Chapter 5.5 - One-Way Analysis of Variance
Read section 5.5

COMPARING MANY POPULATION MEANS

- *Analysis of Variance* is often abbreviated as ANOVA.
- A one-way ANOVA considers $k > 2$ populations. The mean of the $i^{th}$ population is $\mu_i$. The variance of the $i^{th}$ population is $\sigma_i^2$.
- An ANOVA is used to compare the $k$ population means, $\mu_1, \mu_2, ..., \mu_k$.
- “One-way” means that the levels of a single factor define the populations being compared. In other words, the categories of a categorical variable define the populations.

Setting:

1. A SRS of size $n_i$ has been chosen from population $i$ for $i = 1, 2, ..., k$.
2. Each SRS is independent of the others.
3. Homogeneity of Variance: $\sigma^2 = \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$.
4. Each population is normally distributed, so the distribution of the $i^{th}$ population is $N(\mu_i, \sigma)$. If $n_i > 30$ for all $k$ samples, we have learned that we can drop the assumption of normality.

Note about Study Design:

- Completely Randomized Design Experiment
  - The $k$ treatment groups from a CRD are considered to be independent samples from $k$ populations.
  - In a CRD, if there is a difference among the $k$ means, it is appropriate to claim that the factor (treatment) caused the difference in means.
  - If the groups of individuals in the CRD were chosen from a SRS, then conclusions about $k$ means can be extended to the populations from which the individuals were drawn. However, if the individuals were not from a SRS, then conclusions to larger populations are dubious.
- Randomized Block Design Experiment: the data from this experimental design is analyzed via “two-way ANOVA” that includes one factor that describes the populations being compared (i.e., the treatment groups) and a second factor that describes the blocks.
- Observational Study
  - The $k$ samples to be compared in an observational study are considered independent if individuals in each sample were randomly chosen from each respective population.
  - Do not claim that the factor (explanatory variable) caused the difference in means.
The one-way ANOVA Model

\[ X_{ij} = \mu_i + \epsilon_{ij} \quad \text{where} \quad \epsilon_{ij} \sim N(0, \sigma) \]

- \( X_{ij} \) is the response of the \( j^{th} \) individual in the SRS from the \( i^{th} \) population.
- \( \mu_i \) is mean of \( i^{th} \) population.
- \( \epsilon_{ij} \) is the error term, \( x_{ij} - \mu_i \), also called the deviation of \( x_{ij} \) from \( \mu_i \)
- \( \epsilon_{ij} \sim N(0, \sigma) \) is equivalent to the assumption that the \( i^{th} \) population is normal, \( x_{ij} \sim N(\mu_i, \sigma) \).

The “Estimated” one-way ANOVA Model

\[ X_{ij} = \bar{X}_i + e_{ij} \]

- \( \bar{X}_i \) is the sample mean of the SRS from the \( i^{th} \) population, an unbiased point estimator of \( \mu_i \).
- \( e_{ij} = X_{ij} - \bar{X}_i \) is the residual, the deviation of \( X_{ij} \) from \( \bar{X}_i \)

**HYPOTHESIS TEST TO COMPARE \( k \) POPULATION MEANS**

The Overall Test:

**The Idea:** We will compare the variability among the sample means (\( MSG \)) to the variability within each sample (\( MSE \)).

- If the variability among the \( \bar{x}_i \)'s is large relative to the variability within each sample (\( MSG >> MSE \)), then we will claim that there is a difference among the \( \mu_i \)'s.

- If the variability among the \( \bar{x}_i \)'s is not large relative to the variability within each sample, then we will fail to claim that there is a difference among the \( \mu_i \)'s.

- The statistic \( MSE \) is an unbiased estimator of the constant variance \( \sigma^2 \).

1. Hypotheses:
   \[ H_0: \mu_1 = \mu_2 = \ldots = \mu_k \]
   \[ H_a: \mu_i \neq \mu_j \text{ for some } i \text{ and } j \] (at least one of the \( \mu_i \) are different than the others)

2. Check Assumptions:
   (a) Independent SRS’s have been chosen.
   (b) If the samples were taken without replacement, then \( .05N_i \geq n_i \) for each \( i = 1, 2, \ldots, k \).
   (c) Check normal probability plots to determine if the residuals are not normally distributed.
   (d) Assume that the constant variance assumption holds if \( \frac{\text{largest } s}{\text{smallest } s} < 2 \).
Perform Steps 3 and 4 assuming that $H_0$ is true!

