de injustificadas terceras dimensiones en los gráficos de barras y gráficos circulares se pueden encontrar en *www.amstat.org/education/gaise/2*, un sitio web manejado por Carl Schwarz en la Universidad Simon Fraser. Los

El dólar de hoy a la mitad del tamaño con 50% en su longitud

El dólar de hoy a la mitad del tamaño, con los lados en la proporción correcta con respecto al original.

estudiantes en el Nivel B deben tener oportunidades para identificar gráficos que incorrectamente representen información y luego, con la ayuda de software estadístico, dibujar las versiones correctas. Esto les ofrece excelente práctica al calcular áreas y volúmenes.

Hay muchos usos indebidos muy famosos del análisis de datos en la literatura y tres son mencionados aquí. La revista *Literary Digest* se equivocó en 1936 cuando proyectó que Alf Landon derrotaría a Franklin Delano Roosevelt por un margen de 57 a un 43 por ciento sobre la base de las respuestas a su encuesta. Cada encuesta incluyó una subscripción a la revista, y más de 2.3 millones fueron devueltos. Desafortunadamente, incluso las grandes encuestas con respuestas voluntarias son generalmente no representativas de la población, y Roosevelt ganó con 62% del voto. George Gallup correctamente proyectó el ganador, y a partir de entonces empezó una exitosa carrera en el uso de técnicas de muestreo aleatorio para realizar encuestas. Aprender lo que Gallup hizo bien y lo que *Literary Digest* hizo mal da a los estudiantes del Nivel B información valiosa sobre el diseño y análisis de encuestas. Una discusión más detallada de este problema puede ser encontrada en Hollander y Proschan (1984).

La Draft Lottery de 1970 ofrece un ejemplo de una aplicación incorrecta de la aleatoriedad. En el procedimiento que se usó, se ubicaron capsulas que contenían fechas de nacimientos en una caja grande. Aunque hubo un intento de mezclar las capsulas, fue insuficiente para superar el hecho de que las capsulas quedaran dispuestas en la caja en orden de enero a diciembre. Esto dio lugar a que los jóvenes con fechas de nacimiento en los últimos meses fueran más propensos a que seleccionaran sus fechas de nacimiento mucho más pronto que fechas de nacimiento en otro momento del año. Hollander y Proschan (1984) ofrecen una excelente discusión de este problema.

El vuelo 25avo del programa del transbordador espacial de la NASA despegó el 20 de enero de 1986. Justo después del despegue, una nube de humo gris se veía desde el lado derecho del cohete. Setenta y tres segundos en vuelo y el *Challenger* explotó, matando a los siete astronautas a bordo. Se determinó que la causa de la explosión fue una falla en el anillo sellador u obturador, debido al clima frio. El desastre pudo haber sido evitado si los datos disponibles se hubieran representado en un diagrama de dispersión simple y se hubieran interpretado correctamente. El desastre del *Challenger* se ha convertido en un caso de estudio de las posibles consecuencias catastróficas de un pobre análisis de datos.

Resumen del Nivel B

Comprender los conceptos estadísticos del Nivel B permite a los estudiantes empezar a apreciar que el análisis de datos es un proceso de investigación que consiste en formular sus propias preguntas, recolectar información apropiada a través de varias fuentes (censos, encuestas no aleatorias y aleatorias, y experimentos comparativos con asignación aleatoria), analizar información a través de gráficos y medidas de resumen simples, e interpretar los resultados con la mirada puesta en la inferencia a una población basada en una muestra. A medida que comienzan a formular sus propias preguntas, los estudiantes empiezan a ser conscientes que el mundo alrededor de ellos está lleno de información que afecta sus propias vidas, y empiezan a apreciar que la estadística puede ayudarles a tomar decisiones basadas en datos.

En Esta Sección

Nivel C

El Nivel C está diseñado para construir sobre las bases desarrolladas en los Niveles A y B. En particular, los Niveles A y B introdujeron los estudiantes a la estadística como un proceso investigativo, la importancia de usar datos para responder apropiadamente preguntas formuladas, tipos de variables (categóricas versus numéricas), representaciones tabulares (incluyendo tablas de frecuencias de dos entradas para datos categóricos y tablas de frecuencia/frecuencia relativa no agrupada y agrupada para datos numéricos), y resúmenes numéricos (incluyendo conteos, proporciones, media, mediana, rango, cuartiles, rango intercuartílico, DMA y CCC).

Adicionalmente, en los Niveles A y B se abordaron diseños de estudios comunes (incluyendo censos, muestra aleatoria simple, y diseños aleatorizados de experimentos), el proceso de sacar conclusiones de los datos, y el rol de la probabilidad en las investigaciones estadísticas.

En el Nivel C, todas estas ideas son revisadas, pero los tipos de estudios en los que se enfatiza son de una naturaleza estadística más profunda. Los estudios estadísticos a este nivel requieren que los estudiantes recurran a conceptos básicos trabajados previamente, amplíen los conceptos para cubrir un espectro más amplio de temas de investigación, y desarrollen una comprensión profunda del razonamiento inferencial y su conexión con la probabilidad. Los estudiantes también deben haber incrementado la habilidad para explicar su razonamiento estadístico a otros.

En el Nivel C, los estudiantes desarrollan estrategias adicionales para producir, interpretar, y analizar datos que ayudan a responder preguntas de interés. En general, los estudiantes deben estar en condiciones de formular preguntas que puedan ser respondidas con datos; idear un plan razonable para recolectar información apropiada mediante observación, muestreo o experimentación; sacar conclusiones y usar datos que apoyen las conclusiones; y comprender el rol que la variación aleatoria juega en el proceso de inferencia.

Específicamente, al Nivel C las recomendaciones incluyen:

I. Formular Preguntas

→ Los estudiantes deben estar en condiciones de formular preguntas y determinar cómo pueden ser recolectados y analizados los datos para proporcionar una respuesta.

II. Recolección de Información

→ Los estudiantes deben entender lo que constituye una buena práctica al realizar una encuesta por muestreo.

→ Los estudiantes deben entender lo que constituye una buena práctica al realizar un experimento.

→ Los estudiantes deben entender lo que constituye una buena práctica al realizar un estudio observacional.

→ Los estudiantes deben estar en condiciones de diseñar e implementar un plan de recolección de información para estudios estadísticos, incluyendo estudios observacionales, encuestas por muestreo, y experimentos comparativos simples.

III. Análisis de Datos

→ Los estudiantes deben estar en condiciones de identificar formas apropiadas para resumir datos numéricos o categóricos usando tablas, representaciones gráficas y resúmenes estadísticos numéricos.

→ Los estudiantes deben entender cómo las distribuciones muestrales (desarrolladas mediante simulación) son usadas para describir la variabilidad muestra-a-muestra de muestras estadísticas.

→ Los estudiantes deben estar en condiciones de reconocer cuando la relación entre dos variables numéricas es razonablemente lineal, saber que el coeficiente de correlación de Pearson es una medida de fuerza de la relación lineal entre dos variables numéricas y comprender el criterio de mínimos cuadrados en el ajuste de la recta.

IV. Interpretar Resultados

→ Los estudiantes deben entender la significancia estadística y la diferencia entre significancia estadística y significancia práctica.

→ Los estudiantes deben entender el rol de los valores-*p* al determinar significancia estadística.

→ Los estudiantes deben estar en condiciones de interpretar el margen de error asociado con una estimación de una característica de la población.

Un Ejemplo Introductorio—Obesidad en América

¡Los datos y las historias que rodean los datos, deben ser de interés para los estudiantes! Es importante recordar esto al enseñar análisis de datos. También es importante elegir datos e historias que tengan suficiente profundidad para demostrar la necesidad del razonamiento estadístico. El siguiente ejemplo ilustra este punto.

Los estudiantes están interesados en asuntos que afectan sus vidas, y los asuntos relacionados con la salud generalmente pertenecen a esa categoría. Las noticias son un excelente lugar para buscar historias de interés actual, incluyendo noticias de salud. Uno de los tópicos relacionados con la salud que últimamente genera muchas noticias es la obesidad. El siguiente párrafo se relaciona con una noticia que es lo suficientemente rica como para proveer un contexto para muchos de los temas estadísticos cubiertos en el Nivel C.

Un artículo de periódico que apareció en 2004 empieza con las siguiente líneas: "Pregunte a cualquiera: Los americanos están cada vez más gordos. Las campañas de publicidad dicen que están engordando. También lo dicen funcionarios federales y los científicos en quienes ellos confían… En 1991, 23% de los americanos cayeron en la categoría de obesos; ahora 31% lo hacen, un incremento de más del 30%. Pero, el Dr. Jeffrey Friedman, un investigador en obesidad de la Universidad de Rockefeller, argumenta que contrario a la opinión popular, los datos nacionales no muestran que los americanos estén engordando uniformemente. En cambio, él dice, que las estadísticas muestran claramente que mientras los muy obesos están engordando más, la gente delgada permanece más o menos igual… El peso promedio de la población ha incrementado sólo de 7 a 10 libras". La discusión en el artículo se refiere a adultos.

Las siguientes son preguntas sugeridas para explorar con los estudiantes que tienen bases académicas de Nivel B en estadística, pero están pasando al Nivel C.

→ Bosqueje un histograma mostrando cómo usted piensa que la distribución del peso de los americanos adultos pudo haber sido en 1991. Ajuste el bosquejo para mostrar como la distribución de pesos pudo haber sido en 2002, el año del estudio reportado. Antes de hacer sus bosquejos piense sobre la forma, centro y dispersión de su distribución. ¿Será la distribución sesgada o simétrica? ¿Será la mediana menor, mayor o del mismo tamaño de la media? ¿Aumentará la dispersión a medida que se avanza de la distribución de 1991 a la distribución de 2002?

→ ¿Qué suena de más interés periodístico: "La obesidad ha incrementado por más del 30%" o "En promedio, el peso de los americanos ha incrementado por menos de 10 libras"? Explique su razonamiento.

→ El título del artículo es *La Epidemia de la Grasa: Él dice es una ilusión.* [ver *New York Times*, Junio 8, 2004, o *CHANCE*, Vol. 17., No. 4, Otoño 2004, p. 3 para el artículo completo]. ¿Cree que este es un título apropiado? Explique su razonamiento.

→ Los datos en los cuales se basan los porcentajes vienen del Centro Nacional de Estadísticas de Salud, Encuesta de Salud Nacional y Nutrición 2002 [National Center for Health Statistics, National Health and Nutrition Examination Survey 2002]. Esta es una encuesta de aproximadamente 5,800 residentes de los Estados Unidos. Aunque el diseño de la encuesta es más complicado que el de una muestra aleatoria simple, el margen de error calculado como si se tratase de una muestra aleatoria simple es una aproximación razonable. ¿Cuál es el margen de error aproximado asociado con el 31% estimado de obesidad para 2004? Interprete este margen de error para un lector de periódicos quien nunca estudió estadística.

Para los curiosos, la información sobre cómo se define la obesidad se puede encontrar en *www.amstat.org/education/gaise/3*.

Al responder estas preguntas, los estudiantes del Nivel C deben darse cuenta que una distribución de pesos va a ser sesgada hacia los valores grandes. Esto generalmente produce una situación en la cual la media es mayor que la mediana. Como el porcentaje de la población que es obesa aumentó 8% entre 1991 y 2002, pero el peso promedio (o centro) no cambió mucho, la cola superior de la distribución tiene que haber "engordado", indicando una dispersión mayor para los datos de 2002. Los estudiantes tendrán una variedad de respuestas interesantes para las preguntas segunda y tercera. El rol del profesor es ayudar a los estudiantes a entender si sus respuestas están respaldadas por hechos. La última pregunta logra que los estudiantes piensen sobre un concepto de estimación importante estudiado en el Nivel C.

El Proceso Investigativo al Nivel C

Como el Nivel C revisa muchas de las mismas temáticas abordadas en los Niveles A y B, pero en un sentido mucho más profundo y sofisticado, empezamos por describir cómo el proceso investigativo se ve al Nivel C. Esta discusión general es seguida por varios ejemplos.

Formulando Preguntas

Como se estableció al inicio del Nivel A, los datos son más que números. Los estudiantes necesitan entender los tipos de preguntas que pueden ser respondidas con datos. Por ejemplo, la pregunta "¿Está la salud general de los estudiantes de la preparatoria declinando en este país?" es una pregunta muy amplia para responder con una investigación estadística (o con muchas investigaciones estadísticas). Ciertos aspectos de la salud de los estudiantes, sin embargo, pueden ser investigados formulando preguntas mucho más específicas, tales como "¿Cuál es la proporción de obesidad entre los estudiantes de la preparatoria?"; "¿Cuál es el promedio de consumo de calorías diario para un estudiante de preparatoria de último año?"; "¿Es un régimen de ejercicios de tres días semanales suficiente para mantener la frecuencia cardiaca y peso dentro de los límites aceptables?" La formulación de las preguntas, entonces, se convierte en el punto de partida para una investigación estadística.

Recolectando Datos—Tipos de Estudios Estadísticos

La mayoría de las preguntas pueden ser respondidas mediante la recolección de datos y la interpretación requiere datos de un estudio diseñado, bien sea una *encuesta por muestreo* o un *experimento*.

Estos dos tipos de investigaciones estadísticas tienen algunos elementos comunes—cada uno requiere aleatorización para los propósitos de reducir el sesgo y construir una base para la inferencia estadística y cada uno hace uso de mecanismos de inferencia comunes de margen de error para la estimación y valor-*p* en pruebas de hipótesis (ambos se explicarán luego). Pero estos dos tipos de investigaciones tienen objetivos y requerimientos muy diferentes. Las encuestas por muestreo se usan para estimar o tomar decisiones sobre características (parámetros) de la población. Una población muy bien definida y fija es el ingrediente principal de este tipo de estudios. Los experimentos son usados para estimar o comparar los efectos de diferentes condiciones experimentales (tratamientos), y requieren tratamientos bien definidos y unidades experimentales en las cuales estudiar esos tratamientos.

Estimar la proporción de residentes de una ciudad que apoyaría un incremento en los impuestos para la educación requiere una encuesta por muestreo. Si la selección de residentes es aleatoria, entonces los resultados de la muestra pueden ser extendidos para representar la población de la cual se seleccionó la muestra. Una medida de error de muestreo (margen de error) se puede calcular para comprobar qué tan lejos podría estar la estimación del valor real.

Elaborar una prueba para ver si un nuevo medicamento para mejorar la respiración en pacientes con asma produce una capacidad pulmonar mayor que un medicamento estándar, requiere un experimento en el cual un grupo de pacientes que ha dado su consentimiento para participar en el estudio son asignados aleatoriamente al medicamento nuevo o al estándar. Con este tipo de diseño comparativo aleatorizado, un investigador puede determinar, midiendo del grado de incertidumbre, si el nuevo medicamento causa una mejora en la capacidad pulmonar. Los experimentos aleatorizados son, de hecho, el único tipo de estudio estadístico capaz de establecer relación de causa y efecto. Cualquier generalización se extiende sólo a los tipos de unidades usadas en el experimento, puesto que las unidades experimentales usualmente no son muestreadas aleatoriamente de una población mayor. Para generalizar a una clase mayor de unidades experimentales, más experimentos tendrían que ser realizados. Esa es una de las razones por las cuales la réplica es una característica de la buena ciencia.

Los estudios que no tienen selección aleatoria de unidades de muestreo o asignación aleatoria de tratamientos a unidades experimentales se llaman *estudios observacionales* en este documento. Un estudio de cuántos estudiantes en tu escuela preparatoria tienen asma y cómo se descompone por género y grupos de edad sería un ejemplo de este tipo de estudio. Los estudios observacionales no son susceptibles de inferencia estadística en el usual sentido del término, pero pueden proveer información valiosa sobre la distribución de valores medidos y los tipos de asociación entre las variables que se puede esperar.

En el Nivel C, los estudiantes deben entender las características claves de las encuestas por muestreo y de los diseños experimentales, incluyendo cómo configurar versiones simples de ambos tipos de investigaciones, cómo analizar apropiadamente los datos (puesto que el análisis correcto está relacionado con el diseño), y cómo establecer conclusiones claras y precisas para estos diseños de estudios. A continuación se presentan los elementos claves para el diseño e implementación de planes de recolección de datos para estos tipos de estudios.

"Cuando la aleatoriedad se incorpora en el procedimiento de muestreo, la probabilidad ofrece una forma de describir el comportamiento "a largo plazo" de esta variabilidad muestral."

Encuestas por Muestreo

Los estudiantes deben comprender que obtener buenos resultados de una encuesta por muestreo depende de cuatro características básicas: la población, la muestra, el proceso de aleatorización que conecta las dos, y la precisión de las medidas tomadas en los elementos muestreados. Por ejemplo, para investigar una pregunta con respecto a la salud de los estudiantes, puede planearse una encuesta para una escuela preparatoria. ¿Cuál es la población a ser investigada? ¿Son ellos todos los estudiantes en la escuela (la cual cambia a diario)? Quizás la pregunta de interés involucra sólo a los estudiantes de los dos últimos años. Una vez la población se define tan precisamente como sea posible, se tiene que determinar un tamaño de muestra apropiado y un método para la sección aleatoria de la muestra de ese tamaño. ¿Hay, por ejemplo, una lista de estudiantes que pueden ser enumerados para la selección aleatoria? Una vez encontrada la muestra de estudiantes, ¿qué preguntas se harán? ¿Son las preguntas razonables e imparciales (tanto como sea posible)? ¿Pueden los estudiantes responderlas con precisión?

