Javier Saldana

80

UNIT 6 HW

1. **Handicap Study.** Use the Bonferroni method to construct simultaneous confidence intervals for $\mu_2 - \mu_3$, $\mu_2 - \mu_5$, and $\mu_3 - \mu_5$ (to see whether there are differences in attitude toward the mobility type of handicaps).

μ_1, μ_2, μ_3, μ_4, and μ_5, are the mean scores in the none, amputee, crutches, hearing, and wheelchair groups respectively. Be careful when identifying 'k' here. This study is mentioned throughout Chapter 6 of Statistical Sleuth.

```
data Handicap;
infile "/folders/myfolders/Data_Sources/Handicap.csv" firstobs=2 dlm=","; /*call data file*/
input Score Handicap $; /*name variables*/
run;

proc glm data=handicap order=data; /*data already sorted w/ control as first entry set*/
    class Handicap;
    model score = Handicap;
    means Handicap / hovtest=bf bon cldiff; /*looking to compare means of variables*/
    /*perform 1st contrast/estimate*/
    contrast 'Avg. Amputee v. Avg. Crutches' Handicap 0 1 -1 0 0;
    estimate 'Avg. Amputee v. Avg. Crutches' Handicap 0 1 -1 0 0;
    /*perform 2nd contrast/estimate*/
    contrast 'Avg. Amputee v. Avg. Wheelchair' Handicap 0 1 0 0 -1;
    estimate 'Avg. Amputee v. Avg. Wheelchair' Handicap 0 1 0 0 -1;
    /*perform 3rd contract/estimate*/
    contrast 'Avg. Crutches v. Avg. Wheelchair' Handicap 0 0 1 0 -1;
    estimate 'Avg. Crutches v. Avg. Wheelchair' Handicap 0 0 1 0 -1;
    /*this will perform contrasts/estimates on the desired groups of the
    sample data set*/
run;
```

-4

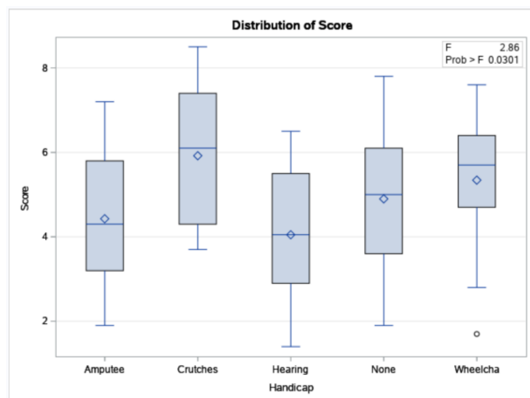
Dependent Variable: Score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	30.5214286	7.6303571	2.86	0.0301
Error	65	173.3214286	2.6664835		
Corrected Total	69	203.8428571			

R-Square	Coeff Var	Root MSE	Score Mean
0.149730	33.13206	1.632937	4.928571

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Handicap	4	30.52142857	7.63035714	2.86	0.0301

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Handicap	4	30.52142857	7.63035714	2.86	0.0301



Bonferroni (Dunn) t Tests for Score

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	65
Error Mean Square	2.666484
Critical Value of t	2.90602
Minimum Significant Difference	1.7936

Comparisons significant at the 0.05 level are indicated by ***.				
Handicap Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Crutches - Wheelcha	0.5786	-1.2150	2.3721	
Crutches - None	1.0214	-0.7721	2.8150	
Crutches - Amputee	1.4929	-0.3007	3.2864	
Crutches - Hearing	1.8714	0.0779	3.6650	***
Wheelcha - Crutches	-0.5786	-2.3721	1.2150	
Wheelcha - None	0.4429	-1.3507	2.2364	
Wheelcha - Amputee	0.9143	-0.8793	2.7079	
Wheelcha - Hearing	1.2929	-0.5007	3.0864	
None - Crutches	-1.0214	-2.8150	0.7721	
None - Wheelcha	-0.4429	-2.2364	1.3507	
None - Amputee	0.4714	-1.3221	2.2650	
None - Hearing	0.8500	-0.9436	2.6436	
Amputee - Crutches	-1.4929	-3.2864	0.3007	
Amputee - Wheelcha	-0.9143	-2.7079	0.8793	
Amputee - None	-0.4714	-2.2650	1.3221	
Amputee - Hearing	0.3786	-1.4150	2.1721	
Hearing - Crutches	-1.8714	-3.6650	-0.0779	***
Hearing - Wheelcha	-1.2929	-3.0864	0.5007	
Hearing - None	-0.8500	-2.6436	0.9436	
Hearing - Amputee	-0.3786	-2.1721	1.4150	

