Multiple Comparisons

When an IV in either ANOVA or regression contains more than two categories or groups, one should routinely make comparisons among the groups to determine precisely how the groups differ. As previously discussed, when multiple comparisons are made, the likelihood of a Type 1 error is increased. The Type 1 error rate may be viewed in several ways, and each are discussed below.

For each distinct comparison, a given Type 1 error rate is selected (usually .10, .05 or .01). Thus, if one wishes to compare attitudes between Black and Hispanic respondents, the error rate per comparison would be, say, $\alpha = .05$. The error rate for the comparison between Black and White respondents would also be .05, and the error per comparison between Hispanic and White respondents would be .05. The error for each unique comparison is called the *error rate per contrast or comparison* (α_{pc}).

The problem with the error rate per comparison is that as more and more comparisons are performed, the overall error rate increases. Typically for one-way ANOVA and simple regression with a single categorical IV, the overall error rate is referred to as the *experimentwise error rate* (α_{ew}). If, for example, each of the three comparisons discussed above were independent, and $\alpha_{pc} = .05$, then the experimentwise error rate would be

$\alpha_{ew} = 1 - (1 - \alpha_{pc})^C$
$\alpha_{ew} = 1 - (1 - .05)^3$
$\alpha_{ew} = 1 - (0.95)^3$
$\alpha_{ew} = 1 - .857375$
$\alpha_{ew} = .142625.$

where C is the number of comparisons being made.

A third type of error rate is the *familywise error rate* (α_{fw}). The familywise error rate refers to comparisons made for each categorical IV in a regression or ANOVA model. While regression models with two or more IVs have not been discussed, the concept can be easily explained. For instance, one may be interested in learning whether two demographic variables or factors, race and location, are associated with attitudes toward increases in millage rates. Race, as illustrated above, may have three groups in which to compare: Black, Hispanic, and White respondents. Location refers to the area in which the respondent lives, and there may be three categories: urban, suburban, and rural. Note that with these two IVs, one may make, at the minimum, six comparisons, as the table below shows.

Race	Location
Black vs. Hispanic	Suburban vs. Rural
Black vs. White	Suburban vs. Urban
Hispanic vs. White	Rural vs. Urban

The set of comparisons for each IV is called a family, and the Type 1 error rate associated with each set or family of comparisons is known as the familywise error rate. The procedures for controlling α_{ew} discussed below also hold for α_{fw}, but will not be illustrated for familywise comparisons until multiple regression and multi-way ANOVAs are presented.

Planned and Post Hoc Comparisons

A planned comparison is one in which the researcher planned or formulated prior to collecting the study data. Such comparisons are sometimes called a priori comparisons. Sometimes researchers may not anticipate that certain groups will differ, and find, after collecting the study data, that statistically significant overall F tests revealed that some differences exist among the groups. These types of follow-up comparisons are known as post hoc or *a posterior* comparisons.

Some researchers argue that post hoc comparisons are equivalent to data snooping (attempting to find things without using theory as a guide), and should thus be penalized in the sense that one should prevent, as far as possible, finding chance differences. As a result, post hoc comparison procedures generally have lower statistical power. The best known and perhaps most versatile procedure used in post hoc comparisons is the Scheffé test. The Scheffé test is very conservative (low in power), but allows one to make any type of comparison or contrast one wishes.

When comparisons are planned prior to data collection, the researcher is rewarded with comparison procedures than are more liberal (higher in statistical power) than the Scheffé test, like, for example, the Bonferroni. Other benefits also accrue, such as not having to calculate the overall F test when comparisons are orthogonal (uncorrelated), but these benefits will not be discussed in this class. See Pedhazur a detailed discussion.

For this class only two comparison procedures will be discussed—the Bonferroni and Scheffé. When one is performing a planned comparison, the Bonferroni is usually the one to use. Note, however, that as the number of comparisons increases, the Scheffé may be the better procedure (see Table 5.5 in Maxwell & Delaney, 1994, p. 191, for information on when to use Scheffé rather than Bonferroni for planned comparisons). As a general rule, if the number of comparison is less than seven, use Bonferroni; if the number is greater than seven use Scheffé. For unplanned or post hoc comparisons, the Scheffé will be stressed. (Note that some statisticians argue that the Bonferroni can be used for post hoc comparisons, but most argue against its use.)

The Bonferroni Procedure for Controlling α_{ew}

As mentioned, the Bonferroni (or Dunn) procedure is most often used for planned comparisons. The Bonferroni procedure simply requires that α be divided by the total number of comparisons to be made. That is,

$$\alpha_{ew} / C = \text{adj. } \alpha_{pc} \qquad (1)$$

where α is the experimentwise error rate, such as .05; C is the total number of comparisons; and *adj.* α_{pc} is the new per comparison error rate.

For example, one may want to have an experimentwise error rate of .05. If the IV contained four groups, then a total of six pairwise comparisons would be possible. The formula for determining the total number of pairwise comparisons for a given number of groups is

$$C = G\,(G-1) / 2 \qquad (2)$$

where G is the number of groups in the IV, and C is the number of pairwise comparisons. For example, if four groups are present, then C = G(G-1)/2 = 4(4-1)/2 = 4(3)/2 = 12/2 = 6, six pairwise comparisons are possible. Using this information, the Bonferroni adjusted error rate per comparison would be

adj. $\alpha_{pc} = \alpha_{EW} / ^{C}$,
adj. $\alpha_{pc} = .05 / 6$,
adj. $\alpha_{pc} = 0.0083$.

When computing pairwise comparisons via the computer, it is possible to obtain p-values for each comparison, and these p-values should be referenced against the adj. α_{pc} to determine statistical significance. The usual decision rule for testing H_0 (no difference between groups) applies:

If p ≤ *adj.* α_{pc} then reject H_0, otherwise fail to reject H_0.

If it is not possible to obtain p-values for each comparison, then one may use critical t values to determine statistical significance. To determine the Bonferroni adjusted critical t value, simply refer to the attached table entitled "Dunn's Critical Values for Bonferroni t." To use this table, three things are needed: (a) α_{ew}, (b) number of comparisons, and (c) df for the t calculated ratio.

As usual, one may set α_{ew} to the conventional level of .10, .05, or .01. Assume .05 was selected as the experimentwise error rate. The number of comparisons is determined using formula (2). Degrees of freedom for the t test are

df = n - k - 1

where n is the total sample size and k is the number of variables or vectors in the regression or ANOVA model.

For example, if four groups are present, then six comparisons are possible. Assume df equals 10. The Bonferroni adjusted critical t value would be

adj. $t_{crit} = 3.264$.

Confidence Intervals Based Upon the Bonferroni Adjustment

Once Bonferroni critical t values are obtained, one may construct confidence intervals for the contrasts. A .95CI would be:

.95CI: $b_1 \pm t_{(\text{Bonferroni adjusted } \alpha/2, df)} SE_{b1}$.

Using the data from "Simple Linear Regression: One Qualitative IV," the difference between Black and White respondents in terms of attitudes toward millage rates is 3.25 points. The standard error of this difference is 0.957. The sample size is 12, and k = 2, df is

df = n - k - 1
= 12 - 2 - 1 = 9.

Since there are three groups in the data, Black, Hispanic, and White, there are three possible comparisons. If α_{ew} is set at .05, the Bonferroni adjusted critical t, taken from "Dunn's Critical Values for Bonferroni t" table and based upon three comparisons, is 2.923. The .95CI for this comparison is:

.95CI: $b_1 \pm t_{(\text{Bonferroni adjusted } \alpha/2, df)} SE_{b1}$.
.95CI: 3.25 ± (2.923)(0.957)
.95CI: 3.25 ± (2.80)
.95CI: 6.05, 0.45

The original non-Bonferroni adjusted CI for this comparison with $\alpha_{pc} = .05$ was:

.95CI: 5.415, 1.085

Note that the Bonferroni adjusted CI is wider than the non-Bonferroni CI which reflects the conservative adjustment to control for inflation of the Type 1 error rate.

Example Application of Bonferroni Adjusted α and Critical t

Recall the data on attitudes toward millage rate increases discussed in the notes dealing with regression with a categorical IV. The data are represented below in Table 1.

The sample regression equation for this data is

$Y_i = b_0 + b_1B + b_2H + e_i$

where B represents the dummy variable for Black respondents and H is the dummy variable for Hispanic respondents. Regression results for this data are

$Y' = 4.25 + 3.25(B) + 3.50(H)$.

Table 1

Y = attitude	Race	B = Black	H = Hispanic
7	B	1	0
9	B	1	0
6	B	1	0
8	B	1	0
9	H	0	1
8	H	0	1
8	H	0	1
6	H	0	1
6	W	0	0
5	W	0	0
3	W	0	0
3	W	0	0

Regression results in tabular form are presented below.

Table 2

Descriptive Statistics and Correlations Between Millage Rate Attitude and Race

Variable		Correlations	
	Attitude	Black	Hispanic
Attitude	---		
Black	.36	---	
Hispanic	.45	-.50	---
Mean	6.50	0.33	0.33
SD	2.07	0.49	0.49

Note. Black (1 = Black respondent; 0 = others) and Hispanic (1 = Hispanic respondent, 0 = others) are dummy variables; n = 12.

*p<.05.

Table 3
Regression Analysis of Millage Rate Attitude by Race

Variable	b	se b	95% CI	t
Black	3.25	0.957	1.08, 5.42	3.395*
Hispanic	3.50	0.957	1.33, 5.67	3.656*
Intercept	4.25	0.677	2.72, 5.78	6.278*

Note. R^2 = .65, adj. R^2 = .57, $F_{2,9}$ = 8.31, MSE = 1.83, n = 12. Black (1 = Black respondent; 0 = others) and Hispanic (1 = Hispanic respondent, 0 = others) are dummy variables.
*$p < .05$.

Table 4
Comparison in Mean Millage Attitudes by Race

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Bonferroni Adjusted CI of Mean Difference
Black vs. White	3.25*	0.957	0.45, 6.05
Hispanic vs. White	3.50*	0.957	0.70, 6.30
Black vs. Hispanic	-0.25	0.957	-3.05, 2.55

*$p < .05$, where p-values are adjusted using the Bonferroni method.

If one compares Table 4 reported here against Table 3 reported in "Simple Linear Regression: One Qualitative IV" on p. 25 (check this page, may be off now), you will note that the CIs have different widths, with the Table 4 widths reported here larger due to the Bonferroni adjustment.

When tables are used one may report results (and interpretations) as follows:

> Analysis of the data (see Table 3) indicates a statistically significant difference in millage rates attitudes among groups of respondents. All pairwise comparisons among respondents were performed, and results, based upon the Bonferroni adjustment for multiple comparisons, are presented in Table 4. Note that, in general, both Black and Hispanic respondents hold a more positive attitude toward rate increases than do White respondents, and there does not appear to be a substantial, or statistically significant, attitude difference between Black and Hispanic respondents.

I recommend that tables be used to report descriptive statistics and pairwise comparisons.

Exercises for Bonferroni Analysis

Instructions: Each example below is taken from "Simple Linear Regression: One Qualitative IV." For each example, calculate the Bonferroni comparison and confidence interval, and report the results as reported in Table 4 above.

(1) According to the leadership literature, there are a number of different leadership styles. Listed below are scores obtained from an instrument designed to measure a particular leadership style, which will be referred to as style X. Of interest is whether X differs by school district type in terms of urbanity. A stratified random sample of school principals were selected from three district types (mostly urban, mostly suburban, and mostly rural).

The scores on style X range from 100 to 0. The closer the score to 100, the more the respondent conforms to style X, while the closer the score to 0, the less the respondent conforms to style X.

Is there any evidence that X differs among the three district types?

District Type	Style X	District Type	Style X	District Type	Style X
Suburban	49	Rural	33	Urban	85
Suburban	48	Rural	95	Urban	98
Suburban	56	Rural	26	Urban	75
Suburban	78	Rural	11	Urban	63
Suburban	35	Rural	33	Urban	91
Suburban	50	Rural	25	Urban	49
		Rural	65	Urban	62

(2) Which of the following small car makers, if any, have the better fuel economy in terms of miles per gallon (MPG)? For this study, a small car is defined as any vehicle with a 2.2 litre (or less), 4-cylinder engine with less than 130 horsepower.

Maker	MPG	Maker	MPG	Maker	MPG	Maker	MPG
Honda	43	Toyota	37	GM	33	Subaru	33
Honda	37	Toyota	38	GM	31	Subaru	36
Honda	29	Toyota	41	GM	29	Subaru	34
Honda	38	Toyota	36	GM	33	Subaru	33
Honda	46	Toyota	33	GM	40	Subaru	36
Honda	35	Toyota	37	GM	33	Subaru	32
Honda	34	Toyota	34	GM	37	Subaru	37
Honda	37	Toyota	33	GM	26	Subaru	37

(3) A researcher is interested in learning whether frequency of reading at home to elementary-aged children produces differential effects on reading achievement. After obtaining information from a randomly selected sample of parents about this behavior, the following classifications and standardized achievement scores were recorded. (Note: frequency classifications as follows: a = less than once per month, b = once to three times per month, c = more than three times per month.)

Freq. of Reading	Achieve.	Freq. of Reading	Achieve.	Freq. of Reading	Achieve.
A	48	B	57	C	61
A	37	B	39	C	55
A	47	B	49	C	51
A	65	B	45	C	30

Is frequency of reading at home related to student reading achievement?

(4) Does a difference in salary for beginning assistant professors exist at GSU by college? Which colleges appear to have difference salary levels from the other colleges?

College	Salary	College	Salary
Education	33500	Education	34000
Business	32900	Business	43000
Health	30000	Health	29000
Arts & Science	26200	Arts & Science	24000
Technology	27250	Technology	29000
Education	35000	Education	32500
Business	37000	Business	44500
Health	49000	Health	31000
Arts & Science	29000	Arts & Science	23500
Technology	27500	Technology	30000

Answers to Exercises provide at end of document.