Cuando se utiliza una muestra de una población, pueden ocurrir errores por varias razones, incluyendo:

→ el procedimiento de muestreo es sesgado

→ la muestra fue seleccionada de una población equivocada

→ algunas de las unidades seleccionadas para estar en la muestra no estuvieron disponibles (no dispuestos) a participar

→ las preguntas fueron mal redactadas

→ las respuestas fueron ambiguas

Estos tipos de errores deben ser considerados cuidadosamente antes que el estudio empiece, de manera que se puedan hacer planes para reducir la probabilidad de ocurrencia tanto como sea posible. Una forma de resolver el sesgo en el procedimiento de muestreo es incorporar la aleatoriedad en el proceso de selección.

Dos muestras de tamaño 50 de la misma población de estudiantes probablemente no darán los mismos resultados de, por ejemplo, la proporción de estudiantes que toman un desayuno saludable. Esta variación de muestra a muestra es llamada *variabilidad muestral*. Cuando la aleatoriedad se incorpora en el procedimiento de muestreo, la probabilidad ofrece una forma de describir el comportamiento "a largo plazo" de esta variabilidad muestral.

Experimentos

En el Nivel C, los estudiantes deben comprender que obtener buenos resultados de un experimento depende de cuatro características básicas: tratamientos bien definidos, unidades experimentales apropiadas a las cuales estos tratamientos se pueden asignar, un sentido del proceso de aleatorización para asignar tratamientos a unidades experimentales y mediciones precisas de los resultados del experimento. Las unidades experimentales generalmente no son seleccionadas aleatoriamente de una población de posibles unidades. Más bien, son las que están disponibles para el estudio. En experimentos con seres humanos, las personas involucradas generalmente son voluntarias que deben firmar un acuerdo diciendo que ellos están dispuestos a participar en el estudio experimental. En experimentos con cultivos agrícolas, las unidades experimentales son las parcelas de campo que están disponibles. En experimentos industriales de mejoramiento de procesos, las unidades pueden ser las líneas de producción en operación durante una semana dada.

Como en una encuesta por muestreo, replicar un experimento producirá diferentes resultados. Una vez más, la asignación aleatoria de unidades experimentales a tratamientos (o viceversa) permite el uso de la probabilidad para predecir el comportamiento en los valores de los resúmenes estadísticos a partir de un gran número de repeticiones del experimento. La aleatorización en experimentos es importante por otra razón. Suponga que un investigador decide asignar el tratamiento A sólo a pacientes mayores de 60 años y el tratamiento B sólo a pacientes menores de 50 años. Si la respuesta al tratamiento difiere, es imposible decir si la diferencia es debida al tratamiento o a la edad de los pacientes. (Este tipo de sesgo en experimentos y otros estudios estadísticos se llama *factor de confusión*.) El proceso de aleatorización, si se hace bien, usualmente balanceará los grupos de tratamiento de forma que este tipo de sesgo es minimizado.

Estudios Observacionales

En el Nivel C, los estudiantes deben comprender que los estudios observacionales son útiles para sugerir patrones en los datos y relaciones entre las variables, pero no proporcionan una base sólida para estimar los parámetros poblaciones o para establecer la diferencia entre tratamientos. El preguntar a los estudiantes de un salón si ellos consumen un desayuno saludable no ayudará a establecer la proporción de los que desayunan saludable en la escuela, porque los estudiantes de un salón en particular podrían no ser representativos de los estudiantes en la escuela. El muestreo aleatorio es la única forma de estar seguros de una muestra representativa para propósitos estadísticos. De forma similar, alimentar tus gatos con la dieta A y los gatos de tu vecino con la dieta B no te permitirá decir que una dieta es mejor que otra en términos del control de peso, porque no hubo asignación aleatoria de las unidades experimentales (gatos) a los tratamientos (dietas). En consecuencia, podrían resultar factores de confusión. Los estudios del tipo sugerido anteriormente son meramente observacionales; ellos pueden sugerir patrones y relaciones, pero no son bases confiables para la inferencia estadística.

Analizando Datos

Cuando se analizan datos de encuestas por muestreo bien diseñadas, los estudiantes al Nivel C deben comprender que un análisis apropiado es el que pueda conducir a afirmaciones inferenciales justificables sobre de los parámetros de la población basados en estimaciones de información de la muestra. La habilidad para sacar conclusiones sobre la población usando información de una muestra depende de información proporcionada por la distribución muestral de la muestra estadística usada para resumir los datos de la muestra. Al Nivel C, los dos parámetros de interés más comunes son la proporción poblacional para datos categóricos y la media poblacional para datos numéricos. Las muestras estadísticas apropiadas usadas para estimar estos parámetros son la proporción muestral y la media muestral, respectivamente. Al Nivel C, la variabilidad de muestra a muestra, como se describe en la distribución muestral para cada uno de estos estadísticos, se estudia en más detalle.

Explorar cómo la información proporcionada por la distribución muestral se usa para generalizar de una muestra a la población mayor permite a los estudiantes del Nivel C sacar conclusiones más sofisticadas de estudios estadísticos. En el Nivel C, se recomienda que las distribuciones muestrales de una proporción muestral y de una media muestral se desarrollen mediante simulación. Los tratamientos mucho más formales de distribuciones muestrales se pueden dejar a la estadística AP y los cursos estadísticos introductorios a nivel universitario.

Como la distribución muestral de una muestra estadística es un tema con el que muchos profesores podrían no sentirse familiarizados, aquí se incluyen varios ejemplos para mostrar cómo la simulación puede ser usada para obtener una distribución muestral aproximada para una proporción muestral y media muestral.

Ejemplo 1: La Distribución Muestral de una Proporción Muestral

Las propiedades de la distribución muestral para proporción de la muestra, pueden ser ilustradas mediante la simulación del proceso de selección de una muestra aleatoria de una población, usando dígitos aleatorios como un dispositivo para modelar varias poblaciones.

Por ejemplo, suponga que se asume que una población tiene 60% "éxitos" (p=.6) y vamos a tomar una muestra aleatoria de n=40 casos de esta población. ¿Hasta dónde podemos esperar que la proporción muestral de éxitos difiera del verdadero valor de la población de .60? Esto puede ser contestado mediante la determinación de una distribución muestral empírica de la proporción muestral.

Una forma de modelar una población con 60% de éxitos (y 40% de fracasos) es utilizar los 10 dígitos 0, 1,…,9. Etiquetar seis de los 10 dígitos como "éxitos" y los otros cuatro como "fracasos". Para simular la selección de una muestra de tamaño 40 de esta población, seleccione aleatoriamente 40 dígitos (con remplazamiento). Registre el número de éxitos de los 40 dígitos seleccionados y convierta este conteo a la proporción de éxitos en la muestra. Note que:

Proporción de Éxitos en la Muestra =
(Número de Éxitos en la Muestra)/(Tamaño de la Muestra)

Al repetir este proceso muchas veces, y determinar la proporción de éxitos para cada muestra, se ilustra la idea de la variabilidad de muestra a muestra en la proporción muestral.

Al simular la selección de 200 muestras aleatorias de tamaño 40 de una población con 60% de éxitos y determinar la proporción de éxitos para cada muestra resulta en la distribución empírica que se muestra en la Figura 25. Esta distribución empírica es una aproximación a la verdadera distribución muestral de la proporción muestral para muestras de tamaño 40 de una población en la cual la proporción real es .60.

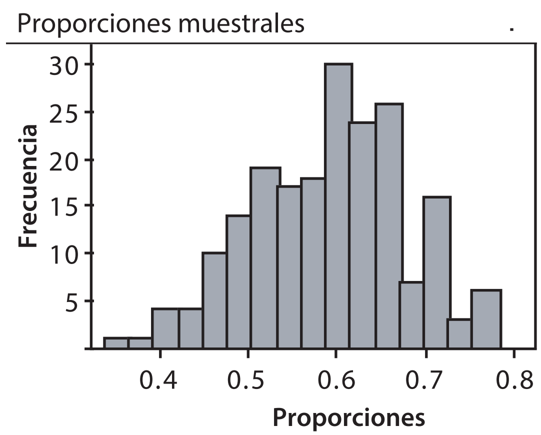


Figura 25: Histograma de proporciones muestrales

Resumiendo la distribución anterior y teniendo en cuenta su forma, centro y dispersión, se puede establecer que esta distribución muestral empírica tiene forma de montículo (aproximadamente normal). Como la media y la desviación estándar de la proporción muestral de 200 es .59 y .08, respectivamente, la distribución empírica que se muestra en la Figura 25 tiene una media de .59 y una desviación estándar de .08.

Mediante el estudio de esta distribución muestral empírica, y otras que puedan ser generadas de manera similar, los estudiantes verán emerger patrones. Por ejemplo, los estudiantes observarán que, cuando el tamaño de la muestra es razonablemente grande (y la proporción poblacional de éxitos no es muy cercana a los extremos de 0 ó 1), las formas resultantes de la distribución muestral empírica son aproximadamente normales. Cada una de las distribuciones muestrales empíricas debe estar centrada cerca del valor de p, la proporción poblacional de éxitos, y la desviación estándar para cada distribución debe estar cerca de:

$$\sqrt{\frac{p(1-p)}{n}}$$

Note que en el Ejemplo 1, la media de la distribución empírica es .59, lo cual es cercano a .6, y la desviación estándar es .08, lo cual está cerca de:

$$\sqrt{\frac{.6(.4)}{40}} \approx .0775$$

Un seguimiento al análisis de estas distribuciones muestrales empíricas pueden revelar a los estudiantes que cerca del 95% de las proporciones muestrales se encuentran dentro de una distancia de:

$$2\sqrt{\frac{.6(.4)}{40}} \approx 0.155$$

del verdadero valor de p. Esta distancia se llama *margen de error*.

Ejemplo 2: La Distribución Muestral de una Media Muestral

Las propiedades de la distribución muestral para la media de la muestra se pueden ilustrar en una forma similar a la usada para proporciones en el Ejemplo 1. La Figura 26 exhibe la distribución de la media de la muestra cuando se seleccionan (con remplazamiento) 200 muestras de 30 dígitos aleatorios y se calcula la media de la muestra. Esto simula el muestreo de una población que tiene una distribución uniforme con números iguales de 0s, 1s,

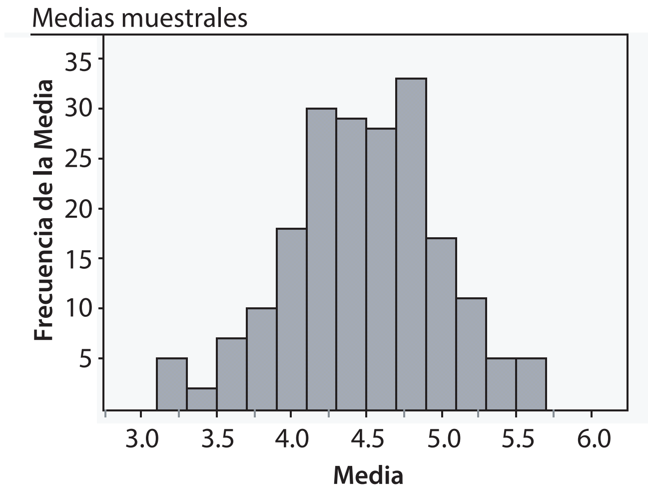


Figura 26: Histograma de las medias muestrales

2s,…,9s. Note que esta población de valores numéricos tienen una media, μ, de 4.5 y una desviación estándar, σ,de 2.9.

La distribución muestral empírica que se exhibe en la Figura 26 se puede describir como aproximadamente normal con media de 4.46 (la media de las 200 medias muestrales de la simulación) y una desviación estándar de 0.5 (la desviación estándar de 200 medias muestrales).

Mediante el estudio de esta distribución muestral empírica, y otras que puedan ser generadas de manera similar, los estudiantes verán emerger patrones. Por ejemplo, los estudiantes observarán que, cuando el tamaño de la muestra es razonablemente grande, las formas de las distribuciones muestrales empíricas son aproximadamente normales. Cada una de las distribuciones muestrales empíricas deben estar centradas cerca del valor de μ, la media de la población, y la desviación estándar para cada distribución debe ser cercana a:

$$\frac{\sigma}{\sqrt{n}}$$

Note que en el Ejemplo 2, la media de la distribución muestral empírica es 4.46, lo cual está cerca de μ=4.5, y la desviación estándar (0.5) está cerca de:

$$\sigma \big/ \sqrt{n} = 2.9 \big/ \sqrt{30} = 0.53$$

El margen de error al estimar una media poblacional usando la media muestral de una muestra aleatoria simple es aproximadamente:

$$2\frac{\sigma}{\sqrt{n}}$$

La media muestral debe estar dentro de esta distancia de la verdadera media poblacional cerca del 95% de las veces en muestreos aleatorios repetidos.

Interpretando Resultados

Generalizando desde Muestras

La clave para la inferencia estadística es la distribución muestral del estadístico muestral, el cual proporciona información acerca del parámetro de la población que está siendo estimado. Como se describió en la sección previa, el conocimiento de la distribución muestral para un estadístico, como la proporción muestral o la media muestral, lleva a un margen de error que proporciona información sobre la distancia máxima probable entre un estimado muestral y el parámetro que está siendo estimado. Otra forma de establecer este concepto clave de inferencia es que un estimador más o menos el margen de error produce un intervalo de posibles valores para el parámetro poblacional. Cualquiera de estos valores posibles puede haber producido el resultado muestral observado como un resultado razonablemente probable.

Generalizando desde Experimentos

¿Difieren los resultados de los experimentos? Al analizar datos experimentales, esta es una de las primeras preguntas que se hacen. Esta pregunta de diferencia generalmente se plantea en términos de las diferencias entre los centros de las distribuciones de los datos (aunque podría ser planteada como diferencia entre los percentiles 90 o cualquier otra medida de localización en una distribución). Como la media es el estadístico más comúnmente usado para medir el centro de una distribución, esta pregunta de diferencias generalmente es planteada como una pregunta sobre una diferencia en las medias. El análisis de datos experimentales, entonces, usualmente involucra una comparación de medias.

A diferencia de las encuestas de muestreo, los experimentos no dependen de muestras aleatorias de una población fija. En cambio, requieren asignación aleatoria de tratamientos a las unidades experimentales preseleccionadas. La pregunta clave, entonces, es: "¿Podría la diferencia observada en las medias de los tratamientos ser debida únicamente a la asignación aleatoria (azar) o puede ser debida al tratamiento administrado?"

Los siguientes ejemplos están diseñados para ilustrar e iluminar aún más los conceptos importantes del Nivel C considerando cuidadosamente las cuatro fases del análisis estadístico—pregunta, diseño, análisis, interpretación—en una variedad de contextos.

Ejemplo 3: Una Encuesta de Preferencias Musicales

En el Nivel A se introdujo una encuesta de preferencias musicales de los estudiantes, donde el análisis consistió en hacer recuentos de las respuestas de los estudiantes y en presentar los datos en diagramas de barras. Al Nivel B, el análisis se amplió para considerar frecuencias relativas de las preferencias y clasificación cruzada de las respuestas para dos tipos de música representado en una tabla de doble entrada. Suponga que la encuesta incluyó las siguientes preguntas:

1. *¿Qué tipo de música te gusta?*

   ¿Te gusta la música country?

   Sí o No

   ¿Te gusta la música rap?

   Sí o No

   ¿Te gusta la música rock?

   Sí o No

2. ¿Cuál de los siguientes tipos de música te gusta más? Selecciona sólo uno.

Country   Rap/Hip Hop   Rock

Para generalizar a todos los estudiantes de la escuela, se necesita una muestra representativa de los estudiantes de la escuela. Esto se puede conseguir seleccionando una muestra aleatoria de 50 estudiantes de la escuela. Los resultados se pueden generalizar a la escuela (pero no mas allá), y la discusión del Nivel C se centrará en principios básicos de generalización—de la inferencia estadística.

El análisis al Nivel C empieza con una tabla de doble entrada que resume los datos de dos preguntas: "¿Te gusta la música Rock?" y "¿Te gusta la música rap?" La tabla proporciona una forma de examinar separadamente las respuestas a cada pregunta y explora conexiones posibles (asociaciones) entre las dos variables categóricas. Suponga que los resultados de 50 estudiantes encuestados se resumen en la Tabla 11.

Como se demostró en el Nivel B, hay una variedad de formas de interpretar datos resumidos en una tabla de doble entrada, como la Tabla 11. Algunos ejemplos basados en los 50 estudiantes encuestados incluyen:

→ A 25 de los 50 estudiantes (50%) les gustan ambos tipos de música: rap y rock.

→ A 29 de los 50 estudiantes (58%) les gusta la música rap.

→ A 19 de los 50 estudiantes (38%) no les gusta la música rock.

Tabla 11: Tabla de Frecuencia de Doble Entrada

		¿Le Gusta la Música Rock?		
		Sí	No	**Totales Fila**
¿Le gusta la música Rap?	Sí	25	4	29
	No	6	15	21
Totales Columna		31	19	50

Un tipo de inferencia estadística está relacionado con conjeturas (hipótesis) hechas antes que los datos sean recolectados. Suponga que un estudiante dice "Yo creo que a más del 50% de los estudiantes de la escuela les gusta la música rap". Como a 58% de los estudiantes en la muestra les gusta la música rap (que es más de 50%), hay evidencia para apoyar la afirmación del estudiante. Sin embargo, como sólo tenemos una muestra de 50 estudiantes, es posible que a 50% de todos los estudiantes les guste el rap (caso en el cual, la afirmación del estudiante no es correcta), pero la variación debida al muestreo aleatorio puede producir 58% (o aún más) que les gusta el rap. La pregunta estadística, entonces, es si el resultado de la muestra de 58% es razonable dada la variación que esperamos que ocurra cuando se selecciona una muestra aleatoria de una población con 50% de éxitos.

Una forma de llegar a una respuesta es configurar una población hipotética que tenga 50% éxitos (tal como los números pares e impares producidos por un generador de números aleatorios) y repetidamente tomar de ella muestras de tamaño 50, cada vez registrar la proporción de dígitos pares.

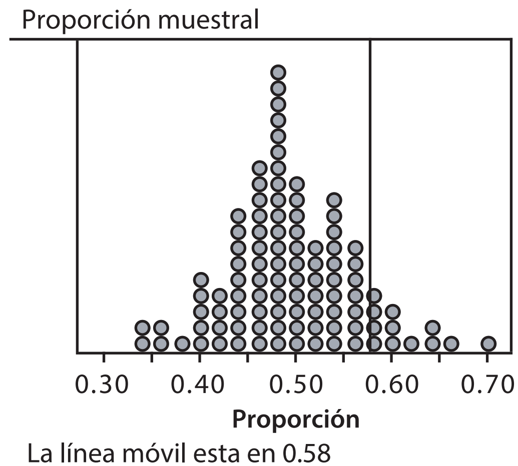


Figura 27: Diagrama de puntos de una proporción muestral de una población hipotética en la cual 50% le gusta la música rap.

La distribución muestral de proporciones así generada será similar a la siguiente.

Basados en la simulación, una proporción muestral mayor o igual al valor observado de .58 ocurrió 12 veces de 100 sólo por variación aleatoria cuando la real proporción de la población es .50. Esto sugiere que el resultado de .58 no es una ocurrencia muy inusual cuando se toman muestras de una población con .50 como la "verdadera" proporción de estudiantes que les gusta la música rap. De manera que un valor poblacional de .50 es posible basado en lo que fue observado en la muestra, y la evidencia que apoya la afirmación del estudiante no es muy fuerte. La fracción de veces que el resultado observado es alcanzado o excedido (.12 en esta investigación) se llama el valor-*p* aproximado. El valor-*p* representa la posibilidad de obtener el resultado observado en la muestra, o un resultado más extremo, cuando el valor hipotético es en efecto correcto. Un valor-*p* pequeño habría apoyado la afirmación del estudiante, porque esto habría indicado que si la proporción poblacional fue .50, habría sido muy improbable que una proporción muestral de .58 hubiera sido observada.

Suponga que otro estudiante conjeturó que a **más del 40%** de los estudiantes en la escuela les gusta la música rap. Para probar esta afirmación, se deben seleccionar repetidamente muestras de tamaño 50 de una población que tiene 40% de éxito. La Figura 28 muestra los resultados de una simulación de estas. El resultado observado de .58 sólo se consiguió una vez de 100, y ninguna

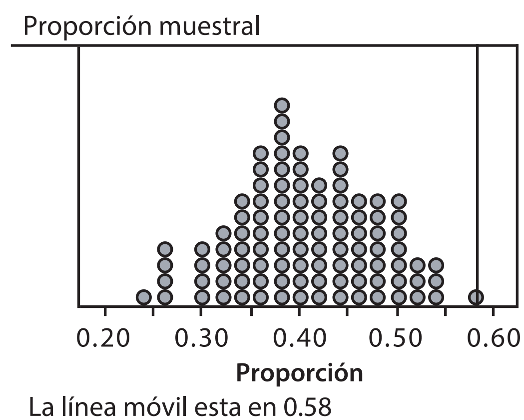


Figura 28: Diagrama de puntos de las proporciones muestrales de una población hipotética en la cual al 40% les gusta la música rap

muestra produjo una proporción mayor que .58. De esta manera, el valor-*p* aproximado es .01, y no es probable que una población en la cual al 40% de los estudiantes les guste la música rap hubiera producido una proporción muestral de 58% en una muestra aleatoria de tamaño 50. Este valor- proporciona fuerte evidencia para apoyar la afirmación del estudiante que a más del 40% de los estudiantes en toda la escuela les gusta la música rap.

Otra forma de expresar lo anterior es que .5 sea un valor posible para la proporción poblacional verdadera, basados en la evidencia de la muestra, pero .4 no lo es. Un conjunto de posibles valores se puede encontrar usando el margen de error introducido en el Ejemplo 1. Como se explicó previamente, el margen de error de la proporción muestral es aproximadamente:

$$2\sqrt{\frac{p(1-p)}{n}}$$

Sin embargo, en este problema, el verdadero valor de *p* es desconocido. Nuestra proporción muestral ($\hat{p}$=.58) es nuestra "mejor estimación" para lo que podría ser *p*, por lo que el margen de error puede ser estimado:

$$2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2\sqrt{\frac{.58(.42)}{50}} \approx .14$$

Así, cualquier proporción entre .58 - .14 = .44 y .58 + .14 =.72 puede ser considerada un valor posible para la verdadera proporción de estudiantes en la escuela que les gusta la música rap. Note que .5 está dentro de este intervalo, pero .4 no lo está.

Otro tipo de pregunta que puede ser formulada sobre las preferencias musicales de los estudiantes sería "¿A quiénes les gusta la música rock también les gusta la música rap?" En otras palabras, ¿hay una asociación entre el gusto por la música rock y el gusto por la música rap? Los mismos datos de la muestra aleatoria de 50 estudiantes pueden ser usados para responder esta pregunta.

De acuerdo a la Tabla 11, a un total de 31 estudiantes en la encuesta les gusta la música rock. Dentro de esos estudiantes, la proporción a quienes también les gusta la música rap es 25/31 = .81. Entre los 19 estudiantes que no les gusta la música rock, 4/19 = .21 es la proporción de quienes les gusta la música rap. Esta gran diferencia entre estas dos proporciones (.60) sugiere que puede existir una fuerte asociación entre el gusto por la música rock y el gusto por la música rap. Pero, ¿podría esta asociación deberse simplemente al azar (una consecuencia sólo del muestreo aleatorio)?

Si no existiera asociación entre los dos grupos, entonces los 31 estudiantes que les gusta el rock se comportarían como una selección aleatoria de los 50 en la muestra. Esperaríamos que la proporción de quienes les gusta la música rap entre esos 31 estudiantes fuera cercana a la proporción de a quienes les gusta la música rap entre los 19 estudiantes que no les gusta la música rock. Esencialmente, esto significa que si no hay asociación, esperaríamos que la diferencia entre estas dos proporciones fuera aproximadamente 0. Como la diferencia en

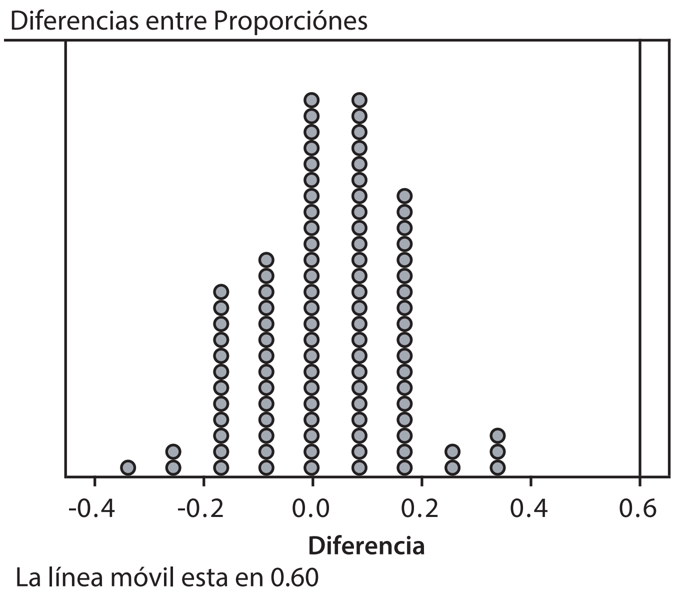


Figura 29: Diagrama de puntos presentando la distribución muestral simulada

nuestra encuesta es .6, esto sugiere que hay una asociación. ¿Puede la diferencia, .6, ser explicada por la variación aleatoria que esperamos cuando seleccionamos una muestra aleatoria?

Para simular esta situación, creamos una población de 29 unos (a quienes les gusta el rap) y 21 ceros (a quienes no les gusta el rap) y los mezclamos entre sí. Luego, aleatoriamente seleccionamos 31 (representando a quienes les gusta el rock) y miramos cuantos unos (a quienes les gusta el rap) conseguimos. Esta es la entrada que va en la celda (sí, sí) de la tabla, y desde este dato se puede calcular la diferencia en proporciones. Repitiendo el proceso 100 veces produce una distribución muestral simulada para la diferencia entre las dos proporciones, como se muestra en la Figura 29.

La diferencia observada en las proporciones de los datos muestrales, .6, nunca se alcanzó en 100 ensayos, indicando que la diferencia observada no puede ser atribuida sólo al azar. Así, hay evidencia convincente de una asociación verdadera entre el gusto por la música rock y el gusto por la música rap.

Ejemplo 4: Un experimento del Efecto de la Luz en el Crecimiento de Plántulas de Rábano

¿Cuál es el efecto de diferente exposición a la luz y a la oscuridad en el crecimiento de plántulas de rábano? Esta pregunta se formuló a estudiantes de una clase de biología que luego se dedicaron a diseñar y a llevar a cabo el experimento para investigar la pregunta. Todas las posibles duraciones de luz y oscuridad no pueden ser investigadas en un experimento, por lo que los estudiantes decidieron

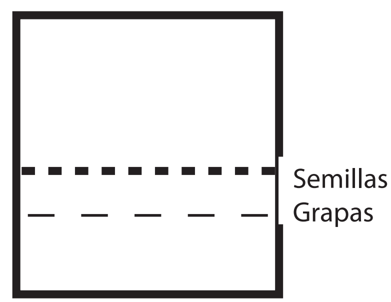


Figura 30: Experimento de semillas

Tabla 12: Longitud de los Plántulas de Rábano

Tratamiento 1 **24 horas de luz**	**Tratamiento 2** **12 horas luz, 12 oscuridad**	**Tratamiento 3** **24 horas de oscuridad**	**Tratamiento 1** **24 horas de luz**	**Tratamiento 2** **12 horas luz, 12 oscuridad**	**Tratamiento 3** **24 horas de oscuridad**
2	3	5 20	10	17	15 30
3	4	5 20	10	20	15 30
5	5	8 22	10	20	15 30
5	9	8 24	10	20	15 31
5	10	8 25	10	20	15 33
5	10	8 25	10	20	15 35
5	10	10 25	10	21	16 35
7	10	10 25	10	21	20 35
7	10	10 25	14	22	20 35
7	11	10 26	15	22	20 35
8	13	10 29	15	23	20 35
8	15	11 30	20	25	20 36
8	15	14 30	21	25	20 37
9	15	14 30	21	27	20 38
					20 40

enfocar la pregunta en tres tratamientos: 24 horas de oscuridad, 12 horas de luz y 12 horas de oscuridad, y 24 horas de oscuridad. Este cubre los casos extremos y uno en la mitad.

Con la ayuda del profesor, la clase decidió usar bolsas plásticas como cámaras de crecimiento. Las bolsas plásticas permitirían a los estudiantes observar y medir la germinación de las semillas sin perturbarlas. Dos capas de toallas papel húmedo se colocaron en una bolsa plástica desechable, con una línea grapada a 1/3 del fondo de la bolsa (ver Figura 30) para mantener la toalla de papel en su lugar y para proporcionar una junta que mantenga las semillas de rábano.

Aunque tres cámaras de crecimiento serían suficientes para examinar los tres tratamientos, esta clase hizo cuatro cámaras de crecimiento, con una designada para el tratamiento 24 horas de luz, una para el tratamiento 12 horas de luz y 12 horas de oscuridad y dos para el tratamiento

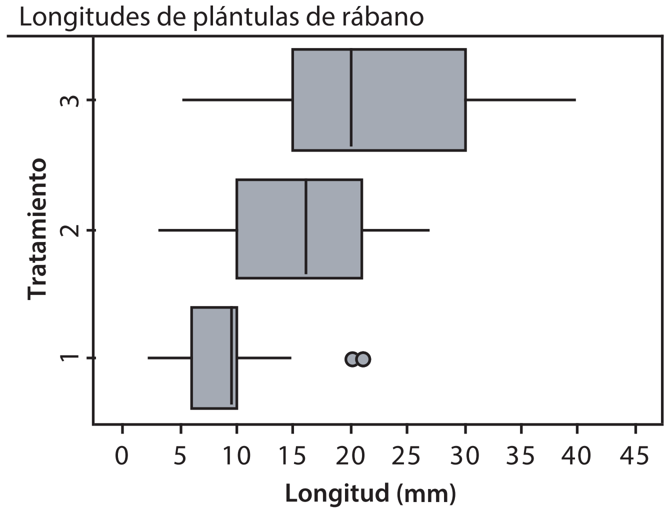


Figura 31: Diagramas de cajas mostrando el crecimiento bajo diferentes condiciones

24 horas de oscuridad. Ciento veinte semillas estuvieron disponibles para el estudio. Treinta semillas se eligieron aleatoriamente y se ubicaron a lo largo de la costura grapada de la bolsa de 24 horas de luz. Treinta semillas se eligieron aleatoriamente de las restantes 90 semillas y se ubicaron en la bolsa 12 horas de luz y 12 horas de oscuridad. Finalmente, 30 de las 60 semillas restantes se eligieron aleatoriamente y se ubicaron en una de las bolsas de 24 horas de oscuridad. Las últimas 30 semillas se ubicaron en la otra bolsa de 24 horas de oscuridad. Después de tres días, se midió y se registró la longitud de los plántulas de rábano de las semillas germinadas. Estos datos se proporcionan en la Tabla 12; las medidas

Tabla 13: Resumen Estadístico de los Tratamientos

Tratamiento	**n**	**Media**	**Mediana**	**Desviación Estándar**
1	28	9.64	9.5	5.03
2	28	15.82	16.0	6.76
3	58	21.86	20.0	9.75

están en milímetros. Note que no todas las semillas en cada grupo germinaron.

Un buen primer paso en el análisis de datos numéricos como estos es hacer gráficas para buscar patrones y alguna desviación inusual de los patrones. Los diagramas de caja son ideales para comparar datos de más de un tratamiento, como se puede ver en la Figura 31. Ambos, los centros y dispersiones aumentan a medida que la cantidad de horas en la oscuridad aumenta. Hay tres datos extremos (uno de 20 mm y dos de 21 mm) en los datos del Tratamiento 1 (24 horas de luz). Por lo demás, las distribuciones son bastante simétricas, lo cual es bueno para la inferencia estadística.

En la Figura 31, el Tratamiento 1 es 24 horas de luz; el Tratamiento 2 es 12 horas de luz y 12 de oscuridad; el Tratamiento 3 es 24 horas de oscuridad.

Los resúmenes estadísticos para estos datos se muestran en la Tabla 13.

Los experimentos son diseñados para comparar efectos de tratamientos, usualmente mediante la comparación de las medias. La pregunta original sobre el efecto de diferentes periodos de luz y oscuridad en el crecimiento

“Los experimentos son diseñados para comparar efectos de tratamientos, usualmente mediante la comparación de las medias.”

