**Chapter 6: Introduction to Inference**

- Confidence interval for mean $\mu$ ($\sigma$ known):
  \[ \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \]
- Sample size for confidence interval for $\mu$ with margin of error $m$:
  \[ n = \left[ \frac{z^* \sigma}{m} \right]^2 \]
- $z$ statistic for $H_0: \mu = \mu_0$ ($\sigma$ known):
  \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \]

**Chapter 7: Inference for Distributions**

- Standard error of $\bar{x}$: $\text{SE}_{\bar{x}} = \frac{s}{\sqrt{n}}$
- Confidence interval for mean $\mu$ ($\sigma$ unknown):
  \[ \bar{x} \pm t^* \frac{s}{\sqrt{n}} , \quad df = n - 1 \]
- One-sample $t$ statistic for $H_0: \mu = \mu_0$ ($\sigma$ unknown):
  \[ t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} , \quad df = n - 1 \]
- Two-sample $z$ statistic for $H_0: \mu_1 = \mu_2$ ($\sigma_1, \sigma_2$ known):
  \[ z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \]
- Two-sample $t$ statistic for $H_0: \mu_1 = \mu_2$ ($\sigma_1, \sigma_2$ unknown):
  \[ t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} , \quad df = \text{minimum of } n_1 - 1, n_2 - 1 \]
- Two-sample confidence interval for $\mu_1 - \mu_2$:
  \[ (\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} , \quad df = \text{minimum of } n_1 - 1, n_2 - 1 \]
- Pooled two-sample estimator of $\sigma^2$:
  \[ s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \]
- Pooled two-sample $t$ statistic for $H_0: \mu_1 = \mu_2$ when $\sigma_1 = \sigma_2$:
  \[ t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} , \quad df = n_1 + n_2 - 2 \]
- Pooled two-sample confidence interval for $H_0: \mu_1 - \mu_2$ when $\sigma_1 = \sigma_2$:
  \[ (\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} , \quad df = n_1 + n_2 - 2 \]
- Two-sample $F$ statistic for $H_0: \sigma_1 = \sigma_2$:
  \[ F = \frac{s_1^2}{s_2^2} , \quad df = n_1 - 1 (\text{numerator}), \quad n_2 - 1 (\text{denominator}) \]

**Chapter 8: Inference for Proportions**

- Sample proportion: $\hat{p} = x/n$, $x =$ number of “successes”
- Wilson estimate of the population proportion:
  \[ \hat{p} = \frac{x + 2}{n + 4} \]
- The standard error of $\hat{p}$:
  \[ \text{SE}_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n + 4}} \]
- Confidence interval for $p$:
  \[ \hat{p} \pm z^* \text{SE}_{\hat{p}} \]
- $z$ statistic for $H_0: p = p_0$
  \[ z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \]
- Sample size for desired margin of error $m$:
  \[ n + 4 = \left( \frac{z^*}{m} \right)^2 p^*(1 - p^*) \quad \text{where } p^* = \text{guessed value} \]
  or
  \[ n + 4 = \left( \frac{z^*}{2m} \right)^2 \quad \text{conservative approach with } p^* = \frac{1}{2} \]
- Wilson estimate of the difference of two population proportions:
  \[ \hat{D} = \hat{p}_1 - \hat{p}_2 = \frac{x_1 + 1}{n_1 + 2} - \frac{x_2 + 1}{n_2 + 2} \]
- Standard error of the difference:
  \[ \text{SE}_{\hat{D}} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1 + 2} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2 + 2}} \]
• Confidence interval for $p_1 - p_2$:
  \[ \hat{D} \pm z \cdot \text{SE}_{\hat{D}} \]
• Test Statistic for $H_0 : p_1 = p_2$:
  \[ z = \hat{P}_1 - \hat{P}_2 \cdot \frac{1}{\text{SE}_{\hat{P}_1}} \]
  where $\hat{P}_1 = \frac{x_1}{n_1}$ and $\hat{P}_2 = \frac{x_2}{n_2}$.

**Chapter 9: Inference for Two-way Tables**
• Expected cell counts:
  \[ \text{Expected cell count} = \frac{\text{row total} \times \text{column total}}{n} \]
• Chi-square test statistic:
  \[ X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]
  \[ \text{df} = (\# \text{ of rows} - 1)(\# \text{ of columns} - 1) \]