3. Test Statistic:

$$F = \frac{MSG}{MSE},$$

where $MSG$ and $MSE$ are from the following ANOVA table:

\[
\begin{array}{cccccc}
\text{Source} & \text{DF} & \text{Sum of Squares (SS)} & \text{Mean Squares (MS)} & F & \text{p-value} \\
\hline
\text{Treatments} & \text{DFG} = k-1 & \text{SSG} = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2 & \text{MSG} = \frac{\text{SSG}}{\text{DFG}} & F^* = \frac{\text{MSG}}{\text{MSE}} & P(F > F^*) \\
\text{Error} & \text{DFE} = N-k & \text{SSE} = \sum_{i=1}^{k} (n_i - 1) s_i^2 & \text{MSE} = \frac{\text{SSE}}{\text{DFE}} \\
\text{Total} & \text{DFT} = N-1 & \text{SST} = \sum_{\text{all } x} (x_{ij} - \bar{x})^2 & & & \\
\end{array}
\]

where $N = \sum n_i$ is the total sample size and $\bar{x} = \sum n_i \bar{x}_i / \sum n_i$ is the grand mean. Observe that $DFT = DFG + DFE$ and $SST = SSG + SSE$. Because the sums of squares add, a useful summary of the sum of squares is $R^2 = \frac{SSG}{SST}$ which is interpreted as the percentage of variance in the data that is explained by the ANOVA model.

- When $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ is true, $\mu_{MSG} = \mu_{MSE}$ and therefore $F = \frac{MSG}{MSE} \approx 1$.
- When $H_0$ is false, $\mu_{MSG} > \mu_{MSE}$ and therefore $F = \frac{MSG}{MSE} >> 1$.
- Large $F = \frac{MSG}{MSE}$ values are strong evidence against $H_0$ and for $H_a$.

4. $p$-value: The $p$-value $= P(F > F^*)$.

- The test statistic has an $F$ distribution, $F \sim F(DFG, DFE)$, when $H_0$ is true.
- An $F$ distribution $F(DFG, DFE)$ has two parameters, the numerator degrees of freedom $DFG$ and the denominator degrees of freedom $DFE$.
- An $F$ distribution is a beautiful right-skewed distribution. Probabilities can be found using R’s $\text{pf}(F, \text{df1} = \#, \text{df2} = \#, \text{lower.tail=}\text{FALSE})$

function. Unfortunately, your textbook does not have a table of $F$ probabilities.

5. and 6. Make a Decision and give a Conclusion.
MULTIPLE COMPARISONS FOLLOW-UP TEST

If we reject $H_0$ in the overall test and conclude that at least one of the $\mu_i$’s is different that the others, then we should ask “Which population means are different?”

Only do a follow-up test if you REJECT $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$!

Tukey’s Method (your book suggests Bonferroni’s Method)

Tukey’s Method calculates a family of CI’s for all possible pairwise differences of the means $\mu_1, \mu_2, \ldots, \mu_k$. The overall family-wise confidence level is held at some confidence level $C = 1 - \alpha$ (which means that the confidence level for each individual CI is more than $C$). For a given pair $\mu_i$ and $\mu_j$ (with $i \neq j$):

- The Tukey’s confidence interval for $\mu_i - \mu_j$ is:

$$\bar{x}_i - \bar{x}_j \pm q_{1-\alpha, k, DFE} \sqrt{\frac{MSE}{n_i} + \frac{MSE}{n_j}}$$

- The Tukey critical value $q_{1-\alpha, k, DFE}$ can be found using R’s

  TukeyHSD( )

- The Tukey’s CI for $\mu_i - \mu_j$ can be used to test the hypothesis $H_0 : \mu_i - \mu_j = 0$ versus $H_a : \mu_i - \mu_j \neq 0$.
  - If 0 is in the CI, then fail to reject $H_0$.
  - If 0 is not in the CI, then reject $H_0$.

EXAMPLE:
An archeologist is interested in studying skull breadths of humans from different epochs. Significant changes in head shape over time would suggest that interbreeding occurred with immigrant human populations. A sample of 27 head breadths were obtained by measuring skulls of Egyptian males from three different epochs: 4000BC, 1850BC, and 150AD. The data are from Ancient Races of the Thebaid, by Thomson and Randall-Maciver).

