Review of Part VII

Math 2200
## Quick guide

<table>
<thead>
<tr>
<th>Response variable</th>
<th>Predictor Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative</td>
<td>Categorical</td>
</tr>
<tr>
<td>Quantitative</td>
<td>Regression</td>
</tr>
<tr>
<td>Categorical</td>
<td>Chi-square tests (Chapter 26)</td>
</tr>
</tbody>
</table>

- F-test (in ANOVA)
- T-test (in linear regression)
Assumptions and conditions

• The independence assumption
• Equal variance assumption
• Nearly normal assumption
• Additive condition
  – If violated, add an interaction term
One-way ANOVA

• This is an extension of the two (independent) sample t-test

• Conditions or Assumptions
  – The data are randomly sampled
  – The variances of each sample are assumed equal
  – The residuals are normally distributed

• The null hypothesis: The means of the treatment groups are all equal
# One-way ANOVA Table

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares</th>
<th>Deg. of freedom</th>
<th>Mean squares</th>
<th>F ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor (Between)</td>
<td>SS</td>
<td>k-1</td>
<td>MS</td>
<td>MS/MS&lt;sub&gt;e&lt;/sub&gt;</td>
</tr>
<tr>
<td>Error (Within)</td>
<td>SS&lt;sub&gt;e&lt;/sub&gt;</td>
<td>n-k</td>
<td>MS&lt;sub&gt;e&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SS&lt;sub&gt;T&lt;/sub&gt;</td>
<td>n-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Two-way ANOVA without interaction

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sums of squares</th>
<th>Deg. of freedom</th>
<th>Mean squares</th>
<th>F ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor₁</td>
<td>SS₁</td>
<td>df₁ = k₁ - 1</td>
<td>MS₁ = SS₁/(k₁ - 1)</td>
<td>MS₁/MSₑ</td>
</tr>
<tr>
<td>Factor₂</td>
<td>SS₂</td>
<td>df₂ = k₂ - 1</td>
<td>MS₂ = SS₂/(k₂ - 1)</td>
<td>MS₂/MSₑ</td>
</tr>
<tr>
<td>Error</td>
<td>SSₑ</td>
<td>dfₑ = N - 1 - (k₁ - 1) - (k₂ - 1)</td>
<td>MSE = SSE/ dfₑ</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SS_T</td>
<td>N - 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Two-way ANOVA with interaction

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sums of squares</th>
<th>Deg. of freedom</th>
<th>Mean squares</th>
<th>F ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>SS1</td>
<td>K1-1</td>
<td>MS1 = SS1/(K1-1)</td>
<td>MS1/MSE</td>
</tr>
<tr>
<td>Factor 2</td>
<td>SS2</td>
<td>K2-1</td>
<td>MS2 = SS2/(K2-1)</td>
<td>MS2/MSE</td>
</tr>
<tr>
<td>Interaction=Factor1*Factor 2</td>
<td>SSI</td>
<td>(K1-1)(K2-1)</td>
<td>MSI = SSI/(K1-1)(K2-1)</td>
<td>MSI/MSE</td>
</tr>
<tr>
<td>Error</td>
<td>SSE</td>
<td>K1<em>K2</em>(L-1)</td>
<td>MSE = SSE/K1<em>K2</em>(L-1)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SST</td>
<td>N-1 = K1<em>K2</em>L-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$L$ is number of replicates at each treatment level
Interaction

• How to interpret the effect of a factor?
  – Interaction plot

• In a two-way ANOVA, interaction term can not be tested without replicates
Simple regression

$E(Y) = \alpha + \beta X$
Least square estimates

\[ y_i = a + bx_i + e_i \quad \text{choose } a, \ b \text{ to minimize } \sum_{i=1}^{n} e_i^2 \]

\[ \beta \approx b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{s_x^2} = r \frac{s_y}{s_x} \]

\[ \alpha \approx a = \bar{y} - bx \]

\[ \sigma \approx s_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n-2}} = \sqrt{\frac{\text{SSE}}{n-2}} \]
Inference

- T-test
- Confidence interval
- Prediction interval

\[
se(\hat{y}_0) = \sqrt{se^2 (b_1) \times (x_0 - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}
\]

\[
\hat{y}_0 \pm t^*_{n-2} \times se(\hat{y}_0)
\]
Multiple regression

• Output is similar to simple regression
• Interaction
  – one predictor is categorical
• Comparing models
  – R-square
  – Adjusted R-square
• Inference
  – T-test
  – Confidence interval