MM 217 Formulae

Advait Risbud

May 21'

Contents

1 Descriptive Statistics

1.1 Measures of central tendency

Mean If $x_1, x_2, x_3, \cdots x_n$ are data points then the mean is given by

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

R Command: `mean(data)`

Median It is the value which divides the data in half. Let n denote the number of data points then if

- n=even, then
$$\tilde{x} = x_{\frac{n}{2}} + x_{\frac{n+2}{2}} \tag{2}$$
- n=odd, the
$$\tilde{x} = x_{\frac{n+1}{2}} \tag{3}$$

  R command: `median(data)`

Mode Data point with the highest frequency.
R command: `mode(data)`

Variance Variance is σ^2 and standard deviation is σ.

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{4}$$

Between one sigma limits of a normal distribution 67% of the data is contained.
R command: `sd(data)`

1.2 Graphical Methods

Empty for now, will add post midsems.

1.3 Correlation Coefficient

Consider two data sets $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_n\}$. Covariance and correlation coefficient are defined as,

$$Cov(x, y) = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \times \frac{1}{n-1} \tag{5}$$

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}} \tag{6}$$

- $-1 \leq Corr(x, y) \leq 1$
- $Corr(x, y) = 1 \implies$ perfect linear relationship with positive slope.
- $Corr(x, y) = -1 \implies$ perfect linear relationship with negative slope.
- Intermediate values indicate different extent of linearity in the two variables.

R command: `cov(x,y)` and `cor(x,y)`.

2 Probability