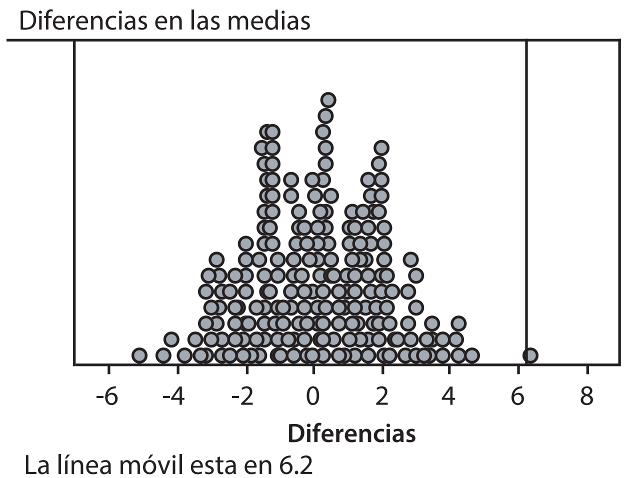


Figura 32. Diagrama de puntos mostrando las diferencias en las medias

de semillas de rábano puede ser transformada en dos preguntas sobre las medias de los tratamientos. ¿Hay evidencia que el grupo de 12 horas de luz y 12 horas de oscuridad (Tratamiento 2) tenga una media significativamente más alta que el grupo de 24 horas de luz (Tratamiento 1)? ¿Hay evidencia que el grupo de 24 horas de oscuridad (Tratamiento 3) tenga la media significativamente más alta que el grupo de 12 horas de luz y 12 horas de oscuridad (Tratamiento 2)? Basados en los diagramas de caja y en el resumen estadístico, es claro que las medias muestrales difieren. *¿Son esas diferencias lo suficientemente grandes para descartar la variación debido al azar como una posible explicación para la diferencia observada?*

La media del Tratamiento 2 es 6.2 mm más grande que la media del Tratamiento 1. Si no hay diferencia real entre los

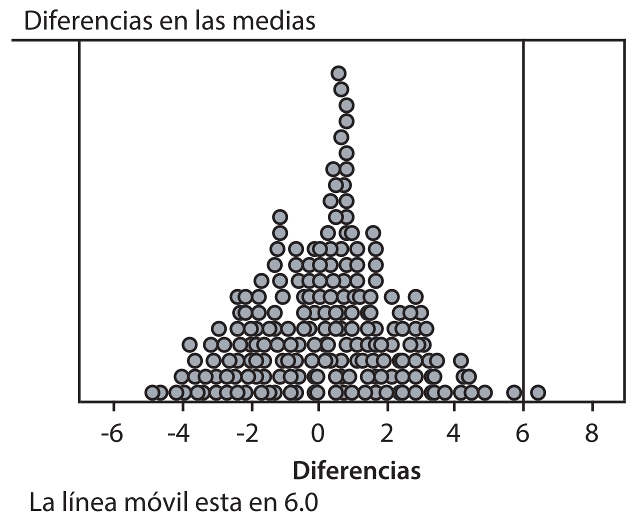


Figura 33. Diagrama de puntos mostrando las diferencias en las medias

dos tratamientos en términos de sus efectos en el crecimiento de la plántula, entonces la diferencia observada puede ser debida a la asignación aleatoria de las semillas a las bolsas; esto es, una bolsa simplemente fue lo suficientemente afortunada de obtener un predominio de semillas buenas y vivas. Pero, si la diferencia es grande (6.2 mm) es probable que sea el resultado sólo de la aleatorización, entonces debemos ver diferencias de esta magnitud muy a menudo si repetidamente asignamos al azar las medidas y calculamos una nueva diferencia en las medias observadas. Esto, sin embargo, no es el caso, como se puede ver en la Figura 32. Este diagrama de puntos se produjo mezclando las medidas de crecimientos de los Tratamientos 1 y 2, dividiéndolas al azar en dos grupos de 28 medidas, registrando la diferencia en medias para los dos grupos, y repitiendo el proceso 200 veces.

La diferencia observada de 6.2 mm fue excedida sólo una vez en 200 ensayos, para un valor-*p* aproximado de 1/200. Esto es muy pequeño, y ofrece evidencia extremadamente fuerte para apoyar la hipótesis que hay una diferencia estadísticamente significativa entre los Tratamientos 1 y 2. La diferencia observada de 6.2 mm es muy improbable para ser debida simplemente a la variación por azar.

En una comparación de las medias de los Tratamientos 2 y 3, el mismo proceso se usa, excepto que las medidas combinadas son divididas en grupos de 28 y 58 cada vez. La diferencia observada de 6 mm fue excedida sólo una vez de 200 ensayos (ver la Figura 33), ofreciendo evidencia extremadamente fuerte de una diferencia estadísticamente significativa entre las medias para los Tratamientos 2 y 3. En resumen, los tres grupos de tratamientos muestran diferencias estadísticamente significativas en el crecimiento medio que no pueden razonablemente ser explicadas por la asignación aleatoria de las semillas a las bolsas. Esto nos da evidencia convincente de un efecto tratamiento—mientras más horas de oscuridad, mayor el crecimiento de los vástagos, al menos para estos tres periodos de luz versus oscuridad.

Los estudiantes deben ser animados a profundizar en las interpretaciones, relacionando con lo que se conoce sobre los fenómenos o problemas bajo estudio. ¿Por qué los vástagos crecen más rápido en la oscuridad? He aquí una explicación de un profesor de biología. Parece ser una adaptación de las plantas para obtener los vástagos de la oscuridad (bajo tierra) donde germinan en la luz (sobre la tierra) lo más rápidamente posible. Obviamente, el vástago no puede fotosintetizarse en la oscuridad y usa la energía guardada en la semilla para potenciar el crecimiento. Una vez que el vástago se expone a la luz, cambia su energía destinada a crecer en longitud para producir clorofila e incrementar el tamaño de sus hojas. Estos cambios permiten que la planta se haga autosuficiente y empiece a producir su propio alimento. Aunque el crecimiento en longitud del tallo es lento, el crecimiento en diámetro del tallo aumenta y el tamaño de las hojas aumenta. Los vástagos que continúan creciendo en la oscuridad son delgados y amarillos, con pequeñas hojas amarillas. Los vástagos cultivados en la luz son ricos, de color verde con hojas gruesas y grandes, y tallos cortos.

Ejemplo 5: Estimando la Densidad de La Tierra—Un Estudio Clásico

¿Cuál es la densidad de la Tierra? Esta es una pregunta que intrigó al gran científico Henry Cavendish quien intentó responder la pregunta en 1798. Cavendish estimó la densidad de la Tierra usando herramientas rudimentarias disponibles en esa época. Él literalmente no tomó una muestra aleatoria; él, como pudo, midió en diferentes días y a diferentes horas. Pero la densidad de la Tierra no cambia con el tiempo, por lo que estas medidas pueden ser consideradas como una muestra aleatoria de todas las medidas que él pudo haber tomado de esta medida constante. La variación en las medidas es debida al error de medida, no al cambio en la densidad de la

Tierra. La densidad de la Tierra es la constante que está siendo estimada.

Este es un ejemplo típico de un problema de estimación que ocurre en la ciencia. No hay "población" real de medidas que puedan ser muestreadas; a cambio, se asume que los datos de la muestra son una selección aleatoria de la población conceptual de todas las medidas que pudieron haber sido tomadas. En este punto, podría existir algo de confusión entre un "experimento" y una "encuesta por muestreo" porque Cavendish realmente llevó a cabo una investigación científica para obtener sus medidas. La clave, sin embargo, es que esencialmente él llevó a cabo la misma investigación muchas veces con el objetivo de estimar una constante, como entrevistar mucha gente para estimar la proporción de los que favorecen un determinado candidato a un cargo. Él no asignó tratamientos aleatoriamente a unidades experimentales con el propósito de comparar los efectos de los tratamientos.

El famoso conjunto de datos de Cavendish contiene sus 29 medidas de la densidad de la Tierra, en gramos por centímetro cúbico. Los datos se muestran a continuación [Fuente: *http://lib.stat.cmu.edu/DASL*]

5.50	5.57	5.42	5.61	5.53	5.47	4.88
5.62	5.63	4.07	5.29	5.34	5.26	5.44
5.46	5.55	5.34	5.30	5.36	5.79	5.75
5.29	5.10	5.86	5.58	5.27	5.85	5.65
5.39						

Uno debe mirar los datos antes de proceder con un análisis. El histograma en la Figura 34 muestra que los datos son aproximadamente simétricos, con un valor pequeño inusual. Si Cavendish estuviera vivo, le podrías preguntar si él cometió un error (y eso es exactamente lo se debe hacer para un conjunto real de datos).

La media de las 29 medidas es 5.42 y la desviación estándar es 0.339. Recuerde que el margen de error para la media muestral es:

$$2\frac{\sigma}{\sqrt{n}}$$

donde σ es la desviación estándar de la población. En este problema, la desviación estándar no es conocida; sin embargo, la desviación estándar de la muestra proporciona una estimación para la desviación estándar de la población. Consecuentemente, el margen de error puede ser estimado:

$$2\frac{s}{\sqrt{n}} = 2\frac{0.339}{\sqrt{29}} = 0.126$$

El análisis muestra que cualquier valor entre 5.420 - 0.126 y 5.420 + 0.126, o en el intervalo (5.294, 5.546), es un valor posible de la densidad de la Tierra. Es decir, cualquier valor en el intervalo es consistente con los datos obtenidos por Cavendish. Ahora, la cuestionable observación baja se debe tomar en cuenta, ya que disminuirá la media e incrementará la desviación estándar. Si esa medida es considerada como un error y se remueve del conjunto de datos, la media de las 28 observaciones restantes es 5.468 y la desviación estándar es

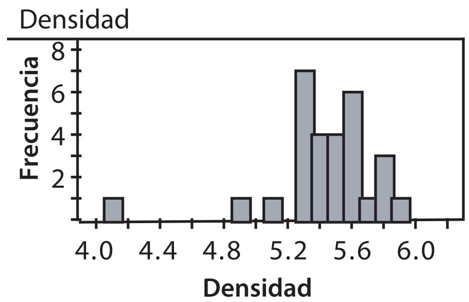


Figura 34: Histograma de las medidas de densidad de la Tierra

0.222, produciendo un margen de error de 0.084 y un intervalo de posibles valores de (5.384, 5.552).

Los estudiantes ahora pueden comprobar qué tan bien lo hizo Cavendish; los métodos modernos están bastante de acuerdo que la densidad promedio de la Tierra es cerca de 5.515 gramos por centímetro cúbico. ¡El gran científico del siglo 18 lo hizo bien!

Ejemplo 6: Análisis de Regresión Lineal—Estatura vs. Longitud de Antebrazo

El análisis de regresión se refiere al estudio de la relación entre variables. Si la "nube" de puntos en un diagrama de dispersión de parejas de datos numéricos tiene una forma lineal, una línea recta puede ser un modelo realístico de la relación entre las variables bajo estudio. La recta de mínimos cuadrados pasa por el centro (en algún sentido) de la nube de puntos. Los residuos se definen

Tabla 14: Estaturas vs. Longitud de Antebrazos

Antebrazo (cm)	Estatura (cm)	Antebrazo (cm)	Estatura (cm)
45.0	180.0	41.0	163.0
44.5	173.2	39.5	155.0
39.5	155.0	43.5	166.0
43.9	168.0	41.0	158.0
47.0	170.0	42.0	165.0
49.1	185.2	45.5	167.0
48.0	181.1	46.0	162.0
47.9	181.9	42.0	161.0
40.6	156.8	46.0	181.0
45.5	171.0	45.6	156.0
46.5	175.5	43.9	172.0
43.0	158.5	44.1	167.0

como la desviación en la dirección *y* entre los puntos del diagrama de dispersión y la recta de mínimos cuadrados; la dispersión es la variación alrededor de la recta de mínimos cuadrados, según lo medido por la desviación estándar de los residuos. Cuando se usa un modelo ajustado para predecir un valor de *y* desde *x*, el margen de error asociado depende de la desviación estándar de los residuos.

Las relaciones entre varias características físicas tales como estatura versus extensión de brazos y tamaño del cuello versus el tamaño del calzado, pueden ser las bases de muchas preguntas interesantes para investigaciones de los estudiantes. Si yo estuviera haciendo una pintura

"El análisis de regresión se refiere al estudio de la relación entre variables."

de una persona, ¿cómo podría obtener correctamente los tamaños relativos de las partes del cuerpo? Esta pregunta pide a los estudiantes llevar a cabo una investigación de una de las posibles relaciones, entre la longitud del antebrazo y la estatura.

Los estudiantes responsables del estudio tomaron una muestra de otros estudiantes en los cuales midieron la longitud del antebrazo y la estatura. Aunque los detalles de cómo la muestra fue realmente seleccionada no son claros, supondremos que es representativa de los estudiantes de la escuela y tiene las características de una muestra aleatoria. Una consideración importante aquí es estar de acuerdo en la definición de "antebrazo" antes de empezar a tomar medidas. Los datos obtenidos por los estudiantes (en centímetros) se dan en la Tabla 14.

Un buen primer paso en todo análisis es graficar los datos, como lo hemos hecho en la Figura 35. La tendencia lineal en la gráfica es bastante fuerte. El diagrama de dispersión, junto con el coeficiente de correlación de Pearson de .8, indica que una recta podría ser un modelo razonable para resumir la relación entre estatura y longitud del antebrazo.

El diagrama de dispersión incluye una gráfica de la recta de mínimos cuadrados:

Estatura Predicha = 45.8 + 2.76 (Longitud del Antebrazo).

La gráfica debajo del diagrama de dispersión muestra los residuos. En el gráfico de residuos hay unos pocos residuos grandes pero no patrones inusuales. La pendiente (aproximadamente 2.8) puede ser interpretada como una

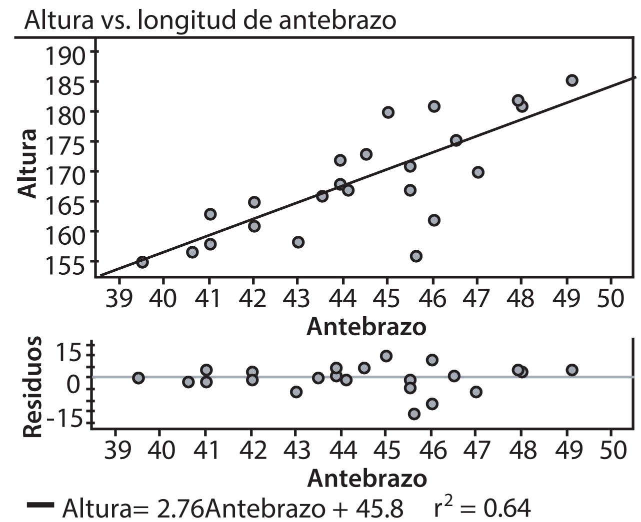


Figura 35: Diagrama de dispersión y gráfico de residuos

estimación de las diferencias promedio en estaturas para dos personas cuyas longitudes de antebrazo son diferentes en 1 cm. ¡El intercepto de 45.8 centímetros no puede ser interpretado como la estatura esperada de una persona con un antebrazo de cero centímetros de longitud! Sin embargo, la recta de regresión puede ser usada razonablemente para predecir la estatura de una persona para quien la longitud del antebrazo es conocida, siempre que la longitud conocida del antebrazo esté en el rango de los datos usados para desarrollar la ecuación de predicción (39 a 50 cm para estos datos). El margen de error para este tipo de predicción es aproximadamente 2 (desviaciones estándar de los residuos). Para estos datos, la desviación estándar de los residuos es 5.8 (no se muestra aquí, pero se proporciona como parte de los resultados de computador), así el margen de error es 2(5.8) = 11.6 cm. La estatura predicha de alguien con una longitud de antebrazo de 42 cm sería:

$$\text{Estatura Predicha} = 45.8 + 2.76\,(42) = 161.7 \text{ cm}$$

Con una confianza del 95%, predeciríamos que la estatura de la gente con longitud de antebrazo de 42 cm está entre 150.1 cm y 173.3 cm (161.7 ± 11.6).

¿Es la pendiente de 2.8 "real", o simplemente un resultado de la variación debida al azar en el proceso de selección aleatoria? Esta pregunta se puede investigar usando simulación. Una descripción de esta simulación se incluye en el Apéndice para el Nivel C.

Ejemplo 7: Comparando Puntajes de Matemáticas—Un Estudio Observacional

Frecuentemente, los datos son presentados en una forma que no requiere mucho análisis, pero requiere algo de perspicacia en los principios estadísticos para la interpretación correcta. Las pruebas estandarizadas frecuentemente caen en esta categoría. La Tabla 15 ofrece información sobre los puntajes promedio estatales en el National Assessment of Educational Progress (NAEP) [Evaluación Nacional de Progreso Educativo] 2000 para Grado 4 en matemáticas, para los estados de Louisiana y Kentucky. Aunque estos puntajes están basados en una muestra de estudiantes, estos son puntajes asignados a los estados, y consecuentemente, desde ese punto de vista pueden ser considerados datos observacionales.

Tabla 15: Puntajes en Matemáticas en el NAEP 2000

	Media Total	Media para los Blancos	Media para los no Blancos	% de Blancos
Louisiana	217.96	229.51	204.94	
Kentucky	220.99	224.17		87

Para ver si los estudiantes entienden la tabla, es informativo pedirles llenar las entradas omitidas.

→ Llenar las dos entradas en la tabla (53% y 199.71)

Preguntas más sustantivas involucran las aparentes contradicciones que pueden ocurrir en datos de este tipo. Podrían ser enunciadas de la siguiente manera:

→ Para los dos estados, compare las medias totales. Compare las medias de los Blancos. Compare las medias de los no Blancos ¿Qué observa?

→ Explique por qué la inversión en la dirección toma lugar una vez las medias se separan en grupos raciales.

Es sorprendentemente genuino para los estudiantes que los resúmenes de los datos (medias en este caso) puedan ir en una dirección en los datos agregados, pero puedan ir en dirección opuesta para cada subcategoría cuando son desagregados. Este fenómeno se llama la Paradoja de Simpson.

