MacEwan University
STAT 161
Formula Sheet – Final Exam

Descriptive Statistics

- Sample Variance: $s^2 = \dfrac{\sum(x_i - \bar{x})^2}{n-1} = \left(\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}\right)/(n-1)$
- Sample Standard Deviation: $s = \sqrt{\text{Sample Variance}} = \sqrt{s^2}$
- Median: Order the data from smallest to largest. The median M is either the unique middle value or the mean of the two middle values.
- Lower Quartile: Order the data from smallest to largest. The lower quartile Q_1 is the median of the smaller half of the values.
- Upper Quartile: Order the data from smallest to largest. The upper quartile Q_3 is the median of the upper half of the values.
- Interquartile Range (IQR) = Upper Quartile – Lower Quartile $= Q_3 - Q_1$
- Outliers: lower fence $= Q_1 - 1.5$ IQR and upper fence$= Q_3 + 1.5$ IQR

Probability Theory

- Addition Rule: $P(A \text{ or } B) = P(A) + P(B) - P(A\&B)$
- Complement Rule: $P(\text{A does not occur}) = P(\text{not } A) = 1 - P(A)$
- General Multiplication Rule: $P(A \text{ \& } B) = P(A\&B) = P(A|B)P(B)$
- Multiplication Rule for **Independent** Events:
  If A and B are **independent**, then $P(A \text{ and } B) = P(A\&B) = P(A)P(B)$
- Conditional Probability of A given B, if $P(B) > 0$: $P(A|B) = \dfrac{P(A\&B)}{P(B)}$

Normal Distribution

- Z-score: $Z = (X - \mu)/\sigma$
- p-th percentile of a normal distribution with mean μ and standard deviation σ:
  $x_p = \mu + \sigma z_p$

Sampling Distribution

- Sampling Distribution of a Sample Mean, $\bar{X}$: $\mu_{\bar{X}} = \mu$, $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

Estimation

Parameter	Estimator	Standard Error	Confidence Interval
μ	$\bar{x}$	$\dfrac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$
$\mu_1 - \mu_2$	$\bar{x}_d$	$\dfrac{\sigma_d}{\sqrt{n}}$	$\bar{x}_d \pm t^* \dfrac{s_d}{\sqrt{n}}$

Choosing the Sample Size

- Estimate a mean μ using a $C\%$ confidence interval within an amount of m.

$$n \geq \left(\frac{z^* \sigma}{m}\right)^2$$

Test Statistics

- Test Statistic for 1-sample t-Test concerning μ

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \qquad df = n - 1$$

- Test Statistic for two–sample t-Test for comparing two population means:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \qquad df = \min(n_1 - 1, n_2 - 1)$$

- Test Statistic for paired t-Test for comparing two population means:

$$t = \frac{\bar{x}_d}{(s_d/\sqrt{n})} \qquad df = n - 1$$

Simple Linear Regression Analysis

- Covariance: $s_{xy} = \dfrac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{n-1}$
- Pearson's Correlation Coefficient: $r = \dfrac{s_{xy}}{s_x s_y}$
- Least Squares Regression Line $\hat{y} = b_0 + b_1 x$, with

$$b_1 = r\left(\frac{s_y}{s_x}\right) = \frac{s_{xy}}{s_x^2}, \text{ and } \; b_0 = \bar{y} - b_1\bar{x}$$

- Coefficient of Determination : r^2
- Estimation of σ

$$\hat{\sigma} = s_e = \sqrt{\frac{SSE}{n-2}}, \text{ with } \; SSE = \sum(\hat{y}_i - y_i)^2 = SS_{yy} - b_1 SS_{xy}$$

- Confidence interval for β_1

$$b_1 \pm t^* \frac{s_e}{\sqrt{(n-1)s_x^2}} = b_1 \pm t^* SE_{b_1}$$

- test statistic for a test about β_1

$$t_0 = \frac{b_1}{s_e/\sqrt{(n-1)s_x^2}} = \frac{b_1}{SE_{b_1}}, \quad df = n-2$$

Multiple Linear Regression Analysis

- Model Utility
  - Adjusted R^2: $R^2_{adj} = 1 - \dfrac{(1-R^2)(n-1)}{n-k-1}$
  - Test Statistic: F with $df_n = k$ and $df = n-k-1$
- Inference for an individual slope
  - Confidence interval for β_i: $b_i \pm t^* SE_{b_1},\ df = n-k-1$
  - Test statistic for a test about β_i: $t_0 = b_i/SE_{b_i}, \quad df = n-k-1$