Numerical Summaries of Centre and Variation:

- Mean: $\bar{x} = \frac{\sum x}{n}$ Standard deviation: $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$ Variance: $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$
- **Note**: $\sum(x-\bar{x})^2 = \sum x^2 - n\,\bar{x}^2$

Empirical Rule:

- $\%(\bar{x} \pm s)$: 68% $\%(\bar{x} \pm 2s)$: 95% $\%(\bar{x} \pm 3s)$: Nearly 100%
- ***z*-Score:** $z = \frac{x-\bar{x}}{s}$ **Interquartile range**: $Q_3 - Q_1$ **Range**: $Maximum - Minimum$

Correlation and Regression:

- Pearson Correlation: $r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$
- Regression line: Predicted $y = b_0 + b_1 x$; slope: $b_1 = r\frac{s_y}{s_x}$; intercept: $b_0 = \bar{y} - b_1\bar{x}$

Probability Rules:

- Rule 1: $0 \le P(A) \le 1$
- Rule 2: $P(A^c) = 1 - P(A)$
- Rule 3: For equally likely outcomes, $P(A) = \frac{Number\ of\ outcomes\ in\ A}{Number\ of\ possible\ outcomes}$
- Rule 4: $P(A \text{ OR } B) = P(A) + P(B) - P(A \text{ AND } B)$
- Rule 4a: $P(A \text{ OR } B) = P(A) + P(B)$ if A, B are mutually exclusive
- Rule 5a: $P(A|B) = \frac{P(A \text{ AND } B)}{P(B)}$
- Rule 5b: $P(A \text{ AND } B) = P(A|B)P(B) = P(B|A)P(A)$
- Rule 5c: For independent events *A* and *B*, $P(A \text{ AND } B) = P(A)P(B)$

Central limit theorem: For large *n*

- For sample proportions: $\hat{p}$ is approximately normal with mean = *p* and

$$SD = \sqrt{\frac{p(1-p)}{n}}$$

- For sample mean: $\bar{x}$ is approximately normal with mean = μ and

$$SD = \frac{\sigma}{\sqrt{n}}$$

INTE 296 Formulae Sheet – Page 2 of 2

Confidence Intervals for proportions and means:

- Proportion: $z = \frac{\hat{p}-p}{SD}$ CI: $\hat{p} \pm z^{*}SD_{est}$; $SD_{est} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$;
- Mean : $z\ or\ t = \frac{\bar{x}-\mu}{SD_{est}}$ CI: $\bar{x} \pm t^{*}SD_{est}$; $SD_{est} = \frac{s}{\sqrt{n}}$

Difference of means:

- **CI:** $(\bar{x}_1 - \bar{x}_2) \pm t^{*}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$; t^{*} obtained from a t-table with df $=$ smaller of $(n_1 - 1)$ and $(n_2 - 1)$

Test for proportions:

Single proportion: $H_0: p = p_0$; Test Statistic: $z = \frac{\hat{p}-p_0}{SD}$ where $SD = \sqrt{\frac{p_0(1-p_0)}{n}}$

- Two Proportions: $H_0: p_1 = p_2$; Test Statistic: $z = \frac{\hat{p}_1-\hat{p}_2}{SD}$ where $SD = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$;

$$\hat{p} = \frac{\text{number of successes in sample 1} + \text{number of successes in sample 2}}{n_1 + n_2}$$

Test for means:

- Single mean: $H_0: \mu = \mu_0$; Test Statistic: $z = \frac{\bar{x}-\mu_0}{SD_{est}}$ where $D_{est} = \frac{s}{\sqrt{n}}$; for large samples
- For small samples and normal population, use Test Statistic: $t = \frac{\bar{x}-\mu_0}{SD_{est}}$ where $SD_{est} = \frac{s}{\sqrt{n}}$, *df=n-1*
- Two means: $H_0: \mu_1 = \mu_2$; for independent samples, Test Statistic: $t = \frac{\bar{x}_1-\bar{x}_2}{SD_{est}}$ where $SD_{est} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- For dependent sample (Paired), Test Statistic: $t = \frac{\bar{x}_{\text{difference}}}{SD_{est}}$ where $D_{est} = \frac{s_{\text{diffference}}}{\sqrt{n}}$; *df = n-1*

Chi-square Test for Testing in Categorical Variable:

- Expected Frequency for a cell $= \frac{(row\ total)\times(Column\ total)}{grand\ total}$
- $\chi^2_{Obs} = \sum_{all\ cells}\frac{(Observed-Expected)^2}{Expected}$

Degrees of Freedom:

- For goodness of fit, df $=$ No. of categories $- 1$;
- For a categorical two-way table, df $=$ (number of rows $- 1$)\times (number of columns $- 1$)

ANOVA

- $SS_{total} = SS_{between} + SS_{within}$; $SS_{total} = \sum(x - \bar{\bar{x}})^2 = (N-1)s^2$; $SS_{within} = \sum(n_i - 1)s_i^2$
- $SS_{between} = \sum n_i(\bar{x}_i - \bar{\bar{x}})^2 = SS_{total} - SS_{within}$
- $df_{between} = k - 1$; $df_{within} = N - k$; $df_{total} = df_{beteween} + df_{within} = N - 1$; $MS = \frac{SS}{df}$; $F = \frac{MSB}{MSW}$