Display Your Data

```r
> D = read.table("http://www.math.montana.edu/parker/courses/STAT401/data/Chapter5.5.headbreadth.txt", header=T)
> attach(D)

# Put the categories in chronological order
> Epoch = factor(as.character(Epoch), levels = c("4000BC","1850BC","150AD"))

# Generate a beautiful side-by-side bixplot
> boxplot(HeadBreadth ~ Epoch,ylab="Head Breadth")
```

![Boxplot of head breadths by epoch](image)
Fit the One-way ANOVA Model

```r
> hb.aov = aov(HeadBreadth~Epoch)
> summary(hb.aov)

Analysis of Variance Table

Response: HeadBreadth
       Df Sum Sq Mean Sq F value    Pr(>F)
Epoch    2 138.74  69.37   4.0497 0.03052 *
Residuals 24 411.11  17.13
---
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

```

Report as a beautiful table:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>138.74</td>
<td>69.37</td>
<td>4.0497</td>
<td>0.03052</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>411.11</td>
<td>17.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>549.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# Check ANOVA table by hand

# Group means
```r
> xbari = tapply(HeadBreadth,Epoch,mean)
> xbari
4000BC 1850BC 150AD
132.6667 134.4444 138.1111
```

# Group variances
```r
> vari=tapply(HeadBreadth,Epoch,var)
> vari
4000BC 1850BC 150AD
17.50000 11.27778 22.61111
```

# Group sample sizes
```r
> ni = tapply(HeadBreadth,Epoch,length)
> ni
4000BC 1850BC 150AD
9 9 9
```

# Grand mean
```r
> gmean = mean(HeadBreadth)
[1] 135.0741
```

# SSG
```r
> sum(ni*(xbari-gmean)^2)
[1] 138.7407
```

# SSE
```r
> sum((ni-1)*vari)
[1] 411.1111
```

# p-value
```r
> 1-pf(4.0497,2,24)
[1] 0.03051911
```
Check Your Assumptions

1. **Independent Random Samples**

2. **Normal Distribution**

```r
> par(mfrow=c(1,2))
> qqnorm(studres(hb.aov))
> qqline(studres(hb.aov))
> hist(studres(hb.aov),freq=FALSE,ylim=c(0,0.45),main="Density Plot of Residuals",
  xlab="Studentized Residuals")
> lines(density(studres(hb.aov)))
> xy=qqnorm(studres(hb.aov))
> cor(xy$x,xy$y)
[1] 0.9722762
```

![Normal Q-Q Plot and Density Plot of Residuals]

3. **Check Constant Variance**

```r
> tapply(HeadBreadth,Epoch,sd)
4000BC  1850BC  150AD
4.183300 3.358240 4.755114
```

\[
\frac{\text{largest } s}{\text{smallest } s} \approx \frac{4.76}{3.35} \approx 1.42 < 2, \text{ so the constant variance assumption appears to hold.}
\]

1. Perform the **overall Hypothesis Test**

   (a) **Hypotheses:**

   (b) **Test statistic value:**

   (c) **Distribution of the test statistic** given that \( H_0 \) is true:

   (d) **\( p \)-value:**

   (e) **Decision at** \( \alpha = .05 \):
2. Give an unbiased estimate of the constant variance $\sigma^2$.

3. Perform Tukey’s Multiple Comparison Test

```r
> TukeyHSD(hb.aov,which="Epoch",conf.level=0.95)
Tukey multiple comparisons of means
  95% family-wise confidence level

Fit: aov(formula = HeadBreadth ~ Epoch)

$\Epoch$

<table>
<thead>
<tr>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850BC-150AD</td>
<td>-3.6667</td>
<td>-8.5390</td>
<td>0.1664</td>
</tr>
<tr>
<td>4000BC-150AD</td>
<td>-5.4444</td>
<td>-0.5722</td>
<td>0.0265</td>
</tr>
<tr>
<td>4000BC-1850BC</td>
<td>-1.7789</td>
<td>3.0945</td>
<td>0.6386</td>
</tr>
</tbody>
</table>

(a) **Conclusions:**

(b) Which epoch appears to have the largest mean head breadth? How much larger is the head breadth during this epoch?

---

**Exercises**

On p269: 15.41 - 15.43 (all), 15.44 (b $DFG = 4$, $DFE = 40$; c The evidence suggests that there is a difference in exam scores, on the average, among the 5 teaching methods), 15.45, 15.47, 15.51