The Scheffé Procedure for Controlling α_{ew}

As previously mentioned, the Scheffé procedure is most often used for post hoc or unplanned comparisons. Recall that the Bonferroni procedure required one to divide α_{ew} by the total number of comparisons to be made. It is generally not possible to specify by how much α_{ew} should be divided when using the Scheffé procedure. Rather than dividing α_{ew}, Scheffé is based upon critical values of F and t.

Scheffé Critical F and t

Critical Scheffé values are simple to calculate. For example, if one wishes to use the F test for testing comparisons, the Scheffé critical F is determined by finding the appropriate critical F for α_{ew}, df_1, and df_2, and then multiplying this critical F by J – 1 (where J is the number of groups involved); that is

Scheffé $F = (J - 1)(F_{\alpha ew,\ J-1,\ n-k-1})$.

For example, the millage rate increase data had degrees of freedom of:

$df_1 = J - 1 = 3 - 1 = 2$, and
$df_2 = n - k - 1 = 12 - 2 - 1 = 9$.

If $\alpha_{ew} = .05$, then the critical F, found in the F Table linked on the course web site is 4.26. The Scheffé adjusted critical F will be

$$
\begin{aligned}
\text{Scheffé } F_{crit} &= (J - 1)(F_{\alpha ew, J-1, n-k-1}); \\
&= (3 - 1)(F_{.05, 3-1, 12-2-1}) \\
&= (2)(4.26). \\
&= 8.52.
\end{aligned}
$$

Seldom in multiple regression will one need to use critical F values for comparisons. A more appropriate critical value is the t. The Scheffé critical t is calculated by simply taking the square-root of the Scheffé F, i.e.,

$$\text{Scheffé } t_{crit} = \sqrt{(J - 1)(F_{\alpha ew, J-1, n-k-1})} \ .$$

For the current example, the Scheffé critical t used to control α_{ew} at .05 is

$$
\begin{aligned}
\text{Scheffé } t_{crit} &= \sqrt{(J - 1)(F_{\alpha ew, J-1, n-k-1})} \\
&= \sqrt{(3 - 1)(F_{.05, 3-1, 12-2-1})} \\
&= \sqrt{(2)(4.26)} \\
&= \sqrt{8.52} \\
&= 2.919.
\end{aligned}
$$

Thus, for each comparison or contrast, the Scheffé critical t value is ± 2.919. The decision rule for using this critical value is the same that has been discussed previously:

If $t \leq$ Scheffé $-t_{crit}$ or $t \geq$ Scheffé t_{crit} then reject H_0, else FTR H_0.

For example, the t ratio for the comparison in millage rate attitudes between Black and White residents was

$$
\begin{aligned}
t &= b_1 / se_{b1} \\
&= 3.25 / 0.957 \\
&= 3.395 \text{ (taken from Tables 2 and 3 above).}
\end{aligned}
$$

Is this difference between Black and White residents statistically significant at the α_{ew} .05 level once the Scheffé procedure is used? Since the Scheffé t_{crit} is ± 2.919, the decision will be to reject H_0 and conclude that attitudes among between Black and White respondents do differ; i.e.,

If $t \leq$ Scheffé $-t_{crit}$ or $t \geq$ Scheffé t_{crit} then reject H_0, else FTR H_0;

so,

If $3.395 \leq -2.919$ or $3.395 \geq 2.919$ then *reject* H_0, else FTR H_0.

The Scheffé t_{crit} will also be used to test whether the difference between Black and Hispanic, and between Hispanic and White respondents are statistically different.

Confidence Intervals Based Upon the Scheffé Adjustment

Once the Scheffé t_{crit} is obtained, one may construct CI for the comparisons. The .95CI would be:

.95CI: $b_1 \pm$ (Scheffé t_{crit})SE_{b1}.

Once again, using the data from "Simple Linear Regression: One Qualitative IV," the difference between Black and White respondents in terms of attitudes toward millage rates is 3.25 points. The standard error of this difference is 0.957. The .95CI for this comparison is:

.95CI: $b_1 \pm$ (Scheffé t_{crit})SE_{b1}.
.95CI: 3.25 ± (2.919)(0.957)
.95CI: 3.25 ± (2.79)
.95CI: 6.04, 0.46.

Recall that the Bonferroni adjusted .95CI for this contrast is 6.05, 0.45. Note that the Scheffé CI above is similar to the Bonferroni .95CI. Usually the Bonferroni adjustment will give more narrower CIs when the number of comparisons is below seven, but with the small sample size in this example, the Scheffé procedure provides similar statistical power.

