1.3 The $F$ test under unbalanced designs

The test for unbalanced designs is very similar - just replacing $J$ with $J_i$. In this case,

$$SSW = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_i)^2$$

$$SSB = \sum_{i=1}^{I} J_i (\bar{Y}_i - \bar{Y})^2$$

$$F = \frac{MSB}{MSW} = \frac{SSB/(I-1)}{SSW/\sum_{i}(J_i-1)} \sim F_{I-1, \sum_{i}(J_i-1)}$$

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>SSB = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_i)^2</td>
<td>I-1</td>
<td>MSB = SSB/(I-1)</td>
<td>MSB/MSW</td>
</tr>
<tr>
<td>Error</td>
<td>SSW = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y}_i)^2</td>
<td>\sum_{i}(J_i-1)</td>
<td>MSW = SSW/\sum_{i}(J_i-1)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SSTO = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y})^2</td>
<td>\sum_{i}(J_i-1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Review about the assumptions we used to construct the $F$ test:

1. $\epsilon_{ij} \sim N(0, \sigma^2)$
2. equal variance
3. independent $\epsilon_{ij}$.

For large samples, assumption (1) is not very important. If the all groups have the same number of observations, violation of (2) does not have a strong impact on results. Assumption (3) is very challenging.

1.4 multiple comparisons

In one way ANOVA we were concerned about whether all the means are the same. If the null hypothesis of equal mean is rejected, we still have no idea about how the means differ. Usually it is interesting to ask whether pairs of a subgroup of treatments show any difference. A naive approach would be to compare all pairs of treatment means using $t$ tests.

Here is the problem:

When considering a pair, with a significance level $\alpha$, we have a type I error rate $\alpha$. But this is for a pair.

If there are multiple comparisons, the overall type I error across all comparisons will be inflated.
For instance, suppose that we perform $K$ independent level $\alpha$ tests and that in each case the null hypothesis is true. The overall type I error rate is defined as

$$\Pr(\text{reject at least one } H_0 | H_0 \text{ true on all tests})$$

$$= 1 - \Pr(\text{no } H_0 \text{ was rejected } | H_0 \text{ true on all tests})$$

$$= 1 - (1 - \alpha)^K$$

For $\alpha = 0.05$ and $K = 10$, the overall type I error rate is $1 - 0.95^{10} = 0.40$.

It is often desirable to control the overall type I error rate at a given level, say 0.05. How to do that? There are many options. Here we focus on a popular method: the Bonferroni correction.

**The Bonferroni Correction**

The Bonferroni Correction comes from the Bonferroni inequality (also known as Bool’s inequality), which states that:

$$P(\bigcup_{i=1}^K A_i) \leq \sum_{i=1}^K P(A_i)$$

for event $A_1, A_2, \ldots, A_K$.

When $K = 2$, it is easy to verify that: From STAT120A, we have $Pr[A_1 \cup A_2] = Pr[A_1] + Pr[A_2] - Pr[A_1 \cap A_2]$. But $Pr[A_1 \cap A_2] \geq 0$. Therefore,

$$Pr[A_1 \cup A_2] \leq Pr[A_1] + Pr[A_2] - Pr[A_1 \cap A_2].$$

For a general $K$, we can use mathematical induction to prove the inequality. (homework assignment)

For test $i$, we define

$A_i = \{ \text{reject } H_i | \text{ all null are true} \}$

Suppose that $\alpha^* = Pr(A_i) = Pr(\text{reject the } i\text{th null } | \text{the } i\text{th null is true})$.

According to the Bonferroni inequality,

overall type I error rate $= Pr(\text{there is at least one positive --- all null hypotheses are true})$

$$= Pr(\bigcup_{i=1}^K A_i)$$

$$\leq \sum_{i=1}^K Pr(A_i)$$

$$= K\alpha^*$$
Therefore, the overall type I error rate is $\leq K\alpha^*$. This implies that if we choose $\alpha^* = \alpha/K$, the overall type I error rate is no greater than $\alpha$. For example, if we have 10 tests and we want to control the overall type I error rate $\leq 0.05$, we reject each null hypothesis at a significance level of 0.005.

Recall that based on Theorem B we have

$$
\frac{(\bar{Y}_{i1} - \bar{Y}_{i2}) - (\alpha_{i1} - \alpha_{i2})}{\sqrt{\frac{S_p^2}{2}(\frac{1}{J} + \frac{1}{J})}} \sim t_{I(J-1)}
$$

Definition. A set of simultaneous 95% confidence intervals for the pairwise comparison is

$$
(\bar{Y}_{i1} - \bar{Y}_{i2}) \pm \sqrt{\frac{S_p^2}{2}} t_{I(J-1),1-0.05/(2K)}
$$

Example. If $I=7$, we have $K = \binom{7}{2} = 21$ pairwise comparisons.