Dependent Variable: Score

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Avg. Amputee v. Avg. Crutches	1	15.60035714	15.60035714	5.85	0.0184
Avg. Amputee v. Avg. Wheelchair	1	5.85142857	5.85142857	2.19	0.1433
Avg. Crutches v. Avg. Wheelchair	1	2.34321429	2.34321429	0.88	0.3520

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Avg. Amputee v. Avg. Crutches	-1.49285714	0.61719220	-2.42	0.0184
Avg. Amputee v. Avg. Wheelchair	-0.91428571	0.61719220	-1.48	0.1433
Avg. Crutches v. Avg. Wheelchair	0.57857143	0.61719220	0.94	0.3520

2. **Handicap Study.**

See what multiple comparison procedures are available within the one-way analysis of variance procedure. Verify the 95% confidence interval half-widths in Display 6.6.

-2

Javier Saldana

DISPLAY 6.6 Summary of 95% confidence interval procedures for differences between treatment means in the handicap study

		Difference with . . .			
Group	Average	Hearing	Amputee	Control	Wheelchair
Crutches	5.921	1.871	1.492	1.021	0.578
Wheelchair	5.343	1.293	0.914	0.443	
Control	4.900	0.850	0.471		
Amputee	4.429	0.379			
Hearing	4.050				

A confidence interval is centered at a difference with half-width given by one of the procedures.

Procedure	95% interval half-width
LSD	1.233
Dunnett	1.545 (for comparisons with control only)
Tukey–Kramer	1.735
Bonferroni	1.794
Scheffé	1.957

Show your work for this problem by simply copying the code and relevant output for each comparison. (Cut and paste your code and relevant output.) The half-width might be found directly from your output. If so, note where it is found. Do this for both R and SAS.

```
data Handicap;
infile "/folders/myfolders/Data_Sources/Handicap.csv" firstobs=2 dlm=","; /*call data file*/
input Score Handicap $; /*name variables*/
run;
```

```
proc glm data=handicap order=data;
    class Handicap;
    model score = Handicap;
    means Handicap / hovtest=bf lsd cldiff;
run;
```

Alpha	0.05
Error Degrees of Freedom	65
Error Mean Square	2.666484
Critical Value of t	1.99714
Least Significant Difference	1.2326

```
proc glm data=handicap order=data;
    class Handicap;
    model score = Handicap;
    means Handicap / hovtest=bf dunnett('None') cldiff;
run;
```

Alpha	0.05
Error Degrees of Freedom	65
Error Mean Square	2.666484
Critical Value of Dunnett's t	2.50316
Minimum Significant Difference	1.5449

```
proc glm data=handicap order=data;
    class Handicap;
    model score = Handicap;
    means Handicap / hovtest=bf tukey cldiff;
run;
```

Alpha	0.05
Error Degrees of Freedom	65
Error Mean Square	2.666484
Critical Value of Studentized Range	3.96804
Minimum Significant Difference	1.7317

```
proc glm data=handicap order=data;
    class Handicap;
    model score = Handicap;
    means Handicap / hovtest=bf bon cldiff;
run;
```

Alpha	0.05
Error Degrees of Freedom	65
Error Mean Square	2.666484
Critical Value of t	2.90602
Minimum Significant Difference	1.7936

```
proc glm data=handicap order=data;
    class Handicap;
    model score = Handicap;
    means Handicap / hovtest=bf scheffe cldiff;
run;
```

Alpha	0.05
Error Degrees of Freedom	65
Error Mean Square	2.666484
Critical Value of F	2.51304
Minimum Significant Difference	1.9568

