Formula Sheet for Quiz 2

STAT 011

For a sample of data

If $\{x_1, x_2, \ldots, x_n\}$ is a data set of n observational units, we have the following:

Sample mean

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

Sample variance

$$Var(x_1, \ldots, x_n) = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Sample standard deviation

$$sd(x_1, \ldots, x_n) = s = \sqrt{s^2}$$

If we want to standardize the data set X, to create a new standardized data set $Z = \{z_1, z_2, \ldots, z_n\}$ we preform

$$z_i = \frac{x_i - \bar{x}}{sd(x_1, \ldots, x_n)}, \text{ for } i = 1, \ldots, n.$$

Simple linear regression notation

The fitted/estimated regression model is $\hat{y}_i = b_0 + b_1 x_i$ where $b_0 = \bar{y} - b_1\bar{x}$ and $b_1 = \frac{s_{xy}}{\sqrt{s_x s_y}} \cdot \frac{s_y}{s_x}$.

$$\text{Residual} = e = y - \hat{y} = \text{observed value} - \text{predicted value}$$

Standard error of the residuals: $s_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}$

Sum of squares terms

$$s_x = \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_y = \sum_{i=1}^{n}(y_i - \bar{y})^2, \quad s_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Correlation coefficient

$$r = \frac{s_{xy}}{\sqrt{s_x s_y}}$$

Coefficient of determination/R-squared

$$R = \left(\frac{s_{xy}}{\sqrt{s_x s_y}}\right)^2$$

Probability

Five Laws of Probability

1) A probability is a number between 0 and 1.

$$0 \leq Pr(A) \leq 1, \quad \text{for } A \in S$$

2) The probability of the set of all possible outcomes of a trial is 1.

$$Pr(S) = 1$$

3) The probability of an event not occuring is equal to 1 minus the probability the event does occur.

$$Pr(A^C) = 1 - Pr(A)$$

4) For any disjoint events in the sample space of a random variable, say, A and B, we have:

$$Pr(A \text{ or } B) = Pr(A) + Pr(B)$$

Generalized version of the addition rule: For any events in the sample space,

$$Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

5) If an event A is independent of another event B, then:

$$Pr(A \cap B) = Pr(A) \times Pr(B).$$

Generalized version of the multiplication rule: For any events in the sample space,

$$Pr(A \cap B) = Pr(A) \times Pr(B \mid A).$$

Definition of conditional probability

$$Pr(B \mid A) = \frac{Pr(A \cap B)}{Pr(A)}$$

Law of total probability

$$Pr(B) = Pr(B \text{ and } A) + Pr(B \text{ and } A^C)$$

Random Variables

For a random variable X,

$$E(X) = \sum_{x \in S} \left[x \times Pr(x)\right], \quad Var(X) = \sum_{x \in S} \left[(x - E(X))^2 \times Pr(x)\right], \quad st.dev(X) = \sqrt{Var(X)}.$$

For two random variables, X and Y:

$$Cov(X,Y) = E\left[(X - E(X)) \cdot (Y - E(Y))\right], \quad Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

Linear transformations of a random Variable

Suppose a is some number between $-\infty$ and $+\infty$. The following are properties of expectation and variance for linear transformations of a random variable X.

- $E(aX) = aE(X), \quad E(a \pm X) = a \pm E(X)$
- $Var(aX) = a^2 Var(X), \quad Var(a \pm X) = Var(X)$

Linear transformations of two random variables

Suppose both X and Y are random variables that may or may not be related to one another. The following are properties of expectation and variance for linear transformations involving both random variables.

- $E(X \pm Y) = E(X) \pm E(Y)$
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$
- If X and Y are independent random variables, then $Cov(X, Y) = 0$.

Normal Random Variable

If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Binomial Random Variable

If $X \sim Bin(n, p)$ then $Pr(X = x) = nCx \cdot p^x \cdot (1-p)^{n-x}$, where $nCx = \frac{n!}{x!(n-x)!}$.