When one is performing post hoc contrasts, Scheffé is the recommended procedure. But when one is performing planned contrasts, Bonferroni provides more power for up to eight comparisons, then Scheffé becomes the more powerful statistical control procedures for more than eight comparisons. Oddly, some combinations of small samples sizes and few comparisons result in Scheffé being more powerful than Bonferroni, but above pattern generally holds.

Example Application of Scheffé Critical t

The sample regression equation for this millage rates data is

$Y_i = b_0 + b_1B + b_2H + e_i$

where B represents the dummy variable for Black respondents and H is the dummy variable for Hispanic respondents. The regression results for this equation are presented below.

Table 5
Descriptive Statistics and Correlations Between Millage Rate Attitude and Race

Variable	Correlations		
	Attitude	Black	Hispanic
Attitude	---		
Black	.36	---	
Hispanic	.45	-.50	---
Mean	6.50	0.33	0.33
SD	2.07	0.49	0.49

Note. Black (1 = Black respondent; 0 = others) and Hispanic (1 = Hispanic respondent, 0 = others) are dummy variables; n = 12.
**p* < .05.

Table 6
Regression Analysis of Millage Rate Attitude by Race

Variable	b	se b	95% CI	t
Black	3.25	0.957	1.08, 5.42	3.395*
Hispanic	3.50	0.957	1.33, 5.67	3.656*
Intercept	4.25	0.677	2.72, 5.78	6.278*

Note. R^2 = .65, adj. R^2 = .57, $F_{2,9}$ = 8.31, MSE = 1.83, n = 12. Black (1 = Black respondent; 0 = others) and Hispanic (1 = Hispanic respondent, 0 = others) are dummy variables.
**p* < .05.

Table 7
Comparison in Mean Millage Attitudes by Race

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Scheffé Adjusted CI of Mean Difference
Black vs. White	3.25*	0.957	0.46, 6.04
Hispanic vs. White	3.50*	0.957	0.71, 6.29
Black vs. Hispanic	-0.25	0.957	-3.04, 2.54

**p* < .05, where p-values are adjusted using the Scheffé method.

If you compare Table 7 reported here against Table 3 reported in "Simple Linear Regression: One Qualitative IV" on p. 25 (check this page number, may be incorrect), you will note that the CIs are of different widths, with the Table 7 widths reported here being larger due to the Scheffé adjustment.

When tables are used one may report results (and interpretations) as follows:

> Analysis of the millage rate data shows that statistical differences in attitudes exist across the three racial categories. All pairwise comparisons between the three groups were performed, and results, based upon the Scheffé adjustment for multiple comparisons, are presented in Table Z. Note that both Black and Hispanic respondents tend to hold, on average, statistically more positive attitudes toward rate increases than do White respondents, and there does not appear to be a statistical difference attitude between Black and Hispanic respondents.

I recommend that tables be used to report descriptive statistics and pairwise comparisons when three or more groups are involved.

Exercises for Scheffé Analysis

Instructions: Each example below is taken from "Simple Linear Regression: One Qualitative IV." For each example, calculate the Scheffé comparison and confidence interval, and report the results as reported in Table 7 above.

(1) According to the leadership literature, there are a number of different leadership styles. Listed below are scores obtained from an instrument designed to measure a particular leadership style, which will be referred to as style X. Of interest is whether X differs by school district type in terms of urbanity. A stratified random sample of school principals were selected from three district types (mostly urban, mostly suburban, and mostly rural).

The scores on style X range from 100 to 0. The closely the score to 100, the more the respondent conforms to style X, while the closer the score to 0, the less the respondent conforms to style X.

Is there any evidence that X differs among the three district types?

District Type	Style X	District Type	Style X	District Type	Style X
suburban	49	rural	33	urban	85
suburban	48	rural	95	urban	98
suburban	56	rural	26	urban	75
suburban	78	rural	11	urban	63
suburban	35	rural	33	urban	91
suburban	50	rural	25	urban	49
		rural	65	urban	62

(2) Which of the following small car makers, if any, have the better fuel economy in terms of miles per gallon (MPG)? For this study, a small car is defined as any vehicle with a 2.2 litre (or less), 4-cylinder engine with less than 130 horsepower.

Maker	MPG	Maker	MPG	Maker	MPG	Maker	MPG
Honda	43	Toyota	37	GM	33	Subaru	33
Honda	37	Toyota	38	GM	31	Subaru	36
Honda	29	Toyota	41	GM	29	Subaru	34
Honda	38	Toyota	36	GM	33	Subaru	33
Honda	46	Toyota	33	GM	40	Subaru	36
Honda	35	Toyota	37	GM	33	Subaru	32
Honda	34	Toyota	34	GM	37	Subaru	37
Honda	37	Toyota	33	GM	26	Subaru	37

(3) A researcher is interested in learning whether frequency of reading at home to elementary-aged children produces differential effects on reading achievement. After obtaining information from a randomly selected sample of parents about this behavior, the following classifications and standardized achievement scores were recorded. (Note: frequency classifications as follows: a = less than once per month, b = once to three times per month, c = more than three times per month.)

