MATH 11: Discussion Week 2

Apr. 2019

The handouts and solution could be found at `https://kimukook.github.io/teaching/math11sp19/`

Properties of Mean and Variance

1. Suppose the temperatures C measured in Celcius in March has mean and variance as $m(C) = 12$ and $\text{var}(C) = 4$. What is the corresponding mean and variance of C transformed in Farenheit? *(Hint: $F = \frac{9}{5}C + 32$.)*

2. ___ In a meeting of a group of executives, the mean salary of the participants is $100, 000$, the median salary is $80, 000$, the upper quartile is $125, 000$, and the lower quartile is $60, 000$. Which of the following gives the best description of the distribution of salaries:
a) The distribution is most likely approximately symmetric.
b) The distribution is most likely skewed to the left.
c) The distribution is most likely skewed to the right.
d) The distribution is most likely bimodal.

Correlation and linear regression

1. Correlation: A statistic that measures the Strength and Direction of a linear association (Form) between two quantitative variables where no Outliers are present.

(1) Determine which of the following statements about the correlation are **true** or **false**:

$$Hint : r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

- ___ The range of correlation r is $-1 \leq r \leq 1$.

- ___ The $+$ sign of correlation r indicates the positive associations.

- ___ The stronger association occurs when r is closer to 1 or -1.

- ___ The choice of predictor/response doesn't matter, i.e. $\text{cor}(X, Y) = \text{cor}(Y, X)$.

- ___ The correlation is unaffected by linear scale changes, i.e. $\text{cor}(X, Y) = \text{cor}(X, 2Y) = \text{cor}(X, Y + 5)$.

- ___ It is improper to measure correlation when even one single outlier is present.

- ___ It is improper to measure correlation for associations that are non-linear.

(2) Could you provide two examples to illustrate that correlation is not causation?
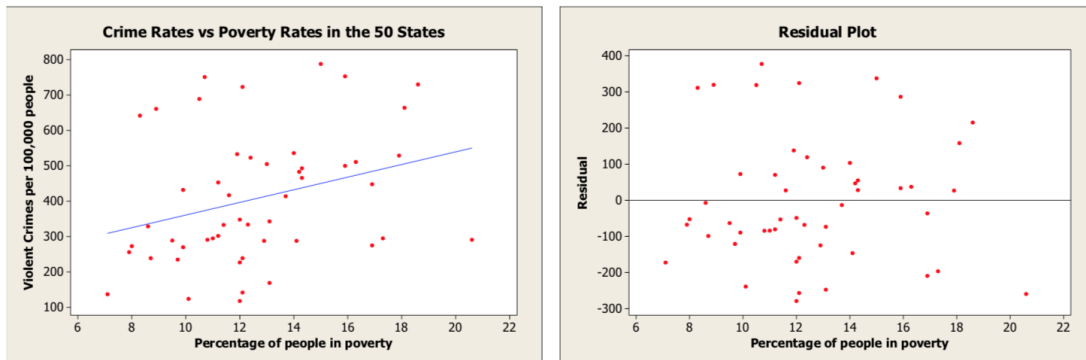
2. Linear regression

We have data on the poverty rate (percentage of people living in poverty) and crime rate (number of violent crimes per 100,000 people) in the 50 states. Below is some Minitab output for a regression in which the crime rate is the response variable and the poverty rate is the explanatory variable:

Regression Analysis: Crime versus Poverty

The regression equation is Crime $= 182 + 17.9$ Poverty

$S = 176.889$, R-Sq $= 9.1\%$, R-sq(adj) $= 7.2\%$.

(1) What would we predict to be the crime rate in a state where 16 percent of people live in poverty?

(2) Is linear regression an appropriate way to predict the crime rate from the poverty rate? Explain your answer.

(3) Explain what the slope of the regression line means in this context.

(4) If a state has a poverty that is one standard deviation above the mean, how many standard deviations above the mean would you predict the state's crime rate to be?