Ejemplo 8: Estudio Observacional—Hacia el Establecimiento de la Causalidad

Los estudios observacionales son la única opción para situaciones en las cuales es imposible o no ético asignar aleatoriamente tratamiento a los sujetos. Estas situaciones son una ocurrencia común en el estudio de las causas de las enfermedades. Un ejemplo clásico de este campo es la relación entre fumar y el cáncer de pulmón, que suscitó intensos debates durante los 1950s y 1960s. La sociedad no consentiría la idea de asignar algunas personas a ser fumadores y otras a ser no fumadores en un experimento para ver si el fumar causa cáncer. Por lo que la evidencia tiene que ser recogida del mundo tal cual es. El proceso de recolección de datos aún puede ser diseñado en formas inteligentes para obtener tanta información como sea posible.

He aquí un ejemplo de los debates sobre fumar versus cáncer de pulmón. Un grupo de 649 hombres con cáncer de pulmón fue identificado de una cierta población en Inglaterra. Un grupo control del mismo tamaño se estableció haciendo coincidir estos pacientes con otros hombres de la misma población que no tenía cáncer de pulmón. La correspondencia fue en variables de antecedentes tales como etnia, edad y estatus socioeconómico. (Este se llama un estudio caso-control). El objetivo, entonces es comparar la proporción de fumadores entre aquellos con cáncer de pulmón con la proporción de aquellos sin cáncer.

Tabla 16: Consumo de Cigarrillos y Cáncer de Pulmón

	Casos de Cáncer de Pulmón	**Controles**	**Totales**
Fumadores	647	622	1269
No Fumadores	2	27	29

Primero, asegúrese que los estudiantes entiendan la naturaleza de los datos en la Tabla 16. ¿Los datos muestran que hubo un alto porcentaje de fumadores en Inglaterra alrededor de 1950? La proporción de fumadores en estos grupos fue (647/649) = .997 para los pacientes con cáncer y (622/649) = .958 para los controles. Si estos datos hubieran resultado de una asignación o selección aleatoria, la diferencia de cerca de 4 puntos porcentuales habría sido estadísticamente significativa (mediante los métodos discutidos anteriormente), los cuales dan al investigador una razón para sospechar de una asociación que no puede ser atribuida sólo al azar. Otra forma de mirar a estos datos es pensar en seleccionar aleatoriamente una persona entre los fumadores y una persona entre los no fumadores. El fumador tiene una probabilidad de 647/1269 = .51 de estar en la columna de cáncer de pulmón, mientras el no fumador tiene sólo una probabilidad de 2/29 = .07 de estar allí. Hay evidencia de una fuerte asociación entre fumar y tener cáncer de pulmón, pero no es evidencia concluyente de que fumar es, en efecto, la causa del cáncer de pulmón. (Este es un buen momento para hacer que los estudiantes especulen sobre otras posibles causas que pudieran haber resultado en datos como estos.)

Tabla 17: Nivel de Tabaquismo y Cáncer de Pulmón

Cigarrillos/ Día	Casos de Cáncer de Pulmón	Controles	Probabilidad
0	2	27	0.07
1–14	283	346	0.45
15–24	196	190	0.51
25+	168	84	0.67

Otro paso para establecer la asociación en estudios observacionales es ver si el aumento en la exposición al factor de riesgo produce un aumento en la incidencia de la enfermedad. Esto se hizo con el mismo estudio caso-control mediante la observación del nivel de tabaquismo para cada persona, produciendo la Tabla 17.

El término "probabilidad" es usado en el mismo sentido descrito anteriormente. Si una persona es seleccionada aleatoriamente del Nivel 1–14, la posibilidad que la persona esté en la columna de cáncer es .45, y así para las otras filas. El resultado importante es que estas "probabilidades" aumentan con el nivel de tabaquismo. Esto es evidencia que un incremento en la tasa de enfermedad está asociada con un incremento en el tabaquismo.

Aun con esta evidencia adicional, los estudiantes deben entender que una relación de causa y efecto no puede ser establecida de un estudio observacional. La razón principal de esto es que estos estudios observacionales están sujetos a sesgos en la selección de pacientes y controles. Otro estudio de este tipo podría haber producido un resultado diferente. (Como se vio, muchos estudios de este tipo produjeron resultados muy similares. Eso, emparejado con experimentos de laboratorio con animales que establecieron un vínculo biológico entre fumar y cáncer de pulmón, eventualmente resolvió la pregunta para la mayoría de personas.)

El Apéndice al Nivel C contiene más ejemplos de los tipos discutidos en esta sección.

El Rol de la Probabilidad en la Estadística

Los profesores y estudiantes tienen que entender que estadística y probabilidad no son lo mismo. La estadística usa probabilidad, como la física usa el cálculo, pero sólo ciertos aspectos de la probabilidad se abren paso en la estadística. Los conceptos de probabilidad que se necesitan para la estadística introductoria (con énfasis en análisis de datos) incluyen interpretaciones de frecuencia relativa de los datos, distribuciones de probabilidad como modelos de poblaciones de medidas, una introducción a la distribución normal como un modelo para la distribución muestral, e ideas básicas de valor esperado y variables aleatorias. Reglas de conteo, distribuciones más especializadas y el desarrollo de teoremas en la matemática de la probabilidad deben dejarse a áreas de matemáticas discretas y/o cálculo.

Comprender el razonamiento de la inferencia estadística requiere una comprensión básica de algunas ideas importantes en probabilidad. Los estudiantes deben estar en condiciones de:

→ Entender la probabilidad como una frecuencia relativa a largo plazo;

→ Entender el concepto de independencia; y

→ Entender como la probabilidad puede ser usada al tomar decisiones y sacar conclusiones

"La probabilidad es un intento por cuantificar la incertidumbre."

Además, como muchos de los procedimientos de inferencia estándar están basados en la distribución normal, los estudiantes deben estar en condiciones de evaluar probabilidades usando la distribución normal (preferiblemente con ayuda de la tecnología).

La probabilidad es un intento por cuantificar la incertidumbre. El hecho que el comportamiento en el largo plazo de un proceso aleatorio sea predecible lleva a la interpretación de la frecuencia relativa a largo plazo de la probabilidad. Los estudiantes deben estar en condiciones de interpretar la probabilidad de un resultado como la proporción a largo plazo de las veces que el resultado podría ocurrir si el experimento aleatorio se repite una gran cantidad de veces. Esta interpretación de la frecuencia relativa a largo plazo de la probabilidad también proporciona la justificación de usar simulaciones para estimar probabilidades. Después de observar un gran número de resultados de azar, la proporción observada de ocurrencia del resultado de interés puede ser usada como una estimación de la probabilidad correspondiente.

Los estudiantes también necesitan entender el concepto de independencia. Dos resultados son independientes si nuestra evaluación de la probabilidad de que un resultado ocurra no es afectada por el conocimiento que el otro resultado ha ocurrido. Particularmente importante para la inferencia estadística es la noción de independencia en la configuración de muestreos. La selección aleatoria (con remplazo) de una población, asegura que las observaciones en una muestra sean independientes. Por ejemplo, conocer el valor de la tercera observación no aporta información sobre del valor de la quinta (o cualquier otra) observación. Muchos de los métodos usados para sacar conclusiones sobre una población basada en datos de una muestra requieren que las observaciones en la muestra sean independientes.

Más importante aún, los conceptos de probabilidad juegan un papel crítico en el desarrollo de métodos estadísticos que hacen posible sacar inferencias basadas en datos muestrales y evaluar nuestra confianza en tales conclusiones.

Para clarificar la conexión entre análisis de datos y probabilidad, regresaremos a las ideas claves presentadas en la sección de inferencia. Suponga que una encuesta de opinión revela que el 60% de los votantes muestreados están a favor de una nueva ley propuesta. Una pregunta estadística básica es, "¿Qué tan alejada podría estar esta proporción muestral de la verdadera proporción poblacional?" Que la diferencia entre el valor estimado y el verdadero valor sea menor que el margen de error aproximadamente el 95% de las veces está basado en una comprensión probabilística de la distribución muestral de la proporción muestral. Para muestras aleatorias grandes, la distribución de frecuencia relativa de la proporción muestral es aproximadamente normal. Así que, los estudiantes deben estar familiarizados con usar tecnología apropiada para encontrar áreas bajo la curva normal.

Suponga que un experimentador divide a los sujetos en dos grupos, uno recibiendo un nuevo tratamiento para una enfermedad y el otro recibiendo un placebo. Si el grupo tratamiento se desempeña mejor que el grupo placebo, una pregunta estadística básica es, "¿Puede la diferencia haber sido el resultado de la variación por azar únicamente?" La aleatorización nos permite determinar la probabilidad que una diferencia sea mayor que la observada bajo el supuesto de no efecto tratamiento. A su vez, esta probabilidad nos permite sacar conclusiones significativas de los datos. (Un modelo propuesto se rechaza como inverosímil, no porque la probabilidad de un resultado observado sea pequeña, sino porque está en la cola de la distribución.) Una respuesta adecuada a la pregunta anterior también requiere conocimiento del contexto en el cual la pregunta es formulada y un sólido diseño experimental. Esta confianza en el contexto y en el diseño es una de las diferencias básicas entre estadística y matemáticas.

Como se mostró previamente, la distribución muestral de una media muestral será aproximadamente normal bajo muestreo aleatorio, siempre y cuando el tamaño de la muestra sea razonablemente grande. Usualmente, la media y la desviación estándar de esta distribución son desconocidas (introduciendo la necesidad de la inferencia), pero algunas veces los valores de estos parámetros pueden ser determinados de información básica sobre la población muestreada. Para calcular los valores de estos parámetros, los estudiantes necesitarán algún conocimiento de *valor esperado*, como se muestra a continuación.

Tabla 18: Distribución del Tamaño de la Familia

Tamaño de la familia, x	**Proporción, p(x)**
2	0.437
3	0.223
4	0.201
5	0.091
6	0.031
7	0.017

De acuerdo al Current Population Survey [Encuesta de la Población Actual] de marzo de 2000 de la U.S. Census Bureau [Oficina de Censo de los Estados Unidos], la distribución del tamaño de la familia está dada en la Tabla 18. (Una familia se define como dos o más personas relacionadas que viven juntas. El número "7" realmente es la categoría "7 o más", pero muy pocas familias son más grandes que 7).

Primero, note la conexión entre los datos y la probabilidad: Estas proporciones (realmente estimados de una encuesta por muestreo muy grande) pueden ser tomadas como probabilidades aproximadas para la siguiente encuesta. En otras palabras, si alguien aleatoriamente selecciona una familia de los Estados Unidos para una nueva encuesta, la probabilidad que tenga tres miembros es aproximadamente .223.

Segundo, note que ahora podemos encontrar la media y la desviación estándar de una variable aleatoria (llámela X), definida como el número de personas en una familia seleccionada aleatoriamente. La media, algunas veces llamada *valor esperado* de X y denotada por E(X), se encuentra usando la fórmula:

$$E(X) = \sum_{\substack{\text{Todos los valores} \\ x \text{ posibles}}} x \cdot p(x)$$

el cual resulta ser 3.11 para esta distribución. Si la siguiente encuesta contiene 100 familias seleccionadas aleatoriamente, entonces se espera que la encuesta, en promedio, presente 3.11 miembros por familia, para un estimado total de 311 personas en las 100 familias.

La desviación estándar de X, DE(X), es la raíz cuadrada de la varianza de X, V(X), dado por:

$$V(X) = \sum_{\substack{\text{Todos los valores} \\ x \text{ posibles}}} [x - E(X)]^2 \cdot p(x)$$

Para los datos del tamaño de la familia, V(X) = 1.54 y DE(X) = 1.24.

Tercero, estos hechos se pueden juntar para describir la distribución muestral esperada del tamaño medio de la familia en una muestra aleatoria de 100 familias que se deberán tomar. Esa distribución muestral será de forma aproximadamente normal, con centro en 3.11 y desviación estándar de $1.24/\sqrt{100} = 0.124$. Esto sería información útil para la persona que diseñe la siguiente encuesta.

En resumen, la definición de frecuencia relativa de probabilidad, la distribución normal, y el concepto de valor esperado son las claves para comprender distribuciones muestrales e inferencia estadística.

Resumen del Nivel C

Los estudiantes al Nivel C deben volverse expertos en usar herramientas estadística como una parte natural del proceso investigativo. Una vez se ha implementado un plan apropiado para recolección de datos y los datos resultantes se tienen a la mano, el siguiente paso usualmente es resumir los datos usando representaciones gráficas y resúmenes numéricos. Al Nivel C, los estudiantes deben estar en condiciones de seleccionar técnicas de resumen apropiadas para el tipo de datos disponibles, producir estos resúmenes y describir en contexto las características importantes de los datos. Los estudiantes usarán resúmenes gráficos y numéricos aprendidos en los Niveles A y B, pero deben estar en condiciones de proporcionar una interpretación mucho más sofisticada que integre el contexto y los objetivos del estudio.

Al Nivel C, los estudiantes también deben estar en condiciones de sacar conclusiones de los datos y respaldar estas conclusiones usando evidencia estadística. Los estudiantes deben ver la estadística como el suministro de herramientas poderosas que les permiten responder preguntas y tomar decisiones informadas. Los estudiantes también deben entender las limitaciones de conclusiones basadas en datos de encuestas muestrales y experimentos, y deben estar en condiciones de cuantificar la incertidumbre asociada con estas conclusiones usando un margen de error y las propiedades relacionadas con las distribuciones muestrales.

Apéndice para el Nivel A

¿Cuál es la Longitud de los Nombres Comunes?

Formular Preguntas

Durante la primera semana de escuela, una profesora de tercer grado intenta ayudar a sus estudiantes a aprender los nombres de sus compañeros mediante varios juegos. Durante uno de los juegos, una estudiante llamada MacKenzie notó que ella y su compañero llamado Zacharius tenían nueve letras en sus nombres. MacKenzie conjeturó que sus nombres eran más largos que los nombres de los demás. La profesora decidió que esta observación hecha por la estudiante ofrecía una excelente apertura para una lección de estadística.

El siguiente día de escuela, la profesora recordó a los estudiantes el comentario del día anterior de MacKenzie y preguntó a la clase que les gustaría saber acerca de los nombres de sus compañeros. La clase generó una lista de preguntas, las cuales la profesora escribió en el pizarrón como sigue:

→ ¿Quién tiene el nombre más largo? ¿El más corto?

→ ¿Hay más nombres de nueve letras o de seis letras? ¿Cuántos más?

→ ¿Cuál es la longitud de nombre más común?

→ ¿Cuántas letras hay en todos nuestros nombres?

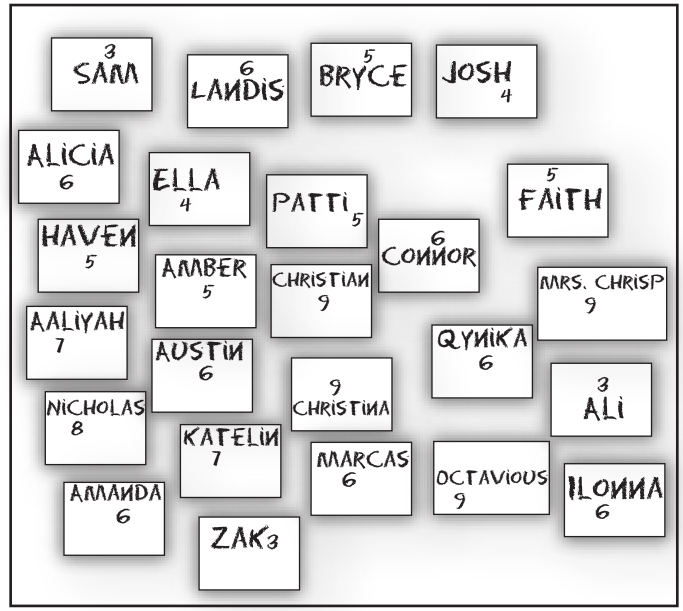


Figura 36: Ubicación aleatoria de nombres

→ Si pones juntos todos los nombres de ocho y nueve letras, ¿habrán tantos como nombres de cinco letras?

Recolección de Datos

La lección estadística empieza con los estudiantes escribiendo sus nombres en notas adhesivas y fijándolas en el pizarrón al frente del salón. Esto es un censo de la clase porque ellos están recogiendo información de todos los estudiantes en la clase.

