

# Formula Sheet for Quiz 2

STAT 011

## Sample Statistics

### For a sample of data

If  $\{x_1, x_2, \dots, x_n\}$  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

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

Sample standard deviation

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

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

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

### 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  $b_1 = \frac{s_{xy}}{\sqrt{s_x s_y}} \cdot \frac{s_y}{s_x}$ .

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

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

# 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 occurring is equal to 1 minus the probability the event does occur.

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

4) For any events in the sample space of a random variable, say,  $A$  and  $B$ , we compute the probability of event  $A$  or event  $B$  or both events  $A$  and  $B$  occurring with the formula:

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

5) If an event  $A$  is independent of another event  $B$ , then the probability that both events occur is the product of the probabilities of the two individual events:

$$Pr(A \text{ and } B) = Pr(A) \times Pr(B).$$

## Definition of conditional probability

$$Pr(B | A) = \frac{Pr(A \text{ and } B)}{Pr(A)}$$

## Law of total probability

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

## Random Variables

For a random variable  $X$ ,

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

For two random variables,  $X$  and  $Y$ :

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

## 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  $X$ .

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

## 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  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ .