Chapter 6 - Linear combinations of means

6.1 CASE STUDY

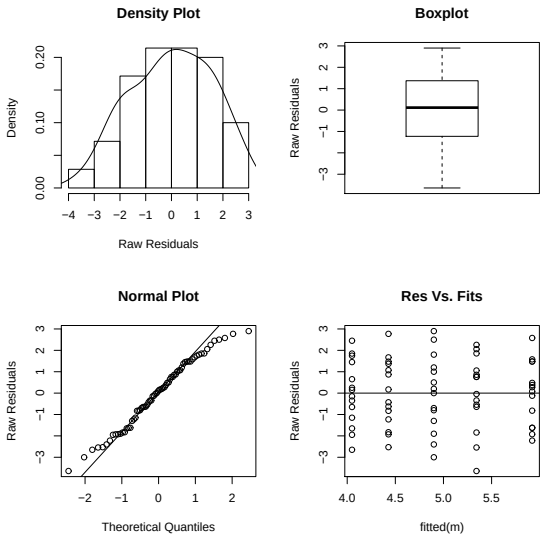
Discrimination Against the Handicapped

Five videotaped job interviews were shown to $n = 70$ undergraduate students, with $n_i = 14$ viewing each interview. Two male actors were involved in each interview (the same two for each of the 5). A set script was followed. The only difference is that the "applicant" in each case appeared with a different handicap; amputee, wheelchair, crutches, hearing, with a control of no apparent handicap. Subjects reviewed the tapes and then rated the applicants suitability for the job on a 0 to 10 point scale. Your book and these notes use Y_{ij} to denote the score of the j^{th} student for the i^{th} interview. To be clear - the only difference between the tapes is the handicap.

Question of Interest: Do subjects systematically evaluate qualifications differently according to the candidate's handicap? If so, which handicaps produce the different evaluations?

```
# Setup
library(Sleuth3)
source("http://www.math.montana.edu/parker/courses/STAT411/diagANOVA.r")

# Import the data
d<-case0601
summary(d)
##     Score             Handicap
## Min.   :1.400   Amputee   :14
## 1st Qu.:3.700   Crutches  :14
## Median :5.050   Hearing   :14
## Mean   :4.929   None      :14
## 3rd Qu.:6.100   Wheelchair:14
## Max.   :8.500

# Create a table that summarizes the data by group
n<-tapply(d$Score,d$Handicap,length)
Mean<-tapply(d$Score,d$Handicap,mean)
SD<-tapply(d$Score,d$Handicap,sd)
cbind(Mean,SD,n)
##               Mean       SD  n
##Amputee    4.428571 1.585719 14
##Crutches   5.921429 1.481776 14
##Hearing    4.050000 1.532595 14
##None       4.900000 1.793578 14
##Wheelchair 5.342857 1.748280 14

# One-way ANOVA using aov() so we can perform Tukey's follow-up (NOT lm()!)
m<-aov(Score~Handicap,data=d)
anova(m)
##Analysis of Variance Table
##Response: Score
##          Df  Sum Sq Mean Sq F value  Pr(>F)
##Handicap   4  30.521  7.6304  2.8616 0.03013 *
```

```
##Residuals 65 173.321  2.6665

diagANOVA(m)
##[1] "In this sample of size n=70, correlation of the residuals in the qq-plot is r=0.991457"
##[1] "In the following table, if r < critical.r, then the qq-plot suggests the residuals
##     are not normal:"
##    n critical.r
##1   5      0.832
##2  10      0.880
##3  15      0.911
##4  20      0.929
##5  25      0.941
##6  30      0.949
##7  40      0.960
##8  50      0.966
##9  60      0.971
##10 75      0.976
```



Identify the group means by μ_i and the sample means by $\overline{Y}_i$ for $i = 1$ through 5:

Group	**Population mean**	**Sample mean**
Amputee	μ_1	$\overline{Y}_1$
Crutches	μ_2	$\overline{Y}_2$
Hearing	μ_3	$\overline{Y}_3$
None	μ_4	$\overline{Y}_4$
Wheelchair	μ_5	$\overline{Y}_5$