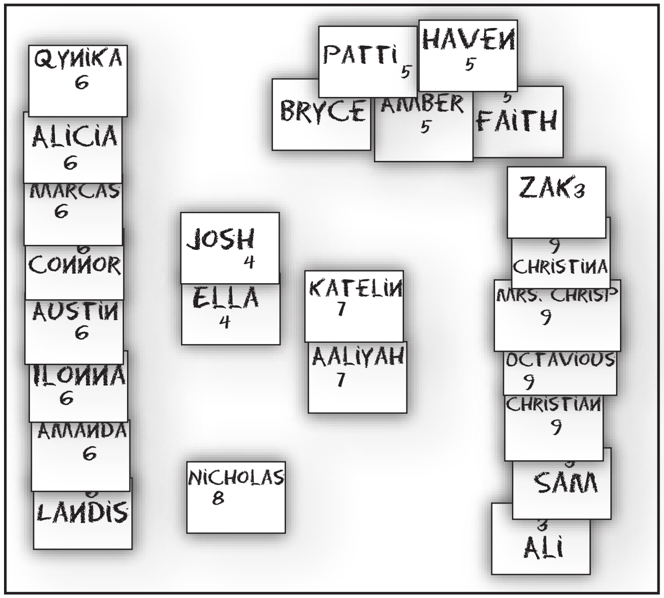


Figura 37: Nombres agrupados por longitud

No se dan instrucciones acerca de cómo organizar las notas, así que los estudiantes arbitrariamente las ubican en el pizarrón.

Para ayudar a los estudiantes a pensar sobre cómo usar herramientas estadísticas para analizar datos, la profesora pregunta a los estudiantes si ahora ellos pueden fácilmente responder cualquiera de las preguntas formuladas mirando a las notas adhesivas, y los estudiantes dicen que no pueden. La profesora entonces sugiere que piensen en formas de organizar mejor las notas. Un estudiante sugiere agrupar los nombres de acuerdo a cuántas letras hay en cada nombre.

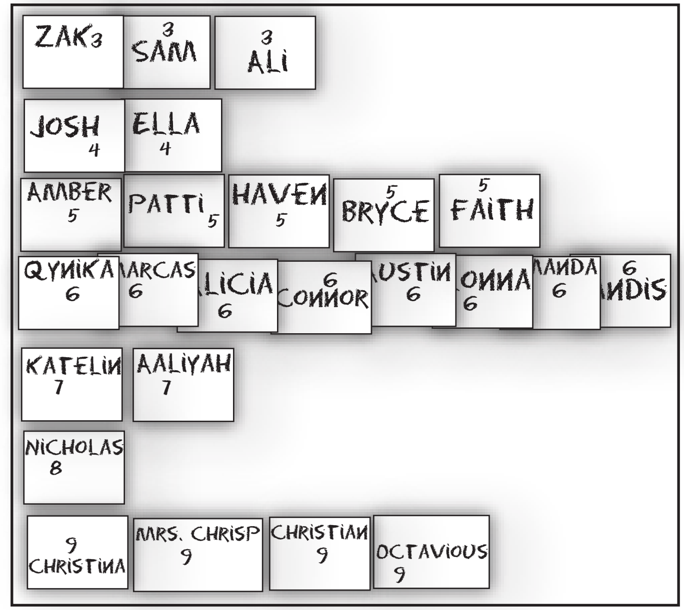


Figura 38: Diagrama de puntos preliminar

La profesora de nuevo pregunta si ellos pueden fácilmente responder las preguntas que han formulado. Los estudiantes dicen que ellos pueden responder *algunas* de las preguntas, pero no fácilmente. La profesora pregunta que podrían hacer para facilitar responder a las preguntas. Puesto que los estudiantes han estado construyendo gráficas desde el preescolar, ellos rápidamente responden "¡hacer una gráfica!" La profesora entonces facilita una discusión sobre qué tipo de gráfica ellos harán, y la clase decide un diagrama de puntos, dado que sus nombres están en notas adhesivas y dado el espacio disponible en el pizarrón. Note que esta representación *no* es un diagrama de barras porque los diagramas de barras se hacen cuando los datos representan variables categóricas (tales como el color favorito). Un diagrama de puntos es apropiado para variables numéricas, tales como el número de letras en el nombre.

La profesora luego usa software de computador para traducir la información en un diagrama de puntos mucho más abstracto, como se muestra en la Figura 39. Esto ayuda a los estudiantes a enfocarse en la forma general de los datos, y no en los nombres particulares de los estudiantes.

Interpretación de Resultados

La profesora facilita la discusión de cada pregunta propuesta por los estudiantes, usando los datos representados en la gráfica para responder las preguntas. Los estudiantes ponen también etiquetas apropiadas y títulos a la gráfica. La profesora ayuda a los estudiantes a usar la palabra "moda" para responder la pregunta sobre la longitud de nombre más común. Ella introduce el término "rango" para ayudar a los estudiantes a responder las preguntas sobre los nombres más cortos y más largos. Los estudiantes visualizan desde el diagrama de puntos que hay variabilidad en la longitud de los nombres de individuo a individuo. El rango da una comprensión de la cantidad de variabilidad en la longitud de los nombres dentro de la clase. Usando el rango, sabemos que si el nombre cualquiera de dos estudiantes se compara, las longitudes de los nombres no pueden diferir por más que el valor del rango.

La profesora luego les dice a los estudiantes que hay otra pregunta útil que ellos pueden responder desde estos datos.

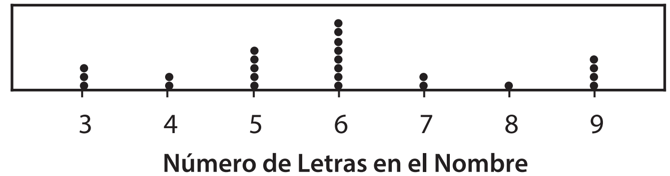


Figura 39: Diagrama de puntos generado por computador

Algunas veces es útil saber "qué tan largos son la mayoría de nombres". Por ejemplo, si estuvieras haciendo tarjetas de ubicación para un almuerzo, podrías querer saber qué tan largo es el nombre típico para decidir qué tamaño de tarjetas comprar. La longitud del nombre típico o promedio se llama la media. Otra forma de pensarlo es "Si todos nuestros nombres fueran de la misma longitud, ¿qué tan largos serían?" Para ilustrar esta idea, la profesora hace que los estudiantes trabajen en grupos de cuatro, y cada niño toma un número de cubos igual al número de letras en su nombre. Luego los cuatro niños colocan todos los cubos en un montón en el centro de la mesa. Ellos cuentan cuantos cubos tienen en total. Luego comparten los cubos equitativamente, cada niño toma uno a la vez hasta que todos los cubos se agoten o hasta que no haya suficiente para compartir. Ellos registran cuántos recibió cada niño. (Los estudiantes de algunas mesas están en condiciones de usar fracciones para mostrar esto, por ejemplo, cuando quedan dos cubos, cada persona puede obtener la mitad de un cubo. En otras mesas los estudiantes, simplemente, dejan los dos cubos restantes sin distribuir.) La profesora luego ayuda a los estudiantes a simbolizar lo que ellos han hecho usando adición para reflejar la acción de poner juntos los cubos

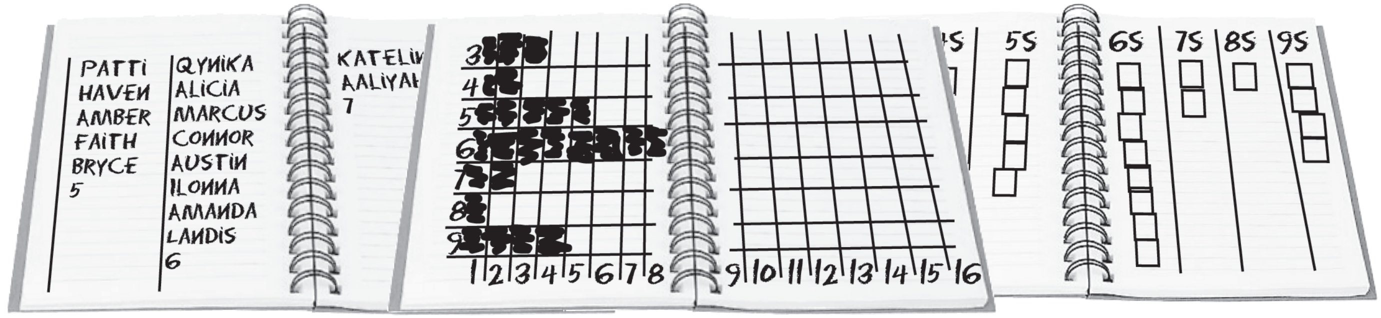


Figura 40: Gráficas dibujadas por los estudiantes

y usando división para reflejar la acción de compartir equitativamente los cubos. Ellos atribuyen las palabras "media" y "promedio" a esta idea.

Finalmente, se les pide a los estudiantes que transfieran los datos de las notas adhesivas en el tablero a sus propios gráficos. La clase ayuda a la profesora a generar preguntas adicionales sobre los datos que se pueden dejar como tarea. Como los gráficos de los estudiantes lucen diferente, el día siguiente la profesora guiará la discusión sobre las características de los diversos gráficos que los estudiantes han construido y las ventajas y desventajas de cada uno.

El Día de San Valentín y los Corazones de Caramelo

Formular Preguntas

Como el día de San Valentín se acerca, una profesora decide planificar una lección en la cual los niños analizarán las características de una bolsa de corazones de caramelo. Para empezar la lección, la profesora levanta una gran bolsa de corazones de caramelo y les pregunta a los niños, en su experiencia previa, qué saben de ellos. Los niños saben que los corazones son de diferentes colores y que tienen algunas palabras escritas en ellos. La profesora les pregunta a los niños que desean saber acerca de la bolsa de corazones. Los niños quieren saber, cuántos corazones hay en la bolsa, qué dicen, y si hay muchos corazones rosados, porque a la mayoría de la gente les gusta más los rosados. La profesora les dice a los niños que ellos podrán responder algunas de esas preguntas sobre sus propias bolsas de corazones.

Recolección de Datos

Cada niño recibe un pequeño paquete de corazones de caramelo. A los estudiantes se les pregunta como ellos pueden organizar sus corazones, y los estudiantes sugieren clasificarlos por color—una variable categórica.

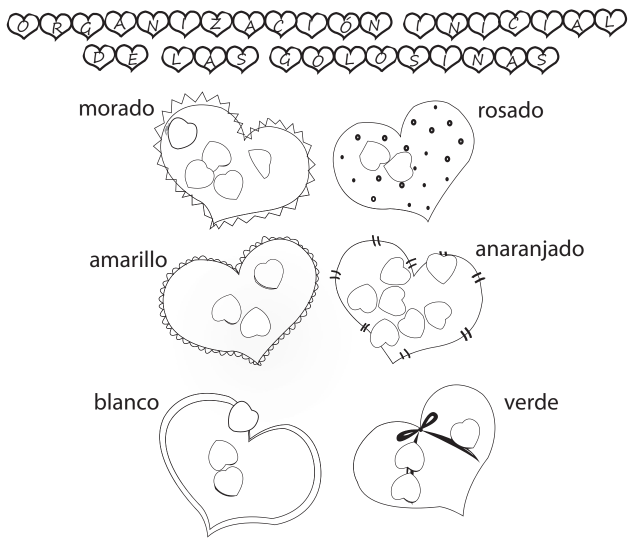


Figura 41: Organización inicial de las golosinas

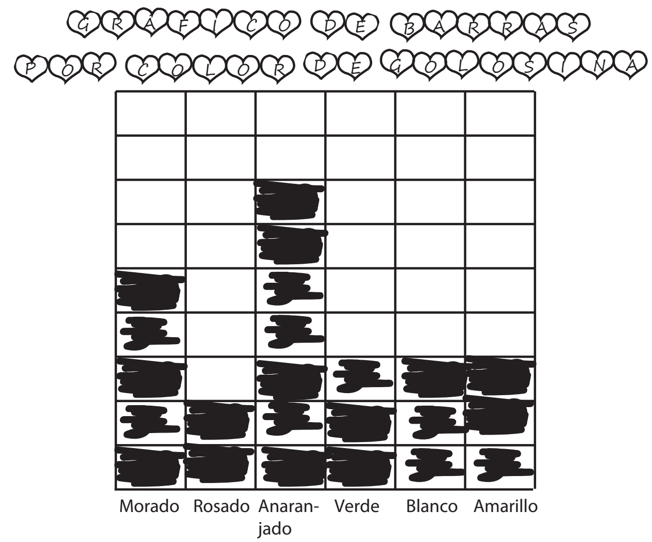


Figura 42: Gráfico de barras por color de golosina

La profesora indaga a los estudiantes que preguntas les ayudará a responder al clasificar los corazones por color, y los estudiantes rápidamente reconocen que esto les dirá cuál color de golosinas aparece más frecuente en la bolsa.

Análisis de Datos

Después de organizar las golosinas en montones y contar y registrar el número de ellas en cada montón, la profesora guía a los estudiantes a hacer un diagrama de barras con sus golosinas en una hoja de papel blanco. Los niños construyen diagramas de barras individuales alineando todas sus golosinas rosadas, todas sus golosinas blancas, etc. La profesora les proporciona una cuadrícula con etiquetas de los colores en el eje-x y etiquetas numéricas en el eje-y para que los estudiantes puedan transferir sus datos de las golosinas reales a un gráfico más permanente.

Interpretación de Resultados

Después que los estudiantes construyen sus gráficos individuales, la profesora distribuye una hoja de registro en la cual cada estudiante registra qué color ocurrió más frecuentemente (la categoría modal) y cuántos obtuvieron de cada color. Esto es seguido por una discusión de clase en la cual la profesora resalta algunos aspectos sobre la variabilidad. Primero, los estudiantes reconocen que el número de cada color varía dentro de un paquete. Los estudiantes también reconocen que sus paquetes de golosinas no son idénticos, notando que algunos estudiantes no tuvieron corazones verdes mientras otros no tuvieron corazones morados. Algunos estudiantes tuvieron más corazones rosados que de otro color, mientras otros tuvieron más corazones blancos. Al Nivel A, los estudiantes están reconociendo la variabilidad entre paquetes—el concepto de variabilidad entre grupos será explorado en detalle en el Nivel B. Los estudiantes conjeturan que estas variaciones entre paquetes fueron debido a como las golosinas fueron empacadas por máquinas. Los estudiantes también notaron diferencias en el número total de golosinas por paquete, pero encontraron esta diferencia pequeña. Los estudiantes con la menor cantidad de golosinas tuvieron 12, mientras los estudiantes con la mayor cantidad tuvieron 15. La profesora preguntó a los estudiantes si ellos habían leído alguna vez la frase "empacados por peso, no por volumen" en el lado del paquete. La clase luego discutió que significa esto y como puede relacionarse al número de golosinas en una bolsa.

(Nota: Las imágenes en este ejemplo fueron adaptadas de *www.littlegiraffes.com/valentines.html*.)

Apéndice para el Nivel B

Muchos cuestionarios piden una respuesta de "Sí" o "No". Por ejemplo, en el documento del Nivel B, exploramos conexiones entre si a los estudiantes les gusta la música rap y si les gusta la música rock. Para investigar posibles conexiones entre estas dos variables categóricas, los datos se resumieron en la siguiente *tabla de frecuencia de doble entrada* o *tabla de contingencia*.

Tabla 4: Tabla de Frecuencia de Dos Entradas

		¿Le Gusta la Música Rap?		
		Si	**No**	**Total de Filas**
¿Le Gusta la Música Rock?	**Si**	27	6	33
	No	4	17	21
Total Columnas		31	23	54

Como el 82 % (27/33) de los estudiantes a quienes les gusta la música rock también les gusta la música rap, a los estudiantes que les gusta la música rock también tienden a gustarles la música rap. Como a los estudiantes que les gusta la música rock tiende a gustarles la música rap, hay una *asociación* entre el gusto por la música rock y el gusto por la música rap.

Al Nivel B, exploramos la asociación entre estatura y extensión de brazos examinando los datos en un diagrama de dispersión, y medimos la fuerza de la asociación con el Cociente de Conteo de Cuadrante, o CCC. Para el problema de la estatura/extensión de brazos, ambas variables son numéricas. También es posible medir la fuerza y la dirección de la asociación entre ciertos tipos de variables categóricas. Recordemos que dos variables numéricas están asociadas positivamente cuando los valores por encima del promedio de una variable tienden a ocurrir con los valores por encima del promedio de la otra variable y cuando los valores por debajo del promedio de una variable tienden a ocurrir con los valores por debajo del promedio de la otra variable. Dos variables numéricas están asociadas negativamente cuando los valores por debajo del promedio de una variable tienden a ocurrir con los valores por encima del promedio de la otra y cuando valores por encima del promedio de una variable tienden a ocurrir con valores por debajo del promedio de la otra variable.

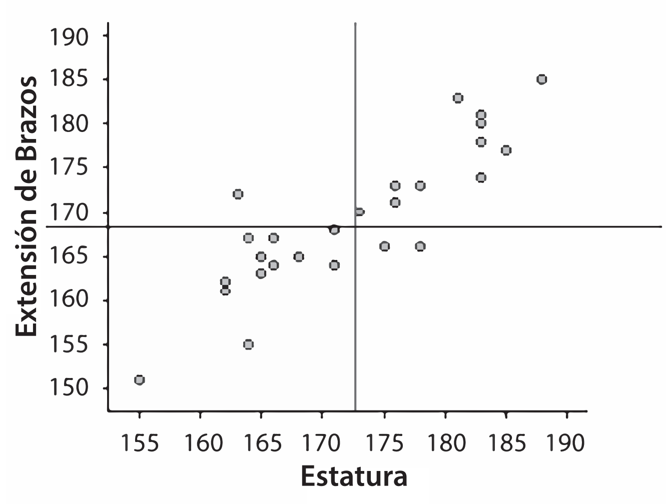


Figura 43: Diagrama de dispersíon para los datos extensión de brazos/estatura

El siguiente diagrama de dispersión para los datos de la estatura/extensión de brazos incluye una línea vertical (x = 172.8) dibujada a través de la estatura media y una línea horizontal (y= 169.3) dibujada a través de la media de la extensión de brazos.

