

ANCOVA: Combining ANOVA and Regression

1. What is ANCOVA?

ANCOVA is a statistical procedure that enables one to compare groups on a quantitative dependent variable while simultaneously controlling for quantitative predictors. Thus, ANCOVA combines both qualitative and quantitative independent variables in the prediction equation.

As compared to ANOVA, ANCOVA is used because inclusion of the **covariate** (quantitative predictor) in the model can (a) increase power to detect group differences and (b) precision of estimates. Both (a) and (b) are possible if the covariate is associated (correlated) with the DV and not highly correlated with the factor (i.e., ANOVA language for qualitative/categorical predictor variables) of interest. A result of this inclusion of the covariate is the reduction of the model mean squared error (MSE) and thus the increase of corresponding F-ratios and reduction of standard errors such as with pairwise comparisons among group means. Additionally, inclusion of the covariate (c) provides estimates of group means on the DV that statistically control, or adjust, for differences on the covariate. For example, suppose one conducts a true experimental study in which students are randomly assigned to treatment and control condition, but the groups exhibit, through random chance, mean differences in motivation to learn. This is illustrated in Table 1 below.

Table 1: Hypothetical ANCOVA Adjustment

Groups	Motivation to Learn (Covariate)	Observed Posttest Achievement (DV Mean)	Adjusted Posttest Achievement (Estimated DV Mean)
Experimental Condition	26.00	85.00	82.00
Control Condition	22.00	75.00	78.00

Since the experimental group began the study with a higher mean motivation than the control group, ANCOVA can be used to provide a statistical estimate of group differences that account for this initial motivation difference. This estimate is known as an **adjusted mean, predicted mean, conditional mean, or marginal mean**.

The amount of adjustment varies according to the size of the difference between groups on the covariate, the strength of the correlation between the covariate and the DV, and the strength of the association between the covariate and the factor. In the above example, the 10-point mean difference in achievement was adjusted to a 4 point mean difference in achievement after taking into account the 4-point difference in motivation between the groups. In effect, researchers use ANCOVA to equalize group differences on covariates to statistically “level the playing field” so they can better estimate differences on the DV that remove differences due to covariates.

It is important to understand limitations on ANCOVA because some estimates provided by ANCOVA may be unrealistic and therefore lead to inappropriate conclusions and inferences. Two sources to help explain issues with ANCOVA are linked below.

Miller, G.A., & Chapman, J.P. (2001). Misunderstanding Analysis of Covariance. *Journal of Abnormal Psychology*, 110, 40-48.

http://www.bwgriffin.com/gsu/courses/edur8132/notes/anova/Miller_Chapman_ANCOVA_Abuses.pdf

Engqvist, L. (2005) The mistreatment of covariate interaction terms in linear mode analyses of behavioral and evolutionary ecology studies. *Animal Behavior*, 70, 967-971.

http://www.bwgriffin.com/gsu/courses/edur8132/notes/anova/Engqvist_ANCOVA_Interaction_Term.pdf

While issues raised about interpretation of statistical estimates in ANCOVA are important, often researchers overlook these issues with other regression analysis models. This is unfortunate because all problems discussed in the readings above also apply to other multivariate statistical analyses.

2. Illustrative Data

The data used to illustrate ANCOVA below were taken from research on student ratings of instruction. A total of 920 students provided ratings of their course, their instructor, and provided responses to other variables too. For this analysis three variables are of interest: overall instructor rating, course difficulty, and instructor reputation.

Instructor Rating was assessed by student responses to item 30 below.

	Poor	Fair	Good	Very Good	Excellent
29. Overall, how would you rate this course?	1	2	3	4	5
30. Overall, how would you rate this instructor?	1	2	3	4	5

Course Difficulty was measured by responses to item 35 below.

	One of Easiest		About Average		One of Most Difficult
35. Course difficulty, relative to other courses was:	1	2	3	4	5

Instructor Reputation was measured by responses to item 39 below.

	Instructor very bad		About Average		Instructor very good	Didn't know about instructor
39. Before taking this course, what did you hear about this instructor?	1	2	3	4	5	6

Responses to item 39 were then sorted into three categories:

- **Negative** reputation if responses were 1 or 3;
- **No information** (or nothing negative) if responses were 3 or 6; and
- **Positive** reputation if responses were 4 or 5.

Research Question and Covariate Rationale

Do overall ratings of instructors differ among those instructors with negative reputations, positive reputations, or no reputation, and does this difference persist once course difficulty is controlled?

Some argue that lower ratings are given to instructors who teach more difficult courses, so it is important to control for course difficulty when evaluating the relation between ratings and reputation.

Data Links

SPSS: <https://www.bwgriffin.com/gsu/courses/edur8132/notes/StudentRatings2003.sav>

JASP: <https://www.bwgriffin.com/gsu/courses/edur8132/notes/StudentRatings2003.jasp>

CSV: <https://www.bwgriffin.com/gsu/courses/edur8132/notes/StudentRatings2003.csv>

The first 10 records of the 920 cases of these data are provided below in Table 2 which is included to offer a visual of the data used for this analysis. For ANCOVA only the first three columns are needed, overall rating, course difficulty, and reputation group. The two indicator variables, negative group and positive group, are used in regression and not needed for ANCOVA.