1. Hypotheses:

2. Check assumptions:

3. Test statistic value:

4. Distribution of the test statistic and p-value given that H_0 is true:

5. Decision at $\alpha = .05$:

6. Conclusion:

7. Construct a follow-up 95% t-CI of the mean score of the interviewee in a wheelchair vs the mean score of the interviewee with no handicap (i.e., $\mu_5 - \mu_4$). *Hint:*

```
qt(.975,65)
##[1]  1.997138
```

```
5.342857 - 4.900000 + c(-1,1)*qt(.975,65)*sqrt(2.6665*(2/14))}
##[1] -0.7897648  1.6754788.
```

6.2 LINEAR COMBINATIONS

- Questions of interest often involve several group means, not just two. How about the difference between the averages of multiple group means.

$$\frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4 + \mu_5}{3}$$

For the handicap example, this linear combination would be interpreted as the mean of the amputee and crutches scores (populations 1 and 2) minus the mean of the Hearing, None and Wheelchair scores (populations 3-5).

- A **linear combination** of group means is written as

$$\gamma = C_1\mu_1 + C_2\mu_2 + \ldots + C_I\mu_I$$

where $C_1, C_2, \ldots, C_I$ are fixed *coefficients* chosen by the researcher.

- The linear combination γ is specified by a Greek letter (for "g") to emphasize that it is a parameter that we want to estimate. We estimate γ in the obvious way - replace the population means with their corresponding sample averages. This estimator is g:

$$g = C_1\overline{Y}_1 + C_2\overline{Y}_2 + \ldots + C_I\overline{Y}_I$$

- The standard deviation of g, under the assumptions that each population has the same variance σ^2 and that the groups are independent, is

$$SD(g) = \sqrt{\sigma^2\left(\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_I^2}{n_I}\right)}$$

- As usual, we will estimate $SD(g)$ by replacing σ^2 with the its pooled estimator $s_p^2 = MSF = \frac{(n_1-1)s_1^2+\cdots+(n_I-1)s_I^2}{n-I}$. The result is the standard error for g:

$$SE(g) = \sqrt{MSF\left(\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_I^2}{n_I}\right)}$$

- A CI for the linear combination γ is:

$$\boxed{g \pm t_{1-\alpha/2,df=DFF}\sqrt{MSF\left(\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_I^2}{n_I}\right)}}$$

For one-way ANOVA, $DFF = n - I$.

- To test $H_0 : \gamma = \gamma_0$ vs. $H_a : \gamma \neq, <, > \gamma_0$, the test statistic is the t-ratio:

$$t = \frac{g - \gamma_0}{SE(g)}$$

with $df = DFF$. For one-way ANOVA, $DFF = n - I$.

EXAMPLE: Compare the wheelchair/crutches mean with the amputee/hearing mean (see page 155 in the text and Display 6.4). Do the data suggest a real difference between these two aggregate groups?

The parameter γ and the estimator g: The linear combination of the 5 group means is:

$$\begin{aligned}\gamma &= C_1\mu_1 + C_2\mu_2 + C_3\mu_3 + C_4\mu_4 + C_5\mu_5 \\ &= -(1/2)\mu_1 + (1/2)\mu_2 - (1/2)\mu_3 + 0\mu_4 + (1/2)\mu_5 \\ &= \frac{(\mu_2+\mu_5)}{2} - \frac{(\mu_1+\mu_3)}{2}.\end{aligned}$$

We will estimate this by $g = \frac{(\overline{Y}_2+\overline{Y}_5)}{2} - \frac{(\overline{Y}_1+\overline{Y}_3)}{2}$.

