

Simple Linear Regression: One Multinomial Categorical IV and Multiple Comparisons

1. Extension of the Binary Predictor: Multiple Group Comparisons with Regression

Dummy variables can represent not only binary variables but also multinomial variables, i.e., categorical variables with more than two categories.

For each distinct group or category in the IV, **c - 1 dummy variables** are needed to represent the variable in the regression equation, where c is the number of categories.

For example, suppose one wishes to model stress differences exercise status with four groups – those who practice yoga, aerobic exercises, weightlifting, and a control group who does none of these. Since there are four groups, ($c - 1 = 4 - 1 = 3$) three dummy variables will be needed to represent this variable in the regression equation. Or if, for instance, a categorical IV had 9 groups, $9 - 1 = 8$ dummy variables would be required in the regression equation. Dummy variable representation of the categorical variable exercise status is illustrated below in Table 1.

Everything we learned earlier in regression continues to apply. Because of this, some of the detailed discussed earlier will not be repeated in this presentation.

2. Fictional Data

Research Question: Is there a difference in self-reported stress (using a 10-point scale, 1 = low to no stress, 10 = very high level of stress) among by exercise status:

- Yoga = those who participate in yoga at least once per week
- Aerobic = those who engage in aerobic exercise at least once per week
- Weight = those who lift weights at least once per week
- Control = those who do not engage in any of the above activities

Table 1: Fictional Data

Y = Stress	Group	Yo = Yoga	A = Aerobic	W = Weights	C = Control
7	Y	1	0	0	0
9	Y	1	0	0	0
6	Y	1	0	0	0
8	Y	1	0	0	0
9	A	0	1	0	0
8	A	0	1	0	0
8	A	0	1	0	0
6	A	0	1	0	0
6	W	0	0	1	0
7	W	0	0	1	0
5	W	0	0	1	0
4	W	0	0	1	0
6	C	0	0	0	1
5	C	0	0	0	1
3	C	0	0	0	1
3	C	0	0	0	1

Note. Stress means by group: Yoga = 7.50, Aerobic = 7.75, Weights = 5.50, and Control = 4.25.

In addition to reporting regression results, the goal is to examine all possible pairwise comparisons among groups to pinpoint which groups differ in stress levels. The table below must be filled with estimates of these differences.

Pairwise Comparisons in Mean Stress Levels by Exercise Status

Comparison	Estimated Mean Difference	Standard Error of Difference	t ratio	95% CI of Difference
Yoga vs. Control	?			
Aerobic vs. Control	?			
Weightlifting vs. Control	?			
Aerobic vs. Yoga	?			
Weightlifting vs. Yoga	?			
Aerobic vs. Weightlifting	?			

* $p < .05$.

3. Regression Equation

The sample regression equation for these data with three dummy variables is

$$Y_i = b_0 + b_1Y_o + b_2A + b_3W + e_i$$

where Y_o represents the dummy variable for Yoga respondents, A represents aerobic respondents, and W is the dummy variable for weightlifters. The control group is the omitted group with no dummy variable.

4. Coefficient Interpretation

As with one dichotomous dummy variable, the intercept, b_0 , represents the mean for the omitted group, which in this example is the control group. How can this be? When the dummy variables for Y_o , A , and W all equal 0, then the regression equation reduces to the intercept:

$$Y'_{\text{Control}} = Y_i = b_0 + b_1Y_o + b_2A + b_3W$$

$$Y'_{\text{Control}} = b_0 + b_1(0) + b_2(0) + b_3(0)$$

$$Y'_{\text{Control}} = b_0$$

where the predicted mean for control respondents is symbolized as Y'_{Control} . The predicted mean for the Yoga group is:

$$Y'_{\text{Yoga}} = Y_i = b_0 + b_1Y_o + b_2A + b_3W$$

$$Y'_{\text{Yoga}} = Y_i = b_0 + b_1(1) + b_2(0) + b_3(0)$$

$$Y'_{\text{Yoga}} = Y_i = b_0 + b_1(1)$$

The predicted mean for Aerobic group is

$$Y'_{\text{Aerobic}} = Y_i = b_0 + b_1Y_o + b_2A + b_3W$$

$$Y'_{\text{Aerobic}} = Y_i = b_0 + b_1(0) + b_2(1) + b_3(0)$$

$$Y'_{\text{Aerobic}} = Y_i = b_0 + b_2(1)$$

and for the weightlifting group is

$$Y'_{\text{Weight}} = Y_i = b_0 + b_1Y_o + b_2A + b_3W$$

$$Y'_{\text{Weight}} = Y_i = b_0 + b_1(0) + b_2(0) + b_3(1)$$

$$Y'_{\text{Weight}} = Y_i = b_0 + b_3(1)$$

Note that the omitted group, the control group, serves as the comparison or reference group. The omitted group, the group without a dummy variable in the regression equation, serves as the group to which other groups are compared.

As a result, the regression coefficients, b_1 , b_2 , and b_3 in this example represent the mean deviation of the Yoga, Aerobic, and Weightlifting respondents' responses, respectively, from the control group's mean.

