Stat 1040, Fall 2009   Name: _____

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. — women live longer, women exercise more!

   diet, lifestyle, gender, health at the start

   eg. people with a healthy lifestyle tend to exercise & do other things (eat right, not smoke, get checkups) & maybe these are the things that make them live longer.

   eg. people who are sick to begin with can't exercise & their illness causes them to die younger.

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?

   It was a controlled experiment because the researchers determined who got the vaccine & who got the placebo.

   (b) (1 point) What is a placebo?

   Something that resembles the treatment (in this case, a shot) but lacks the active ingredient.

   (c) (6 points) Give 2 reasons why a placebo is used in a study like this one.

   So that the subjects don't know who is in the treatment group.
   "   "   doctors   "   "   "   "   "   "

   So that people don't change their behavior because of which group they are in.
   To try to make sure psychosomatic effects are the same for both groups.

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. Due to the fact that she was originally
ii. close to 0.8. using an ecological correlation (artificially
iii. larger than 0.8. strong).

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

F 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. They can all sell for less than listing price.

T 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.

F iii. (2 points) The correlation coefficient tells us that the 50 points on the scatterplot are football shaped. We need to see the plot.

T iv. (2 points) The correlation coefficient tells us that states with higher average listing prices tend to have higher average selling prices, on average. and vice versa.

F v. (2 points) The correlation coefficient tells us that a house sells for approximately 80% of its listing price, on average. It's a measure of how tight the points are to a line, but it doesn't control what the line is.

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 We took out some expensive ones

(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. We took out unusually large ones.

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

Width	Number of Years in Research	Percentage of Respondents	height
5	0–5	11	$\frac{11}{5} = 2.2$
2	5–7	10	$\frac{10}{2} = 5$
3	7–10	17	$\frac{17}{3} = 5.7$
5	10–15	21	$\frac{21}{5} = 4.2$
5	15–20	24	$\frac{24}{5} = 4.8$
10	20–30	17	$\frac{17}{10} = 1.7$

(a) (8 points) Draw a histogram. Be sure to label the axes.
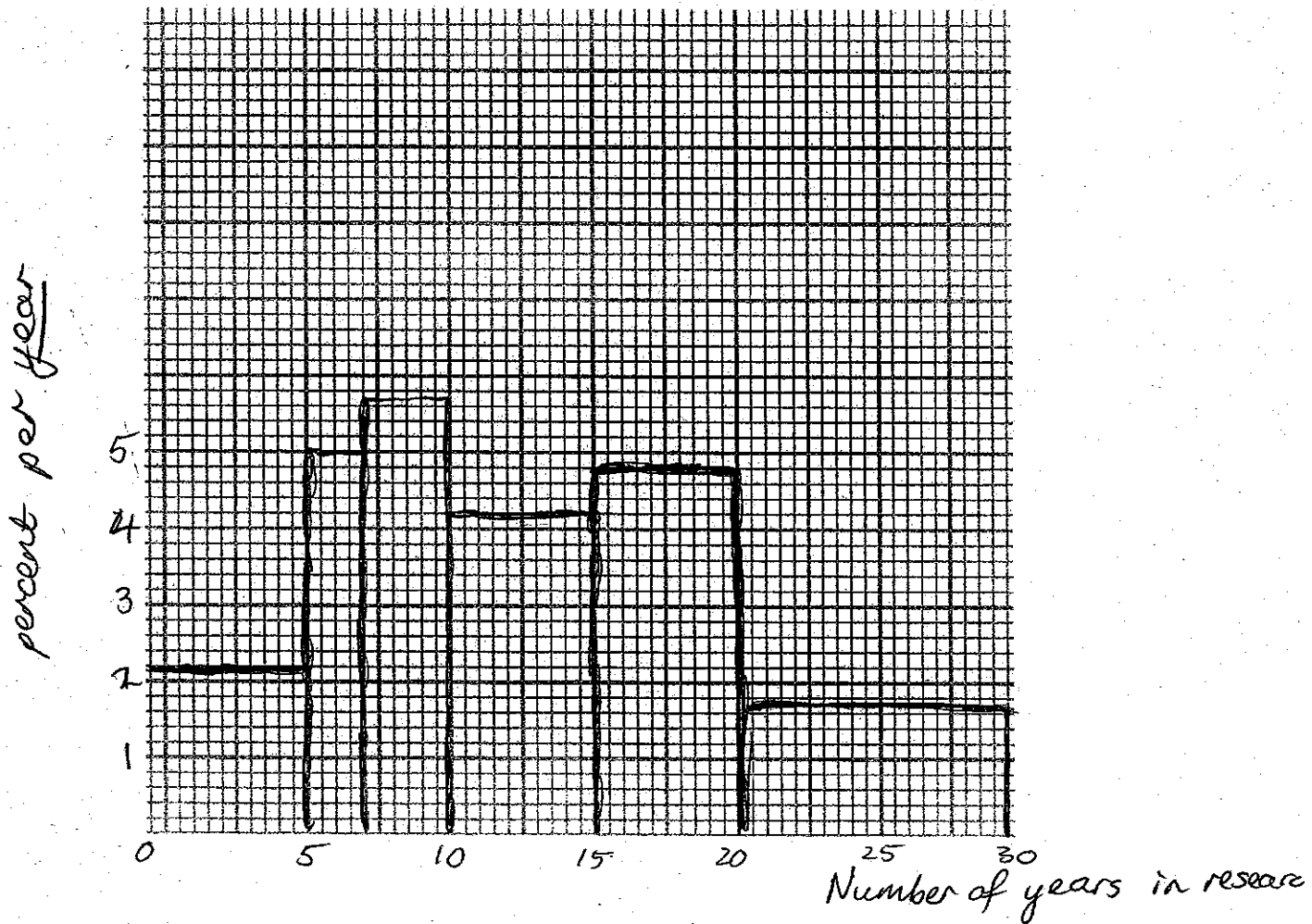
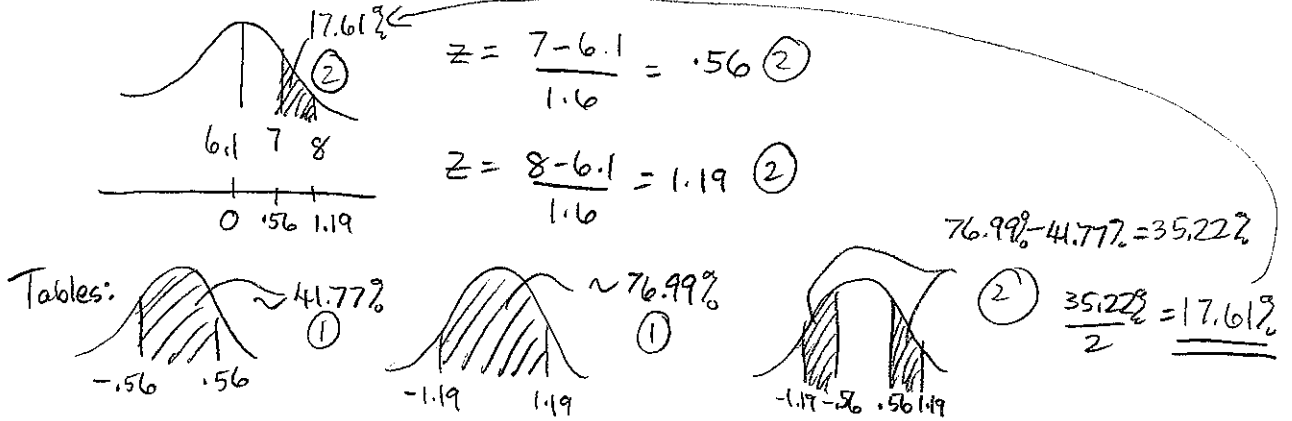(b) (2 points) Which interval contains the median?