Table 2: Student Ratings Data – Sample of 10 records from 920 total cases

Overall Rating	Course Difficulty	Reputation Group	Negative Group Indicator	Positive Group Indicator
5.00	5.00	2.00	.00	.00
5.00	5.00	2.00	.00	.00
4.00	3.00	1.00	1.00	.00
5.00	3.00	3.00	.00	1.00
5.00	3.00	2.00	.00	.00
2.00	3.00	2.00	.00	.00
5.00	3.00	2.00	.00	.00
4.00	3.00	3.00	.00	1.00
3.00	3.00	3.00	.00	1.00
3.00	5.00	3.00	.00	1.00

3. Regression Analysis and Results of Ratings Data

As mentioned previously, both regression and ANOVA are mathematically equivalent, but they look different. To help confirm the similarities between these statistical models, results from regression is provided below and will be compared to ANCOVA results later in the presentation.

Assume we are interested in learning whether overall ratings differ by instructor reputation controlling for course difficulty. The regression equation is

$$Y_i = b_0 + b_1\text{Negative}_{1i} + b_2\text{Positive}_{2i} + b_3\text{Difficulty}_{3i} + e_i, \quad (1)$$

where Negative (1 = Negative reputation, 0 = others) and Positive (1 = Positive reputation, 0 = others) are indicators, or dummy, variables, and difficulty is the covariate. The reference group are instructors with no reputation or a reputation that lies in the middle of the reputational ratings. Results of this regression analysis with multiple comparisons, and in APA style, are provided below.

Note: Show both regression analysis in both SPSS and JASP

Table 3: Descriptive Statistics and Correlations Among Ratings, Reputation, and Difficulty

Variable	Correlations			
	Overall Rating	Negative	Positive	Difficulty
Overall Rating	---			
Negative	-0.3660*	---		
Positive	0.1994*	-0.2732*	---	
Difficulty	0.1272*	0.1097*	0.0479	---
Mean	3.951	0.184	0.247	3.227
SD	1.138	0.388	0.431	0.903

Note. Negative (1 = Negative reputation, 0 = others) and Positive (1 = Positive reputation) are dummy variables; n = 920. *p<.05.

Table 4: Regression of Overall Ratings on Reputation and Difficulty

Variable	b	se	ΔR^2	95%CI	F	t
Reputation			0.154		85.152*	
Negative	-1.049	0.092		-1.231, -0.868		11.354*
Positive	0.247	0.083		0.084, 0.409		2.989*
Difficulty	0.204	0.038	0.025	0.129, 0.279		5.335*
Intercept	3.424	0.128		3.173, 3.675		26.743*

Note. $R^2 = .170$, adj. $R^2 = .167$, $F_{4,7} = 62.728^*$, $MSE = 1.078$, $n = 920$. ΔR^2 represents the squared semi-partial multiple correlation or the increment in R^2 due to adding the respective variable. Negative (1 = Negative reputation, 0 = others) and Positive (1 = Positive reputation, 0 = others) are indicator variables.

* $p < .05$.

Table 5: Comparisons of Adjusted Mean Overall Ratings Among Reputation Groups

Comparisons	Estimated Adjusted Mean Difference	Standard Error of Difference	Bonferroni Adjusted .95CI
Negative vs. No Info	-1.050*	0.092	-1.272, -0.828
Positive vs. No Info	0.247*	0.083	0.049, 0.446
Positive vs. Negative	1.297*	0.105	1.044, 1.550

Note. Overall Rating comparisons adjusted based upon Course Difficulty.

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

Written inference and interpretation

Course difficulty is positively and significantly related to overall ratings (converse of critics' expectations), and there are significant mean differences in overall ratings by reputation groups. Pairwise comparisons adjusted using the Bonferroni method show significant mean differences among all groups. Those who reported a positive reputation rated their instructor highest (adj. $M = 4.33$), while those who reported negative reputation rated their instructor lowest (adj. $M = 3.03$). Those who reported no information or no negative or positive information rated their instructor between the positive and negative groups (adj. $M = 4.08$).

Calculation of Adjusted Means in Regression

To better compare regression and ANCOVA, adjusted means for each of the three groups are calculated below using the regression equation and the mean of course difficulty.

