ANOVA: Equivalence of ANOVA and Regression

1. Review of ANOVA

The purpose of this presentation is to demonstrate the equivalence of ANOVA and regression in the case of one categorical predictor. The summary presentation on one-way ANOVA, linked below, examined miles per gallon (MPG) of vehicles from Europe, Japan, and North America. The data for this analysis are also linked below.

Summary Presentation: https://www.bwgriffin.com/gsu/courses/edur8132/anova-01.htm
Data File: http://www.bwgriffin.com/gsu/courses/edur8131/data/cars.sav

Recall that a one-way ANOVA refers to the number of predictors, independent variables, not the number of categories in the predictor. So, a two-way ANOVA means there are two categorical predictors, a three-way ANOVA has three categorical predictors, etc.

Note: Show ANOVA in SPSS

Results from SPSS for a one-way ANOVA comparing mean mpg by vehicle origin are shown below.

Descriptive Statistics

Dependent Variable: Miles per Gallon

Country of Origin	Mean	Std. Deviation	N
American	20.08	6.415	244
European	27.60	6.580	68
Japanese	30.45	6.090	79
Total	23.48	7.781	391

Tests of Between-Subjects Effects

Dependent Variable: Miles per Gallon

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7817.309[a]	2	3908.655	96.030	.000	.331
Intercept	194029.594	1	194029.594	4767.050	.000	.925
origin	7817.309	2	3908.655	96.030	.000	.331
Error	15792.466	388	40.702			
Total	239224.740	391				
Corrected Total	23609.775	390				

a. R Squared = .331 (Adjusted R Squared = .328)

Pairwise Comparisons

Dependent Variable: Miles per Gallon

(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
American	European	-7.524*	.875	.000	-9.244	-5.804
	Japanese	-10.372*	.826	.000	-11.996	-8.748
European	American	7.524*	.875	.000	5.804	9.244
	Japanese	-2.848*	1.055	.007	-4.923	-.773
Japanese	American	10.372*	.826	.000	8.748	11.996
	European	2.848*	1.055	.007	.773	4.923

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

The ANOVA shows a significant main effect for origin ($F = 96.03$, $df = 2$, $p < .001$) with mean mpg of 20.08 for American, 27.60 for European, and 30.45 for Japanese cars. Mean comparisons, with unadjusted confidence intervals (i.e., no Tukey, Bonferroni, or Scheffé adjustments for inflated Type 1 error rates), are presented in the pairwise comparisons table.

2. Regression Analysis of the MPG and Origin

As mentioned previously, both regression and ANOVA are mathematically equivalent, but they look different. Comparison of results between the two will help to show the similarity. Below in Table 1 are six cases of the 391 in the cars data file to show the indicator variable (dummy variable) coding used for the regression analysis.

Table 1: MPG and Origin – Sample of 6 records from 391 cases

MPG	Origin	American Indicator	European Indicator	Japanese Indicator
13	1 = American	1.00	.00	.00
15	1 = American	1.00	.00	.00
17	2 = European	.00	1.00	.00
20	2 = European	.00	1.00	.00
18	3 = Japanese	.00	.00	1.00
21	3 = Japanese	.00	.00	1.00

The regression model includes two indicator variables to represent the three categories of origin. The reference group are Japanese cars. The regression equation is

$$Y_i = b_0 + b_1 American_{1i} + b_2 European_{2i} + e_i, \quad (1)$$

where American (1 = American cars, 0 = others) and European (1 = European cars, 0 = others) are the indicator variables. Regression results from JASP are provided below.

Note: Show regression in JASP

Model Summary - mpg

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000000000	0.000000000	0.000000000	7.780609559
H_1	0.575417038	0.331104768	0.327656854	6.379830101

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	7817.309131369	2	3908.654565684	96.030472105	< .001
	Residual	15792.466060447	388	40.702232115		
	Total	23609.775191816	390			

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	95% CI Upper
H_0	(Intercept)	23.482864450	0.393482338		59.679589543	< .001	22.709252466	24.256476434
H_1	(Intercept)	30.450632911	0.717786966		42.422939353	< .001	29.039394199	31.861871624
	origin (European)	−2.847691735	1.055357932		−2.698318408	0.007273416	−4.922627671	−0.772755799
	origin (American)	−10.371944387	0.825851412		−12.559092636	< .001	−11.995648270	−8.748240504

[a] Standardized coefficients can only be computed for continuous predictors.