For example, Control respondents' mean on stress is 4.25 and Yoga respondents' mean is 7.5. The difference between these Yoga and Control is

$$7.50 - 4.25 = 3.25.$$

This difference of 3.25 points will be reflected in the regression equation by the regression coefficient associated with the dummy for Yoga participants, b_1 . Similarly, the mean difference between Aerobic and Control groups is

$$7.75 - 4.25 = 3.50,$$

and this difference is represented in the regression equation by b_2 . Lastly, the mean difference between the Weightlifting and Control groups is b_3 and will equal

$$5.50 - 4.25 = 1.25.$$

The table below summarizes the mean differences and the dummy variable coefficients.

Group	Stress Mean	Mean Comparison	Mean Difference	Model Coefficient
Weightlifting	5.50	Weight vs. Control	$5.50 - 4.25 = 1.25$	b_3
Aerobic	7.75	Aerobic vs. Control	$7.75 - 4.25 = 3.50$	b_2
Yoga	7.50	Yoga vs. Control	$7.50 - 4.25 = 3.25$	b_1
Control	4.25	NA	$M = 4.25$	b_0

5. Obtaining Regression Results

Note: Run regression results with both SPSS and JASP. Illustrate order control of categories in JASP so groups are ordered C, Yo, A, and W.

JASP regression results with the Control group as the comparison group are shown below. Note that the coefficients agree with the calculations presented above.

Model Summary - Stress				
Model	R	R ²	Adjusted R ²	RMSE
H ₀	0.000	0.000	0.000	1.915
H ₁	0.780	0.609	0.511	1.339

ANOVA						
Model		Sum of Squares	df	Mean Square	F	p
H ₁	Regression	33.500	3	11.167	6.233	0.009
	Residual	21.500	12	1.792		
	Total	55.000	15			

Note. The intercept model is omitted, as no meaningful information can be shown.

Coefficients								
Model		Unstandardized	Standard Error	Standardized ^a	t	p	95% CI	
							Lower	Upper
H ₀	(Intercept)	6.250	0.479		13.056	< .001	5.230	7.270
H ₁	(Intercept)	4.250	0.669		6.350	< .001	2.792	5.708
	Group (Y)	3.250	0.946		3.434	0.005	1.188	5.312
	Group (A)	3.500	0.946		3.698	0.003	1.438	5.562
	Group (W)	1.250	0.946		1.321	0.211	-0.812	3.312

^a Standardized coefficients can only be computed for continuous predictors.

For the regression equation in JASP, focus on the one labeled Model H₁. Note that the regression coefficients provided in JASP output above match the mean differences calculated in the table above. Namely:

$$b_0 = 4.25 \text{ (mean of the control group)}$$

$$b_1 Y_0 = 3.25 \text{ (mean difference between Yoga and Control)}$$

$$b_2 A = 3.50 \text{ (mean difference between Aerobic and Control)}$$

$$b_3 W = 1.25 \text{ (mean difference between Weightlifting and Control)}$$

Each of the exercise groups had higher mean stress than the control group. Are these differences large enough to be judged real? Inferential procedures are needed to detect random differences from inferred real differences.

6. Inferential Procedures for Coefficients and Pairwise Comparisons

Statistical inference of regression is a multi-step process.

First, one should assess the overall model fit (i.e., does the model predict more variance in the DV than expected by chance?

Second, if the answer to model fit is yes, reject H₀, then next step is to assess predictor variable contribution to the model. For quantitative variables this means assessing t-tests for each coefficient. But for categorical variables, it may sometimes mean assessing an F-test of the ΔR^2 for that variable if there are more than two groups.

Lastly, if a categorical variable makes a significant contribution to model fit, then the next step is to compare group means using multiple comparison procedures.

6a. Model Fit

Interpretations for Multiple R, R^2 , adj. R^2 , SEE (standard error of estimate), and MSE (mean squared error or mean squared residual) provided in previous notes on regression continue to hold here. The linkage between regression and ANOVA also holds. In fact, the regression of Y (stress) on a qualitative IV is exactly equal to one-way ANOVA, as will be illustrated later.

Below are the fit statistics for the estimated model.

$$Y_i = b_0 + b_1Y_0 + b_2A + b_3W + e_i$$

Note that the model fit information for regression, in terms of MSE (33.45) and F ratios are identical to ANOVA results, presented below, for data in Table 1.

JASP model fit statistics provided below.

Model Summary - Stress				
Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	1.915
H_1	0.780	0.609	0.511	1.339

ANOVA						
Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	33.500	3	11.167	6.233	0.009
	Residual	21.500	12	1.792		
	Total	55.000	15			

Note. The intercept model is omitted, as no meaningful information can be shown.

The regression sums of squares, SSR, equals 33.50, and the overall model F ratio is 6.233 with an associated p-value, .009. In short, regression with a qualitative IV is identical to ANOVA.