Freq. of Reading	Achieve.	Freq. of Reading	Achieve.	Freq. of Reading	Achieve.
a	48	b	57	c	61
a	37	b	39	c	55
a	47	b	49	c	51
a	65	b	45	c	30

Is frequency of reading at home related to student reading achievement?

(4) Does a difference in salary for beginning assistant professors exist at GSU by college? Which colleges appear to have difference salary levels from the other colleges?

College	Salary	College	Salary
Education	33500	Education	34000
Business	32900	Business	43000
Health	30000	Health	29000
Arts & Science	26200	Arts & Science	24000
Technology	27250	Technology	29000
Education	35000	Education	32500
Business	37000	Business	44500
Health	49000	Health	31000
Arts & Science	29000	Arts & Science	23500
Technology	27500	Technology	30000

Answers to Exercises

(1) According to the leadership literature, there are a number of different leadership styles. Listed below are scores obtained from an instrument designed to measure a particular leadership style, which will be referred to as style X. Of interest is whether X differs by school district type in terms of urbanity. A stratified random sample of school principals were selected from three district types (mostly urban, mostly suburban, and mostly rural).

The scores on style X range from 100 to 0. The closer the score to 100, the more the respondent conforms to style X, while the closer the score to 0, the less the respondent conforms to style X. Is there any evidence that X differs among the three district types?

Table 1a
Descriptive Statistics and Correlations among Style X and District Type

Variable	Correlations		
	Style X	Urban	Suburban
Style X	---		
Urban	.55*	---	
Suburban	-.10	-.48*	---
Mean	56.35	0.35	0.30
SD	25.07	0.49	0.47

Note. n = 20; Urban (1 = urban, 0 = otherwise) and Suburban (1 = suburban, 0 = otherwise) are dummy variables.
* $p < .05$

Table 1b
Regression of Style X on District Type

Variable	b	se	95%CI	t
Urban	33.57	11.51	9.29, 57.85	2.92*
Suburban	11.52	11.98	-13.75, 36.80	0.96
Intercept	41.14	8.14	23.97, 58.31	5.06*

Note. R^2 = .34, adj. R^2 = .26, F = 4.38*, MSE = 463.63; df = 2,17; n = 20; Urban (1 = urban, 0 = otherwise) and Suburban (1 = suburban, 0 = otherwise) are dummy variables.
*$p < .05$.

Table 1c
Comparison of Style X by District Type

Comparison	Estimated Mean Difference	Standard Error of Difference	95% Bonferroni Adjusted CI of Mean Difference
Urban vs. Rural	33.57*	11.51	2.90, 64.24
Suburban vs. Rural	11.52	11.98	-20.41, 43.45
Urban vs. Suburban	22.05	11.98	-9.88, 53.98

* $p < .05$, where p-values are adjusted using the Bonferroni method.

Regression results show that Style X does appear to vary, statistically, among principals from different district types. As shown in the table of comparisons, urban principals show a statistically higher mean of Style X than do rural principals, but there appears to be no statistical difference in Style X between suburban and rural principals, or between urban and suburban principals.
Table below shows Scheffé results should those be needed.

Table 1d
Comparison of Style X by District Type

Comparison	Estimated Mean Difference	Standard Error of Difference	95% Scheffé Adjusted CI of Mean Difference
Urban vs. Rural	33.57*	11.51	2.73, 64.41
Suburban vs. Rural	11.52	11.98	-20.58, 43.62
Urban vs. Suburban	22.05	11.98	-10.05, 54.15

* $p < .05$, where p-values are adjusted using the Scheffé method.

(2) Which of the following small car makers, if any, have the better fuel economy in terms of miles per gallon (MPG)? For this study, a small car is defined as any vehicle with a 2.2 litre (or less), 4-cylinder engine with less than 130 horsepower.

Table 2a
Descriptive Statistics and Correlations among MPG and Make of Automobile

Variable	Correlations			
	MPG	Honda	Subaru	Toyota
MPG	---			
Honda	.31	---		
Subaru	-.07	-.33	---	
Toyota	.13	-.33	-.33	---
Mean	35.25	0.25	0.25	0.25
SD	4.02	0.44	0.44	0.44

Note. n = 32; Honda (1 = Honda, 0 = otherwise), Subaru (1 = Subaru, 0 = otherwise), and Toyota (Toyota = 1, 0 = otherwise) are dummy variables.
* $p < .05$

Table 2b
Regression of MPG on Maker

Variable	b	se	95%CI	t
Honda	4.63	1.91	0.72, 8.53	2.42*
Subaru	2.00	1.91	-1.91, 5.91	1.05
Toyota	3.38	1.91	-0.53, 7.28	1.77
Intercept	32.75	1.35	29.99. 35.51	24.27*

Note. $R^2 = .18$, *adj.* $R^2 = .10$, $F_{3,28} = 2.16$, MSE = 14.56, n = 32. Honda (1 = Honda, 0 = otherwise), Subaru (1 = Subaru, 0 = otherwise), and Toyota (Toyota = 1, 0 = otherwise) are dummy variables.
* $p < .05$.

