

APPENDIX D: EXAMPLES OF NAKED, REALISTIC, AND REAL DATA

Naked Data (not recommended)

Find the least squares line for the data below. Use it to predict Y when $X=5$.

X	1	2	3	4	6	8
Y	3	4	6	7	14	20

Critique: Made-up data with no context (not recommended). The problem is purely computational with no possibility of meaningful interpretation.

Realistic Data (better, but still not the best)

The data below show the number of customers in each of six tables at a restaurant and the size of the tip left at each table at the end of the meal. Use the data to find a least squares line for predicting the size of the tip from the number of diners at the table. Use your result to predict the size of the tip at a table that has five diners.

Diners	1	2	3	4	6	8
Tip	\$3	\$4	\$6	\$7	\$14	\$20

Critique: A context has been added that makes the exercise more appealing and shows students a practical use of statistics.

Real Data (recommended)

The data below show the quiz scores (out of 20) and the grades on the midterm exam (out of 100) for a sample of eight students who took this course last semester. Use these data to find a least squares line for predicting the midterm score from the quiz score. Assuming the quiz and midterm are of equal difficulty this semester and the same linear relationship applies this year, what is the predicted grade on the midterm for a student who got a score of 17 on the quiz?

Critique: Data are from a real situation that should be of interest to students taking the course.

Quiz	20	15	13	18	18	20	14	16
Midterm	92	72	72	95	88	98	65	77