Stats 10
Homework 3
Yongqian Li
004997466
1A

(Part B)

4 2. a. The number of square feet in the home, because there are less scatter in part B than in part A.

b. By knowing the number of square feet, because there is a stronger relationship between the number of square feet and the value than the number of fireplaces and the value.

6. The negative trend shows that there is a negative correlation between and start year: Salary. A point in the top-left corner of this scatter plot represent the professor who has the highest salary and the start year is 1985; a point near the bottom-right corner represents the professor who receives the lowest salary and whose start year is 2003.

10. It has very little trend, because there is a lot of random scatter.

14. No, it would not make sense to find the correlation for this data set because the trend is obviously not linear. According to the graph, at approximately the age of 27 the highest fertility rate occur.

18. -0.903 B
0.374 A
0.777 C

30. a. The independent variable is Mother's Height; the dependant variable is Daughters Height.

b. Predicted Daughter's Height = 29.92 + 0.5417 · Mother's Height
= 29.92 + 0.5417 · 60
= 62.422

b. From the graph, the predicted Daughter's Height is about 62 inches.

d. The slope implies that for each additional inch in mother's Weight, there is on average 0.5417 inch additional daughter's height.

e. There are many other factors such as father's height and family's income.

32. a. from the graph, the predicted price is about 320 thousand dollars.

b. Predicted Price = -11.77 + 0.1146 Sqft
= -11.77 + 0.1146 · 3000
= 332.03 (thousand dollars)

34. a. Predicted Foot = 5.67 + 0.998 Hand

b. $b = r\frac{s_y}{s_x}$
$= 0.948 \cdot \frac{1.230}{1.168}$
≈ 0.998

c. $a = \bar{y} - b\bar{x}$
$= 23.318 - 0.998 \cdot 17.682$
≈ 5.67

d. $\hat{y} = 5.67 + 0.998x$
$= 5.67 + 0.998 \cdot 18$
$= 23.634$ (cm)

40. estimation: $r \approx -1$ and the correlation is negative. The reason is that all points in the graph almost fall on the same straight line.

42. a. x is the Weights of nuggets of gold and y is their market value.
Reason: We want to predict the value based on the weight given.

b. x is the time and y is the concentration of Chlorine in the pool.
Reason: We want to predict the Chlorine concentration in the pool based on the time known.

c. y is the age of Oak trees and x is the circumference of oak trees.
Reason: We want to predict the age of ~~Oak~~ Oak trees based on the circumference given.

54. a. Extrapolation is to make a prediction for a x value outside the range of the data.
It is a bad idea because the linear model may no longer hold outside that range.

b. The closer the coefficient of correlation is to 1 or -1, the stronger the correlation is. When the coefficient is in [-1, 0), the correlation is negative; when it is in (0, 1], the correlation is positive.
It shows the strength of linear association between two numerical variables.

c. No, because correlation does not imply causation.

68. a. The trend is positive. It means that as number of people per household increases, the weight of trash also increases.

b. $r^2 = 76.9\% = 0.769$
$r = \sqrt{r^2} = \sqrt{0.769} \approx 0.877$

c. ~~$b_1 = \frac{s_y}{s_x} \cdot 0.877$~~
The slope is 11.30 (Pounds/person)
This means that for each additional person in the house, there are on average 11.30 additional pounds of ~~trash~~ trash.

d. It is not appropriate to interpret the ~~the~~ intercept because it does not make sense to have a ~~family~~ house with 0 people
The slope is only used to adjust the height of the line in this case.

70. The regression is not appropriate because most data are in the left side of the graph, and ~~the~~ the existance of a few outliers have a heavy influence on the regression. If the outliers are removed, the distribution is not linear. Therefore, ~~it is impossible for~~ We can not raise life expectancy by buying more TVs.