Regression Equation

$$\text{Predicted } Y_i = b_0 + b_1\text{Negative}_{1i} + b_2\text{Positive}_{2i} + b_3\text{Difficulty}_{3i} + e_i$$

$$\text{Predicted } Y_i = 3.424 + -1.049(\text{Negative}) + 0.247(\text{Positive}) + 0.204(\text{Difficulty})$$

Negative Reputation Group

$$\text{Predicted } Y_i = 3.424 + -1.049(1) + 0.247(0) + 0.204(3.227) = \mathbf{3.033}$$

No info Reputation Group

$$\text{Predicted } Y_i = 3.424 + -1.049(0) + 0.247(0) + 0.204(3.227) = \mathbf{4.082}$$

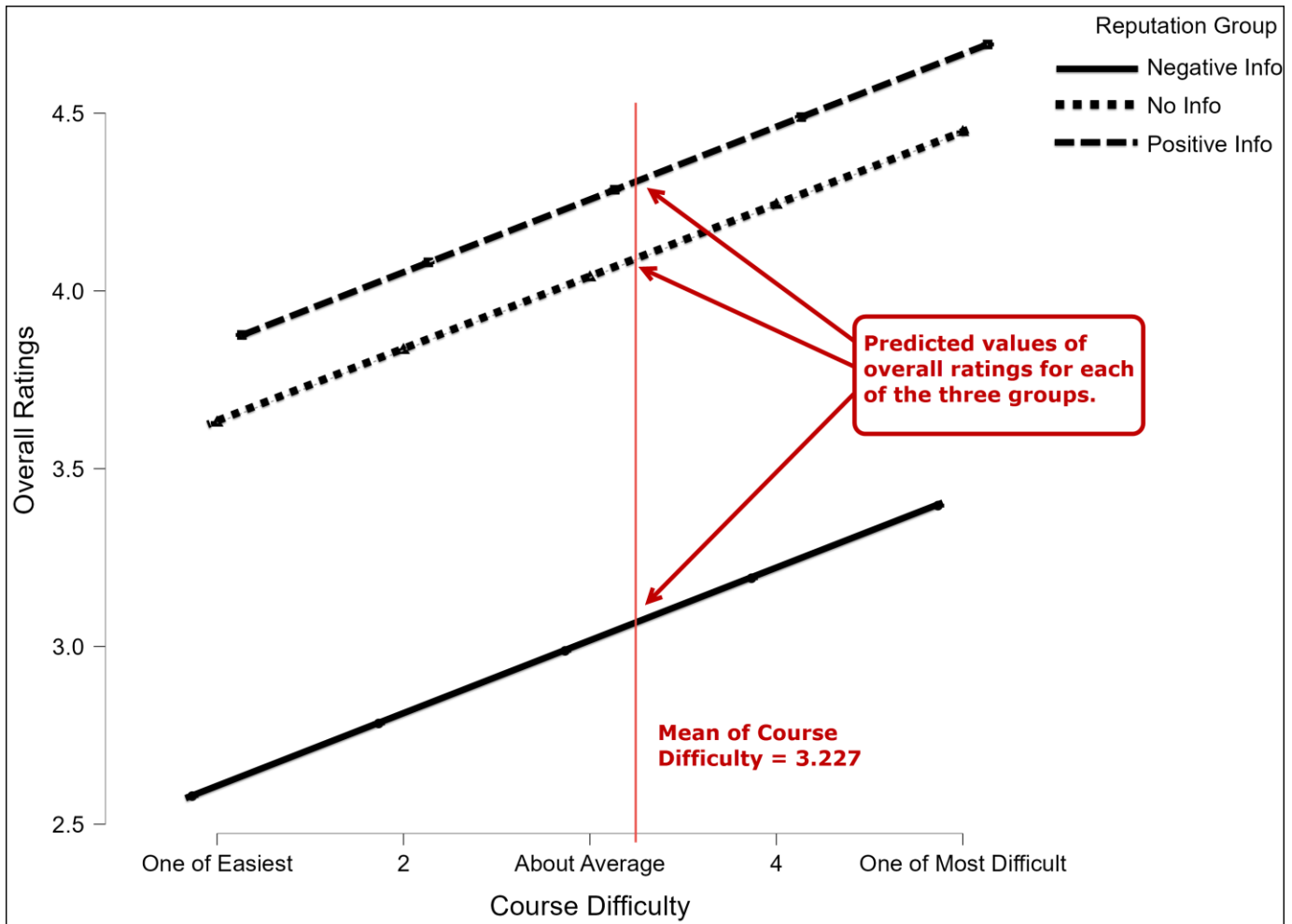
Positive Reputation Group

$$\text{Predicted } Y_i = 3.424 + -1.049(0) + 0.247(1) + 0.204(3.227) = \mathbf{4.329}$$

Visual Display of Parallel Slopes and Adjusted Means

ANCOVA assumes homogeneity of regression, i.e., all groups have the same regression slope for the relation between ratings and course difficulty. The graph below shows parallel slopes – the same regression line for each group but shifted up or down on Y axis of overall ratings. The vertical red line is the mean for course difficulty (3.227) and where that line crosses each group's regression slope is the predicted mean for overall ratings on the Y axis.

Note: The assumption of homogeneity of regression does not mean ANCOVA is invalid if the slopes are not parallel, instead, if homogeneity of regression does not hold then one must include in ANCOVA an interaction between the factor and covariate to obtain correct adjusted means (which now vary across levels of the covariate and factor), interferences, and interpretations. Interactions are discussed in separate presentations.