11% less than 5
11 + 10 = 21% less than 7
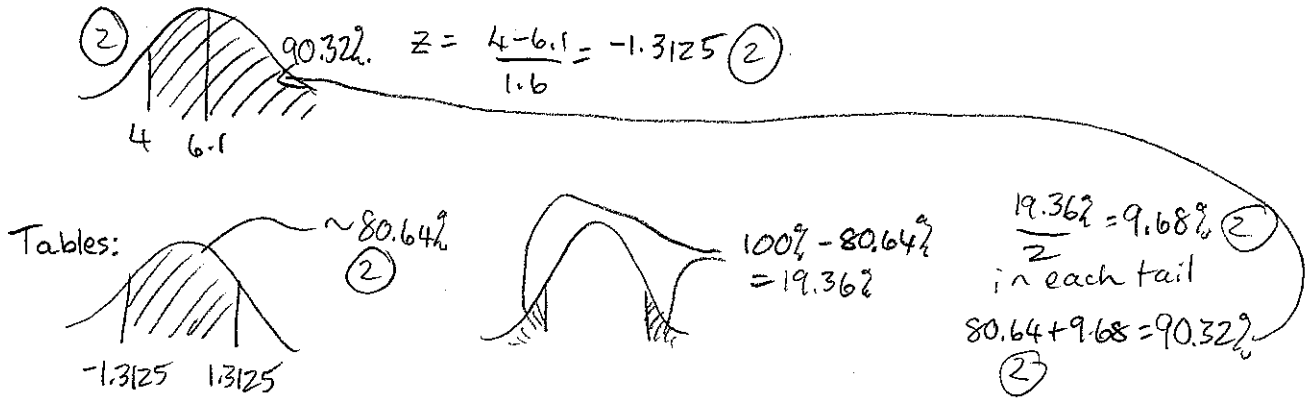21 + 17 = 38% less than 10
38 + 21 = 59% less than 15

The median is in the 10–15 year interval.

