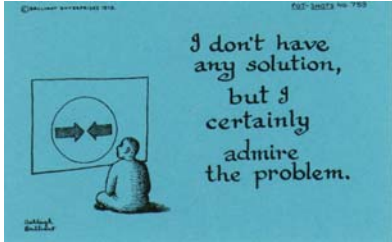


Midterm I Review



Chapters 1-2

- Randomized controlled experiments
- Observational studies
- Confounding factors

A study of baseball players shows that left-handed players have a higher death rate than right-handed ones. A critic of the study explained this as “due to confounding: baseball players are more likely to be left-handed than the general population, and the players also have higher death rates.”

This is a (circle one):
controlled experiment observational study

Is this a case of confounding and is it a good explanation for the data? Answer yes or no and explain briefly.

Circle one: *Yes* *No*

Histograms

In a density histogram, the area of each block is equal to the percentage of observations in that class interval.

Symmetric and skewed histograms

The following table gives (hypothetical) data on the prices of single family residences in Berkeley that sold during one year.

Price	Percent
\$100,000 --\$200,000	40
\$200,000--\$250,000	30
\$250,000--\$300,000	20
\$300,000--\$400,000	6
\$400,000--\$600,000	4

(a) Draw a histogram for this distribution.

(b) Is the normal curve a good approximation for this histogram?

The Average, the Median, and the Standard Deviation

- The average and median indicate the center
- The SD indicates the spread
- Definitions of each
- Properties of the SD
- Rules of thumb for the SD

(c) The median price is (circle one):

- Less than \$200,000
- Between \$200,000 and \$250,000
- Greater than \$250,000
- Can't tell

Price	Percent
\$100,000 --\$200,000	40
\$200,000--\$250,000	30
\$250,000--\$300,000	20
\$300,000--\$400,000	6
\$400,000--\$600,000	4

(d) The average price is (circle one):

- Less than the median
- Equal to the median
- Greater than the median

The Normal Curve

- Standard units and finding areas under the normal curve
- Finding percentiles under the normal curve
- Approximating histograms with the normal curve

An aerobic study involves 1,040 men, age 18 to 24. The histogram of systolic blood pressure for these men followed the normal curve closely. The average is 120 mm, and 600 of the men have blood pressures between 110 and 130 mm. How many men have blood pressure between 115 and 125 mm?

An aerobic study involves 1,040 men, age 18 to 24. The histogram of systolic blood pressure for these men followed the normal curve closely. The average is 120 mm, and 600 of the men have blood pressures between 110 and 130 mm. Only about 5% of the men have blood pressure greater than _____.

Correlation and Regression

Chapters 8 and 9: Correlation

- scatter plots
- correlation coefficient & its properties
- common errors (eg correlation and causation).
- ecological correlation

Chapters 10 and 11: Regression: estimating the value of one variable based on the value of another.

- The regression line, the graph of averages and SD line.
- Residuals and the RMS error.
- The normal curve can be used in vertical strips if the scatter diagram is football shaped.

A study is done of the correlation of income and education for a large number of families in all the 50 states. Correlation is computed in two different ways. In the first method, the average income and the average education are determined for each state and then the correlation of the resulting 50 pairs of numbers is calculated. In the second method, the correlation of income and education is calculated for each state separately and then the 50 correlation coefficients are averaged to produce an overall correlation. Which method produces the larger correlation? Explain your answer briefly.

A large study of human growth obtained the following results for males:

average height at age 4 = 41 inches

SD = 1.5 inches

average height at age 18 = 70 inches

SD = 2.5 inches

Correlation = 0.75

(a) Predict the height at 18 for someone in the study who was 38 inches tall at age 4.

average height at age 4 = 41 inches

SD = 1.5 inches

average height at age 18 = 70 inches

SD = 2.5 inches

Correlation = 0.75

(b) Of those in the study who were 38 inches tall at age 4, what percent had added more than 30 inches to their height by age 18?

As part of a longitudinal study of personality change, 2500 normal subjects were given a psychological test twice; once at age 21, and the second time at age 35. At both ages the average score was 50, and the SD was 10. There was a strong positive association between the two scores.

For each of the 2500 subjects, the investigators used the regression method to predict score at age 35 from score at age 21. The predicted score turned out to be within 10 points of the actual score for _____ of the 2500 subjects.

Check the best option below to fill in the blank. Explain your choice.

_____ more than 68%

_____ around 68%

_____ less than 68%

As part of a longitudinal study of personality change, 2500 normal subjects were given a psychological test twice; once at age 21, and the second time at age 35. At both ages the average score was 50, and the SD was 10. There was a strong positive association between the two scores.

An individual scored 38 at age 21. At age 35, you would expect him to score (circle one):

- Less than 38
- About 38
- Higher than 38

Midterm scores in a large statistics class have an average of 60 and an SD of 15. The final exam scores in this class have an average of 56 and an SD of 17. The correlation between the midterm and final exam scores is 0.6, and the scatter diagram of the two variables is football shaped.

Students complain that the exams were too hard, so the professor gives each student 5 more points on the midterm and 10 more points on the final.

(a) Find the average and SD of the new midterm scores, as well as those of the new final scores.

(b) If possible, find the correlation between the new midterm scores and the new final scores, and justify your answer. If it is not possible to find the correlation, explain why not.