Instructor's note: Used predicted ratings as DV when creating graph above.

4. Comparison of ANOVA to ANCOVA: Linear Model Representations

As noted above some benefits of ANCOVA include increased power (e.g., larger F-ratios, smaller CI), increased precision (e.g., smaller standard errors, smaller CI), and statistical control/adjustments. To help illustrate the benefits of a covariate, both an ANOVA and ANCOVA will be fitted to the data and compared.

The ANOVA model includes the DV (overall ratings) and the fixed factor (reputational groups). Symbolically it looks like the following equation.

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

where

Y_{ij} = is overall rating for student i in group j ,

μ = grand mean across all persons and groups in the sample,

α_j = the mean difference from μ , or effect, for reputation group j , and

ε_{ij} = is the error term, or how far each score deviates from $\mu + \alpha_j$. (Glass & Hopkins, 1984)

The ANOVA summary table shown in the first presentation on one-way ANOVA lists the sources of variation in the data and calculation of MS and the F ratio. That table is reproduced below.

Source	SS (Sums of Squares)	df (Degrees of Freedom)	MS (Mean Square, i.e., variance)	F (F-ratio)
Between (group, regression)	SSb	df between	MSb = SSb/df b	MSb/MSw
Within (error, residual)	SSw	df within	MSw = SSw/df w	
Total	SSt	df total	SSt/df t = DV variance	

The ANCOVA model includes the DV (overall ratings), the covariate (course difficulty), and the fixed factor (reputational groups). Symbolically it looks like the following equation.

$$Y_{ij} = \mu + \alpha_j + \beta_1 X_{ij} + \varepsilon_{ij}$$

where

Y_{ij} = is overall rating for student i in group j ,

μ = grand mean across all persons and groups in the sample,

α_j = the mean difference from μ , or effect, for reputation group j ,

X_{ij} = is the covariate i 'th score from the j 'th group

β_1 = is the within group slope for the covariate, and

ε_{ij} = is the error term, or how far each score deviates from $\mu + \alpha_j$.

The model equations above are examples of ANOVA and ANCOVA representations of linear models. One could also use the regression equation, provided earlier, to represent the ANOVA/ANCOVA models.

The ANCOVA summary table, shown below, is adjusted to reflect the addition of the covariate.

Source	SS (Sums of Squares)	df (Degrees of Freedom)	MS (Mean Square, i.e., variance)	F (F-ratio)
Treatment (group, regression)	SSb	df between	MSb = SSb/df b	MSb/MSw
Covariate	SSc	df covariate	MSc = SSc/df c	MSc/MSw
Within (error, residual)	SSw	df within	MSw = SSw/df w	
Total	SSt	df total	SSt/df t = DV variance	

5. Comparison of ANOVA to ANCOVA: Results

Below are screenshots showing the ANOVA summary tables and pairwise comparisons tables from both ANOVA and ANCOVA models for the overall rating data. The difference is that the ANOVA model does not include the covariate, course difficulty.

Note: Run both ANOVA and ANCOVA in SPSS and JASP

Summary Table Results

SPSS ANOVA Summary Results

Tests of Between-Subjects Effects					
Dependent Variable: Overall_Instructor_Rating					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	172.252 ^a	2	86.126	77.539	.000
Intercept	10828.303	1	10828.303	9748.745	.000
Reputation_Group	172.252	2	86.126	77.539	.000
Error	1018.547	917	1.111		
Total	15553.000	920			
Corrected Total	1190.799	919			

a. R Squared = .145 (Adjusted R Squared = .143)

SPSS ANCOVA Summary Results

Tests of Between-Subjects Effects					
Dependent Variable: Overall_Instructor_Rating					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	202.947 ^a	3	67.649	62.729	.000
Intercept	620.893	1	620.893	575.733	.000
Course_Difficulty	30.695	1	30.695	28.463	.000
Reputation_Group	183.664	2	91.832	85.152	.000
Error	987.852	916	1.078		
Total	15553.000	920			
Corrected Total	1190.799	919			

a. R Squared = .170 (Adjusted R Squared = .168)

Differences between ANOVA and ANCOVA

F-ratios

ANOVA Reputation F = 77.53

ANCOVA Reputation F = 85.15

Larger F-ratio means more power to detect group differences with ANCOVA.

MSE (error variance)

ANOVA MSE = 1.111

ANCOVA MSE = 1.078

Note the smaller error variance with ANCOVA, so this will lead to larger F-ratios and smaller standard errors.

Model Fit

ANOVA R² = .145

ANCOVA R² = .170

The inclusion of covariate shows that the ANCOVA model predicts more of the variance in overall ratings than does the ANOVA model. This also means the error variance (shown above) and standard error of estimates will be smaller for the ANCOVA model.