Analysis of the data (see Table 2b) indicates that no statistical differences in average MPG were observed among the four automobile makers compared based upon the overall model F = 2.16 (p>.05).

(Note that while the overall model was not statistically significant, there was one statistically significant contrast--Honda vs. GM. Odds are this one rejection of the null is a Type 1 error. If multiple comparison procedures, such as Scheffé or Bonferroni, were used to control for inflation of the Type 1 error rate, then it is doubtful that Honda would be different from GM.)

Since the overall model was not statistically significant, the follow tables of multiple comparisons are not needed but are provided below as a calculation check. Note that none of the comparisons are statistically significant whether the Bonferroni or Scheffé procedure was used.

Table 2c
Comparisons of Mean Levels of MPG by Auto Maker

Comparison	Estimated Mean Difference	Standard Error of Difference	95% Bonferroni Adjusted CI of Mean Difference
Honda vs. GM	4.63	1.91	10.01, -0.75
Subaru vs. GM	2.00	1.91	7.38, -3.38
Toyota vs. GM	3.38	1.91	8.76, -2.00
Honda vs. Toyota	1.25	1.91	6.63, -4.13
Subaru vs. Toyota	-1.38	1.91	4.00, -6.76
Honda vs. Subaru	2.63	1.91	8.01, -2.75

* $p < .05$, where p-values are adjusted using the Bonferroni method.

Table 2d
Comparisons of Mean Levels of MPG by Auto Maker

Comparison	Estimated Mean Difference	Standard Error of Difference	95% Scheffé Adjusted CI of Mean Difference
Honda vs. GM	4.63	1.91	10.21, -0.95
Subaru vs. GM	2.00	1.91	7.58, -3.58
Toyota vs. GM	3.38	1.91	8.96, -2.20
Honda vs. Toyota	1.25	1.91	6.83, -4.33
Subaru vs. Toyota	-1.38	1.91	4.20, -6.96
Honda vs. Subaru	2.63	1.91	8.21, -2.95

* $p < .05$, where p-values are adjusted using the Scheffé method.

(3) A researcher is interested in learning whether frequency of reading at home to elementary-aged children produces differential effects on reading achievement. After obtaining information from a randomly selected sample of parents about this behavior, the following classifications and standardized achievement scores were recorded. (Note: frequency classifications as follows: a = less than once per month, b = once to three times per month, c = more than three times per month.)
Is frequency of reading at home related to student reading achievement?

Table 3a
Descriptive Statistics and Correlations among Reading Achievement and Frequency of Reading

Variable	Correlations 1	2	3
1. Achievement	---		
2. Once to 3 Per Month	-.09	---	
3. More the 3 Per Month	.04	-.50	---
Mean	48.67	0.33	0.33
SD	10.13	0.49	0.49

Note. n = 12; Once to 3 Per month (1 = 1 to 3, 0 = otherwise) and More than 3 per month (1 = 3 or more, 0 = otherwise) are dummy variables.
* $p < .05$

Table 3b
Regression of Reading Achievement on Frequency of Reading Per Month

Variable	b	se	95%CI	t
One to three times	-1.75	7.89	-19.60, 16.10	-0.22
Three times or more	-0.00	7.89	-17.85, 17.85	-0.00
Intercept	49.20	5.58	36.63, 61.87	8.83*

Note. R^2 = .01, *adj.* R^2 = -.21, $F_{2,9}$ = 0.03, MSE = 124.50, n = 12. Once to 3 Per month (1 = 1 to 3, 0 = otherwise) and More than 3 per month (1 = 3 or more, 0 = otherwise) are dummy variables.
* $p < .05$.

The analysis of the data (see Table 3b) indicates that no statistical differences in average levels of achievement exists among students exposed to different levels of reading at home (according to the model F = 0.03, p>.05).

(Note that neither Table 3c nor Table 3d is needed since no statistical difference was found, but both are reported here to allow one to check accuracy of statistical estimates.)

Table 3c
Contrasts in Mean Levels of Achievement by Frequency of Reading at Home

Comparison	Estimated Mean Difference	Standard Error of Difference	95% Bonferroni Adjusted CI of Difference
One to Three vs. None	-1.75	7.89	21.31, -24.81
Three + vs. None	-0.00	7.89	23.06, -23.06
One to Three vs. Three +	-1.75	7.89	21.31, -24.81

* p < .05, where p-values are adjusted using the Bonferroni method.