The R^2 may be calculated for both ANOVA and regression using

$$R^2 = SSR/SST = 33.5/55 = 0.609.$$

The adj. R^2 is .57 and is the proportional reduction in error variance due to the regression. JASP reports the RMSE with its root mean squared error. Squaring these figures provides the variance of Stress and the variance of Stress errors after regression. The amount of reduction provides the adjusted R^2 .

$$\text{adj. } R^2 = (1.915^2 - 1.339^2) / 1.915^2 = .511 \text{ (proportional reduction in error predicting Stress)}$$

Conceptually, $\text{adj. } R^2 = (\text{Variance of Stress} - \text{Variance of residuals of Stress after regression}) / (\text{Variance of Stress})$.

6b. Overall Model Fit and Statistical Inference

With any regression or ANOVA model the initial question of interest is whether the model, i.e., the combination of predictors (Xs or vectors), explains more variation in Y than would be expected by chance. You should test the overall hypothesis before testing hypotheses for individual coefficients. The overall hypothesis in regression is denoted as

$$H_0: R^2 = 0.00 \text{ (the model predicts no variance in the dependent variable)}$$

or

$$H_0: \beta_j = 0.00; \text{ (all coefficients equal 0.00, i.e., have no relationship with Y)}$$

note that both hypotheses are equivalent.

The alternative hypotheses would simply indicate that R^2 is not equal to 0.00 (thus some variation in Y is being explained or predicted by the predictors), or that at least one of the regression coefficients is not likely to equal zero (for the multiple regression interpretation). In short, if the null hypothesis is rejected, then one may conclude that some aspect of the model used, i.e., the IVs selected, is statistically related to Y (or at least predicts Y). If the overall hypothesis is rejected, the next step in the analysis is to examine and test the individual regression coefficients.

Also note that the overall fit hypothesis for regression, when using dummy variables for a categorical variable, tests the same thing as the overall ANOVA hypothesis of

$$H_0: \mu_C = \mu_YO = \mu_A = \mu_W,$$

or

$$H_0: \mu_{\text{Control}} = \mu_{\text{Yoga}} = \mu_{\text{Aerobics}} = \mu_{\text{Weightlifting}}.$$

As before, the overall F test is used to test the H_0 . F is calculated, like ANOVA, using any of the following formulae:

$$F = \frac{SSR/df_r}{SSE/df_e} = \frac{SSR/k}{SSE/(n - k - 1)} = \frac{MS_{\text{reg}}}{MSE}$$

where:

- SSR = regression sums of squares;
- SSE = residual sums of squares;
- df_r = regression degrees of freedom;
- df_e = residual degrees of freedom;
- k = number of independent variables (vectors) in the model;
- n = sample size (or number of observations in sample);
- MS_{reg} = mean square (same as ANOVA) due to regression (e.g., between);
- MSE = mean square error (same as ANOVA mean square within).

The overall F test may also be calculated using R^2 as the basis:

$$F = \frac{R^2/k}{(1 - R^2)/(n-k-1)}$$

The F test has two degrees of freedom, one due to regression (explained variation) denoted df_r or df_1 , and one due to residuals or error which is denoted as df_e or df_2 . The more commonly used symbols for any F test are df_1 and df_2 . The formulas for F degrees of freedom are

$$df_1 = k$$

and

$$df_2 = n - k - 1$$

where k is the number of IVs (or vectors) in the model and n is the sample size or total number of observations in the sample. For the example data, the model F is

$$F = \frac{R^2/k}{(1 - R^2)/(n-k-1)} = \frac{.609/3}{(1 - .609)/(16 - 3 - 1)} = 6.23$$

or

$$F = \frac{SSR/df_r}{SSE/df_e} = \frac{33.5/3}{21.5/12} = 6.23.$$

With $df_1 = 3$ and $df_2 = 12$, the .05 level critical F value is

$$.05F_{2,9} = 3.49.$$

Since the calculated F ratio, 6.23, is larger than the critical F ratio, 3.49, the null hypothesis of no explained or predicted variation is rejected at the .05 level of significance. In short, the model appears to provide more explanatory power for Y (stress) than expected by chance.

Note also that the corresponding p-value for F is .009. Since this value is less than .05, the null is rejected.

6c. Reported Pairwise Comparisons

The null hypothesis for each regression coefficient is expressed, symbolically, as previously outlined in an earlier presentation:

$$H_0: \beta_0 = 0.00,$$

$$H_0: \beta_1 = 0.00,$$

$$H_0: \beta_2 = 0.00, \text{ and}$$

$$H_0: \beta_3 = 0.00.$$

The null for b_0 indicates the b_0 does not differ from 0.00. Since b_0 represents the mean of the control group, an equivalent null is

$$H_0: \mu_{\text{Control}} = 0.00.$$

Since b_1 represents a pairwise comparison of Control to Yoga groups, an equivalent null for $H_0: \beta_1 = 0.00$ is

$$H_0: \mu_{\text{Control}} = \mu_{\text{Yoga}}$$

and for $H_0: \beta_2 = 0.00$, the pairwise comparison of Control to Aerobic, the equivalent null is:

$$H_0: \mu_{\text{Control}} = \mu_{\text{Aerobic}}$$

For $H_0: \beta_3 = 0.00$, the pairwise comparison of Control to Weightlifters, the equivalent null is:

$$H_0: \mu_{\text{Control}} = \mu_{\text{Weight}}$$

As mentioned, b_1 represents the mean difference between Control and Yoga respondents. When one simply compares one group to another, such contrasts are often referred to as **pairwise comparisons**. The ratio of b_1 to its standard error, se_{b_1} , forms the usual t-ratio for the comparison. Also, a p-value is presented for $H_0: \beta_1 = 0.00$. Similarly, b_2 is the mean difference between Control and Aerobic respondents. As with b_1 , ratio b_2 / se_{b_2} forms a t-ratio that may be used to test the tenability of $H_0: \beta_2 = 0.00$. The same is true for b_0 and b_3 , that is, the ratio of each to its standard error forms a t-ratio that can be used for hypothesis testing their respective null hypotheses.

The procedure for testing H_0 for dummy IVs is identical to the procedure outlined for regression with one quantitative IV. Since the ratio of the coefficient to its standard error forms a t-ratio, one may choose to use critical t-values for determining statistical significance. But with statistical software, p-values are usually reported, so one may simply employ the usual decision rule to test H_0 :

Decision Rule: If $p \leq \alpha$, reject H_0 , otherwise fail to reject H_0 .

For example, the se of b_1 is 0.946, so the ratio is

$$t_{b_1} = b_1 / se_{b_1} = 3.25 / 0.957 = 3.434$$

with a corresponding p-value of .005. The null hypothesis of no difference, i.e.,

$$H_0: \beta_1 = 0.00 \text{ or } H_0: \mu_{\text{Control}} = \mu_{\text{Yoga}}$$

is rejected since .005 is less than .05.

Similarly, the p-value for b_2 is .003, which is smaller than .05, so the null is rejected, i.e., reject $H_0: \beta_2 = 0.00$, or $H_0: \mu_{\text{Control}} = \mu_{\text{Aerobic}}$. For the comparison between control and weightlifting, however, is not significant, with a p-value of .211 which is greater than alpha so one would conclude that stress is similar for control and weightlifting respondents.

6d. Unreported Pairwise Comparisons

The regression results above provided comparison details for three pairwise comparisons:

1. Yoga vs. Control
2. Aerobic vs. Control
3. Weightlifting vs. Control

Unreported, however, are the other possible pairwise comparisons:

4. Aerobic vs. Yoga
5. Weightlifting vs. Yoga
6. Aerobic vs. Weightlifting

The goal is to provide statistics on each possible pairwise comparison in the regression equation. The additional comparisons must be modeled and included.

Pairwise Comparisons in Mean Stress Levels by Exercise Status

Comparison	Estimated Mean Difference	Standard Error of Difference	t ratio	95% CI of Difference
Yoga vs. Control	3.25*	0.946	3.434	1.18, 5.31
Aerobic vs. Control	3.50*	0.946	3.698	1.43, 5.56
Weightlifting vs. Control	1.25	0.946	1.32	-0.81, 3.31
Aerobic vs. Yoga	?			
Weightlifting vs. Yoga	?			
Aerobic vs. Weightlifting	?			

*p < .05.

Modeling Note: To obtain the correct direction of the mean difference using the regression estimates, the group listed second in the pairwise comparison must be the omitted or reference group (i.e., the group in which b_0 provides the mean). Thus, the Control group was omitted so the mean differences above represent Yoga M – Control M, Aerobic M – Control M, and Weight M – Control M. To obtain the mean difference of Aerobic M – Yoga M, Yoga must be the reference or omitted group in the regression equation.

Inferential Note: Recall that the confidence interval provides a way to test the null of no mean difference. If 0 lies within the interval, the difference is not significant. The value 0.00 is not within the interval for Y vs. C or A vs. C, but 0.0 does lie within the interval for W vs. C so that comparison is not significantly different.

6e. Assessing Pairwise Comparisons using the Software Approach

In the readings for regression with categorical predictors I explained how to perform tests of pairwise comparisons using a manual calculation approach and software approach. Illustrated below is the software approach.

The goal is to obtain the t-ratio for the mean difference between each unreported pairwise comparison listed above. To obtain the comparisons between Yoga and both Aerobic and Weightlifting groups, create a dummy variable for the Control group and insert it into the regression equation to replace the Yoga dummy variable. This makes the Yoga group the comparison group in the regression equation. Thus, the new equation looks like this:

$$Y_i = b_0 + b_1C + b_2A + b_3W + e_i$$

Since Yoga is now the comparison group, the regression coefficients take the following interpretation:

- b_0 = mean of the Yoga group
- b_1C = mean difference between Control and Yoga
- b_2A = mean difference between Aerobic and Yoga
- b_3W = mean difference between Weightlifting and Yoga

So b_2 and b_3 provide the mean differences, and standard errors, needed to perform the comparisons between Aerobics and Yoga, and between Weightlifting and Yoga.