Pairwise Comparisons

SPSS ANOVA Comparisons

Pairwise Comparisons						
Dependent Variable: Overall_Instructor_Rating						
(I) Instruction Reputation	(J) Instruction Reputation	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
Negative Info	No Info	-.987*	.093	.000	-1.210	-.764
	Positive Info	-1.270*	.107	.000	-1.526	-1.014
No Info	Negative Info	.987*	.093	.000	.764	1.210
	Positive Info	-.283*	.084	.002	-.484	-.083
Positive Info	Negative Info	1.270*	.107	.000	1.014	1.526
	No Info	.283*	.084	.002	.083	.484

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

SPSS ANCOVA Comparisons

Pairwise Comparisons						
Dependent Variable: Overall_Instructor_Rating						
(I) Instruction Reputation	(J) Instruction Reputation	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
Negative Info	No Info	-1.050*	.092	.000	-1.272	-.828
	Positive Info	-1.297*	.105	.000	-1.550	-1.044
No Info	Negative Info	1.050*	.092	.000	.828	1.272
	Positive Info	-.247*	.083	.009	-.446	-.049
Positive Info	Negative Info	1.297*	.105	.000	1.044	1.550
	No Info	.247*	.083	.009	.049	.446

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Differences between ANOVA and ANCOVA

Standard errors of the mean difference are smaller with ANCOVA, so tighter CIs are the result, and this will provide more power to detect group differences, if differences exists.

The values for the mean differences are also different between the two models reflecting the adjustment of the means for the overall ratings due to differences in course difficulty.

6. Statistical Inference: Overall Model Fit

Like with regression one may perform statistical inferences on the overall model (although this is less a focus in ANOVA literature), predictors, and group comparisons. In this section overall model fit is examined.

SPSS ANCOVA

SPSS provides a test of the overall model fit; look for the line called Corrected Model in the SPSS ANOVA summary table below. In regression this F test assesses the plausibility of the following null.

$$H_0: R^2 = 0.00$$

As explained in the presentation on two-way ANOVA, model fit in the ANOVA literature is usually assessed, if at all, with eta-squared, η^2 , which is mathematically the same as R^2 . Therefore, the model fit null would be expressed in terms of η^2 .

$$H_0: \eta^2 = 0.00$$

This null implies that none of the model factors or covariates are related to the DV. If $H_0: \eta^2 = 0.00$ is rejected, that suggests at least one of the model factors or covariates is statistically related to the DV.

ANCOVA Summary Results

Tests of Between-Subjects Effects					
Dependent Variable: Overall_Instructor_Rating					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	202.947 ^a	3	67.649	62.729	.000
Intercept	620.893	1	620.893	575.733	.000
Course_Difficulty	30.695	1	30.695	28.463	.000
Reputation_Group	183.664	2	91.832	85.152	.000
Error	987.852	916	1.078		
Total	15553.000	920			
Corrected Total	1190.799	919			

a. R Squared = .170 (Adjusted R Squared = .168)

The ANCOVA summary results from JASP are provided below.

JASP ANCOVA

ANCOVA - Overall_Instructor_Rating					
Cases	Sum of Squares	df	Mean Square	F	p
Reputation_Group	183.663700147	2	91.831850074	85.152447684	< .001
Course_Difficulty	30.695350889	1	30.695350889	28.462720272	< .001
Residuals	987.851517550	916	1.078440521		

Note. Type III Sum of Squares

Model F Test

Note that JASP does not provide an overall model F test to assess $H_0: \eta^2 = 0.00$, nor does it provide the overall model SS, df, MS, or R^2 (η^2) value. It is possible, however, to calculate the statistics needed to perform a F-test of the model.

Values needed for Model F Test:

- Model sums of squares (SS);
- Model df which is the sum of factors and covariates df which is $2 + 1 = 3$; and
- error (residual) mean squared error which is given by JASP and is 1.07844.

To calculate the model SS requires the

- total sample size minus one, $n - 1$, so for this sample $n = 920$, and $n - 1 = 919$;
- error SS (or residual SS or within SS) which provided by JASP above as 987.8515; and
- variance of the dependent variable, which is the SD squared.

The variance, or SD, of the DV can be found with most any descriptive statistics command in JASP. The **Descriptives** command reports the variance to be 1.295755.

To calculate the model SS, follow these two steps:

1. Calculate total SS = (variance of the DV) \times ($n - 1$) = $1.295755 \times 919 = 1190.798845$ (which agrees with SPSS)
2. Calculate the model SS = total SS – error SS = $1190.798845 - 987.8515 = 202.9473$ (also agrees with SPSS)

Now to calculate the F ratio for the model SS, follow these two steps:

1. Calculate model MS = model SS / model df = $202.9473 / 3 = 67.64911$ (agrees with SPSS)
2. Calculate model F ratio = model MS / MSE = $67.64911 / 1.07844 = 62.7286$ (agrees with SPSS).

One would then compare this calculated F ratio to the critical F value with 3 and 916 df to determine statistical significance.

Faster Approach

If desired, one could use regression to perform this test since the regression SS = corrected model SS produced by SPSS. Below is the regression output from JASP showing the regression ANOVA summary that is equal to the corrected model SS, df, MS, and F produced by SPSS. Additionally, it provides model fit statistics, also highlighted.