Table 3d
Contrasts in Mean Levels of Achievement by Frequency of Reading at Home

Comparison	Estimated Mean Difference	Standard Error of Difference	95% Scheffé Adjusted CI of Difference
One to Three vs. None	-1.75	7.89	21.29, -24.79
Three + vs. None	-0.00	7.89	23.04, -23.04
One to Three vs. Three +	-1.75	7.89	21.29, -24.79

* p < .05, where p-values are adjusted using the Scheffé method.

(4) Does a difference in salary for beginning assistant professors exist at GSU by college? Which colleges appear to have difference salary levels from the other colleges?

Table 4a
Descriptive Statistics and Correlations among Salary and College of Employment

Variable	Correlations				
	1	2	3	4	5
1. Salary	---				
2. Health	.18	---			
3. Technology	-.30	-.25	---		
4. Arts & Science	-.51*	-.25	-.25	---	
5. Business	.53*	-.25	-.25	-.25	---
Mean	32392.50	0.20	0.20	0.20	0.20
SD	6701.07	0.41	0.41	0.41	0.41

Note. n = 20; Health (1 = College of Health, 0 = otherwise), Technology (1 = College of Technology, 0 = others), Arts & Science (1 = College of Arts and Science, 0 = others), and Business (1 = College of Business, 0 = others) are dummy variables.
* p < .05

Table 4b
Regression of Salary on College of Employment

Variable	b	se	95%CI	t
Health	1000.00	3591.12	-6654.29, 8654.29	0.28
Technology	-5312.50	3591.12	-12966.79, 2341.79	-1.48
Arts	-8075.00*	3591.12	-15729.29, -420.71	-2.25*
Business	5600.00	3591.12	-2054.29, 13254.29	1.56
Intercept	33750.00	2539.31	28337.60, 39162.40	13.29

Note. R^2 = .55, adj. R^2 = .43, $F_{4,15}$ = 4.52*, *MSE* = 25792291.66, n = 20. Health (1 = College of Health, 0 = otherwise), Technology (1 = College of Technology, 0 = others), Arts & Science (1 = College of Arts and Science, 0 = others), and Business (1 = College of Business, 0 = others) are dummy variables.
*p < .05.

Table 4d
Comparison of Mean Salary by College

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Scheffé Adjusted CI of Mean Difference
Health vs. Educa.	1000.00	3591.12	13568.92, -11568.92
Tech. vs. Educa	-5312.50	3591.12	7256.42, -17881.42
Art vs. Educa	-8075.00	3591.12	4493.92, -20643.92
Bus vs. Educa	5600.00	3591.12	18168.92, -6968.92
Health vs. Bus	-4600.00	3591.12	7968.92, -17168.92
Tech. vs. Bus	-10912.50	3591.12	1656.42, -23481.42
Art vs. Bus	-13675.00*	3591.12	-1106.08, -26243.92
Healt vs. Art	9075.00	3591.12	21643.92, -3493.92
Tech vs. Art	2762.00	3591.12	15330.92, -9806.92
Health vs. Tech	6312.50	3591.12	18881.42, -6256.42

* p < .05, where p-values are adjusted using the Scheffé method.

[Note, since there are 10 comparison, Scheffe should be used for comparisons.]

> The data analysis (see Tables 4b and 4d) indicates that differences in salary do appear to exist among the various colleges at Georgia Southern. In specific, the College of Arts and Sciences has the lowest average salary, and the College of Business had the highest salary, on average. The only mean salary difference to be statistically significant was the mean difference between these two colleges. No other statistically significant differences were observed.

Table 4c provided as a computational check.

Table 4c
Comparisons of Mean Salary by College

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Bonferroni Adjusted CI of Mean Difference
Health vs. Educa.	1000.00	3591.12	12760.92, -10760.92
Tech. vs. Educa	-5312.50	3591.12	6448.42, -17073.42
Art vs. Educa	-8075.00	3591.12	3685.92, -19835.92
Bus vs. Educa	5600.00	3591.12	17360.92, -6160.92
Health vs. Bus	-4600.00	3591.12	7160.92, -16360.92
Tech. vs. Bus	-10912.50	3591.12	848.42, -22673.42
Art vs. Bus	-13675.00*	3591.12	-1914.08, -25435.92
Healt vs. Art	9075.00	3591.12	20835.92, -2685.92
Tech vs. Art	2762.00	3591.12	14522.92, -8998.92
Health vs. Tech	6312.50	3591.12	18073.42, -5448.42

* $p < .05$, where p-values are adjusted using the Bonferroni method.