**Chapter 10: Inference for Regression**
• Sample variance of $x$'s:
  \[ S_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \]
• Sample variance of $y$'s:
  \[ S_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2 \]
• Sample correlation:
  \[ r = \frac{1}{n-1} \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y} = \rho \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \]
• Regression line:
  \[ \hat{y} = b_0 + b_1x \]
• Slope: $b_1 = \frac{S_y}{S_x}$
• Intercept: $b_0 = \bar{y} - b_1 \bar{x}$
• Mean squared error:
  \[ S^2 = \frac{1}{n-2} \sum e_i^2 \], where $e_i = y_i - \hat{y}_i$
• Standard error of $b_0$:
  \[ \text{SE}_{b_0} = S \sqrt{\frac{1 + \frac{x^2}{n \sum (x_i - \bar{x})^2}}{n}} \]
• Level C confidence interval for $b_0$:
  \[ b_0 \pm t \cdot \text{SE}_{b_0}, \text{ df} = n - 2 \]
• Standard error of $b_1$:
  \[ \text{SE}_{b_1} = \frac{S}{\sqrt{\sum (x_i - \bar{x})^2}} \]
• Level C confidence interval for $b_1$:
  \[ b_1 \pm t \cdot \text{SE}_{b_1}, \text{ df} = n - 2 \]
• Test statistic for $H_0 : \beta_1 = 0$:
  \[ t = \frac{b_1}{\text{SE}_{b_1}}, \text{ df} = n - 2 \]
• Estimate for mean response $\mu_y$ when $x = x^*$:
  \[ \hat{\mu}_y = b_0 + b_1x^* \]
• Standard error of $\hat{\mu}_y$ when $x = x^*$:
  \[ \text{SE}_{\hat{\mu}_y} = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{\hat{\rho}^2 \sigma^2 \sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \]
• Level C confidence interval for $\mu_y$ when $x = x^*$:
  \[ \hat{\mu}_y \pm t \cdot \text{SE}_{\hat{\mu}_y}, \text{ df} = n - 2 \]
• Estimate for future observation of $y$ when $x = x^*$
  \[ \hat{y} = b_0 + b_1x^* \]
• Standard error of $\hat{y}$ when $x = x^*$:
  \[ \text{SE}_{\hat{y}} = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{\hat{\rho}^2 \sigma^2 \sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \]
• Level C confidence interval for $\hat{y}$ when $x = x^*$
  \[ \hat{y} \pm t \cdot \text{SE}_{\hat{y}}, \text{ df} = n - 2 \]
• Sum of Squares:
  \[ \text{SST} = \sum (y_i - \bar{y})^2 \]
  \[ \text{SSM} = \sum (\hat{y}_i - \bar{y})^2 \]
  \[ \text{SSE} = \sum (y_i - \hat{y}_i)^2 \]
• SST = SSM + SSE
• MS = Sum of squares / degrees of freedom
• The ANOVA $F$ test for $H_0 : \beta_j = 0$:
  \[ F = \frac{\text{MSM}}{\text{MSE}}, \text{ df} = (1, n-2) \]
• The test for $H_0 : \rho = 0$:
  \[ t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}, \text{ df} = n - 2. \]

**Chapter 11: Multiple Regression**
• Multiple Regression Model:
  \[ y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} + \varepsilon_i \]
• Least squares estimates of $\beta_0, \beta_1, \ldots, \beta_p$:
  \[ b_0, b_1, \ldots, b_p \]
• Estimate of $\sigma$: $s = \sqrt{\text{MSE}}$
• Level C confidence interval for $\beta_j$:
  \[ b_j \pm t \cdot \text{SE}_{b_j}, \text{ df} = n - p - 1 \]
• Test statistic for $H_0 : \beta_j = 0$:
\[
t = \frac{b_j}{SE_{b_j}}, \text{ df } = n - p - 1
\]
- Sum of squares SS: \( SST = SSM + SSE \)
- Degrees of freedom DF: \( DFT = DFM + DFE \)
  \( DFT = n - 1, DFM = p, DFE = n - p - 1 \)
- Mean Square Model: \( MSM = SSM/DFM \)
- Mean Square Error: \( MSE = SSE/DFE \)
- Test statistic for \( H_0 : \beta_1 = \ldots = \beta_p = 0 \):
  \( F = MSM/MSE, \text{ df } = (p, n - p - 1) \)
- Squared Multiple Correlation: \( R^2 = \frac{SSM}{SST} \)

----------------------------------------

**Chapter 12: One-way ANOVA**

- One-way ANOVA model: \( x_{ij} = \mu_i + \epsilon_{ij} \)
  for \( i = 1, \ldots, I \) and \( j = 1, \ldots, n_i \) and \( N = n_1 + \ldots + n_I \)
- Pooled-sample variance:
  \[
s^2_p = \frac{(n_1 - 1)s_1^2 + \ldots + (n_I - 1)s_I^2}{(n_1 - 1) + \ldots + (n_I - 1)} = MSE
\]
- Sum of squares (SS): \( SST = SSG + SSE \)
- Degrees of freedom (DF):
  \( DFT = DFG + DFE \), where \( DFT = N - 1, DFG = I - 1, DFE = N - 1 \).
- Mean square (MS):
  \( MSG = \frac{SSG}{DFG}, MSE = \frac{SSE}{DFE} \)
- Test statistic for \( H_0 : \mu_1 = \mu_2 = \ldots = \mu_I : \)
  \( F = MSG/MSE, \text{ df } = (I - 1, N - I) \).
- Coefficient of determination: \( R^2 = \frac{SSG}{SST} \)
- Population contrast: \( \psi = \sum a_i \mu_i \), where \( \sum a_i = 0 \)
- Sample contrast: \( c = \sum a_i \bar{x}_i \)
- Standard error of \( c \):
  \[
  SE_c = s_p \sqrt{\frac{\sum a_i^2}{n_I}}
  \]
- Test statistic for \( H_0 : \psi = 0 : \)
  \[
  t = \frac{c}{SE_c}, \text{ df } = n - I
  \]
- Level C confidence interval for \( \psi : \)
  \( c \pm t^* SE_c, \text{ df } = N - I \)
- Multiple Comparison t statistic:
  \[
  t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}, \text{ df } = N - I
  \]
- Simultaneous Confidence Interval for Mean Differences:
  \[
  (\bar{x}_i - \bar{x}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}},
  \]
  The critical values \( t^{**} \) are the same as those used for the multiple comparisons procedure chosen.

