TWO-WAY Anova Homework
Due Wednesday, September 30th

Gender Differences in Performance on Mathematics Achievement Tests Data set on 861 ACT Assessment Mathematics Usage Test scores from 1987. The test was given to a sample of high school seniors who met one of three profiles of high school mathematics course work: (a) Algebra I only; (b) two Algebra courses and Geometry; and (c) two Algebra courses, Geometry, Trigonometry, Advanced Mathematics and Beginning Calculus.
These data were generated from summary statistics for one particular form of the test as reported by Doolittle (1989). *Source:*Ramsey, F.L. and Schafer, D.W. (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis (2nd ed)*, Duxbury.

Summary statistics, side-by-side boxplots, and interaction plots are given for these data.

```
options(show.signif.stars = F)
require(Sleuth2)
require(mosaic)
math <- ex1320
names(math)

## [1] "Sex"        "Background" "Score"
```

1. Is the design balanced? See the output below.

```
favstats(Score~Sex+Background, data = math)

##   Sex.Background min Q1 median   Q3 max  mean   sd   n missing
## 1       female.a   0  6    9.5 12.0  18  9.07 4.19  82       0
## 2         male.a   2  7   12.0 15.0  25 11.46 5.09  48       0
## 3       female.b   0 11   14.0 17.0  28 13.96 5.00 387       0
## 4         male.b   2 12   16.0 19.0  29 15.57 4.89 223       0
## 5       female.c  15 21   25.0 28.0  34 24.63 4.85  54       0
## 6         male.c  13 22   25.0 30.5  36 25.43 5.55  67       0
```

2. What do the side-by-side boxplots tell you about the effects of *Sex* and *Background*? In other words, describe the relationships you see in the boxplots.

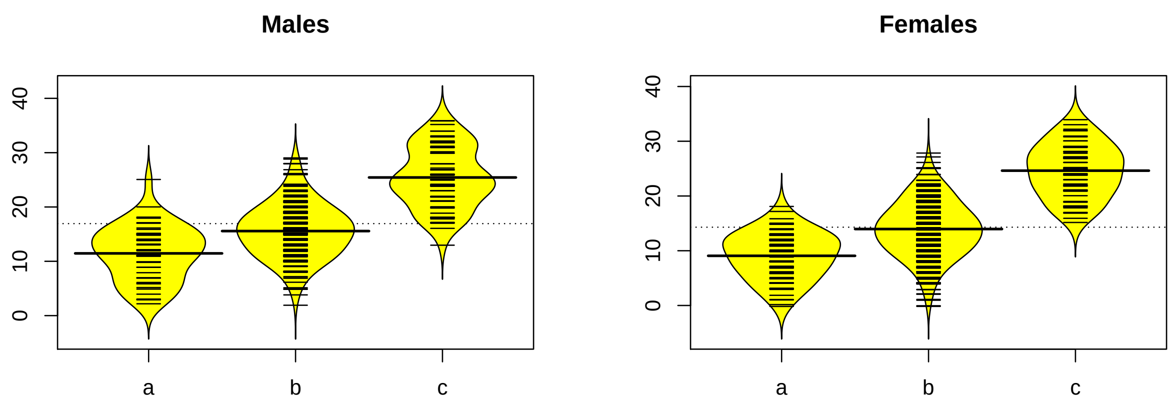
   (a) For both males and females, the average score is highest for background a, and males have higher average scores than females for backgrounds a and b. The average scores for males and females with background c are about the same.

   (b) For both males and females, the average score is highest for background c, and females have higher average scores than males for backgrounds a and b. The average scores for males and females with background c are about the same.

   (c) For both males and females, the average score is highest for background c, and males have higher average scores than females across all backgrounds.

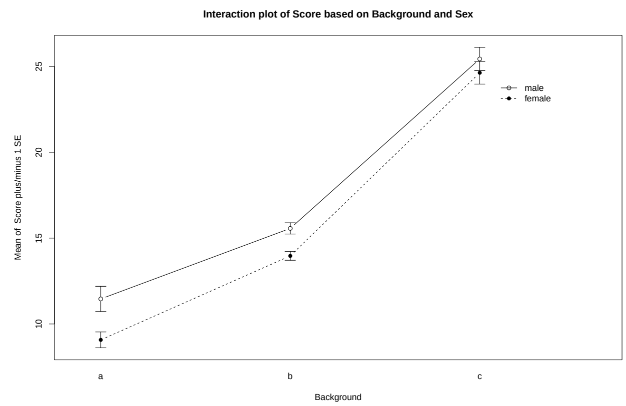
   (d) For both males and females, the average score is highest for background c, and males have higher average scores than females for backgrounds a and b. The average scores for males and females with background c are about the same.

```
require(beanplot)
par(mfrow=c(1,2))
beanplot(Score~Background, data = subset(math, Sex=="male"), method = "jitter", log = "", col = 7, m
beanplot(Score~Background, data = subset(math, Sex=="female"), method = "jitter", log = "", col = 7,
```



3. Interpret an interaction in this context.
   (a) There is an interaction if the difference in mean scores between males and females changes across backgrounds.
   (b) There is an interaction if males have higher scores than females on average.
   (c) There is an interaction if the true mean score for background c is higher than the true mean scores for backgrounds a and b.
   (d) There is an interaction if the interaction plot shows parallel lines.
4. Does the plot below suggest that there is an interaction between *Sex* and *Background*?
   (a) Yes because males appear to be higher than females
   (b) Yes because background c appears to be higher than backgrounds a or b.
   (c) No because the lines are parallel. The difference in average scores between males and females does not change much across levels of background.
   (d) No because it would not make sense to have an interaction in this context.

```
require(HH)
source("~/Documents/Stat217Fall2015/exams/exam1/intplot.R")
intplot(Score~Sex*Background, data=math)
```



5. Write out the two-way anova interaction model for this scenario in terms of y_{ijk}'s, τ_j's, γ_k's and ω_{jk}'s. Define all of the parameters and don't forget the errors!

6. Below is output for the two-way ANOVA model with an interaction.

```
fit.Math <- lm(Score~Background*Sex, data = math)
anova(fit.Math)

## Analysis of Variance Table
##
## Response: Score
##                 Df Sum Sq Mean Sq F value  Pr(>F)
## Background       2  15619    7809  319.83 < 2e-16
## Sex              1    517     517   21.16 4.9e-06
## Background:Sex   2     38      19    0.77    0.46
## Residuals      855  20877      24
```

(a) Write the null and alternative hypothesis for the test you should look at first.

i. $H_0 : \tau_1 = \tau_2 = 0$ vs. at least one $\tau_j \neq 0$

ii. $H_0 : \omega_{11} = \omega_{12} = \omega_{13} = \omega_{21} = \omega_{22} = \omega_{23} = 0$ vs. H_A : at least one $\omega_{jk} \neq 0$

iii. $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$ vs. H_A : at least one $\gamma_k \neq 0$

iv. H_0 : there is no gender effect vs. H_A : there is a gender effect

(b) What is the distribution of the F statistic under the null hypothesis for this test?

(c) What is the value of the F-statistic? What is the p-value?

(d) What would you conclude about the interaction effect?

i. There is strong evidence of an interaction effect (p-value=0.46 from F-stat=0.77 on 2 and 855 df).

ii. There is no evidence that the difference in true mean scores between males and females depends on the level of background (p-value=0.46 from F-stat=0.77 on 2 and 855 df).

iii. There is strong evidence that the difference in true mean scores between males and females depends on the level of background (p-value=0.46 from F-stat=0.77 on 2 and 855 df).

iv. There is no evidence that the difference in the average scores in our sample between males and females depends on the level of background (p-value=0.46 from F-stat=0.77 on 2 and 855 df).

(e) Would you use an additive model or an interaction model for these data?

7. Which anova function should you use to fit an additive model? (circle the correct choice) `anova` or `Anova`

8. The type II sums of squares ANOVA table is given below:

```
fit.Math.add <- lm(Score~Background + Sex, data = math)
Anova(fit.Math.add)

## Anova Table (Type II tests)
##
## Response: Score
##              Sum Sq  Df F value   Pr(>F)
## Background  14705   2   301.3 < 2e-16
## Sex           517   1    21.2 4.8e-06
## Residuals   20914 857
```

For *Background:*

(a) Write the null and alternative hypothesis for testing the effect of background.

i. $H_0 : \tau_1 = \tau_2 = 0$ vs. at least one $\tau_j \neq 0$

ii. $H_0 : \omega_{11} = \omega_{12} = \omega_{13} = \omega_{21} = \omega_{22} = \omega_{23} = 0$ vs. H_A : at least one $\omega_{jk} \neq 0$

iii. $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$ vs. H_A : at least one $\gamma_k \neq 0$

iv. H_0 : there is no gender effect vs. H_A : there is a gender effect

(b) What is the distribution of the F statistic under the null hypothesis for this test?

(c) What would you conclude about the effect of *Background*?

i. There is strong evidence of an interaction effect (p-value=0.46 from F-stat=0.77 on 2 and 855 df).

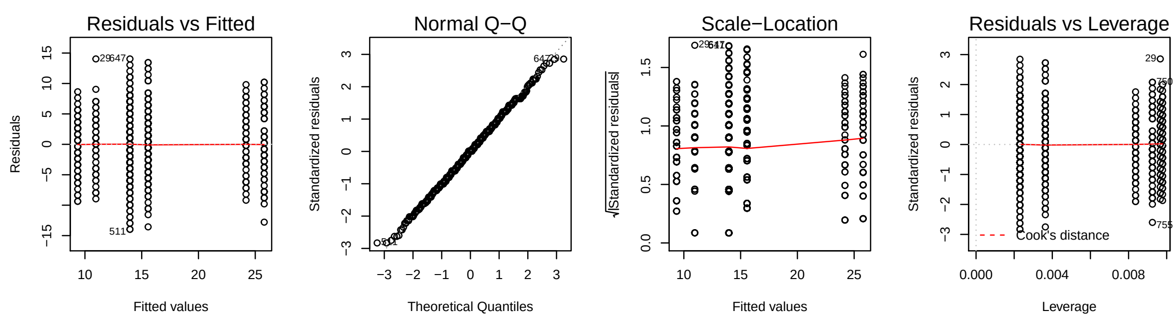
ii. There is no evidence of a difference in the true mean scores across levels of background, after accounting for sex (p-value$<$ 0.0001 from F-stat=301.3 on 2 and 857 df).

iii. There is strong evidence of a difference in the true mean scores across levels of background, after accounting for sex (p-value$<$ 0.0001 from F-stat=301.3 on 2 and 857 df).

iv. There is strong evidence of a difference in the true mean scores across levels of background (p-value$<$ 0.0001 from F-stat=301.3 on 2 and 857 df).

(d) Use the following plots to assess the assumptions and conditions required for the ANOVA.

```
par(mfrow = c(1,4))
plot(fit.Math.add)
```



The normality assumption...

i. Is not met because the horizontal line in the Scale-Location plot is not perfectly flat.

ii. Is not met because there are several points in the Residuals vs. Leverage plot that have a large Cook's Distance.

iii. Is met because the points in the Residuals vs. Fitted Values plot show approximately equal spread across all groups.

iv. Is met because the points in the normal Q-Q plot lie on the diagonal line.

The equal variance assumption...

i. Is not met because the horizontal line in the Scale-Location plot is not perfectly flat.

ii. Is not met because there are several points in the Residuals vs. Leverage plot that have a large Cook's Distance.

iii. Is met because the points in the Residuals vs. Fitted Values plot show approximately equal spread across all groups.

iv. Is met because the points in the normal Q-Q plot lie on the diagonal line.

(e) Use the output below to find the estimated means for each combination of background and gender.

```
summary(fit.Math.add)

##
## Call:
## lm(formula = Score ~ Background + Sex, data = math)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.964  -3.363   0.036   3.435  14.037
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.363      0.452   20.72  < 2e-16
## Backgroundb    4.601      0.477    9.64  < 2e-16
## Backgroundc   14.825      0.627   23.63  < 2e-16
## Sexmale        1.601      0.348    4.60  4.8e-06
##
## Residual standard error: 4.94 on 857 degrees of freedom
## Multiple R-squared:  0.436,Adjusted R-squared:  0.434
## F-statistic:  220 on 3 and 857 DF,  p-value: <2e-16
```

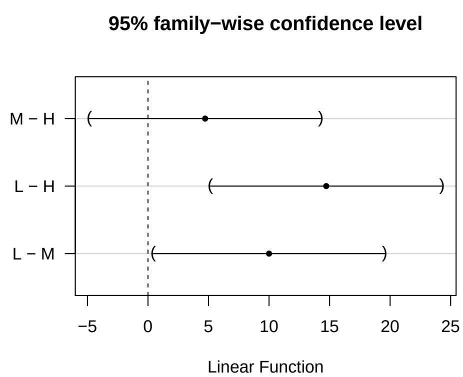
(f) Which combination of *Background* and *Sex* had the highest mean score?

9. The "warpbreak" data(manipulated below) gives the number of warp breaks per loom, where a loom corresponds to a fixed length of yarn. The tension refers to the tension of the loom, set at low, medium, or high.

```
require(multcomp)
tension <- factor(warpbreaks$tension, levels=c("H", "M", "L"))
amod <- aov(warpbreaks$breaks ~ tension)
ps <- glht(amod, linfct = mcp(tension = "Tukey"))
confint(ps)

##
##   Simultaneous Confidence Intervals
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = warpbreaks$breaks ~ tension)
##
## Quantile = 2.41
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##            Estimate lwr    upr
## M - H == 0  4.722   -4.839 14.283
## L - H == 0 14.722    5.161 24.283
## L - M == 0 10.000    0.439 19.561

plot(confint(ps))
```



Choose the correct interpretation of the confidence interval(s) above.

(a) We are 95% confident that the true difference in mean breaks between those yarns at a low tension and those yarns at a high tension is between 24.282 and 5.163 breaks.

(b) We are 95% confident that the high group has more breaks than the low group.

(c) We are 95% confident that all of the intervals above contain their respective true differences in mean breaks.

(d) We are more than 95% confident that all of the intervals above contain their respective true differences in mean breaks.

10. Based on the output and plot above, which groups (if any) are found to be significantly different?