Javier Saldana

```
library(agricolae)
library(multcomp)

Handicap <- read.csv(file.choose())

#Least Significant Difference
LSD.test(aov(lm(Score ~ Handicap, data=Handicap)), "Handicap")

$`statistics`
   MSerror Df     Mean       CV  t.value      LSD
  2.666484 65 4.928571 33.13206 1.997138 1.232618

$parameters
        test p.ajusted   name.t ntr alpha
  Fisher-LSD      none Handicap   5  0.05

$means
              Score      std  r      LCL      UCL Min Max   Q25  Q50   Q75
Amputee    4.428571 1.585719 14 3.556979 5.300164 1.9 7.2 3.300 4.30 5.725
Crutches   5.921429 1.481776 14 5.049836 6.793021 3.7 8.5 4.500 6.10 7.150
Hearing    4.050000 1.532595 14 3.178407 4.921593 1.4 6.5 3.025 4.05 5.300
None       4.900000 1.793578 14 4.028407 5.771593 1.9 7.8 3.725 5.00 6.050
Wheelchair 5.342857 1.748280 14 4.471265 6.214450 1.7 7.6 4.725 5.70 6.350

$comparison
NULL

$groups
              Score groups
Crutches   5.921429      a
Wheelchair 5.342857     ab
None       4.900000    abc
Amputee    4.428571     bc
Hearing    4.050000      c

attr(,"class")
[1] "group"
#Dunnett
Handicap$Handicap = relevel(Handicap$Handicap, ref = "None")
fit = aov(Score ~ Handicap, data = Handicap)
gfit = glht(fit, linfct = mcp(Handicap = "Dunnett"))
summary(gfit)
confint(gfit)

#Tukey-Kramer
HSD.test(aov(lm(Score ~ Handicap, data=Handicap)), "Handicap")
```

Javier Saldana

```
$`statistics`
   MSerror Df     Mean       CV      MSD
  2.666484 65 4.928571 33.13206 1.731733

$parameters
   test   name.t ntr StudentizedRange alpha
  Tukey Handicap   5         3.968034  0.05

$means
              Score      std  r Min Max   Q25  Q50   Q75
Amputee    4.428571 1.585719 14 1.9 7.2 3.300 4.30 5.725
Crutches   5.921429 1.481776 14 3.7 8.5 4.500 6.10 7.150
Hearing    4.050000 1.532595 14 1.4 6.5 3.025 4.05 5.300
None       4.900000 1.793578 14 1.9 7.8 3.725 5.00 6.050
Wheelchair 5.342857 1.748280 14 1.7 7.6 4.725 5.70 6.350

$comparison
NULL

$groups
              Score groups
Crutches   5.921429      a
Wheelchair 5.342857     ab
None       4.900000     ab
Amputee    4.428571     ab
Hearing    4.050000      b

attr(,"class")
[1] "group"
#Bonferroni
LSD.test(aov(lm(Score ~ Handicap, data=Handicap)), "Handicap", p.adj=c("bonferroni"))
```

Javier Saldana

```
$`statistics`
   MSerror Df     Mean       CV  t.value     MSD
  2.666484 65 4.928571 33.13206 2.906015 1.79357

$parameters
        test  p.ajusted   name.t ntr alpha
  Fisher-LSD bonferroni Handicap   5  0.05

$means
             Score      std  r      LCL      UCL Min Max   Q25  Q50   Q75
Amputee   4.428571 1.585719 14 3.556979 5.300164 1.9 7.2 3.300 4.30 5.725
Crutches  5.921429 1.481776 14 5.049836 6.793021 3.7 8.5 4.500 6.10 7.150
Hearing   4.050000 1.532595 14 3.178407 4.921593 1.4 6.5 3.025 4.05 5.300
None      4.900000 1.793578 14 4.028407 5.771593 1.9 7.8 3.725 5.00 6.050
Wheelchair 5.342857 1.748280 14 4.471265 6.214450 1.7 7.6 4.725 5.70 6.350

$comparison
NULL

$groups
             Score groups
Crutches   5.921429      a
Wheelchair 5.342857     ab
None       4.900000     ab
Amputee    4.428571     ab
Hearing    4.050000      b

attr(,"class")
[1] "group"
#Scheffe
scheffe.test(aov(lm(Score ~ Handicap, data=Handicap)), "Handicap"))
```