R-code:

```
# Create a vector containing the coefficients
LC.vec <- c(-1/2, 1/2, -1/2, 0, 1/2)

# Estimate the linear combination g using LC.vec and the vector Mean from above
g<-sum(Mean*LC.vec)
g
##[1] 1.392857

# Get the pooled variance = MSF
DFF <- m$df.residual
SSF <- sum(m$residuals^2)
MSF <- SSF/DFF
MSF = sum((n - 1)*SD^2)/DFF  # Or calc. the pooled variance = MSF this way
cbind(DFF,SSF,MSF)   # Compare with the "Residuals" row of the ANOVA table above
##     DFF      SSF      MSF
##[1,]  65 173.3214 2.666484

# Now we can get SE(g)
se.g <- sqrt(MSF*sum(LC.vec^2*1/n))
se.g
##[1] 0.4364208

# CI for true LC gamma
g + c(-1,1)*qt(.975,DFF)*se.g
##[1] 0.5212646 2.2644497   # Same as in Display 6.4

# Test Ho: gamma = gamma0 vs Ha: gamma not equal gamma0
```

```
gamma0=0                    # value of gamma under Ho
t.stat <- (g-gamma0)/se.g   # test stat
2*(1-pt(abs(t.stat),DFF))       # two-sided p value
##[1] 0.002180647

# Let R do all the work
require(gmodels)  # You may have to install this from CRAN
m2 = aov(Score ~ Handicap - 1,data=d)  # The "-1" refits the model without an intercept
estimable(m2,LC.vec,conf.int=.95)
##                       Estimate Std. Error  t value DF    Pr(>|t|)  Lower.CI Upper.CI
##(-0.5 0.5 -0.5 0 0.5) 1.392857  0.4364208 3.191546 65 0.002180647 0.5212646  2.26445
```

6.3 SIMULTANEOUS INFERENCE

Simultaneous testing

- Multiple tests performed simultaneously are a **family of tests**.
- **Individual significance level:** α_1: the chance that 1 test incorrectly rejects H_0.
- **Family-wise significance level** α: the chance that at least one test incorrectly rejects H_0 when a family of tests are performed.
- **Compound uncertainty**: the increased chance of making at least one mistake when performing more than one test.
  - When you perform a family of many independent tests each at a significance level $\alpha_1 = 5\%$, then $\alpha_1 = 5\%$ of these tests will incorrectly reject H_0 when H_0 is true.
  - When you perform a family of tests, each at $\alpha_1 = 5\%$, the chance that *at least one* of the tests incorrectly rejects H_0 is LARGER than $\alpha_1 = 5\%$ (familywise $\alpha > \alpha_1$).
  - The greater the number of tests performed using an individual significance level α_1, the higher the chance that a low p-value $< \alpha_1$ will be found for at least one of the tests when H_0 is true.
  - You'll "discover" differences that aren't real more often than α_1.
  - If the tests performed at an individual significance level α_1 are independent, then the actual familywise significance level over k tests is $\alpha = 1 - (1 - \alpha_1)^k$.
    - For $k = 2$ independent tests each at $\alpha_1 = 5\%$, familywise $\alpha = 9.8\%$; there's a 10% chance that at least one of the 2 tests will incorrectly reject H_0.
    - For $k = 10$ independent tests each at $\alpha_1 = 5\%$, $\alpha = 40.1\%$; there's a whopping 40% chance that at least one of the 10 tests will incorrectly reject H_0. So your tests have a high chance of "working" for each test individually but the family of tests does not have a high chance of working on all 10 parameters at the same time.

Simultaneous CIs

- Multiple CIs constructed simultaneously are a **family of CIs**.
- **Individual confidence level** $C_1 = 1 - \alpha_1$: probability that a single CI captures its parameter.
- **Family-wise confidence level** $C = 1 - \alpha$: The chance that all of the CIs in a family of CIs capture their parameters.

- **Compound uncertainty**: the increased chance of making at least one mistake when constructing more than one CI.
  - When you construct a family of independent 95% CIs, then 95% of these CIs will correctly capture their parameters.
  - When you construct a family of 95% CIs, the chance that *all* the CIs *simultaneously* capture their parameters is SMALLER than 95% ($C < C_1$).
  - The greater the number of CIs, the higher the chance that one of the CIs will not capture the parameter.
  - If the family of 95% CIs are independent, then the actual familywise confidence level over k CIs is $C = 0.95^k$. In general, $C = C_1^k$.
    - For $k = 2$ independent 95% CIs, the familywise confidence level is $C = 90.3\%$; there's a 90% chance that both CIs capture their parameters.
    - For $k = 10$ independent 95% CIs, the familywise confidence level is $C = 59.9\%$; there's a 60% chance that all 10 CIs will capture their parameters. So your CIs have a high chance of "working" for each parameter individually but the family of CIs do not have a high chance of working on all 10 parameters at the same time.

