

Helpful Comments for Project Entries

7–9 Project Competition

Selecting a Question

Selecting a good question for a statistics project is important. Not only should the question be interesting, it should give rise to data that lend themselves to statistical treatment. For example, if the question leads to a categorical response (i.e., What is your favorite color?), one may be left with nothing more than a few counts (one for each category). This limits both the graphical and statistical analyses that can be used. Be sure the question can be answered with the data collected. Questions need to be stated clearly. If more than one question is posed, each should be answered. Finally, upon completion of the project, it should be reviewed to be certain the question being posed was actually answered.

Collecting Data

Collecting data properly is challenging. Students who find data that have already been compiled often do not realize the pitfalls and potential errors of data collection. As a consequence, they miss an opportunity to understand this vital phase of any project. For this reason, the scoring rubric emphasizes data collection by the students, making projects in which students collect data ‘from scratch’ more highly viewed than those in which students use existing data.

The data collection process should be described clearly, and the student’s role in the data collection should be clear. The variables in the study should be defined clearly in terms of what is to be measured and how. If a random sample is taken, the randomization process should be given. Haphazard or other unplanned sampling is not random sampling and can lead to biased results.

Replication is important in any study. For example, the purpose of a study may be to compare

the growth of a corn plant with and without fertilizer. Suppose two pots are used and two corn seeds are planted in each pot. Then it is randomly determined which pot gets which treatment (fertilizer or no fertilizer). Even though there are two plants under each treatment, there is no replication. The reason for this is that treatments were assigned randomly to pots (not plants). More than one pot would have to be used for each treatment for there to be true replication.

If a survey is conducted, a copy of the survey should be included in an appendix. For all projects, raw data should be included as an appendix.

Graphs

Graphical displays provide insights into data. Many projects fail to take advantage of this important statistical tool. In projects using at least one graphical display, the graphs often are only the most rudimentary pie and bar charts. Stem-and-leaf, dot plots, box plots, and scatter plots are some of the methods that might provide more insight into the data. Displaying sample means with error bars also may be helpful. Care should be taken to use appropriate graphs. For example, line plots and scatter plots are used sometimes when bar charts would be better. Replication permits variability to be captured by the data; appropriate graphs make it visible.

Inference

If data are collected on all members in the population, a census is taken. Because inferential methods are used to draw conclusions about the population based on the sample, these methods are inappropriate if all population values have been observed. For example, parameters can be computed and do not need to be estimated. However, some thought should be given to whether a census actually is achieved. If the goal

is to survey everyone in a school, some students may be absent or refuse to respond.

When a sample is drawn, inferential statistics are usually needed to answer a question. While useful, graphs alone are not sufficient at this level of competition. Estimates of the spread and the center of the distribution are important. Students should understand fully the methods they use. Sometimes, the conclusions do not follow from the analysis presented. It is better to use informal (but appropriate) methods correctly than to apply more sophisticated procedures improperly.

Presentation

Font size should be at least 12 pt., and complete sentences and standard grammar should be used. The emphasis in writing should be on the statistical aspects of the study. Background information should lead to a precise statement of the question to be considered. Some projects benefit from a more detailed description of the data collection phase of the study. Details of the statistical analysis should be presented. The statistical methods should be clearly outlined and discussed. The analysis should serve as the foundation for any conclusions drawn.

A “reflection on the process” should be a realistic self-evaluation of the work. Simply stating all went well raises concerns, as few studies ever have everything go right.

10–12 Project Competition

Selecting a Question

Selecting a good question for a statistics project is important. Not only should the question be interesting, it should give rise to data that lend themselves to statistical treatment. For example, if the question leads to a categorical response (i.e., What is your favorite color?), one may be left with nothing more than a few counts (one for each category). This limits both the graphical and statistical analyses that can be used. Be

sure the question can be answered with the data collected. Questions need to be stated clearly. If more than one question is posed, each should be answered. Finally, upon completion of the project, it should be reviewed to be certain the question being posed was actually answered.

Collecting Data

Collecting data properly is challenging. Students who find data that have already been compiled often do not realize the pitfalls and potential errors of data collection. As a consequence, they miss an opportunity to understand this vital phase of any project. For this reason, the scoring rubric emphasizes data collection by the students, making projects in which students collect data ‘from scratch’ more highly viewed than those in which students use existing data.

The data collection process should be described clearly, and the student’s role in the data collection should be clear. The variables in the study should be defined clearly in terms of what is to be measured and how. If a random sample is taken, the randomization process should be given. Haphazard or other unplanned sampling is not random sampling and can lead to biased results.

Replication is important in any study. For example, the purpose of a study may be to compare the growth of a corn plant with and without fertilizer. Suppose two pots are used and two corn seeds are planted in each pot. Then it is randomly determined which pot gets which treatment (fertilizer or no fertilizer). Even though there are two plants under each treatment, there is no replication. The reason for this is that treatments were assigned randomly to pots (not plants). More than one pot would have to be used for each treatment for there to be true replication.

If a survey is conducted, a copy of the survey should be included in an appendix. For all projects, raw data should be included as an appendix.

Graphs

Graphical displays provide insights into data. Many projects fail to take advantage of this important statistical tool. In projects using at least one graphical display, the graphs often are only the most rudimentary pie and bar charts. Stem-and-leaf, dot plots, box plots, and scatter plots are some of the methods that might provide more insight into the data. Displaying sample means with error bars also may be helpful. Care should be taken to use appropriate graphs. For example, line plots and scatter plots are used sometimes when bar charts would be better. Replication permits variability to be captured by the data; appropriate graphs make it visible.

Inference

If data are collected on all members in the population, a census is taken. Because inferential methods are used to draw conclusions about the population based on the sample, these methods are inappropriate if all population values have been observed. However, some thought should be given to whether a census actually was achieved. If the goal was to survey everyone in a school, some students may be absent or refuse to respond.

When a sample is drawn, inferential statistics usually are needed to answer a question. While useful, graphs and descriptive statistics alone are not sufficient in this instance. When using formal inferential statistical tests, the assumptions for any method should be checked. For example, variances should not be pooled if they are substantially different (which can be tested) and the sample sizes are reasonably large. Students should fully understand the methods they use, otherwise inappropriate statistical terminology may be used. It is better to use simpler (but appropriate) methods correctly than to apply more sophisticated procedures improperly.

Care should be taken to state the null and alternative hypotheses appropriately for hypothesis

tests. Remember that in a subject-matter area, the hypothesis is what the researcher wants to prove. In statistics, this usually becomes the alternative hypothesis, as the strongest conclusions can be drawn from rejecting the null in favor of the alternative. Note the null hypothesis is never 'accepted.' Instead, it is traditional to say "we failed to reject the null hypothesis," which gives the proper impression that it is not known with certainty that the null is true but that the data do not refute it. The reason for this is the probability of a type II error is not known.

Confidence intervals can be misinterpreted. For example, a confidence interval cannot confirm a test statistic because the test statistic is, by construction, the center of any confidence interval. Note that r^2 represents the amount of variability in the response variable explained (removed) by the explanatory variable, not the fraction of the response variable explained.

Presentation

Font size should be at least 12 pt., and complete sentences and standard grammar should be used. The writing emphasis should be on the statistical aspects of the study. Background information should lead to a precise statement of the question to be considered. Some projects benefit from a more detailed description of the data collection phase. Details of the statistical analysis should be presented. The statistical methods should be outlined and discussed clearly. The analysis should serve as the foundation for any conclusions drawn.

A "reflection on the process" should be a realistic self-evaluation of the work. Simply stating that all went well raises concerns, as few studies ever have everything go right.