Model Summary - Overall_Instructor_Rating				
Model	R	R ²	Adjusted R ²	RMSE
H ₀	0.000000000	0.000000000	0.000000000	1.138312380
H ₁	0.412831214	0.170429611	0.167712677	1.038479909

ANOVA						
Model		Sum of Squares	df	Mean Square	F	p
H ₁	Regression	202.947395494	3	67.649131831	62.728662817	< .001
	Residual	987.851517550	916	1.078440521		
	Total	1190.798913043	919			

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

Model Fit Statistics: R² and η^2 , and Adj. R² and ε^2

As noted above, in ANOVA literature one will see reference to η^2 (eta squared) which is analogous to R². The formula for R² and η^2 is the ratio Model SS / Total SS (the “corrected” versions of SS as reported by SPSS).

Using the corrected SS from SPSS, the values are

$$R^2 = \eta^2 = \frac{\text{corrected model SS}}{\text{corrected total SS}} = \frac{202.947}{1190.799} = .1704.$$

which agrees with the R² in the SPSS output above.

Calculating R^2 and adj R^2 from JASP Results

JASP (version 0.18.3, 11 Jan 2024) does not provide measures of model fit for ANOVA or ANCOVA, nor does it provide the corrected model SS or corrected total SS. As noted above, one could use regression to obtain these fit measures (see regression output above). However, when these values are not provided it is possible to calculate both R^2 and the adjusted R^2 . In ANOVA parlance the value for adjusted R^2 is called epsilon squared, ε^2 .

R^2 and η^2

Three values are needed to calculate R^2 :

- variance of the dependent variable, $\text{var}(DV)$;
- total sample size minus one, $n - 1$; and
- error SS, also called residual SS and within SS.

Values for the current data:

- variance = SD^2 ; SD for overall rating is 1.13831, so $SD^2 = 1.13831^2 = 1.29575$;
- total sample size is 920, so $n - 1 = 919$; and
- error SS is provided by JASP and called the residual SS = 987.8515.

With these values it is now possible to calculate both the corrected total SS and the corrected model SS.

- Corrected Total SS = $\text{var}(DV) * n-1 = 1.29575 * 919 = 1190.794$
- Corrected Model SS = Corrected Total SS – Error SS = $1190.794 - 987.8515 = 202.9425$

$$R^2 = \eta^2 = \frac{\text{corrected model SS}}{\text{corrected total SS}} = \frac{202.947}{1190.799} = .1704$$

Adjusted R^2 and ε^2

Two values are needed to calculate adj. R^2 , the MSE and the variance of the dependent variable. The formula for adj. R^2 and ε^2 is

$$\text{adj. } R^2 = \varepsilon^2 = 1 - \frac{MSE}{\text{Var}(DV)}$$

where MSE is the mean squared error (or MS residuals, or MS within) from the ANOVA table, and $\text{Var}(DV)$ is the variance of the DV. As noted above, the SD for overall ratings is 1.13831, so the variance is 1.29575. The ANCOVA summary tables from both SPSS and JASP provide the MSE which is 1.07844. The adj R^2 is calculated below.

$$\text{adj. } R^2 = \varepsilon^2 = 1 - \frac{MSE}{\text{Var}(DV)} = 1 - \frac{1.07844}{1.29575} = .1677$$

The value .1677, rounded, agrees with the adj. R^2 of .168 reported in the SPSS ANCOVA output.

ANCOVA Compared to Regression

How do the results for model fit from ANCOVA compare with model fit with regression? Below is the table of regression results for the overall ratings data.

Table 4: Regression of Overall Ratings on Reputation and Difficulty

Variable	b	se	ΔR^2	95%CI	F	t
Reputation			0.154		85.152*	
Negative	-1.049	0.092		-1.231, -0.868		11.354*
Positive	0.247	0.083		0.084, 0.409		2.989*
Difficulty	0.204	0.038	0.025	0.129, 0.279		5.335*
Intercept	3.424	0.128		3.173, 3.675		26.743*

Note. $R^2 = .170$, adj. $R^2 = .167$, $F_{4,7} = 62.728^*$, $MSE = 1.078$, $n = 920$. ΔR^2 represents the squared semi-partial multiple correlation or the increment in R^2 due to adding the respective variable. Negative (1 = Negative reputation, 0 = others) and Positive (1 = Positive reputation, 0 = others) are indicator variables.
* $p < .05$.

Below is the SPSS summary table with several aspects of model fit highlighted in red.

Tests of Between-Subjects Effects					
Dependent Variable: Overall_Instructor_Rating					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	202.947 ^a	3	67.649	62.729	.000
Intercept	620.893	1	620.893	575.733	.000
Course_Difficulty	30.695	1	30.695	28.463	.000
Reputation_Group	183.664	2	91.832	85.152	.000
Error	987.852	916	1.078		
Total	15553.000	920			
Corrected Total	1190.799	919			

a. R Squared = .170 (Adjusted R Squared = .168)

In the table below, note the similarity of results for both ANCOVA and regression.

