STT2640 Formula Sheet

1. Descriptive Statistics

   (a) Sample Mean: $\bar{x} = \frac{\sum x_i}{n}$

   (b) Sample Variance: $s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{1}{n-1}\{\sum x_i^2 - n \cdot \bar{x}^2\}$

   (c) z-score: $z = \frac{x-\bar{x}}{s}$ or $z = \frac{x-\mu}{\sigma}$

   (d) Interquartile Range: $IQR = Q_U - Q_L$

2. Probability Properties and Rules

   (a) Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

   (b) Complementary Rule: $P(A^c) = 1 - P(A)$

   (c) Conditional Probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

   (d) Multiplication Rule: $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

   (e) Complementary Rule for Conditional Probability: $P(A^c|B) = 1 - P(A|B)$

3. Binomial Distribution: $p(x) = \binom{n}{x} p^x q^{n-x}$, $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, $EX = n \cdot p$

4. Poisson Distribution: $p(x) = \frac{\lambda^x \exp^{-\lambda}}{x!}$

5. Properties of the sample mean $\bar{X}$ from a population with mean μ and S.D. σ:
   $\mu_{\bar{x}} = \mu$, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

6. Inferences based on one sample and two independent samples:

TARGET PARAMETER

μ = Mean or Average

Sample Size

- **Large ($n \geq 30$)**
  - **Population has Any Distribution**
  - **$(1-\alpha)$ 100% Confidence Interval**
  - Known: $\bar{x} \pm (z_{\alpha/2})\sigma/\sqrt{n}$
  - Unknown: $\bar{x} \pm (z_{\alpha/2})s/\sqrt{n}$
- **Small ($n < 30$)**
  - **Population has Normal Distribution**
  - **$(1-\alpha)$ 100% Confidence Interval**
  - $\bar{x} \pm (t_{\alpha/2})s/\sqrt{n}$

TARGET PARAMETER

μ = Mean or Average

Sample Size

- **Large ($n \geq 30$)**
  - **Population has Any Dist'n**
  - **Test Statistic**
  - **σ Known:** $z = \sqrt{n}(\bar{x} - \mu_0)/\sigma$
  - **σ Unknown:** $z \approx \sqrt{n}(\bar{x} - \mu_0)/s$
- **Small ($n < 30$)**
  - **Population has Normal Dist'n**
  - **Test Statistic**
  - $t = \sqrt{n}(\bar{x} - \mu_0)/s$

Target Parameter

$(\mu_1 - \mu_2)$

Independent Samples

μ_1 Mean for Pop'n = 1

μ_2 Mean for Pop'n = 2

Sample Size

- **Large** ($n_1 \geq 30$ and $n_2 \geq 30$) Populations have any Dist'n
  - **Test Statistic**
  - $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
  - **Confidence Interval**
  - $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- **Small** ($n_1 < 30$ or $n_2 < 30$) Both Pop'ns have Normal Dist'n
  - **Test Statistic**
  - $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
  - **Confidence Interval**
  - $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

7. Analysis of Variance (ANOVA)

(a)

$$SST = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2, \quad SSE = \sum_{i=1}^{k}(n_i - 1)s_i^2$$

(b) Test Statistic: $F_{obs} = \frac{MST}{MSE} = MST/MSE$

8. Simple Linear Regression

(a) The least squares estimate of β_0 and β_1 are:

$$b_1 = \hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}, \quad b_0 = \hat{\beta}_0 = \bar{y} - b_1\bar{x},$$

where

$$SS_{xx} = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n \cdot \bar{x}^2, \quad SS_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}.$$

(b) $s^2 = \hat{\sigma^2} = SSE/(n-2)$ is an estimate of σ^2 with $SSE = \sum_{i=1}^{n} e_i^2$,
$s = \sqrt{s^2}$ is the estimated standard deviation.

(c) Inferences on β_1, the slope of the regression line

i. $100(1-\alpha)\%$ confidence interval for β_1 is: $b_1 \mp t_{\alpha/2}\frac{s}{\sqrt{SS_{xx}}}$

ii. The test statistic for $H_0 : \beta_1 = 0$ is $t_{obs} = \frac{b_1}{s/\sqrt{SS_{xx}}}$.