Note: Run regression results with both SPSS and JASP to illustrate order control of categories in JASP so groups are ordered Yo, C, A, and W.

JASP Regression Results with Yoga as the Comparison Group

Coefficients								
Model		Unstandardized	Standard Error	Standardized ^a	t	p	95% CI	
							Lower	Upper
H ₀	(Intercept)	6.250	0.479		13.056	< .001	5.230	7.270
H ₁	(Intercept)	7.500	0.669		11.206	< .001	6.042	8.958
	Group (C)	-3.250	0.946		-3.434	0.005	-5.312	-1.188
	Group (A)	0.250	0.946		0.264	0.796	-1.812	2.312
	Group (W)	-2.000	0.946		-2.113	0.056	-4.062	0.062

^a Standardized coefficients can only be computed for continuous predictors.

Recall stress means by group:

- Yoga = 7.50
- Aerobic = 7.75
- Weights = 5.50
- Control = 4.25

Listed below are the mean differences between Aerobic and Yoga and between Weightlifting and Yoga. The regression results confirm these and also present standard errors, t-ratios, and 95% confidence intervals.

- 4. Aerobic vs. Yoga = $b_{2A} = 7.75 - 7.50 = 0.25$
- 5. Weightlifting vs. Yoga = $b_{3W} = 5.50 - 7.50 = -2.00$

Now the table of pairwise comparisons can be updated.

Pairwise Comparisons in Mean Stress Levels by Exercise Status

Comparison	Estimated Mean Difference	Standard Error of Difference	t ratio	95% CI of Difference
Yoga vs. Control	3.25*	0.946	3.43	1.18, 5.31
Aerobic vs. Control	3.50*	0.946	3.69	1.43, 5.56
Weightlifting vs. Control	1.25	0.946	1.32	-0.81, 3.31
Aerobic vs. Yoga	0.25	0.946	0.26	-1.81, 2.31
Weightlifting vs. Yoga	-2.00	0.946	-2.11	-4.06, 0.06
Aerobic vs. Weightlifting	?			

* $p < .05$.

The last comparison is between Aerobic and Weightlifting. Since Weightlifting is listed second in this comparison, this implies the mean difference should be $7.75 - 5.50 = 2.25$. Thus, Weightlifting must be the omitted group in the next regression run to obtain the correct direction of the mean difference.

JASP regression results with Weightlifting as the comparison group provided below.

Coefficients								
Model		Unstandardized	Standard Error	Standardized ^a	t	p	95% CI	
							Lower	Upper
H ₀	(Intercept)	6.250	0.479		13.056	< .001	5.230	7.270
H ₁	(Intercept)	5.500	0.669		8.218	< .001	4.042	6.958
	Group (C)	-1.250	0.946		-1.321	0.211	-3.312	0.812
	Group (Y)	2.000	0.946		2.113	0.056	-0.062	4.062
	Group (A)	2.250	0.946		2.377	0.035	0.188	4.312

^a Standardized coefficients can only be computed for continuous predictors.

The last comparison in the regression table shows the coefficient of interest, Aerobic vs. Weightlifting, and as expected the mean difference is 2.25. Now the pairwise comparison table can be completed.

Pairwise Comparisons in Mean Stress Levels by Exercise Status

Comparison	Estimated Mean Difference	Standard Error of Difference	t ratio	95% CI of Difference
Yoga vs. Control	3.25*	0.946	3.43	1.18, 5.31
Aerobic vs. Control	3.50*	0.946	3.69	1.43, 5.56
Weightlifting vs. Control	1.25	0.946	1.32	-0.81, 3.31
Aerobic vs. Yoga	0.25	0.946	0.26	-1.81, 2.31
Weightlifting vs. Yoga	-2.00	0.946	-2.11	-4.06, 0.06
Aerobic vs. Weightlifting	2.25*	0.946	2.37	0.18, 4.31

*p < .05.

7. Multiple Comparisons and a Faster Method for Obtaining Comparisons

7a. Multiple Comparisons

When more than one comparison is performed, as in the case of this current example with six comparisons, it is necessary to control the familywise Type 1 error rate. Family refers to collect of tests, or related comparisons performed. The table below shows the family of comparisons for stress by exercise status.

Pairwise Comparisons in Mean Stress Levels by Exercise Status

Comparison	Estimated Mean Difference	Standard Error of Difference	t ratio	95% CI of Mean Difference
Yoga vs. Control	3.25*	0.946	3.43	1.18, 5.31
Aerobic vs. Control	3.50*	0.946	3.69	1.43, 5.56
Weightlifting vs. Control	1.25	0.946	1.32	-0.81, 3.31
Aerobic vs. Yoga	0.25	0.946	0.26	-1.81, 2.31
Weightlifting vs. Yoga	-2.00	0.946	-2.11	-4.06, 0.06
Aerobic vs. Weightlifting	2.25*	0.946	2.37	0.18, 4.31

*p < .05.