Model Fit	Regression	ANCOVA
Mean Squared Error	1.078	1.078
Model F-ratio	62.728	62.729
R^2	.170	.170
Adj R^2	.167	.168

7. Statistical Inference: Factors and Covariates

For most researchers the primary interest lies in tests for model factors and covariates. For factors in ANCOVA the null states that adjusted group means are the same, i.e.,

$$H_0: \mu_{1 \text{ adj}} = \mu_{2 \text{ adj}} = \dots = \mu_{j \text{ adj}}$$

where $\mu_{j \text{ adj}}$ is the covariate adjusted mean on the DV for the j 'th group. For covariates, the null holds that each covariate has no relation to the dependent variable, i.e.,

$$H_0: \beta_1 = 0.00.$$

Partial F-tests are used in ANCOVA to assess the statistical significance of each variable. ANCOVA results from both SPSS and JASP are provided below. The F ratio for reputation factor is 85.15 ($p < .001$) and for the covariate, course difficulty, it is 28.46 ($p < .001$). Both predictors are significant at the .05 level.

Note that inferential tests for regression and ANCOVA provide the same information, albeit in somewhat different forms. When a quantitative variable is tested in regression it requires only 1 degree of freedom (df) so a t-ratio will suffice. The t-ratio can be converted to an F-ratio by squaring the t value, i.e., $t^2 = F$ (but only for 1 df tests). This does not apply to categorical variables that require more than one dummy (indicator) variable. For categorical variables the F-test should be used to assess the global effect of that variable in regression (i.e., the F test of variable contribution using the ΔR^2). Note that both ANCOVA and regression provide the same F-ratio results when appropriate t-ratios are squared. The dummy variables Negative and Positive are both required to represent the reputation variable in regression, so reputation requires more than 1 df and therefore it is not appropriate to convert t-ratios to F-ratios for the Negative or Positive indicators.

Summary of Reported Inferential Tests for Predictors

Predictor	Regression F-ratio	Regression t-ratio	ANCOVA F-ratio	Regression ΔR^2	ANCOVA η^2 (JASP)
Course Difficulty	---	5.335 ($t^2 = 28.46$)	F = 28.463	.025	.025
Reputation	F = 85.152	---	F = 85.152	0.154	0.154
Negative Dummy	---	11.354	NA	NA	NA
Positive Dummy	---	2.989	NA	NA	NA

SPSS does not appear to provide η^2 for factors and covariates, but SPSS does provide partial η^2 values, η_p^2 , which have a different interpretation from η^2 . These can be calculated using information found in the ANCOVA summary table. Details are provided in the supplemental material below.

Supplemental: Eta squared (η^2) and Partial Eta squared (η_p^2)

Eta squared (η^2)

- For the ANCOVA or regression models $\eta^2 = R^2$ = proportion of variance in DV predicted by the model
- For factors or covariates, or groups of predictors, $\eta^2 = \Delta R^2$ = proportion of total variance predicted by addition of that IV or group of IVs; also known as a part (or semi-partial) correlation which is the contribution to overall model R or R^2 uniquely above other predictors
- Formula = SS effect / SS **corrected** total (SS = sums of squares; effect is predictor or total model SS)
- Example: IV = course difficulty = 30.695 / 1190.799 = .0257 (regression ΔR^2 , η^2 in JASP)
- Example: IV = Reputation = 183.664 / 1190.799 = .1542 (regression ΔR^2 , η^2 in JASP)
- Example: Total model $R^2 = \eta^2 = 202.947 / 1190.799 = .1704$ (regression R^2)

Partial Eta squared (η_p^2)

- η_p^2 = Proportion of partial variance attributed to factor or covariate; it is a squared partial correlation; a partial correlation is the Pearson r between X and Y controlling for other variables
- Formula = SS effect / (SS effect + SS error)
- Example: IV = course difficulty = 30.695 / (30.695+987.852) = .030136 (SPSS and JASP ANCOVA output)
- Example: IV = Reputation = 183.664 / (183.664+987.852) = .156775 (SPSS and JASP ANCOVA output)

Additional discussion of these is provided in the presentation on two-way ANOVA.

For information on effect size Omega Squared (ω^2), see Maxwell, et al. (1981) and Kroes & Finley (2023).

8. Statistical Inference: Multiple Pairwise Comparisons

If a factor with more than two groups is significant, one should explore group differences using a multiple comparisons procedure that controls for inflation of the Type 1 error rate. Three procedures, the Bonferroni, Scheffé, and Tukey honest significant difference (HSD), have been reviewed in previous presentations and won't be elaborated upon here.

Since a covariate is present in ANCOVA, adjusted means on the DV are compared across the groups when performing pairwise tests. For each pairwise comparison the null specifies that the two adjusted means are the same:

$$H_0: \mu_{j \text{ adj}} = \mu_{k \text{ adj}} \text{ (where } j \text{ and } k \text{ refer to different groups).}$$

Note: Illustrate obtaining multiple pairwise comparisons with JASP and SPSS.