Javier Saldana

```
$`statistics`
   MSerror Df       F     Mean       CV  Scheffe CriticalDifference
  2.666484 65 2.51304 4.928571 33.13206 3.170514           1.956817

$parameters
     test   name.t ntr alpha
  Scheffe Handicap   5  0.05

$means
              Score      std  r Min Max   Q25  Q50   Q75
Amputee    4.428571 1.585719 14 1.9 7.2 3.300 4.30 5.725
Crutches   5.921429 1.481776 14 3.7 8.5 4.500 6.10 7.150
Hearing    4.050000 1.532595 14 1.4 6.5 3.025 4.05 5.300
None       4.900000 1.793578 14 1.9 7.8 3.725 5.00 6.050
Wheelchair 5.342857 1.748280 14 1.7 7.6 4.725 5.70 6.350

$comparison
NULL

$groups
              Score groups
Crutches   5.921429      a
Wheelchair 5.342857      a
None       4.900000      a
Amputee    4.428571      a
Hearing    4.050000      a

attr(,"class")
[1] "group"
```

3. **Education and Future Income.** Reconsider the data problem of Exercise 5.25 concerning the distributions of annual incomes in 2005 for Americans in each of five education categories. (a) Use the Tukey–Kramer procedure to compare every group to every other group. Which pairs of means differ and by how many dollars (or by what percent)? (Use *p*-values and confidence intervals in your answer.) (b) Use the Dunnett procedure to compare every other group to the group with 12 years of education. Which group means apparently differ from the mean for those with 12 years of education and by how many dollars (or by what percent)? (Use *p*-values and confidence intervals in your answer.)

This question is obviously from the book, but assume you are starting this problem from scratch. Show all parts:

(1) Discussion of Assumptions (This could result in the inferences no longer being about the means. IF that happens, you should still compare the groups, just use the appropriate parameters when making inferences. Remember that you already did the work for addressing assumptions in prior homeworks.)

Problem: What is the strength of the evidence that at least one of the 5 group distributions of education has a different mean income than any of the others?

Assumptions: ANOVA assumptions are that the data is normally distributed and the groups have similar variances, along with being independent of each other and within. The data appears to be right skewed, this indicating there is evidence the data is not normally distributed. Considering the groups

Javier Saldana
have a large sample size, we can exact the CLT to make ANOVA robust to this assumption. However, since the data does not share equal variance; a log transformation helps address this issue. After transforming the data, there is greater evidence of similar distributions. With respect to independence, it appears they data may not be entirely independent. Since any viable members of the random sample of homes were included in the study, there is concern about cluster data being an issue. However, for the sake of the tests we will assume the data is independent among and within the groups.

```
data Edu_Income_2005;
infile "/folders/myfolders/Data_Sources/ex0525.csv" firstobs=2 dlm=",";
input Subj $ Ed $ Income; /*name variables*/
run;

*sort to make control group (12) first;
proc sort data=edu_income_2005;
by Ed;
run;

*plot raw data by ed level;
proc sgplot data=edu_income_2005;
scatter x=Ed y=Income;
run;

*plot data to check ANOVA assumptions;
proc univariate data=edu_income_2005;
by Ed;
histogram Income;
qqplot Income;
run;

*transform raw data;
data Edu_Income1; set Edu_Income_2005;
logincome = log(Income);
run;

*plot logged data by education level;
proc sgplot data=edu_income1;
scatter x=Ed y=logincome;
run;

*plot logged data to check ANOVA assumptions;
proc univariate data=Edu_Income1;
by Ed;
Histogram logincome;
qqplot logincome;
run;
```

(2) Selection and Execution of Tests

In order to ensure there is a difference among the groups, we will perform an ANOVA test.
Ho: All means are equal
Ha: At least 1 mean is different from the rest.

```
*ANOVA;
proc glm data=edu_income1;
class Ed;
model logincome = Ed;
run;
```

Javier Saldana

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	217.653784	54.413446	62.87	<.0001
Error	2579	2232.120383	0.865498		
Corrected Total	2583	2449.774168			

R-Square	Coeff Var	Root MSE	logincome Mean
0.088846	8.913094	0.930322	10.43770

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Ed	4	217.6537844	54.4134461	62.87	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Ed	4	217.6537844	54.4134461	62.87	<.0001

```
> Ed.Income$log.income <- log(Ed.Income$Income2005)
> ed.aov <- aov(log.income ~ Educ, data = Ed.Income)
> summary(ed.aov)
              Df Sum Sq Mean Sq F value Pr(>F)
Educ           4  217.7   54.41   62.87 <2e-16 ***
Residuals   2579 2232.1    0.87
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

Based on the results of the ANOVA test, there is not sufficient evidence to support the null hypothesis that all groups have the same mean (p value <0.0001). So we will reject that ANOVA null and proceed with the Tukey and Dunnett tests.

In order to compare every group to each other, we will perform a Tukey-Kramer test.
Ho: All means are equal
Ha: At least one mean is different from the rest