----------------------------------------

**Chapter 13: Two-way ANOVA**

- Two-way ANOVA model: \( x_{ijk} = \mu_{ij} + \epsilon_{ijk} \)
  for \( i = 1, \ldots, I \) and \( j = 1, \ldots, J \) and \( k = 1, \ldots, n_{ij} \)
- Pooled-sample variance:
  \[
  s^2_p = \frac{\sum(n_{ij} - 1)s_{ij}^2}{\sum(n_{ij} - 1)} = MSE
  \]
- Sum of squares (SS):
  \( SST = SSA + SSB + SSAB + SSE \)
- Degrees of freedom (DF):
  \( DFT = DFA + DB + DFAB + DFE \)
- Mean square (MS): For the factors A and B, for the interaction AB, and for the error E:
  \( MS = \frac{SS}{DF} \)
- Test statistic for \( H_0 : \text{Main effect of A is 0} : \)
  \( F = \frac{MSA}{MSE}, \text{ df } = (I - 1, N - I) \).
- Test statistic for \( H_0 : \text{Main effect of B is 0} : \)
  \( F = \frac{MSB}{MSE}, \text{ df } = (J - 1, N - I) \).
- Test statistic for $H_0$: Interaction effect of A and B is 0:
  \[ F = \frac{\text{MSAB}}{\text{MSE}}, \text{df} = (I-1)(J-1), N - IJ. \]

**Additional Formulas**

**Chapter 6: Intro. To Inferences**

<table>
<thead>
<tr>
<th>Z*</th>
<th>1.645</th>
<th>1.960</th>
<th>2.575</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
</tr>
</tbody>
</table>

Type I error:
Reject $H_0$ when $H_0$ is true.

Type II error:
Fail to reject $H_0$ when $H_0$ is false.

Power: the probability that a fixed level $\alpha$ significance test will reject $H_0$ for a given alternative parameter value

**Chapter 8: Inference for Proportions**

95% confidence interval for the relative risk $p_1 / p_2$:

\[ RR \leq \frac{p_1}{p_2} \leq RR \times 1.96^V \]

where $RR$ is the estimate of $p_1 / p_2$ from sample proportions, and

\[ V^2 = \frac{1-p_1}{n_1p_1} + \frac{1-p_2}{n_2p_2} \]

(For other confidence %'s, replace the 1.96 by the appropriate $z^*$.)

**Chapter 10: Inference for Regression Shortcut formulas:**

\[ S_x^2 = \frac{1}{n-1}(\sum x_i^2 - n\bar{x}^2) \]

\[ r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)S_x S_y} \]

**Chapter 12: One-way ANOVA**

\[ SSG = \sum n_i (\bar{x}_i - \bar{x})^2 \]

\[ SSE = \sum (n_i - 1)s_i^2 \]

\[ SST = \sum \sum (x_{ij} - \bar{x})^2 \]

$t^**$ for simultaneous confidence interval:

Tukey-Kramer-Hayter simultaneous interval on all pairs $\mu_i - \mu_j$.

\[ t^** = q^*/\sqrt{2} \]

where $q^*$ comes from tables of the Studentized Range $q_k,v$ with $K = I$ and $v = DFE$; for confidence 95%, use the table labeled .05.

Bonferroni simultaneous intervals on $r$ pairs $\mu_i - \mu_j$, set $\alpha = .05/r$ and find $t^*$ with df=DFE for this $\alpha$ (half in each tail, as is done for one pair). For confidence other than 95%, replace .05 with the appropriate number.

Scheffé simultaneous confidence interval on all contrasts $\psi$ are

\[ c \pm S*S\psi_c \]

where $S = \sqrt{(I-1)F(I-1, DFE)}$

and for 95% confidence we use the S that comes from the $F(I-1, DFE)$ with probability .05 in the upper tail.

The S-method is equivalent to the ANOVA test of $H_0$ in the sense that the ANOVA F-test rejects $H_0$ if and only if at least one of the S-method intervals (for some contrast) does not include the value zero.