

Midterm 1, September 24, 1:30 p.m. - 2:45 p.m.

*Show your work. The test is out of 100 points and you have 75 minutes.*

1. (6 points) The following information came from **YAHOO! NEWS**, September 14 2009:

CHICAGO Even in the “oldest old,” a little physical activity goes a long way, extending life by at least a few years for people in their mid- to late 80s, Israeli researchers found.

The three-year survival rate was about three times higher for active 85-year-olds compared with those who were inactive. Getting less than four hours of exercise weekly was considered inactive; more than that was active.

The study involved 1,861 Jerusalem residents who were 70 years old in 1990. Participants filled out questionnaires about their health and activity levels through 2008.

In the first sentence the researchers suggest that physical activity causes people to live longer (“extends life”). Suggest a plausible confounding factor and clearly explain why it might make you doubt their conclusion.

2. The following information comes from **CNN**, September 24 2009:

A vaccine to prevent HIV infection has shown modest results for the first time, researchers have found.

In what is being called the world’s largest HIV vaccine trial ever, researchers found that people who received a series of inoculations of a prime vaccine and booster vaccine were 31 percent less likely to get HIV, compared with those on a placebo.

- (a) (2 points) Was the study a controlled experiment or an observational study? How do you know?
- (b) (1 point) What is a placebo?
- (c) (6 points) Give 2 reasons why a placebo is used in a study like this one.

3. A realtor has information on the average listing price and the average selling price of homes in each of the 50 states. She plots these 50 points and computes the correlation coefficient, which turns out to be 0.8.

(a) (5 points) If she computed the correlation coefficient between the individual listing prices and selling prices of homes in the whole country, it would be (choose one, no explanation is required):

- i. smaller than 0.8.
- ii. close to 0.8.
- iii. larger than 0.8.

(b) In each case, say whether the statement is TRUE or FALSE. No explanation is required.

- i. (2 points) The regression effect says that expensive houses will tend to sell for less than their listing price and inexpensive houses will tend to sell for more than their listing price.
- ii. (2 points) The average listing price in Idaho is \$338,000. If we knew the 5-number summary for the scatterplot, we could use regression to predict the average selling price in Idaho.
- iii. (2 points) The correlation coefficient tells us that the 50 points on the scatterplot are football shaped.
- iv. (2 points) The correlation coefficient tells us that states with higher average listing prices tend to have higher average selling prices, on average.
- v. (2 points) The correlation coefficient tells us that a house sells for approximately 80% of its listing price, on average.

4. (8 points) At the present time, there are almost 30,000 homes for sale in Utah and the average listing price of these homes is \$453,000. However, the homes that are for sale in Summit County have an average listing price of \$1,464,000, which is almost double the average listing price in any other County in Utah. If these homes in Summit County were removed, the average for the remaining Utah homes would be (choose one, no explanation is required):

- (a) somewhat less than \$453,000
- (b) exactly \$453,000
- (c) between \$453,000 and \$1,464,000
- (d) somewhat more than \$453,000

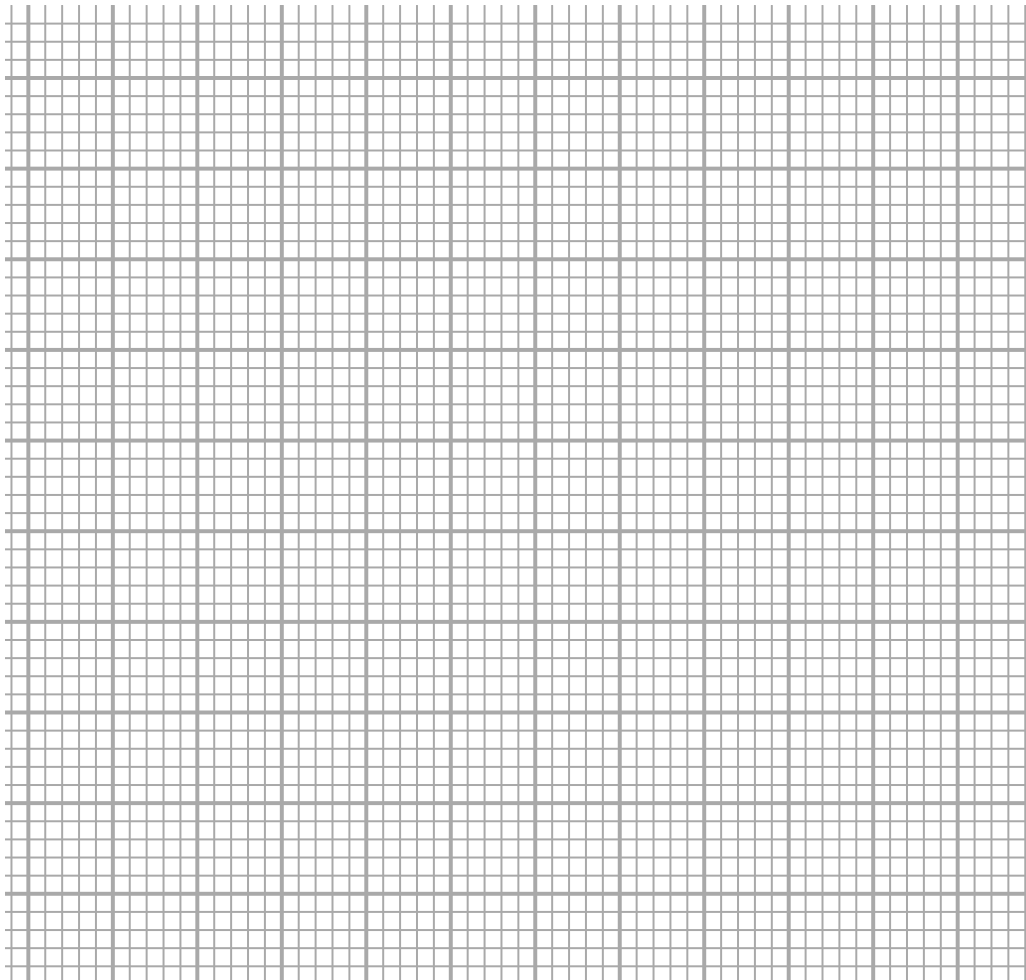
and we would expect the SD to be (choose one, no explanation is required):