Una forma alternativa de resumir los datos habría sido hacer a cada estudiante las siguientes preguntas:

¿Es tu estatura por encima del promedio?

¿Es tu extensión de brazos por encima del promedio?

Note que para estos datos, la respuesta a cada pregunta es "Sí" o "No".

Los 12 individuos en el diagrama de dispersión con estatura y extensión de brazos por debajo del promedio (Cuadrante 3) respondieron "No" a ambas preguntas. Como sus respuestas a ambas preguntas son la misma, estas 12 respuestas *concuerdan*. Los 11 individuos en el diagrama de dispersión con estatura y extensión de brazos por encima del promedio (Cuadrante 1) respondieron "Sí" a ambas preguntas. Como sus respuestas a ambas preguntas son la misma, estas 11 respuestas concuerdan. Cuando las respuestas a dos preguntas de "Sí/No" son la misma (No/No) o (Sí/Sí), las respuestas concuerdan.

El individuo con estatura por debajo del promedio y extensión de brazos por encima del promedio (Cuadrante 2) respondió "No" a la primera pregunta y "Sí" a la segunda pregunta, (No/Sí). Como su respuesta a las dos preguntas es diferente, estas dos respuestas están en *desacuerdo*. Los dos individuos con estatura por encima del promedio y extensión de brazos por debajo del promedio (Cuadrante 4) respondieron "Sí" a la primera pregunta y "No" a la segunda pregunta (Sí/No). Como sus respuestas a las dos preguntas son diferentes, sus respuestas están en desacuerdo. Cuando las respuestas a dos preguntas de "Sí/No" son diferentes (No/Sí) o (Sí/No), las respuestas están en desacuerdo.

Para los datos en el diagrama de dispersión en la Figura 43, los resultados a las dos preguntas anteriores pueden ser resumidos en la siguiente tabla de frecuencias de doble entrada 2x2.

Tabla 19: Tabla de Frecuencia de Doble Entrada 2x2

		¿Estatura por encima del Promedio?		**Total Filas**
		No	Sí	
¿Extensión de Brazos por encima del Promedio?	No	12	2	14
	Si	1	11	12
Total Columnas		13	13	26

Note que hay un total de 23 respuestas que concuerdan (12 No/No y 11 Sí/Sí para las preguntas de estatura/extensión de brazos), y que estos corresponden a los puntos en los Cuadrantes 3 y 1, respectivamente, en el diagrama de dispersión. También hay un total de tres respuestas en desacuerdo (dos Sí/No y un No/Sí), y estos corresponden a los puntos en los Cuadrantes 4 y

2, respectivamente. Recordemos que el CCC es determinado como sigue:

((El número de datos en los Cuadrantes 1 y 3) - (El número de datos en el Cuadrantes 2 y 4)) / (Número de Puntos en los Cuatro Cuadrantes)

Re-expresado en términos de la Tabla 19:

CCC = ((Número de Puntos que Concuerdan) - (Número de Puntos en Desacuerdo)) / (Número de Puntos en los Cuatro Cuadrantes)

Basados en esto, podemos decir que las dos variables categóricas "Sí/No" están asociadas positivamente cuando las respuestas tienden a concordar—mientras más observaciones concuerdan más fuerte será la asociación positiva. La asociación negativa entre dos variables categóricas "Sí/No" ocurre cuando las respuestas tienden a estar en desacuerdo—mientras más observaciones en desacuerdo, más fuerte la asociación negativa.

Las respuestas a dos preguntas de "Sí/No" se pueden resumir en una tabla de frecuencia de doble entrada como la siguiente:

Tabla 20: Tabla de Frecuencia de Doble Entrada

		Pregunta 1		**Total Filas**
		No	Sí	
Pregunta 2	No	a	b	$r_1 = a + b$
	Si	c	d	$r_2 = c + d$
Total Columnas		$c_1 = a + c$	$c_2 = b + d$	$T = a + b + c + d$

Nota: a = el número que quienes respondieron No/No; b = el número de quienes respondieron No/Sí; c = el número de quienes respondieron Sí/No; d = el número de quienes respondieron Sí/Sí.

Conover (1999) sugirió la siguiente medida de asociación basada en una tabla de 2x2 como la presentada anteriormente.

$$((a + d) - (b + c)) / T$$

Llamemos a esta medida *Cociente de Acuerdo-Desacuerdo* (ADC). Note que esta medida de asociación es análoga al coeficiente de correlación CCC para dos variables numéricas.

El ADC para los datos de estatura/extensión de brazos es:

$$ADC = ((12 + 11) - (2 + 1)) / 26 = .77$$

Un ADC de 0.77 indica una asociación positiva fuerte entre las medidas de estatura y la extensión de los brazos.

Retomemos el ejemplo de los datos de la música, los cuales fueron resumidos como sigue:

Tabla 21: Tabla de Frecuencia de Doble Entrada

		¿Le Gusta la Música Rap?		
		No	Sí	**Total Filas**
¿Le Gusta la Música Rock?	No	17	4	21
	Si	6	27	33
Total Columnas		23	31	54

El ADC para los datos de la música rap/rock es:

$$ADC = ((17 + 27) - (4 + 6)) / 54 = .63$$

Un ADC de .63 indica una asociación bastante fuerte entre el gusto por la música rock y el gusto por la música rap.

Otra pregunta presentada al Nivel B fue:

¿A los estudiantes que les gusta la música country tienden a gustar o disgustar la música rap?

Los datos recolectados de 54 estudiantes se resumen en la siguiente tabla de frecuencia de doble entrada:

Tabla 22: Tabla de Frecuencia de Doble Entrada

		¿Le Gusta la Música Rap?		
		No	Si	**Total Filas**
¿Le Gusta la Música Country?	No	10	22	32
	Si	13	9	22
Total Columnas		23	31	54

Para estos datos,

$$ADC = ((10 + 9) - (22 + 13)) / 54 = -.30$$

Un ADC de -.30 indica una asociación negativa entre el gusto por la música country y el gusto por la música rap.

El CCC y el ADC son de naturaleza aditiva, se basan en "cuántos" valores hay en cada cuadrante o celda. Conover (1999) sugiere *el coeficiente phi* como otra posible medida de asociación para los datos resumidos en una tabla 2x2.

$$Phi = (ad - bc) / \sqrt{(r_1\, r_2\, c_1\, c_2)}$$

Conover señala que Phi es análogo al coeficiente de Correlación de Pearson para datos numéricos. Ambos coeficientes de correlación Phi y Pearson son multiplicativos, y el coeficiente de correlación de Pearson está basado en "qué tan lejos" están los puntos en cada cuadrante del punto central.

Recordemos que en el ejemplo 6 del Nivel C, los estudiantes investigaron la relación entre estatura y la longitud del antebrazo. Los datos observados se muestran de nuevo aquí como en la Tabla 14, y los gráficos resultantes y los análisis de regresión se presentan en la Figura 35.

Análisis de Regresión: Estatura vs. Antebrazo

La ecuación de regresión es:

Estatura Predicha = 45.8 + 2.76 (Antebrazo)

Tabla 14: Estaturas vs. Longitud de Antebrazos

Antebrazo (cm)	Estatura (cm)	Antebrazo (cm)	Estatura (cm)
45.0	180.0	41.0	163.0
44.5	173.2	39.5	155.0
39.5	155.0	43.5	166.0
43.9	168.0	41.0	158.0
47.0	170.0	42.0	165.0
49.1	185.2	45.5	167.0
48.0	181.1	46.0	162.0
47.9	181.9	42.0	161.0
40.6	156.8	46.0	181.0
45.5	171.0	45.6	156.0
46.5	175.5	43.9	172.0
43.0	158.5	44.1	167.0

¿Es la pendiente de 2.8 "real", o simplemente un resultado de la variación por el azar del proceso de selección aleatoria? Esta pregunta puede ser investigada usando simulación.

Si no hubiera relación real entre la estatura y la longitud del antebrazo, entonces cualquiera de los valores de estatura podría ser emparejado con cualquiera de los valores del antebrazo sin pérdida de información. En el espíritu de la comparación de medias en el experimento de los rábanos, podrías aleatoriamente mezclar las estaturas

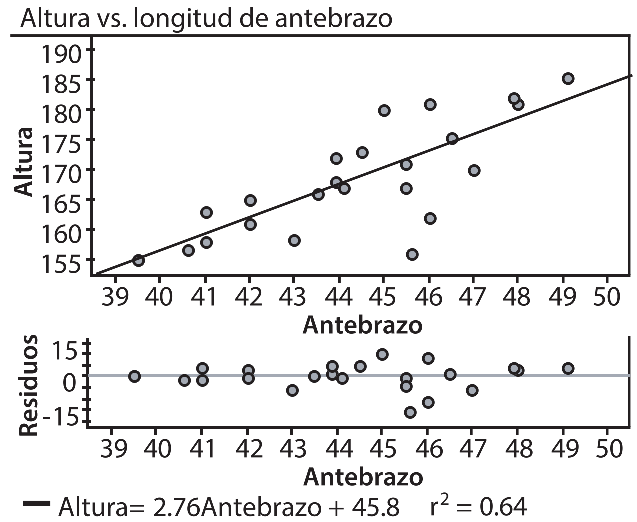


Figura 35: Diagrama de dispersión y gráfico de residuos

(dejando las longitudes de los antebrazos como son), calcular una nueva pendiente, y repetir este proceso muchas veces para ver si la pendiente observada podría ser generada simplemente por la aleatorización. El resultado de estas 200 aleatorizaciones se muestran en la Figura 44. Una pendiente tan grande como 2.8 nunca se rechaza por la aleatorización, lo cual provee una fuerte evidencia que la pendiente observada no es debida simplemente a la variación por azar. Una conclusión apropiada es que hay evidencia significativa de una relación lineal entre la longitud del antebrazo y la estatura.

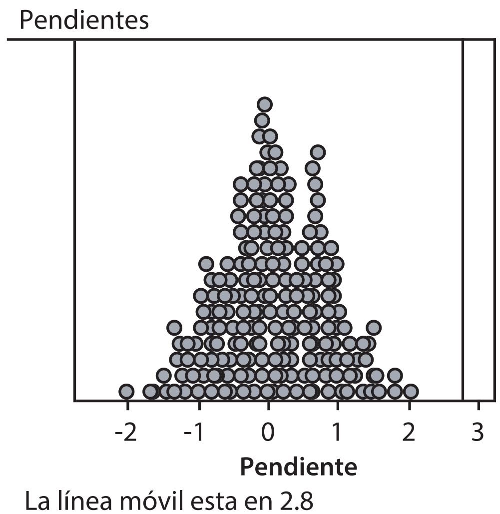


Figura 44: Diagrama de puntos mostrando la asociación

Ejemplo 1: Encuesta de Estilo de Vida Saludable

Una clase de la escuela preparatoria interesada en el estilo de vida saludable, llevó a cabo una encuesta para investigar varias preguntas que pensaban estaban relacionadas con esa cuestión. A una muestra aleatoria de 50 estudiantes de los que asisten a la preparatoria en un día particular se les hizo una variedad de preguntas relacionadas con la salud, incluyendo estas dos:

¿Crees que tienes un estilo de vida saludable?

¿Desayunas al menos tres veces a la semana?

Los datos se presentan en la Tabla 23.

Tabla 23: Resultados de la Pregunta de Estilo de Vida

	Desayuna		
Estilo de Vida Saludable	Sí	No	Total
Sí	13	15	34
No	5	11	16
Total	24	26	50

De estos datos, recogidos en una encuesta por muestreo bien diseñada, es posible estimar la proporción de estudiantes de la escuela que piensan que tienen un estilo de vida saludable y la proporción de quienes desayunan al menos tres veces a la semana. También es posible evaluar el grado de asociación entre estas dos variables categóricas.

Por ejemplo, en la encuesta de estilo de vida descrita previamente, 24 estudiantes en una muestra aleatoria de 50 estudiantes que asisten a una escuela preparatoria particular reportaron que ellos desayunaban al menos tres veces por semana. Basados en esta encuesta por muestreo, se estima que la proporción de estudiantes en esta escuela que desayunan al menos tres veces por semana es 24/50 = .48 con un margen de error de:

$$2\sqrt{(((.48)(.52)) / 50)} = .14$$

Usando el anterior margen de error resultante de (.14), el intervalo de posibles valores para la proporción poblacional de estudiantes que desayunan al menos tres veces a la semana es (0.34, 0.62). Cualquier proporción poblacional en este intervalo es consistente con los datos de la muestra en el sentido que la muestra resultante podría razonablemente haber venido de una población que tiene esta proporción de estudiantes que desayunan.

Para comprobar si las respuestas a las preguntas sobre desayuno y estilo de vida están asociadas entre sí, usted puede comparar la proporción de respuestas *sí* en la pregunta de estilo de vida saludable para aquellos que regularmente toman desayuno con aquellos que no, muy parecido a la comparación de medias para un experimento aleatorizado. En efecto, si un 1 se registra para cada respuesta *sí* y un 0 para cada respuesta *no*, la proporción muestral de respuestas *sí* es precisamente la media muestral. Para los datos observados, hay un total de 34 1s y 16 0s. Re aleatorizando estas 50 observaciones a los grupos de tamaño 24 y 26 (correspondiente a los grupos de sí y no en la pregunta del desayuno) y calculando la diferencia en las proporciones resultantes dan los resultados en la Figura 45. La diferencia observada en las proporciones muestrales (19/24) – (15/26) = 0.21 fue equiparada o excedida 13 veces de 200 veces, para un valor-*p* estimado de 0.065. Este es moderadamente pequeño, así que hay cierta evidencia que la diferencia entre las dos proporciones podría no ser el resultado de variación por azar. En otras palabras, la respuestas a las preguntas sobre estilo de vida saludable y el consumo de desayuno, parecen estar relacionadas en el sentido que aquellos que piensan que tienen un estilo de vida saludable también tienen una tendencia de tomar el desayuno regularmente.

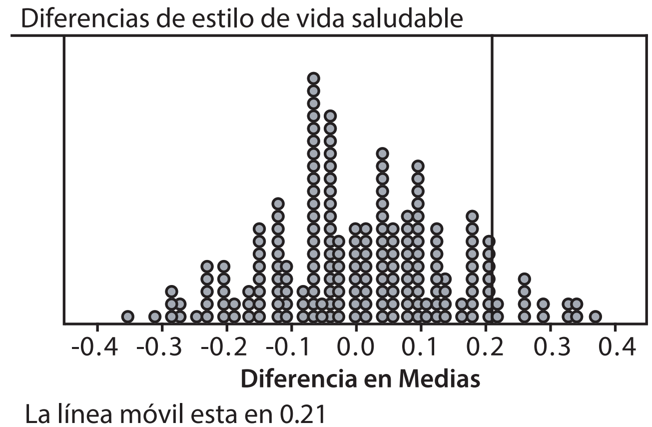


Figura 45: Diagrama de puntos mostrando las diferencias en las proporciones muestrales

Ejemplo 2: Una Investigación Experimental de Frecuencia de Pulso

En otra situación relacionada con la salud, un estudiante decidió responder la pregunta de si estar de pie por algunos minutos incrementa el pulso de las personas (ritmos cardiacos) por una cantidad apreciable. Los sujetos disponibles para este estudio fueron 15 estudiantes de una clase particular. El tratamiento "estar sentado" fue aleatoriamente asignado a ocho de los estudiantes; a los restantes siete se les asignó el tratamiento "estar de pie". La medida registrada fue el conteo del pulso durante 30 segundos, el cual fue luego duplicado para aproximar el conteo en un minuto. Los datos organizados por tratamiento están en la Tabla 24. De estos datos, es posible probar la hipótesis que estar de pie no incrementa la frecuencia del pulso, en promedio, o estimar la diferencia en pulso medio entre aquellos de pie o aquellos sentados. La asignación aleatoria a los tratamientos está intencionada a balancear las variables no medidas o incontrolables que puedan afectar los resultados, tales como género y condiciones de salud. Este se llama un *diseño completamente aleatorizado*.

Tabla 24: Datos de Pulso

	Pulso	Grupo	Categoría
1	62	1	Sentado
2	60	1	Sentado
3	72	1	Sentado
4	56	1	Sentado
5	80	1	Sentado
6	58	1	Sentado
7	60	1	Sentado
8	54	1	Sentado
9	58	2	De Pie
10	61	2	De Pie
11	60	2	De Pie
12	73	2	De Pie
13	62	2	De Pie
14	72	2	De Pie
15	82	2	De Pie

Sin embargo, asignar aleatoriamente 15 estudiantes a dos grupos podría no ser la mejor forma de balancear información antecedente que pueda afectar los resultados. Podría ser mejor poner en *bloque* una variable relacionada con el pulso. Como la gente tiene diferentes frecuencias de pulso en reposo, los estudiantes en el experimento fueron puestos en bloque por la frecuencia de pulso en reposo emparejando los dos estudiantes con la frecuencia de pulso más baja, luego los dos siguientes más bajos y así sucesivamente. Una persona en cada pareja fue aleatoriamente asignada a sentarse y la otra a estar de pie. Los datos de las parejas equiparadas están en la Tabla 25. Como en el diseño completamente aleatorizado, la diferencia media de la frecuencia de pulso entre estar sentado y de pie puede ser estimada. La principal ventaja de poner en bloque es que la variación en las diferencias (la cual ahora forma la base del análisis) es mucho menos que la variación entre las medidas de pulso que forman la base del análisis para el diseño completamente aleatorizado.

Tabla 25: Datos de Pulso de Parejas Equiparadas

	Pulso Medio Sentado	Pulso Medio de Pie	Diferencia
=			
1	68	74	6
2	56	55	-1
3	60	72	12
4	62	64	2
5	56	64	8
6	60	59	-1
7	58	68	10

En el primer experimento de frecuencia de pulso (Tabla 24), los tratamientos de “estar sentado” o “estar de pie” fueron aleatoriamente asignados a los estudiantes. Si no hay diferencia real en las frecuencias de pulso entre estos dos tratamientos, entonces la diferencia observada en las medias (4.1 latidos por minuto) es debido al proceso de aleatorización. Para probar esto, los datos resultantes del experimento pueden ser re aleatorizados (reasignados a sentarse o estar de pie después del hecho) y registrar una nueva diferencia en medias. Hacer la re aleatorización muchas veces generará una distribución de diferencias en las medias muestrales debido solo al azar. Usando esta distribución, uno puede evaluar la probabilidad de la diferencia observada original. La Figura 46 muestra los resultados de estas 200 re aleatorizaciones. La diferencia observada de 4.1 fue equiparada o excedida 48 veces, lo cual da un valor-*p* estimado de 0.24 de ver un resultado de 4.1 o más solo por azar. Como este es un valor-*p* bastante grande, se puede concluir que hay muy poca evidencia de una real diferencia en las medias de las frecuencias de pulso entre las posiciones sentado y de pie basados en los datos observados.

En el diseño emparejado, la aleatorización ocurre en cada par—una persona es asignada aleatoriamente a sentarse mientras la otra permanece de pie. Para evaluar si la diferencia observada puede ser debida sólo al azar y no debida a la diferencia en tratamientos, la re aleatorización debe ocurrir dentro de la pareja. Esto implica que la re aleatorización es simplemente una cuestión de asignar aleatoriamente un signo más o menos a los valores numéricos de las diferencias observadas. La Figura 47 en la siguiente página muestra la distribución de la diferencia de las medias para estas 200 re aleatorizaciones; la diferencia media de 5.14 fue equiparada o excedida ocho veces. Por lo tanto, la probabilidad estimada de conseguir una diferencia media de 5.1 o mayor por sólo azar es 0.04. Esta probabilidad tan pequeña proporciona evidencia que la diferencia media puede ser atribuida a algo más que sólo azar (inducida por el proceso de aleatorización inicial). Una mejor explicación es que estar de pie aumenta la frecuencia de pulso, en promedio, sobre el estar sentado. La diferencia media parece ser significativa aquí, mientras que no lo fue en el diseño completamente aleatorizado, porque el emparejamiento reduce la variabilidad. Las diferencias en el diseño emparejado tienen menos variabilidad que las medidas individuales en el diseño completamente aleatorizado, haciendo más fácil detectar una diferencia en el pulso medio para los dos tratamientos.

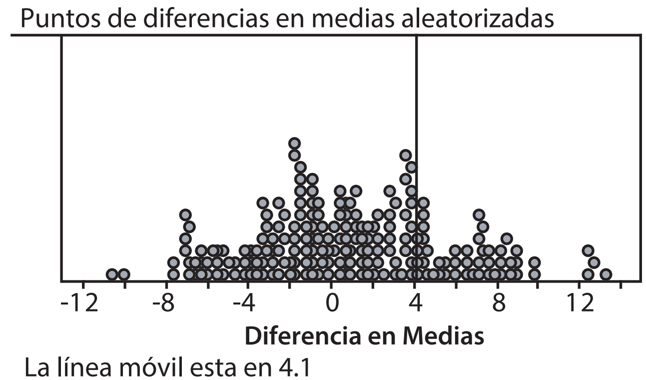
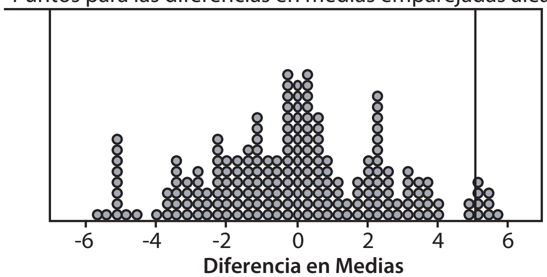


Figura 46: Diagrama de puntos de diferencias en medias aleatorizadas

Puntos para las diferencias en medias emparejadas aleatorizadas



La línea móvil esta en 5.1

Figura 47: Diagrama de puntos para las diferencias en medias emparejadas aleatorizadas

Ejemplo 3: Estudio Observacional—Tasas en el Tiempo

Las estadísticas de vida son un buen ejemplo de datos observacionales que la gente usa cada día en diversos ámbitos de la vida. La mayoría de estas estadísticas se reportan como tasas, una comprensión de tasa es una habilidad crítica para graduados de la preparatoria. La Tabla 26 muestra la población de los Estados Unidos (en miles) de 1990–2001. La Tabla 27 muestra las tasas de muerte por secciones de la población de los Estados Unidos en un periodo de tiempo de 12 años. Frecuentemente, se refiere a los datos registrados en el tiempo como series de tiempo.

La comprensión de los estudiantes de las tasas en la Tabla 27 se puede establecer mediante la formulación de problemas tales como:

→ Explicar cuidadosamente el significado del número 1,029.1 en la celda inferior del lado izquierdo.

Tabla 26: Población de los Estados Unidos (en Miles)

Año	Total Personas	Masculino	Femenino
1990	249,623	121,714	127,909
1991	252,981	123,416	129,565
1992	256,514	125,247	131,267
1993	259,919	126,971	132,948
1994	263,126	128,597	134,528
1995	266,278	130,215	136,063
1996	269,394	131,807	137,587
1997	272,647	133,474	139,173
1998	275,854	135,130	140,724
1999	279,040	136,803	142,237
2000	282,224	138,470	143,755
2001	285,318	140,076	145,242

→ Dé al menos dos razones de por qué las celdas de Hombres Blancos y Hombres Negros no suman la celda Todas las Razas.

→ ¿Puedes decir cuánta gente murió en 2001 basado sólo en la Tabla 27?

Con optimismo, los estudiantes pronto se darán cuenta que no pueden cambiar de tasas de muerte a frecuencias de muerte desconociendo los tamaños poblacionales. La Tabla 26 proporciona los tamaños de población generales, como también las categorías masculino y femenino.

Al notar que las cifras de la población están en miles pero las tasas están por cada 100,000, los estudiantes

Tabla 27: Tasa de Mortalidad en los Estados Unidos (Muertes por Cada 100,000 de la Población)

Año	Todas las Razas		Blancos		Negros	
	Masculino	**Femenino**	**Masculino**	**Femenino**	**Masculino**	**Femenino**
1990	1202.8	750.9	1165.9	728.8	1644.5	975.1
1991	1180.5	738.2	1143.1	716.1	1626.1	963.3
1992	1158.3	725.5	1122.4	704.1	1587.8	942.5
1993	1177.3	745.9	1138.9	724.1	1632.2	969.5
1994	1155.5	738.6	1118.7	717.5	1592.8	954.6
1995	1143.9	739.4	1107.5	718.7	1585.7	955.9
1996	1115.7	733.0	1082.9	713.6	1524.2	940.3
1997	1088.1	725.6	1059.1	707.8	1458.8	922.1
1998	1069.4	724.7	1042.0	707.3	1430.5	921.6
1999	1067.0	734.0	1040.0	716.6	1432.6	933.6
2000	1053.8	731.4	1029.4	715.3	1403.5	927.6
2001	1029.1	721.8	1006.1	706.7	1375.0	912.5

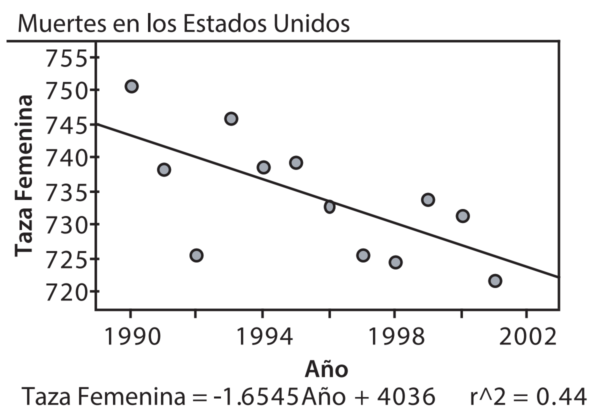


Figura 48: Diagrama de dispersión de tasa de mortalidad

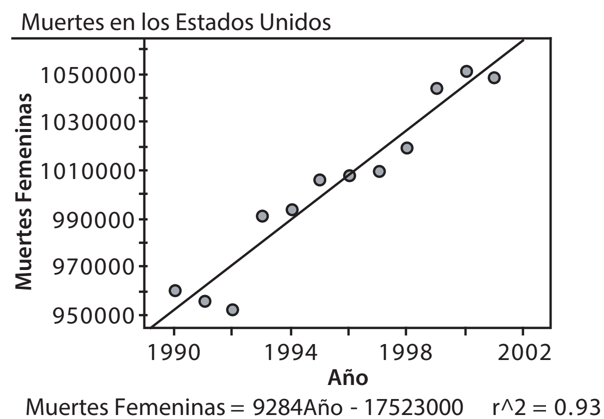


Figura 49: Diagrama de dispersión de muertes reales

necesitan pensar un poco para ir de tasas a conteos haciendo los cómputos que se muestran en la fórmula:

Muertes Femeninas = (Tasa de Muertes Femeninas) ×
((Población Femenina) / 100)

Algunas preguntas sobre series de tiempo pueden ser exploradas ahora. Por ejemplo, ¿cómo el patrón de la tasa de mortalidad femenina en el tiempo se compara con el patrón real de muertes femeninas? Las gráficas de las Figuras 48 y 49 proporcionan una idea visual. Las tasas de mortalidad muestran una tendencia a la baja en el tiempo, con variación considerable, pero las muertes reales están aumentando.

Los estudiantes descubrirán que la situación para los hombres es muy diferente, lo cual puede conducir a interesantes discusiones.

Ejemplo 4: Gráficas: ¿Distorsión de la Realidad?

Examine la gráfica representada en la Figura 50. ¿Ves debilidades en esta representación gráfica? Si es así, descríbelas y explica cómo podrían ser corregidas.

He aquí algunos posibles gráficos para corregir errores de interpretación, y para plantear otras preguntas. Mejores presentaciones empiezan con una tabla de datos, tal como la Tabla 28, y luego con una representación gráfica más estándar de los datos.

La gráfica en la Figura 51 muestra las matrículas totales y las de Afroamericanos en la misma escala. Cuando se mira de esta forma, uno puede ver que las últimas son una pequeña parte de las primeras, con pocos cambios, en comparación, en los años.

Figura 50: Gráfico distorsionado [Fuente: *Athens Banner-Herald*]

Tabla 28: Datos de Matrículas

Año	Total de Estudiantes	Afroamericanos
1996	29404	2003
1997	29693	1906
1998	30009	1871
1999	30912	1815
2000	31288	1856
2001	32317	1832
2002	32941	1825
2003	33878	1897
2004	33405	1845

Mirando las matrículas de los Afroamericanos por sí solas, uno puede ver que la marcada disminución entre 1996 y 2002 puede estar invirtiéndose—o estabilizándose.

Sin embargo, la fracción de matrículas de Afroamericanos con respecto al total es aún decreciente!

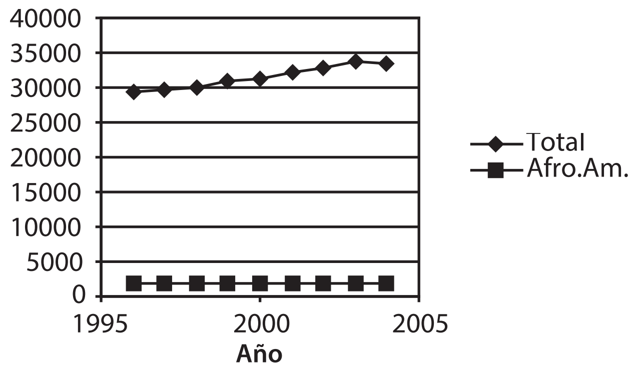


Figura 51: Diagrama de matrículas de Afroamericanos vs. total

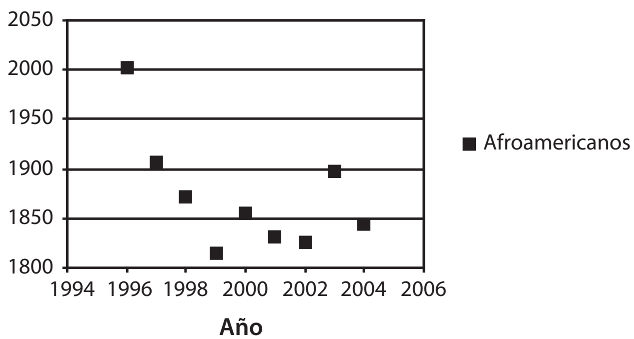


Figura 52: Diagrama de matrícula de solo Afroamericanos

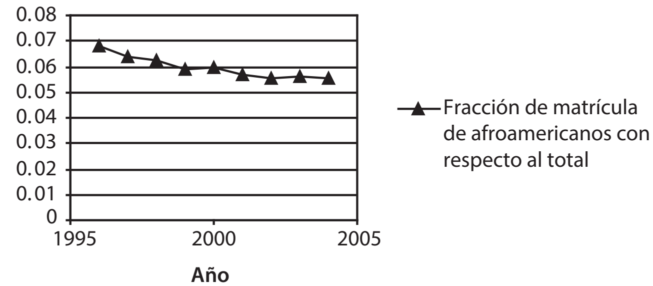


Figura 53: Fracción de matrícula de Afroamericanos con respecto al total

Referencias

Cobb, G. and Moore, D. (2000). "Statistics and Mathematics: Tension and Cooperation," *American Mathematical Monthly*, pp. 615-630.

College Board (2006). *College Board Standards for College Success™: Mathematics and Statistics.*

College Entrance Examination Board (2004). *Course Description: Statistics*. New York: College Board.

Conference Board of the Mathematical Sciences (2001). *The Mathematical Education of Teachers*. Providence, RI, and Washington, DC: American Mathematical Society and Mathematical Association of America.

Conover, W. J. (1999). *Practical Nonparametric Statistics*. John Wiley and Sons, Page 235 (Equation 17).

Consumer Reports (June 1993) Hot dogs. 51(6), 364-367.

Data-Driven Mathematics Series (1998), New York: Pearson Learning (Dale Seymour Publications).

Gnanadesikan, Mrudulla, Richard L. Scheaffer, James M. Landwehr, Ann E. Watkins, Peter Barbella, James Kepner, Claire M. Newman, Thomas E. Obremski, and Jim Swift (1995). *Quantitative Literacy Series,* New York: Pearson Learning (Dale Seymour Publications).

Hollander, Miles and Proschan, Frank (1984). *The Statistical Exorcist: Dispelling Statistics Anxiety*. Marcel Dekker, Pages 83–88 and 121–130.

Holmes, Peter (2001). "Correlation: From Picture to Formula," *Teaching Statistics,* 23(3):67–70.

Kader, Gary (1999). "Means and MADS," *Mathematics Teaching in the Middle School,* 4(6):398–403.

Kader, Gary and Perry, Mike (1984). "Learning Statistics with Technology," *Mathematics Teaching in the Middle School,* 1(2):130–136.

Moore, D. and Cobb, G. (1997). "Mathematics, Statistics, and Teaching," *American Mathematical Monthly*, 104, 801–823.

National Assessment Governing Board (2004). *Mathematics Framework for 2005 National Assessment of Educational Progress*. Available: *www.amstat.org/education/gaise/4*.

National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: The Council.

National Council of Teachers of Mathematics (2002–2004). *Navigating through Data Analysis and Probability Series*. Reston, VA: The Council.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: The Council.

Steen, Lynn, ed. (2001). *Mathematics and Democracy: The Case for Quantitative Literacy*. National Council on Education and the Disciplines. Princeton: Woodrow Wilson Foundation.

U.S. Census Bureau. (2005). *Statistical Abstract of the United States 2004–2005*, Table No. 70. Live Births, Deaths, Marriages, and Divorces: 1950 to 2002.

Utts, Jessica A. (1999). *Seeing Through Statistics*. Pacific Grove, CA: Duxbury, 2nd ed.