```
*perform Tukey-Kramer;
proc glm data=edu_income1 order=data;
    class Ed;
    model logincome = Ed;
    means Ed / hovtest=bf tukey cldiff;
run;
```

```
> ed.tukey <- HSD.test(aov(lm(log.income ~ Educ, data=Ed.Income)), "Educ")
> ed.tukey
$`statistics`
    MSerror   Df    Mean       CV
  0.8654984 2579 10.4377 8.913094

$parameters
   test name.t ntr StudentizedRange alpha
  Tukey   Educ   5         3.860388  0.05

$means
      log.income       std    r      Min      Max       Q25      Q50      Q75
<12      9.89934 0.9988809  136 5.857933 11.51293  9.546813 10.06453 10.51867
>16     10.89790 1.0665910  374 4.143135 13.46402 10.596635 11.01036 11.47210
12      10.22721 0.8539854 1020 5.703782 12.92393  9.902234 10.34174 10.77896
13-15   10.39121 0.9288173  648 6.061457 12.45794 10.085809 10.54534 10.96820
16      10.79709 0.9581051  406 5.298317 13.16031 10.373491 10.94196 11.39639

$comparison
NULL

$groups
      log.income groups
>16     10.89790      a
16      10.79709      a
13-15   10.39121      b
12      10.22721      c
<12      9.89934      d

attr(,"class")
[1] "group"
```

Javier Saldana

Tukey's Studentized Range (HSD) Test for logincome

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	2579
Error Mean Square	0.865498
Critical Value of Studentized Range	3.86039

Comparisons significant at the 0.05 level are indicated by ***.				
Ed Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
>16 - 16	0.10082	-0.08119	0.28283	
>16 - 13-15	0.50669	0.34178	0.67160	***
>16 - 12	0.67069	0.51717	0.82420	***
>16 - <12	0.99856	0.74427	1.25285	***
16 - >16	-0.10082	-0.28283	0.08119	
16 - 13-15	0.40588	0.24514	0.56661	***
16 - 12	0.56987	0.42085	0.71889	***
16 - <12	0.89775	0.64614	1.14935	***
13-15 - >16	-0.50669	-0.67160	-0.34178	***
13-15 - 16	-0.40588	-0.56661	-0.24514	***
13-15 - 12	0.16400	0.03642	0.29157	***
13-15 - <12	0.49187	0.25235	0.73139	***
12 - >16	-0.67069	-0.82420	-0.51717	***
12 - 16	-0.56987	-0.71889	-0.42085	***
12 - 13-15	-0.16400	-0.29157	-0.03642	***
12 - <12	0.32787	0.09605	0.55970	***
<12 - >16	-0.99856	-1.25285	-0.74427	***
<12 - 16	-0.89775	-1.14935	-0.64614	***
<12 - 13-15	-0.49187	-0.73139	-0.25235	***
<12 - 12	-0.32787	-0.55970	-0.09605	***

Based on the Tukey test results, there is sufficient evidence to support that there is a statistically significant difference in means between all the groups except >16 and 16 (-0.08119, 0.28283).

/5

In order to compare groups to the control group (12 years of education), we will perform a Dunnett test.

Ho: Mean of 12 is equal to all other means

Ha: Mean of 12 is not equal to at least one other mean

```
*perform Dunnett;
proc glm data=edu_income1 order=data;
    class Ed;
    model logincome = Ed;
    means Ed / hovtest=bf dunnett('12') cldiff;
run;
```

```
> Ed.Income$Educ <- as.factor(Ed.Income$Educ)
> Ed.Income$Educ = relevel(Ed.Income$Educ, ref = "12")
> fit = aov(log.income ~ Educ, data = Ed.Income)
> gfit = glht(fit, linfct = mcp(Educ = "Dunnett"))
> summary(gfit)

         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Dunnett Contrasts


Fit: aov(formula = log.income ~ Educ, data = Ed.Income)

Linear Hypotheses:
                Estimate Std. Error t value Pr(>|t|)
<12 - 12 == 0   -0.32787    0.08493  -3.861 0.000463 ***
>16 - 12 == 0    0.67069    0.05624  11.926  < 1e-05 ***
13-15 - 12 == 0  0.16400    0.04674   3.509 0.001802 **
16 - 12 == 0     0.56987    0.05459  10.439  < 1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> confint(gfit)