- (a) exactly the same as it was before.
- (b) somewhat larger than it was before.
- (c) somewhat smaller than it was before.

5. The following information comes from the June 2007 issue of “Genome Technology”. As usual, all intervals include the left endpoint but not the right.

Number of Years in Research	Percentage of Respondents
0–5	11
5–7	10
7–10	17
10–15	21
15–20	24
20–30	17

(a) (8 points) Draw a histogram. Be sure to label the axes.



(b) (2 points) Which interval contains the median?

6. In November 2008, unemployment rates for the 50 States followed the normal curve with an average of 6.1 and an SD of 1.6.

(a) (10 points) What percentage of the States had unemployment rates between 7.0 and 8.0?

(b) (10 points) What percentage of the States had unemployment rates higher than 4.0?

(c) (10 points) If I tell you that Oklahoma was at the 20th percentile, what was the unemployment rate in Oklahoma?

7. In a study on female athletes, researchers measured the number of 60-pound bench presses each woman performed (before fatigue), obtaining an average of 11.0, with an SD of 6.0. They also measured the maximum weight each woman could bench press, obtaining an average of 79.9 pounds with an SD of 14.4 pounds. The correlation between the number of 60-pound bench presses and the maximum weight the women could bench press is 0.8. The scatter diagram is football-shaped.

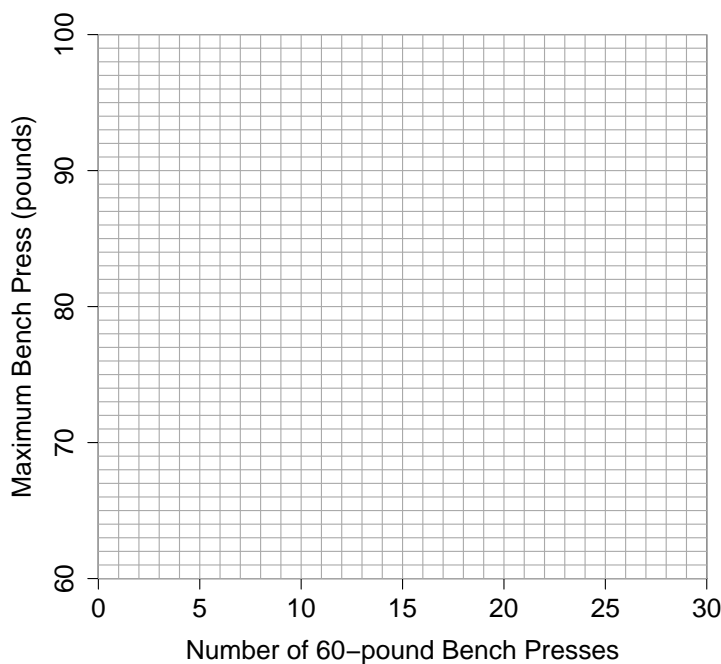
(a) (6 points) Find the equation of the regression line for predicting the maximum weight a female athlete can bench press from the number of 60-pound bench presses she can do.

(b) (6 points) Estimate the maximum weight a female athlete can bench press if she can do 16 60-pound bench presses.

(c) (2 points) Put a give-or-take number on your answer in part (b).

(d) (4 points) Would you be surprised if someone told you the woman in part (b) could lift a maximum of 80 pounds? Explain.

(e) (4 points) Draw the regression line on the scatter diagram provided below.



Please note that these are provided for your convenience, but it is your responsibility to know how and when to use them.

$$\text{height} = \frac{\text{percentage}}{\text{width}}$$

$$z = \frac{x - \text{ave}}{SD}$$

$$x = \text{ave} + z(SD)$$

$$\text{rms error} = \sqrt{1 - r^2}(SD_Y)$$

$$\text{slope} = r \left( \frac{SD_Y}{SD_X} \right)$$

$$\text{intercept} = \text{ave}_Y - (\text{slope})(\text{ave}_X)$$