When multiple tests are performed, the probability of a Type 1 error across that family of tests increases beyond the set standard of .05 or .01. The potential increase in the probability of a Type 1 error can be calculated using the formula below.

$$\text{Familywise error rate} = 1 - (1 - \alpha)^C$$

Assume $\alpha = .05$ per test and there are six group comparisons across stress, so $c = 6$. The familywise Type 1 error rate for this family of six tests is .2649 which means there is a 26.49% chance of falsely claiming there is a significant group difference when, in fact, that differences does not exist in the population.

$$\begin{aligned}\text{Familywise error rate} &= 1 - (1-\alpha)^c \\ \text{Familywise error rate} &= 1 - (1-.05)^6 \\ \text{Familywise error rate} &= 1 - (.95)^6 \\ \text{Familywise error rate} &= 1 - .73509 \\ \text{Familywise error rate} &= .2649\end{aligned}$$

There are many procedures for controlling this inflation of the Type 1 error rate. I have presented this material in discussion of ANOVA and the logic and tests – mean differences – are the same as performed in regression with dummy variables.

Rather than repeat that material, I will instead link below the ANOVA presentation on multiple comparisons. I cover the Bonferroni and Scheffé methods. JASP provides confidence intervals only for the Tukey honest significant difference method. I provide a link to a brief discussion of that approach too.

Multiple Comparisons:

Presentations, Videos, Data, and Supplemental Material within the context of ANOVA (same results, however)

I cover two methods of controlling the familywise error rate for multiple comparisons: Bonferroni and Scheffé. Both are presented in the summary presentation and the detailed presentation in the link below. I also provide an Excel worksheet that calculates confidence intervals for both; the Excel sheet is located at the bottom of the page.

<https://www.bwgriffin.com/gsu/courses/edur8132/anova-01.htm>

The Tukey honest significant difference approach is explained in the following chapter.

<https://www.bwgriffin.com/gsu/courses/edur8132/notes/2010-Abdi-Newman-Keuls-Test-and-Tukey-Test.pdf>

7b. Faster Way to Obtain Comparisons (and Perform Multiple Comparison Error Rate Correction)

SPSS

- Analyze ->
- General Linear Model ->
- Univariate
 - move DV (Stress) to Dependent Variable box
 - move group identifying variable (Group) to Fixed Factor box
 - select Post Hoc
 - move grouping variable (Group) to Post Hoc Texts for box
 - select Type 1 error rate control option (e.g., Tukey, or Bonferroni, or Scheffe)
 - Continue, then OK to obtain results
- Results will include group mean differences, se, p-values, and Tukey, Bonferroni, or Scheffe adjusted confidence intervals at the 95% level of confidence

JASP

- ANOVA ->
- ANOVA or ANCOVA ->
 - move DV (Stress) to Dependent Variable box
 - move group identifying variable (Group) to Fixed Factor box
 - scroll down click on Post Hoc Tests
 - move grouping variable (Group) to right box
 - select Tukey as correction
 - check Confidence Intervals (95%)
- Results will include group mean differences, se, p-values, and Tukey adjusted confidence intervals at the 95% level of confidence (unfortunately JASP does not provide Cis for Bonferroni or Scheffe corrections).

8. Results in APA Style

An acceptable APA styled result is presented below. There are four tables. The first is a table of correlations and descriptive statistics (M, SD, n). Correlations are presented because it is possible to perform regression with only correlations and descriptive statistics. This allows others to replicate the reported regression results without the raw data. The second table presents the dependent variable descriptive statistics by group, and the third table presents regression results. The fourth table presents pairwise comparisons. This table is needed only if the F ratio is significant for that categorical predictor because that signals some mean differences. If the F ratio is not significant, then group means are not different, so a table of comparisons is unnecessary.

JASP out provided throughout this presentation was used construct the regression and multiple comparison tables, so that JASP output will not be presented again here. Correlations and descriptive statistics are needed from JASP to construct the correlation and descriptive statistics table and are reported below.

JASP correlations among Stress and the three dummy variables for Yoga, Aerobic, and Weightlifting

Variable		Stress	Y	A	W	C
1. Stress	Pearson's r	—				
	p-value	—				
2. Y	Pearson's r	0.389	—			
	p-value	0.136	—			
3. A	Pearson's r	0.467	-0.333	—		
	p-value	0.068	0.207	—		
4. W	Pearson's r	-0.234	-0.333	-0.333	—	
	p-value	0.384	0.207	0.207	—	
5. C	Pearson's r	-0.623	-0.333	-0.333	-0.333	—
	p-value	0.010	0.207	0.207	0.207	—

JASP Descriptive Statistics for Variables Included in Regression Analysis

	stress	Yo	A	W
Valid	16	16	16	16
Missing	0	0	0	0
Mean	6.250	0.250	0.250	0.250
Std. Deviation	1.915	0.447	0.447	0.447
Minimum	3.000	0.000	0.000	0.000
Maximum	9.000	1.000	1.000	1.000

JASP descriptives for Stress, and by groups

	Stress			
	W	C	Y	A
Valid	4	4	4	4
Missing	0	0	0	0
Mode	4.000	3.000	6.000	8.000
Median	5.500	4.000	7.500	8.000
Mean	5.500	4.250	7.500	7.750
Std. Deviation	1.291	1.500	1.291	1.258
Minimum	4.000	3.000	6.000	6.000
Maximum	7.000	6.000	9.000	9.000

• The mode is computed assuming that variables are discreet.