         Simultaneous Confidence Intervals

Multiple Comparisons of Means: Dunnett Contrasts


Fit: aov(formula = log.income ~ Educ, data = Ed.Income)

Quantile = 2.4802
95% family-wise confidence level


Linear Hypotheses:
```

Javier Saldana

Dunnett's t Tests for logincome

est controls the Type I experimentwise error for comparisons of all treatments aga

Alpha	0.05
Error Degrees of Freedom	2579
Error Mean Square	0.865498
Critical Value of Dunnett's t	2.48068

Comparisons significant at the 0.05 level are indicated by ***.				
Ed Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
>16 - 12	0.67069	0.53118	0.81020	***
16 - 12	0.56987	0.43445	0.70530	***
13-15 - 12	0.16400	0.04806	0.27993	***
<12 - 12	-0.32787	-0.53855	-0.11720	***

-5

Based on the results from the Dunnet test, there is sufficient evidence to suggest there is clearly a difference in means of income when compared against the control group.

(3) Interpretation and Conclusion.
In short, perform a complete analysis like you usually do. Provide and interpret all the confidence intervals that suggest a significant difference in incomes; provide your SAS and R code as well. (Generate your statistics using both softwares.)
Finally, you should first test to see if any of the groups are different before you consider pairwise comparisons.

Based on the results from the Tukey test, there is evidence to suggest the mean income between all the education levels are different, except the groups 16 and >16 [confidence intervals -0.08119, 0.28283; -0.28283, 0.08119]. The contrast between 16 and >16 is the only confidence interval that contains zero, which would mean the difference in the means is not significantly different. In other words, according to these results education past 16 years didn't doesn't result in a difference than the education of those with 16 years. However, all the other confidence intervals (95%, alpha 0.05) supports that all other means are different. A 12 year education earns anywhere between 0.096 – 0.560 times more than someone with less than a 12 year education. Someone with some college earns 0.0364 – 0.291 times more than someone with a 12 year education. Someone with a college degree earns 0.245 – 0.567 times more than someone with some college. The table is below.

	<12	12	13-15	16	>16
<12	0				
12	0.096, 0.560	0			
13-15	0.492, 0.252	0.0364, 0.291	0		
16	0.646, 1.149	0.421, 0.719	0.245, 0.567	0	
>16	0.744, 1.253	0.517, 0.824	0.342, 0.672	-0.0812, 0.283*	0

Javier Saldana

With respect to the Dunnett test, it appears all of the education levels have a different mean income than that of a high school graduate only. Those with a graudate degree (>16) earned anywhere between .53 - .81 times more. A college graduate earned .43 - .71 times more. Those with some college earn .05 – 0.28 times more. Whereas, someone without a high school diploma earns 0.54 – 0.12 times less.

However, since there is concern about the cluster of the data, the results may not be applied outside of these groups. Furthermore, the study was observational and makes it difficult to establish a casual relationship between education and income (95% confidence level, alpha 0.05).

Bonus: Max 5 pts

Equity in Group Learning. [Continuation of Exercise 5.22.] (a) To see if the performance of low-ability students increases steadily with the ability of the best student in the group, form a linear contrast with increasing weights: -3 = Low, -1 = Low–Medium, $+1$ = Medium–High, and $+3$ = High. Estimate the contrast and construct a 95% confidence interval. (b) For the High-ability students, use multiple comparisons to determine which group composition differences are associated with different levels of test performance.

DISPLAY 5.24 Achievement test scores of Low ability students who worked in different study groups

	Highest ability level in the study group			
	Low	Low-medium	Medium-high	High
Average:	0.26	0.37	0.36	0.47
St. Dev.:	0.14	0.21	0.17	0.21
n:	17	24	25	14

(c) Give the levels of ability a quantitative representation (Low = 1, Low-Medium = 2, etc.) for the low ability students. After completing the questions above, conduct a linear regression (we haven't studied this yet!) of the **AVERAGE** performance against the level variable you just created. Be sure and address the assumptions. Defend the ones you can and assume the others are met. Include a scatterplot and residual plot. Is there evidence of linear trend? Is this inferred from the contrast? Assume the levels are equidistant in ability from each other.

Note: the data for Part b above is in Display 5.25 in your textbook.