- **Multiple Comparison Procedures**: ways of constructing individual CIs so that the familywise confidence level (C) is controlled at a specified level when needed. This involves making the individual confidence levels HIGHER (and the CIs wider!) to ensure the familywise confidence level is at least $C\%$.

Planned vs. Unplanned comparisons

- **One or two Planned comparisons**: Researcher knows BEFORE SEEING THE DATA that one or two comparisons will be performed $\longrightarrow$ use *individual* significance and confidence levels.

- **Many Planned comparisons**: Researcher knows BEFORE SEEING THE DATA that many comparisons will be performed $\longrightarrow$ use *familywise* significance and confidence levels over the family of planned tests/CIs.

- **Unplanned comparisons**: Examine many possible pairs of differences $\longrightarrow$ use *familywise* significance and confidence levels over the family of all pairwise comparisons.

- **Data snooping**: The comparison chosen originates from looking at the data. For example, looking at a plot of the groups or the individual group means and *then* choose the two most different to test $\longrightarrow$ use *familywise* significance and confidence levels over the family of all pairwise comparisons.

  After data snooping you may only want to make a single comparison - e.g., compare the means associated with the largest and smallest sample averages. Nonetheless, you should control the familywise significance and confidence levels as if you were performing all pairwise comparison!

EXAMPLE: One *planned test* of interest to the researchers who conducted the handicap study was whether those with a handicap received lower scores, on average, compared to those without a handicap.

1. Write out the linear combination of interest. *Hint:* Recall the table:

Group	Population mean	Sample mean
Amputee	μ_1	$\overline{Y}_1$
Crutches	μ_2	$\overline{Y}_2$
Hearing	μ_3	$\overline{Y}_3$
None	μ_4	$\overline{Y}_4$
Wheelchair	μ_5	$\overline{Y}_5$

2. Hypotheses:

3. Check assumptions:

4. Test statistic value:

5. Distribution of the test statistic and p-value given that H_0 is true:

6. Decision at $\alpha = .05$:

7. Conclusion:

```
# Relevant R-code
estimable(m2,c(1/4,1/4,1/4,-1,1/4),conf.int=.95)
##                            Estimate Std. Error    t value DF  Pr(>|t|)   Lower.CI Upper.CI
##(0.25 0.25 0.25 -1 0.25) 0.03571429  0.4879333 0.07319502 65 0.9418757 -0.9387558 1.010184
```

6.4 MULTIPLE COMPARISON FOLLOW-UP PROCEDURES

If you reject

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$$

after the ANOVA F-test and conclude that at least one of the μ_i's is different than the others, then you can "follow-up" with a multiple comparison procedure to ask "*Which* pair(s) of population means are different?"

These notes consider two common follow-up tests: **Tukey's** (called **Tukey-Kramer** when the sample sizes are unequal) and **Bonferroni's** (your book considers others).

Regardless of the multiple comparison procedure, the overall familywise confidence level is held at $C = 1 - \alpha$ by setting significance level for each individual test at $\alpha_1 < \alpha$ and/or setting the confidence level for each individual CI at $C_1 > C$.

The difference between the multiple comparison procedures is how the individual significance level α_1 and the individual confidence level C_1 are set.

6.4.1 Tukey's Honest Significant Difference (HSD) Procedure

Tukey's Method controls the familywise significance level of the family of tests for all pairwise differences of the means $\mu_1, \mu_2, ...\mu_I$. Equivalently, Tukey's controls the familywise confidence level of the family of CIs for all pairwise differences of the means $\mu_1, \mu_2, ...\mu_I$.

Setting:

- An ANOVA has been performed, so you have an estimate of σ^2 (the pooled sample variance $s_p^2 = MSF$) and estimates of the group means μ_i (the sample means $\overline{Y}_i$). Tukey's procedure requires MSF, DFF and $\overline{Y}_i$.
- Perform **Tukey's** follow-up tests when you want to maintain a familywise significance and confidence level over the family of all pairwise comparisons among group means.
- Only do Tukey's follow-up procedure if ANOVA REJECTS $H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$! However, some statisticians argue that you can perform Tukey's regardless of the outcome of the F-test. If you do this, then you could face the conundrum that the ANOVA found no difference in means but Tukey's does! It is also possible that ANOVA led you to reject H_0 but then Tukey's follow-up tests fail to find any pairwise difference in means. The conclusion then is that there are no pairwise differences in means!

