How can I compare regression coefficients between two groups in SPSS?

Authored by

stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I compare regression coefficients between two groups in SPSS?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=162371

Comparing regression coefficients between two groups in SPSS can be done using the "Compare Means" function. This allows for the comparison of regression coefficients from linear regression models for two separate groups. To do this, first run the linear regression analysis for each group separately. Then go to the "Analyze" menu and select "Compare Means" and then "Means." In the dialogue box, select the two groups you want to compare and click "Define Groups." Next, click on the "Options" button and select "Estimated Marginal Means" under "Means and Standard Deviations." Finally, click "OK" to run the analysis. This will generate a table with the regression coefficients for each group, and also provide the results of statistical tests to determine if there is a significant difference between the coefficient estimates.

How can I compare regression coefficients between two groups? | SPSS FAQ

Sometimes your research hypothesis may predict that the size of a
regression coefficient should be bigger for one group than for another. For example, you
might believe that the regression coefficient of height predicting
weight
would be higher for men than for women. Below, we have a data file with 10 fictional
females and 10 fictional males, along with their
height in inches and their weight in pounds.

data list free
/ id * gender (A8) height * weight.
begin data.

1 F 56 117

2 F 60 125

3 F 64 133

4 F 68 141

5 F 72 149

6 F 54 109

7 F 62 128

8 F 65 131

9 F 65 131

10 F 70 145

11 M 64 211

12 M 68 223

13 M 72 235

14 M 76 247

15 M 80 259

16 M 62 201

17 M 69 228

18 M 74 245

19 M 75 241

20 M 82 269

end data.

execute.

We analyzed their data separately using the regression

commands below. Note that we have to do two regressions, one with the data for females only and one with the data for males only. We can use the split file command to split the data file by gender and then run the regression. The parameter estimates (coefficients) for females and males are shown below, and the results do seem to suggest that height is a stronger predictor of weight for males (3.18) than for females (2.09).


```
sort cases by gender.
split file by gender.
regression
/dep weight
/method = enter height.
split file off.
```

Variables Entered/Removed[b]

gender	Model	Variables Entered	Variables Removed	Method
F	1	height[a]	.	Enter
M	1	height[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: weight

Model Summary

gender	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
F	1	.989[a]	.978	.976	1.91504
M	1	.994[a]	.988	.987	2.40738

a. Predictors: (Constant), height

ANOVA[b]

gender	Model		Sum of Squares	df	Mean Square	F	Sig.
F	1	Regression	1319.561	1	1319.561	359.812	.000[a]
		Residual	29.339	8	3.667		
		Total	1348.900	9			
M	1	Regression	3882.536	1	3882.536	669.926	.000[a]
		Residual	46.364	8	5.795		
		Total	3928.900	9			

a. Predictors: (Constant), height

b. Dependent Variable: weight

Coefficients[a]

gender	Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
F	1	(Constant)	-2.397	7.053		-.340	.743
		height	2.096	.110	.989	18.969	.000
M	1	(Constant)	5.602	8.930		.627	.548
		height	3.190	.123	.994	25.883	.000

a. Dependent Variable: weight

We can compare the regression coefficients of males with
females to test the null hypothesis Ho: Bf =
Bm,
where Bf is the regression coefficient for females, and Bm
is the regression coefficient for males. To do this analysis, we first make a dummy
variable called female that is coded 1 for female and 0 for male,
and a variable femht
that is the product of female and height. We then use female, height and femht as predictors in the regression equation.

split file off.

compute female = 0.

```
if gender = "F" female = 1.
compute femht = female*height.
execute.

regression
/dep weight
/method = enter female height femht.
```

The output is shown below.

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	femht, height, female[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: weight

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999[a]	.999	.999	2.17518

a. Predictors: (Constant), femht, height, female

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60327.097	3	20109.032	4250.111	.000[a]
	Residual	75.703	16	4.731		
	Total	60402.800	19			

a. Predictors: (Constant), femht, height, female

b. Dependent Variable: weight

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.602	8.069		.694	.497
	female	-7.999	11.371	-.073	-.703	.492
	height	3.190	.111	.421	28.646	.000
	femht	-1.094	.168	-.638	-6.520	.000

a. Dependent Variable: weight

The term femht tests the null hypothesis Ho: Bf = Bm. The T value is -6.52 and is significant, indicating that the regression coefficient Bf is significantly different from Bm.

Let's look at the parameter estimates to get a better understanding of what they mean and how they are interpreted. First, recall that our dummy variable

female is 1 if female and 0 if male; therefore, males are the omitted group. This is needed for proper interpretation of the estimates.

```
Parameter
Variable Estimate
INTERCEP 5.601677 : This is the intercept for the males (omitted group)
This corresponds to the intercept for males in
the separate groups analysis.
FEMALE -7.999147 : Intercept Females - Intercept males
This corresponds to differences of the
intercepts from the separate groups analysis.
and is indeed -2.397470040 - 5.601677149
HEIGHT 3.189727 : Slope for males (omitted group), i.e. Bm.
FEMHT -1.093855 : Slope for females - Slope for males
(i.e. Bf - Bm).
From the separate groups, this is indeed
2.095872170 - 3.189727463 .
```

It is also possible to run such an analysis using glm, using syntax like that below. Note that other statistical packages, such as SAS and Stata, omit the group of the dummy variable that is coded as zero. However, SPSS omits the group coded as one. Therefore, when you compare the output from the different packages, the results seem to be different. To make the SPSS results match those from other packages, you need to create a new variable that has the opposite coding (i.e., switching the zeros and ones). We do this with the male variable. We do not know of an option in SPSS

glm to easily change which group is the omitted group. (Please note that you can use the contrast subcommand to get the contrast coefficient for female using 0 as the reference group; however, the coding of female in the interaction is such that 1 is used as the reference group, so the use of the contrast subcommand is not very helpful in this situation.)

compute male = not female.

glm weight by male with height
/design = male height male by height
/print = parameter.

Between-Subjects Factors

		N
male	.00	10
	1.00	10

Tests of Between-Subjects Effects

Dependent Variable: weight

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	60327.097[a]	3	20109.032	4250.111	.000
Intercept	.376	1	.376	.079	.782
male	2.342	1	2.342	.495	.492
height	4695.831	1	4695.831	992.480	.000
male * height	201.115	1	201.115	42.506	.000
Error	75.703	16	4.731		
Total	733114.000	20			
Corrected Total	60402.800	19			

a. R Squared = .999 (Adjusted R Squared = .999)

Parameter Estimates

Dependent Variable: weight

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	5.602	8.069	.694	.497	-11.504	22.707
[male=.00]	-7.999	11.371	-.703	.492	-32.104	16.105
[male=1.00]	0[a]	.	.	.	.	.
height	3.190	.111	28.646	.000	2.954	3.426
[male=.00] * height	-1.094	.168	-6.520	.000	-1.450	-.738
[male=1.00] * height	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

As you see, the glm output corresponds to the output obtained by regression. The parameter estimates appear at the end of the glm output. They also correspond to the output from regression.