Descriptive Statistics ▼	
Stress	
Valid	16
Missing	0
Mode	6.000 ^a
Median	6.000
Mean	6.250
Std. Deviation	1.915
Minimum	3.000
Maximum	9.000

• The mode is computed assuming that variables are discreet.

Table 1: Descriptive Statistics and Correlations Among Stress and Indicator Variables of Group Membership

Variable	Correlations			
	Stress	Yoga	Aerobic	Weightlifting
Stress	---			
Yoga Group	.389	---		
Aerobic Group	.467	-.333	---	
Weightlifting Group	-.234	-.333	-.333	---
Mean	6.25	.25	.25	.25
SD	1.91	.44	.45	.45
n	16	16	16	16

Note. Yoga (1 = yoga group, 0 = others); Aerobic (1 = aerobic group, 0 = others); Weightlifting (1 = weightlifting group, 0 = others), and Control (1 = Control group, 0 = others) are dummy variables; n = 16.

*p < .05.

Table 2: Descriptive Statistics for Stress by Group

Stress by Group	Mean	SD	n
Yoga	7.50	1.29	4
Aerobic	7.75	1.26	4
Weightlifting	5.50	1.29	4
Control	4.25	1.50	4

Table 3: Regression Analysis of Stress by Group Membership

Variable	b	se b	95% CI	t
Yoga	3.25	0.946	1.19, 5.31	3.43*
Aerobic	3.50	0.946	1.44, 5.56	3.70*
Weightlifting	1.25	0.946	-0.81, 3.31	1.32
Intercept	4.25	0.669	2.79, 5.71	6.35*

Note. $R^2 = .61$, adj. $R^2 = .51$, $F_{3,12} = 8.31^*$, $MSE = 1.79$, $n = 16$.

*p < .05.

Table 4: Pairwise Comparisons in Mean Stress Levels by Exercise Status

Comparison	Estimated Mean Difference	Standard Error of Difference	t ratio	Tukey Adjusted 95% CI of Difference
Yoga vs. Control	3.25*	0.946	3.43	0.44, 6.06
Aerobic vs. Control	3.50*	0.946	3.69	0.69, 6.31
Weightlifting vs. Control	1.25	0.946	1.32	-1.56, 4.06
Aerobic vs. Yoga	0.25	0.946	0.26	-2.56, 3.06
Weightlifting vs. Yoga	-2.00	0.946	-2.11	-4.81, 0.81
Aerobic vs. Weightlifting	2.25*	0.946	2.37	-0.56, 5.06

Note: Confidence interval adjustments based upon Tukey's honest significant difference test.

*p < .05.

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

Regression analysis of the data indicated that mean stress levels varied by exercise group, and these differences were significant at the .05 level. All pairwise comparisons among the four groups were assessed, and results are presented in Table 4. Note that control group respondents reported lower levels of stress than both yoga and aerobic exercise groups. Also, the aerobic group had significantly higher stress than weightlifters. None of the other comparisons, however, were significant, so this suggests similar levels of stress between weightlifters and both control and yoga groups, and between aerobic and yoga groups.

9. Exercises

(1) According to the leadership literature, there are several 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 was 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
Urban	85
Urban	98
Urban	75
Urban	63
Urban	91
Urban	49
Urban	62
Suburban	49
Suburban	48
Suburban	56
Suburban	78
Suburban	35
Suburban	50
Rural	33
Rural	95
Rural	26
Rural	11
Rural	33
Rural	25
Rural	65

(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	Achievement
A	48
A	37
A	47
A	65
B	57
B	39
B	49
B	45
C	61
C	55
C	51
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

10. Answers to Exercises

NOTE: These tables have not been altered to agree in format with tables presented above. Use the tables presented below only for checking your results, not for creating APA style tables.

(1) According to the leadership literature, there are several 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 was 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?

Answer

Table 1.X
Descriptive Statistics for Style by Location

Location	Mean	SD	N
Urban	74.71	17.67	7
Suburban	52.67	14.19	6
Rural	41.14	28.87	7

Table 1.Y
Regression of Style on Location

Variable	B	SE B	t
Rural	-33.57	11.51	-2.92*
Suburban	-22.05	11.98	-1.84
Intercept	74.71	8.14	9.18*

Note. $R^2 = .34$, adj. $R^2 = .26$, $p < .05$, $n = 20$, $F_{2,17} = 4.38^*$, $MSE = 463.62$.

* $p < .05$.