Tukey's multiple comparison procedure is:

- Check the assumptions! These are the same as for ANOVA: RSs from each group; groups are independent of each other; groups are normal; groups have constant variance.
- Tukey's test statistic to test $H_0 : \mu_i = \mu_j$ vs. $H_a : \mu_i \neq \mu_j$ is $\boxed{q = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{MSF}{2}(\frac{1}{n_i} + \frac{1}{n_j})}}}$ The p-value is calculated using the upper tail of a **studentized range distribution**. The p-value can be found using R's `ptukey()` function.
- The Tukey's family of CIs for all pairwise comparisons $\mu_i - \mu_j$ that maintain a familywise confidence level of $C = 1 - \alpha$ is:

$$\boxed{\bar{x}_i - \bar{x}_j \pm q_{1-\alpha,I,DFF}\sqrt{\frac{MSF}{2}(\frac{1}{n_i} + \frac{1}{n_j})}}$$

  where $q_{1-\alpha,I,DFF}$ is the $1-\alpha$ percentile from a studentized range distribution (as opposed to a z- or a t-distribution as we have done up until now); $q_{1-\alpha,I,DFF}$ can be found by R's `qtukey()` function.
- The Tukey p-values and CIs for all pairwise comparisons can be found using R's `TukeyHSD( )` function.

EXAMPLE: For the Handicap experiment case study from section 6.1, recall the *Question of Interest*: Do subjects systematically evaluate qualifications differently according to the candidate's handicap? If so, which handicaps produce the different evaluations?

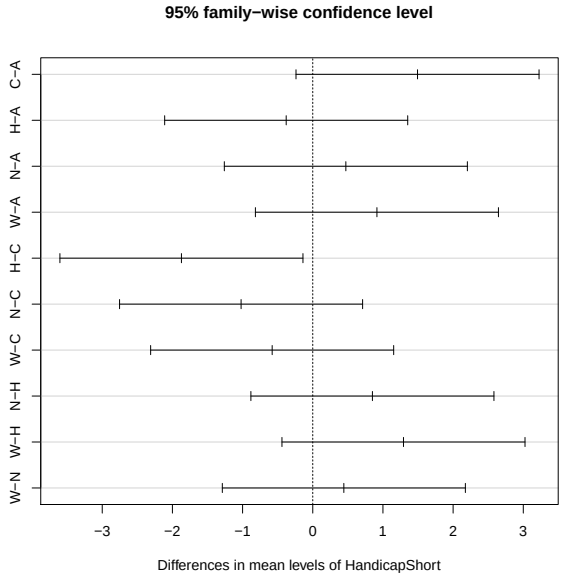
Why is it appropriate to test for any pairwise difference in the group means?:

Explain why it is appropriate to use a multiple comparison procedure such as Tukey's when answering the question of interest?

