## TYPES of ERRORS

<table>
<thead>
<tr>
<th>Reality</th>
<th>Ho true (Ha false)</th>
<th>Ho false (Ha true)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Our Decision</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reject Ho</td>
<td>TYPE I Error</td>
<td>Correct Decision</td>
</tr>
<tr>
<td>(Accept Ha)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not Reject Ho</td>
<td>Correct Decision</td>
<td>TYPE II Error</td>
</tr>
<tr>
<td>(Accept Ho)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TYPE I Error** - Reject Ho, but Ho is actually true

**TYPE II Error** - Fail to reject Ho, but Ho is actually false

$$\alpha = P(\text{TYPE I Error})$$

$$\beta = P(\text{TYPE II Error})$$

Relation between $\alpha$, $\beta$, $n$

Relation between C.I. and Hyp. Test
Ch 10 - \( \chi^2 \) tests

**Multinomial Experiment:**
1. Fixed number of trials
2. Each trial results in exactly one of \( k \) possible outcomes
3. \( p_i = P(\text{outcome } i \text{ on single trial}) \)
   \[ p_1 + p_2 + \ldots + p_k = 1 \]
4. Trials are independent

**Expected cell frequencies:**
\[ E_i = n p_i \quad \text{for outcome } i \]

**Observed cell frequencies:** \( O_i \)

**Goodness of Fit Tests:**
- \( H_0: \) all the \( p_i \) are equal
- \( H_a: \) not all \( p_i \) are equal
\[ \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad \text{(test statistic)} \]
\[ df = k - 1 \quad \chi^2 \text{ table } C \]

Reject \( H_0 \) if \( p\)-value \( \leq \alpha \)
or if \( \chi^2 > \text{critical value} \)
Chi-Square Association of Categorical Variables

Contingency Table

χ² Test for Independence (2x2 Contingency Table)

\[ E = \frac{\text{row total} \times \text{col total}}{n} \]

\[ \chi^2 = \sum \frac{(|O - E| - .5)^2}{E} \]

\[ \text{df} = (r-1)(c-1) \]

Critical Value - Table C

Reject Ho if \( \chi^2 > \) critical value or if p-value ≤ \( \alpha \)
$\chi^2$ Test for Indep. 
(r x c Contingency Table)

$\chi^2 = \sum \frac{(\theta - E)^2}{E}$

$E = \frac{\text{rowtotal} \times \text{col total}}{n}$

df = $(r-1)(c-1)$

Crit. Value - Table C

Reject $H_0$ if $\chi^2 > \text{critical value}$

or if $p-value \leq \alpha$
ch 11 - Correlation
Recall from ch. 3

\[ r = \frac{SS(xy)}{\sqrt{SS(x) \cdot SS(y)}} \]

where

\[ SS(xy) = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n} \]

\[ SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \]

\[ SS(y) = \Sigma y^2 - \frac{(\Sigma y)^2}{n} \]
Hyp. Test for $\rho$

$H_0: \rho = 0$

$$t^* = r \sqrt{\frac{n-2}{1-r^2}}$$

$p$-Value from $t$-dist with $df = n-2$
Chapter 11: Regression Analysis

Recall from Chapter 3:

Regression (Prediction) Equation

\[ y = \beta_0 + \beta_1 x \]

\[ b_1 = \frac{SS(xy)}{SS(x)} \]

\[ b_0 = \bar{y} - b_1 \bar{x} \]

Predicting Values of \( y \)

Point Estimator: \( \hat{y} = \beta_0 + \beta_1 x_0 \)
Hypothesis Test on $\beta_i$

$H_0: \beta_i = 0$

Test Statistic: $(n \leq 30)$

$$T = \frac{b_i}{S_r / \sqrt{SS(x)}}$$

where $p$-Value from $t$-dist with $df = n - 2$ and

$$S_r = \sqrt{\frac{SS(y) - b_i^2 \cdot SS(x)}{n - 2}}$$

If $n \geq 30$, the test statistic is

$$Z = \frac{b_i}{S_r / \sqrt{SS(x)}}$$

$p$-Value from standard normal

Connection between Correlation, Regression.
Confidence Interval for mean Predicted Value of $y$ at some value $X = X_0$

$$(b_0 + b_1X_0) \pm t \cdot S_T \cdot \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{SS(x)}}$$

where

$$S_T = \sqrt{\frac{SS(y) - b_1^2 \cdot SS(x)}{n-2}}$$

$$SS(x) = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$SS(y) = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

and $t$ is from a $t$-dist with $df = n-2$ with $\frac{1}{2}$
(1-\(\alpha\))100\% Prediction Interval for an individual \(\hat{y}\) at some \(x = x_0\)

\[
(b_0 + b_1 x_0) \pm t^* \cdot \sqrt{\frac{S_r^2}{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS(x)}}}
\]

where \(S_r, SS(x), SS(y), t\) as before

(1-\(\alpha\))100\% C.I. on \(b_1\)

\[
b_1 + t^* \cdot \frac{S_r}{\sqrt{SS(x)}}
\]

where \(S_r, SS(x), t\) as before - on previous page
Relevance of Statistics to Understanding and Learning About the World

1) Observe
2) Construct a conjecture (hypothesis)
3) Design an experiment
4) Collect data
5) Analyze data (statistical methods)
6) Conclusion