Table 1.Z
Contrasts in Mean Levels of Style X by Location of School

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Confidence Interval of Difference
Rural vs. Urban	-33.57*	11.51	-57.85, -9.29
Suburban vs. Urban	-22.05	11.98	-47.32, 3.23
Suburban vs. Rural	11.52	11.98	-13.75, 36.80

* $p < .05$.

Analysis of the data (see Tables 1.Y and 1.Z) indicates that differences in Style X varies by location of school. All pairwise comparisons between the three locations were performed, and results are presented in Table 1.Z. In general, principals in schools located in mostly urban areas tend to hold the highest level of Style X, and this level is statistically greater than the level exhibited by principals located in mostly rural areas. There is no statistical difference between levels of Style X found in suburban areas and the levels observed in either urban or rural areas.

(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.

Answer

Table 2.X

Descriptive Statistics for MPG by Auto Maker

Maker	Mean	SD	N
Honda	37.38	5.26	8
Toyota	36.13	2.75	8
GM	33.75	4.37	8
Subaru	34.75	1.98	8

Table 2.Y

Regression of MPG on Maker

Variable	B	SE B	t
Honda	4.63	1.91	2.42
Subaru	2.00	1.91	1.05
Toyota	3.38	1.91	1.77
Intercept	32.75	1.35	24.27

Note. $R^2 = .19$, adj. $R^2 = .10$, $p = .12$, $n = 32$, $F_{3,28} = 2.16$, $MSE = 14.56$. * $p < .05$.

Table 2.Z

Contrasts in Mean Levels of MPG by Auto Maker

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Confidence Interval of Difference
Honda vs. GM	4.63*	1.91	0.72, 8.53
Subaru vs. GM	2.00	1.91	-1.91, 5.91
Toyota vs. GM	3.38	1.91	-0.53, 7.28
Honda vs. Toyota	1.25	1.91	-2.66, 5.16
Subaru vs. Toyota	-1.38	1.91	-5.28, 2.53
Honda vs. Subaru	2.63	1.91	-1.28, 6.53

* $p < .05$.

Analysis of the data (see Table 2.Y) indicates that no statistical differences in average MPG were observed among the four automobile makers compared.

(EDUR 8132 Class: 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.)

(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?

Answer

Table 3.X
Descriptive Statistics for Achievement by Frequency of Reading

Frequency	Mean	SD	N
Less than Once per Month	49.25	11.62	4
One to three times per Month	47.50	7.55	4
More than three times per Month	49.25	13.48	4

Table 3.Y
Regression of Achievement on Frequency of Reading at Home

Variable	B	SE B	t
One to three times	-1.75	7.89	-0.22
Three times or more	-0.00	7.89	-0.00
Intercept	49.20	5.58	8.83

Note. $R^2 = .007$, adj. $R^2 = -.21$, $p = .97$, $n = 12$, $F_{2,9} = 0.03$, $MSE = 124.50$.

* $p < .05$.

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

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Confidence Interval of Difference
One to Three vs. None	-1.75	7.89	-19.60, 16.10
Three + vs. None	-0.00	7.89	-17.85, 17.85
One to Three vs. Three +	-1.75	7.89	-19.60, 16.10

* $p < .05$.

The analysis of the data (see Table 3.Y) indicates that no statistical differences in average levels of achievement exists among students exposed to different levels of reading at home.

(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?

Answer

Table 4.X

Descriptive Statistics for Salary by College

College	Mean	SD	N
Education	33750.00	1040.83	4
Business	39350.00	5384.24	4
Health	34750.00	9535.02	4
Arts & Science	25675.00	2507.82	4
Technology	28437.50	1297.03	4

Table 3.Y

Regression of Achievement on Frequency of Reading at Home

Variable	B	SE B	t
Health	1000.00	3591.12	0.28
Technology	-5312.50	3591.12	-1.48
Arts	-8075.00*	3591.12	-2.25
Business	5600.00	3591.12	1.56
Intercept	33750.00	2539.31	13.29

Note. $R^2 = .55$, adj. $R^2 = .43$, $p = .013$, $n = 20$, $F_{4,15} = 4.52$, $MSE = 25792291.66$.

* $p < .05$.

Table 3.Z

Contrasts in Mean Levels of Salary by College

Contrast	Estimated Mean Difference	Standard Error of Difference	95% Confidence Interval of Difference
Arts vs. Education	-8075.00	3591.12	-15729, -420
Business vs. Technology	10912.50	3591.12	3258, 18566
Arts vs. Business	-13675.00	3591.12	-21329, -6020
Arts vs. Health	-9075.00	3591.12	-16729, -1420

Note. All possible pairwise comparisons were performed; only those that were statistically significant at the .05 level were reported.

The data analysis (see Tables 4.Y and 4.Z) indicates that differences in salary do appear to exist among the various colleges at Georgia Southern. In general, the College of Arts and Sciences has the lowest average salary, and this salary is statistically less than the average salaries for faculty in the Colleges of Education, Business, and Health. While the College of Business has the highest average salary, this salary is only statistically higher than the colleges of Arts and Sciences, and Technology. No other statistically significant differences were observed.