The following **R-code** performs Tukey's at a familywise significance level of $\alpha = 0.05$ to determine which, if any, of the handicap treatment group means are different.

```
# Recall the ANOVA:
m<-aov(Score~Handicap,data=d)

# We will need these results from the calculations above
Mean
##   Amputee   Crutches    Hearing       None Wheelchair
##  4.428571   5.921429   4.050000   4.900000   5.342857
cbind(SSF,MSF)
##          SSF      MSF
##[1,] 173.3214 2.666484

# Tukey test of  Ha: muCrutches - muAmputee not equal 0
# test stat
q=1.4928571/sqrt(MSF/2*2/14)
q
##[1] 3.420683

# p-value - Like an F-test, the p-value is an upper tail
1-ptukey(q,5,DFF)
##[1] 0.1232819

# Tukey's two-sided 95% CI for muCrutches - muAmputee
Mean[2]-Mean[1]+ c(-1,1)*qtukey(.95,5,DFF)*sqrt(MSF/2*2/14)
##[1] -0.2388756  3.2245899

###
# Let R do the tests and CIs for all pairwise comparisons
TukeyHSD(m,conf.level=.95)
##  Tukey multiple comparisons of means
##
##    95% family-wise confidence level
##
##$Handicap
##                           diff        lwr        upr     p adj
##Crutches-Amputee     1.4928571 -0.2388756  3.2245899 0.1232819   # agrees with calcs above
##Hearing-Amputee     -0.3785714 -2.1103042  1.3531613 0.9724743
##None-Amputee         0.4714286 -1.2603042  2.2031613 0.9399911
##Wheelchair-Amputee   0.9142857 -0.8174470  2.6460185 0.5781165
##Hearing-Crutches    -1.8714286 -3.6031613 -0.1396958 0.0277842
##None-Crutches       -1.0214286 -2.7531613  0.7103042 0.4686233
##Wheelchair-Crutches -0.5785714 -2.3103042  1.1531613 0.8812293
##None-Hearing         0.8500000 -0.8817328  2.5817328 0.6442517
##Wheelchair-Hearing   1.2928571 -0.4388756  3.0245899 0.2348141
##Wheelchair-None      0.4428571 -1.2888756  2.1745899 0.9517374

# Generate a lovely plot of the CIs
# You could try plot(TukeyHSD(m,conf.level=.95)) but the group names are too long
# Instead let's assign short names to the groups then plot

levels(d$Handicap)             # Check the alpha-numeric order of the group levels
```

```
##[1] "Amputee"    "Crutches"   "Hearing"    "None"       "Wheelchair"

d$HandicapShort = d$Handicap    # New variable
levels(d$HandicapShort) = c("A","C","H","N","W")    # New names in same alpha-numeric order
m2 = aov(Score ~ HandicapShort,data=d)              # refit ANOVA
plot(TukeyHSD(m2))

# Performing all pairwise comparisons without any familywise significance level
#   p-values are much smaller than Tukey's p-values above
pairwise.t.test(d$Score,d$Handicap,p.adjust.method="none")
##        Pairwise comparisons using t tests with pooled SD
##
##           Amputee Crutches Hearing None
##Crutches   0.0184  -        -       -
##Hearing    0.5418  0.0035   -       -
##None       0.4477  0.1028   0.1732  -
##Wheelchair 0.1433  0.3520   0.0401  0.4756
##
##P value adjustment method: none
```



1. Does any group appear to have a statistically significantly larger mean score than *all* of the others?

2. Which groups are statistically significantly different?

3. Make sure that you understand how to perform the tests of $H_0 : \mu_i = \mu_j$ vs. $H_a : \mu_i \neq \mu_j$ using either the p-values or the CIs. You will get the same results using either approach!

4. Conclusion:

5. Scope of Inference:

6.4.4 Bonferroni's Multiple Comparison Procedure

Bonferroni's is a very general method that can be applied to control the familywise significance or confidence level for ANY family of tests or CIs, even when the tests or CIs pertain to groups that are NOT independent. Tukey's on the other hand is specifically for pairwise comparisons among independent groups.

Setting:

- You might want to use **Bonferroni's instead of Tukey's after ANOVA** to maintain a familywise significance and confidence levels for a family of pairwise comparisons if there are only a few pairwise comparisons among group means.
  - For example, when there are only 3 groups, then there are only 3 pairwise comparisons, and Bonferroni's is a good approach to maintaining the familwise significance and confidence level.
  - To perform Bonferroni's after ANOVA, you'll need an estimate of σ^2 (the pooled sample variance $s_p^2 = MSF$), estimates of the group means μ_i (the sample means $\overline{Y}_i$), and DFF.
  - If there are many pairwise comparisons and for some reason you still don't want to use Tukey's, consider **Dunnett's** in section 6.4.2.
- Bonferroni's is a good choice to maintain a familywise signficance and confidence level for a family of a few linear combinations. If there are many linear combinations, use **Scheffe's** (in 6.4.3).
- Consider using Bonferroni's to maintain familywise significance and confidence levels for any family of tests, not just follow-up tests to an ANOVA. For example, you can apply Bonferroni's to:
  - Permutation or randomization tests and CIs
  - Regression tests and CIs
  - χ^2 tests
  - Likelihood ratio tests
  - Bayesian credible intervals

Bonferroni's Method is simple to implement:

- Bonferroni's maintains a familywise significance level of α for a family of k tests by performing each of the tests at an individual significance level of $\alpha_1 = \alpha/k$. In other words:
  - Calculate each test statistic as you normally would (e.g., a t, z, F, χ^2 or randomization test statistic)
  - Calculate each individual p-value as you normally would (e.g., from a t, z, F, χ^2 or randomization distributon)

  - When it is time to decide whether to REJECT H_0, compare $k\times$(individual p-value) to the familywise significance level α. If $k\times$(individual p-value) $< \alpha$ then REJECT H_0. Otherwise FTR H_0. The quantity $k\times$(individual p-value) is called a *Bonferroni-adjusted p-value*.