3. ANOVA: Linear Model Representations

Like regression, the ANOVA model can be displayed symbolically in linear equation form, as shown below.

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

where

Y_{ij} = is mpg for car i in origin j,
μ = grand mean across all cars and origins in the sample,
α_j = the mean difference from μ, or effect, for origin j, and
ε_{ij} = is the error term, or how far each mpg deviates from $\mu + \alpha_j$. (Glass & Hopkins, 1984)

The origin factor is tested with an F-test shown in the ANOVA summary table.

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/df b
Within (error, residual)	SSw	df within	MSw = SSw/df w	
Total	SSt	df total	SSt/df t = DV variance	

4. Comparison between ANOVA and Regressions

Below are screenshots showing various comparisons between ANOVA and regression.

Model Fit and F-ratio Components

- Same R^2 (.331) and adjusted R^2 (.328) values;
- same SS for model, origin factor, error, and total;
- same degrees of freedom;
- same mean squared values; and
- same model F ratios (96.03) and p-values.

SPSS ANOVA

Dependent Variable: Miles per Gallon

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7817.309[a]	2	3908.655	96.030	.000	.331
Intercept	194029.594	1	194029.594	4767.050	.000	.925
origin	7817.309	2	3908.655	96.030	.000	.331
Error	15792.466	388	40.702			
Total	239224.740	391				
Corrected Total	23609.775	390				

a. R Squared = .331 (Adjusted R Squared = .328)

JASP Regression

Model Summary - mpg

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000000000	0.000000000	0.000000000	7.780609559
H_1	0.575417038	0.331104768	0.327656854	6.379830101

ANOVA

Model		Sum of Squares	df	Mean Square	F	p
H_1	Regression	7817.309131369	2	3908.654565684	96.030472105	< .001
	Residual	15792.466060447	388	40.702232115		
	Total	23609.775191816	390			

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

Mean Comparisons

- Same mean differences for
- (a) American vs. Japanese = -10.372;
- (b) European vs. Japanese = -2.847;
- (c) American vs. European = -7.524; and
- same standard errors and confidence intervals (unadjusted for inflation of Type 1 error).

SPSS ANOVA Mean Comparisons

(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
American	European	-7.524*	.875	.000	-9.244	-5.804
	Japanese	-10.372*	.826	.000	-11.996	-8.748
European	American	7.524*	.875	.000	5.804	9.244
	Japanese	-2.848*	1.055	.007	-4.923	-.773
Japanese	American	10.372*	.826	.000	8.748	11.996
	European	2.848*	1.055	.007	.773	4.923

Based on estimated marginal means

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

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

JASP Regression

Model		Unstandardized	Standard Error	Standardized[a]	t	p	95% CI Lower	95% CI Upper
H_0	(Intercept)	23.482864450	0.393482338		59.679589543	< .001	22.709252466	24.256476434
H_1	(Intercept)	30.450632911	0.717786966		42.422939353	< .001	29.039394199	31.861871624
	origin (European)	−2.847691735	1.055357932		−2.698318408	0.007273416	−4.922627671	−0.772755799
	origin (American)	−10.371944387	0.825851412		−12.559092636	< .001	−11.995648270	−8.748240504

[a] Standardized coefficients can only be computed for continuous predictors.

JASP Regression: American vs. European

Model		Unstandardized	Standard Error	Standardized[a]	t	p	95% CI Lower	95% CI Upper
H_0	(Intercept)	23.482864450	0.393482338		59.679589543	< .001	22.709252466	24.256476434
H_1	(Intercept)	27.602941176	0.773668041		35.678016578	< .001	26.081834857	29.124047496
	origin (American)	−7.524252652	0.874856939		−8.600552064	< .001	−9.244306148	−5.804199155
	origin (Japanese)	2.847691735	1.055357932		2.698318408	0.007273416	0.772755799	4.922627671

[a] Standardized coefficients can only be computed for continuous predictors.

As the above comparisons show, ANOVA and regression produce the same inferential tests, same mean comparisons, same standard errors, and the same confidence intervals for mean differences.

The primary difference between the two is that regression provides a prediction equation that both describes the nature of the mean differences and allows one to predict group means for mpg. ANOVA also has a linear model, but it typically is not shown in research reports or often emphasized in reports of ANOVA.