Pairwise Comparisons

Multiple pairwise comparisons with correction procedures are shown below from SPSS and JASP.

SPSS ANCOVA Pairwise Comparisons

Pairwise Comparisons						
Dependent Variable: Overall_Instructor_Rating						
(I) Instruction Reputation	(J) Instruction Reputation	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
Negative Info	No Info	-1.050*	.092	.000	-1.272	-.828
	Positive Info	-1.297*	.105	.000	-1.550	-1.044
No Info	Negative Info	1.050*	.092	.000	.828	1.272
	Positive Info	-.247*	.083	.009	-.446	-.049
Positive Info	Negative Info	1.297*	.105	.000	1.044	1.550
	No Info	.247*	.083	.009	.049	.446

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

JASP ANCOVA Pairwise Comparisons

Post Hoc Comparisons - Reputation_Group							
		Mean Difference	95% CI for Mean Difference		SE	t	P _{Tukey}
			Lower	Upper			
Negative Info	No Info	-1.049841304	-1.266902901	-0.832779708	0.092464077	-11.354045119	< .001
	Positive Info	-1.297120997	-1.544443193	-1.049798801	0.105354512	-12.311964325	< .001
No Info	Positive Info	-0.247279693	-0.441451199	-0.053108187	0.082713338	-2.989598779	0.008063302

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).

These comparison results will agree with the multiple comparisons performed with regression – the mean difference and the standard error (SE) will all agree (compare results above with Table 5 above). Where differences may arise stem from which Type 1 error rate correction procedure used (e.g., Bonferroni, Scheffé, or Tukey). If the same correction procedure was used, then the confidence intervals and inferential tests (p-values, t-ratios, inference decisions) will also agree.

9. Reporting in APA Style

Using the SPSS tables posted above it is possible now to form a results presentation that is suitable for publication. One exception, however, is the test for the homogeneity of regression slopes. That will be explained in a separate presentation on interactions.

Note: Illustrate in both JASP and SPSS how to obtain the following:

- ANOVA summary table,
- group descriptive statistics,
- marginal means,
- parameter estimates for slope(s) of covariate(s),
- model fit information (calculate adj. R^2 for JASP), and
- multiple comparison table.
- Provide brief review of interaction to test homogeneity of regression.

Table 6: ANCOVA Results and Descriptive Statistics for Overall Ratings by Instructor Reputation and Course Difficulty

Reputation Group	Achievement				
	Observed Mean	Adjusted Mean	SD	n	
Negative	3.0764	3.0172	1.060	170	
No Information	4.0632	4.0865	1.129	522	
Positive	4.3464	4.3400	0.849	228	
Source	SS	df	MS	F	η^2
Course Difficulty	30.6593	1	30.6953	28.4627*	0.025
Reputation	183.6637	2	91.8318	85.1524*	0.154
Residuals	987.8515	916	1.0784		

Note. $\eta^2 = .170$, $\epsilon^2 = .168$, adjustments based on Course Difficulty mean = 3.227. Homogeneity of regression tested and not significant: $F = 1.793$, $p = .167$. Course Difficulty regression coefficient = 0.204*.

* $p < .05$

Table 7: Multiple Comparisons and Mean Differences in Achievement by Instruction Type Controlling for Motivation

Comparison	Mean Difference	s.e.	Tukey HSD Adjusted 95% CI
Negative vs. No Info	-1.0498*	0.0924	-1.266, -0.832
Negative vs. Positive	-1.2971*	0.1053	-1.544, -1.049
No Info vs. Positive	-0.2472*	0.0827	-0.441, -0.053

Note. Comparisons based upon ANCOVA adjusted means controlling for Course Difficulty with mean of 3.227.

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

ANCOVA results show that there are statistically significant reputational differences in overall ratings of instructors and that course difficulty is positively associated with overall ratings. Multiple comparisons show that all groups differ, statistically, in overall ratings. Both the observed and adjusted means show that students in the positive reputation group provided the highest ratings, followed by student in the no information group, with students in negative reputation group ratings instructors lowest. Results also show that the higher rated course difficulty, the greater overall ratings of the instructor.

Comment: It is important to test for homogeneity of regression (interaction between covariates and factors), and if the interaction is not statistically significant, the interaction term should be removed from the model and the model re-estimated. In the above example the interaction term between Reputation and Course Difficulty was tested and removed since it was not statistically significant. Homogeneity of regression and interactions are examined in separate videos.

References

- Glass, G.V., & Hopkins, K.D. (1984). *Statistical methods in education and psychology*. 2nd ed. Allyn and Bacon. London.
- Hays, W.L. (1963). *Statistics*. Holt, Rinehart, and Winston.
- Kroes, A. D., & Finley, J. R. (2023). Demystifying omega squared: Practical guidance for effect size in common analysis of variance designs. *Psychological Methods*.
- Maxwell, S.E., Camp, C.J., & Arvey, R.D. (1981). Measures of strength of association: A comparative examination. *Journal of Applied Psychology*, 66(5), 525-534.