- Bonferroni's maintains a familywise confidence level of $C = 1 - \alpha$ for a family of k CI's by constructing each of the CIs at an individual confidence level of $C_1 = 1 - \alpha/k$. In other words, calculate each CI as you normally would except that you need to use a confidence level of $1 - \alpha/k$ for each CI!

EXAMPLE:
Recall the 27 skull breadths of Egyptian males from three different epochs: 4000BC, 1850BC, and 150AD from Chapter 5 notes.

The researcher wants to determine which Epoch's headbreadths are different on the average. Why is it appropriate to test for any pairwise difference in the group means?:

Explain why it is appropriate to use a multiple comparison procedure such as Tukey's when answering the researcher's question of interest?

The following **R-code** performs Tukey's at a familywise significance level of $\alpha = 0.05$ to determine which, if any, of the Epoch headbreadth means are different.

```
D = read.table("http://www.math.montana.edu/parker/courses/STAT411/Chapter5.skulls.txt",
      header=T)
D$Epoch = factor(D$Epoch, levels = c("4000BC","1850BC","150AD"))
hb.aov=aov(HeadBreadth~Epoch,data=D)
anova(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

# Perform Tukey's Multiple Comparison Tests and CIs at familywise confidence 95%
plot(TukeyHSD(hb.aov))   # This useful plot not shown
TukeyHSD(hb.aov)
##  Tukey multiple comparisons of means
##
##    95% family-wise confidence level
##                   diff        lwr       upr     p adj
##1850BC-4000BC 1.777778 -3.0945467  6.650102 0.6386242
##150AD-4000BC  5.444444  0.5721199 10.316769 0.0264650
##150AD-1850BC  3.666667 -1.2056579  8.538991 0.1664028

###
# Construct Bonferroni t-CIs for all pairwise comparisons at familywise confidence 95%
MSF = 17.13
```

```
DFF = 24
alpha=0.05
k = 3  # number of pairwise comparisons
1.777778 + c(-1,1)*qt(1-alpha/(2*k),DFF)*sqrt(MSF*(1/9 + 1/9))  #1850BC-4000BC
##[1] -3.243571  6.799127

5.444444 + c(-1,1)*qt(1-alpha/(2*k),DFF)*sqrt(MSF*(1/9 + 1/9))  #150AD-4000BC
##[1]  0.4230952 10.4657928

3.666667 + c(-1,1)*qt(1-alpha/(2*k),DFF)*sqrt(MSF*(1/9 + 1/9))  #150AD-1850BC
##[1] -1.354682  8.688016

###
# Perform Bonferroni t-tests for all pairwise comparisons at familywise signficance 5%
# The reported Bonferroni-adjusted p-values are simply k*(individual p-value)
pairwise.t.test(D$HeadBreadth,D$Epoch,p.adjust.method="bonferroni")
##        Pairwise comparisons using t tests with pooled SD
##
##       4000BC 1850BC
##1850BC 1.00   -
##150AD  0.03   0.22
##P value adjustment method: bonferroni

# Perform t-tests for all pairwise comparisons without any familywise significance level
# The reported p-values are individual p-values
pairwise.t.test(D$HeadBreadth,D$Epoch,p.adjust.method="none")
##        Pairwise comparisons using t tests with pooled SD
##
##       4000BC 1850BC
##1850BC 0.371  -
##150AD  0.010  0.072
##P value adjustment method: none

# Perform Bonferroni correction that would work for any set of 3 tests:
# This agrees with Bonferroni results above
p.adjust(c(0.371,.01,.072),method="bonferroni")
##[1] 1.000 0.030 0.216
```

1. Which epoch appears to have the largest mean head breadth? How much larger is the head breadth during this epoch?

2. **Conclusions:**

3. **Scope of Inference::**