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?

$$z = \frac{7-6.1}{1.6} = .56$$ ②

$$z = \frac{8-6.1}{1.6} = 1.19$$ ②

17.61%

6.1 7 8

0 .56 1.19

Tables: ~41.77% ①   ~76.99% ①

−.56 .56   −1.19 1.19   −1.19 −.56 .56 1.19

$76.99\% - 41.77\% = 35.22\%$

② $\frac{35.22\%}{2} = 17.61\%$

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

② 90.32% $z = \frac{4-6.1}{1.6} = -1.3125$ ②

4 6.1

Tables: ~80.64% ②

−1.3125 1.3125

$100\% - 80.64\% = 19.36\%$

$\frac{19.36\%}{2} = 9.68\%$ ② in each tail

$80.64 + 9.68 = 90.32\%$ ②

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

20% ②   so   20%   20%

Middle = $100\% - 40\%$
      = 60% ②

Tables: 60% ②   → $z = -.85$ ②

z
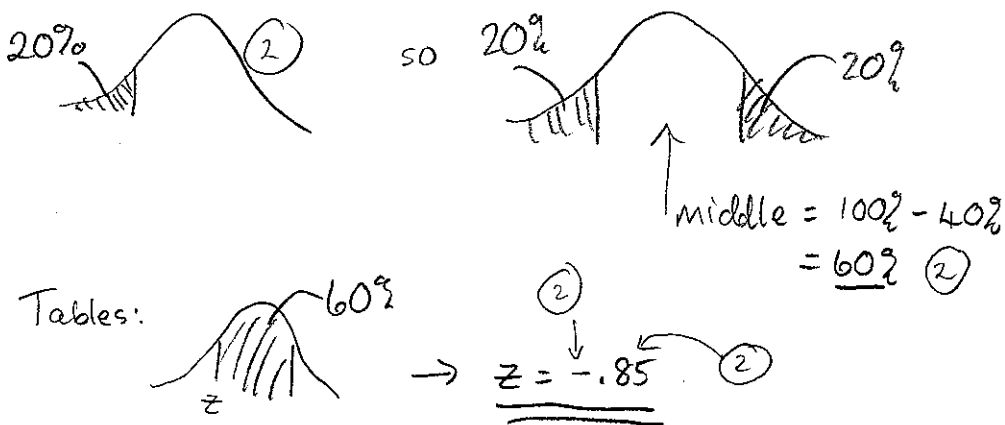
$X = ave + (-.85)(SD) = 6.1 + (-.85)(1.6) = 7.46$ ②

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.

$$slope = r\left(\frac{SD_y}{SD_x}\right) = .8\left(\frac{14.4}{6.0}\right) = 1.92$$

$$intercept = ave_y - slope(ave_x) = 79.9 - 1.92(11) = 58.78$$

$$\left. \right\} y = 58.78 + 1.92x$$

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

$$Y = 58.78 + 1.92(16)$$
$$= 89.50$$

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

$$rms = \left(\sqrt{1-r^2}\right)(SD_y) = \left(\sqrt{1-.8^2}\right)(14.4) = 8.64 \qquad \text{wrong SD} \rightarrow -1$$

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

$\pm 2(SD) \rightarrow -2$

$\pm 1\ rms \rightarrow -1$

$$89.50 - 2(8.64) = 72.22$$
$$89.50 + 2(8.64) = 106.78$$
$\left. \right\}$ 80 is in this interval so we would **not** be surprised.

(e) (4 points) Draw the regression line on the scatter diagram provided below.
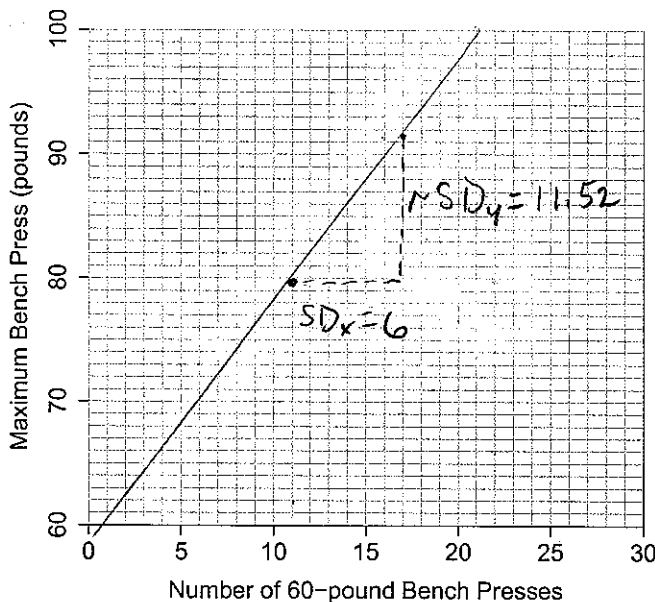
$SD_x = 6$

$r\ SD_y = .8(14.4)$
$\qquad = 11.52$

↑

missed the r (so they correctly drew the SD line)
$\rightarrow -2.$

Number of 60−pound Bench Presses