Lab 6: Chapter 4

1. A teacher decided to calculate z-scores for the scores (out of 100 points) on a test that students earned on an exam. The mean score on the exam was 84 points, and the standard deviation was 6.0 points. The teacher also calculated the z-scores for the points earned by students on homework assignments, which had a total of 250 points. The points earned on homework had a mean of 217 with a standard deviation of 9.8 points.

Student	x = homework points	z-score for x	y = exam score	z-score for y
1	181	-3.67	75	-1.50
2	144	-7.44	73	-1.83
3	212	-0.51	94	1.66

For each of the following pairs of variables in the next 3 exercises, identify which is likely to represent the independent (x) variable and which represents the dependent (y) variable.

2. Variable 1: Days without smoking for a participant in a cessation program y
   Variable 2: Number of sessions attended by a smoker in a cessation program x

3. Variable 1: Time spent by a student studying for a final exam x
   Variable 2: Score on the same final exam y

4. Variable 1: Annual salary for a baseball player y
   Variable 2: Number of home runs hit per season by a baseball player x

For each of the following pairs of variables, indicate whether you would expect a positive correlation, a negative correlation, or a correlation close to 0. Explain your choice.

What happens to y when x increases?

① If y decreases, then there is a negative correlation

② If y increases, then there is a positive correlation

③ If neither ① or ② seems correct, the correlation is close to 0

5. x = daily hours of sunlight (in minutes) — positive
   y = daily growth of plants (in mm)

6. x = number of wolves per square mile
   y = number of elks per square mile — negative

7. x = height of a student in high school
   y = grade point average for a student in high school — close to zero

8. Interest rate and number of loan applications — negative

9. Height and IQ — pos.

10. Height and shoe size — pos.

11. Minimum daily temperature and cooling cost — pos.

(1)

$\bar{x} = 217$	$\bar{y} = 84$
$s_x = 9.8$	$s_y = 6$
homework	exam

Use $z_x = \frac{x - \bar{x}}{s_x}$ and $z_y = \frac{y - \bar{y}}{s_y}$

Student 1 $z_x = \frac{181 - 217}{9.8} = -3.67$

$z_y = \frac{75 - 84}{6} = -1.50$

Student 2 $z_x = \frac{144 - 217}{9.8} = -7.44$

$z_y = \frac{73 - 84}{6} = -1.83$

Student 3 $z_x = \frac{212 - 217}{9.8} = -0.51$

$z_y = \frac{94 - 84}{6} = 1.66$

For each of the lines below, identify values for the slope (b) and y-intercept (a).

	slope	y-intercept
12. $y = 98.9 + 2.3x$	2.3	98.9
13. $y = 0.727 + 0.178x$	0.178	0.727
14. $y = 23 + 4x$	4	23

15. Use the information in the table below to answer parts a through f. The ages (in years) of 10 men and their systolic blood pressures (in millimeters of mercury) are listed in the table.

minimum maximum

Age, x	16	25	39	45	49	64	70	29	57	22
Systolic Blood Pressure, y	109	122	143	132	199	185	199	130	175	118

(a) What is the sample correlation coefficient, r?

(b) Describe the type of correlation

(c) Interpret the meaning of the correlation in the context of the data.

(d) Find the equation that represents the least squares regression line for the data.

(e) Use the regression equation to predict the value of y for $x = 42$.

(f) Use the regression equation to predict the blood pressure for a man aged 67.

(g) Use the regression equation to predict the blood pressure for a man aged 80.

Since the correlation is significant, the regression line is a good predictor of y for x values between 16 and 70 and including years.

16. Use the information in the table below to answer parts a through g. The square footages and sale prices (in thousands of dollars) of seven homes are shown in the table at the left. (*Source: Howard Hanna*)

Square footage, x	1924	1592	2413	2332	1552	1312	1278
Sale price, y	174.9	136.9	275.0	219.9	120.0	99.9	145.0

(a) What is the sample correlation coefficient, r?

(b) Describe the type of correlation

(c) Interpret the meaning of the correlation in the context of the data.

(d) Find the equation that represents the least squares regression line for the data.

(e) Use the regression equation to predict the home's sale price when the square footage is 1450 square feet.

(f) Use the regression equation to predict the home's sale price when the square footage is 2720 square feet.

(g) Use the regression equation to predict the home's sale price when the square footage is 2175 square feet.

(15) (a) $r = 0.908$

(b) There is a strong, positive correlation between age and systolic pressure

(c) As age increases, systolic blood pressure increases.

(d) $\hat{y} = 80.25 + 1.71x$

(e) $\hat{y} = 80.25 + 1.71 \cdot (42) = 152$

(f) $\hat{y} = 80.25 + 1.71(67) = 195$

(g) Since the correlation is significant, we can use the regression line to predict systolic blood pressure amounts for ages 16 (the minimum x value) through 70 (the maximum x value). We cannot predict a y value for x values outside of this range, therefore we cannot predict a y value for $x = 80$.

Don't extrapolate!

(16) (a) $r = 0.92$

(b) There is a strong, positive correlation between square footage and price.

(c) As square footage increases, price also tends to increase.

(d) $\hat{y} = -50.02 + 0.12x$

(e) $\hat{y} = -50.02 + 0.12(1450) = 123.98$

(f) Since the correlation is significant, we can use the regression line to predict sale price for square footages between and including 1278 and 2413. Since 2720 is outside this range, it is not meaningful to use the regression line to make a prediction here.

(g) $\hat{y} = -50.02 + 0.12(2175)$