

How can I compare regression coefficients across three (or more) groups in SPSS?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I compare regression coefficients across three (or more) groups in SPSS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=162395>

The process of comparing regression coefficients across three or more groups in SPSS involves conducting a multiple regression analysis and examining the coefficients for each group separately. This allows for a comparison of the relationships between the independent and dependent variables for each group. Additionally, the use of dummy coding or ANOVA can also aid in comparing the coefficients across groups. Careful interpretation and consideration of the data is necessary in order to accurately compare the coefficients and make meaningful conclusions.

How can I compare regression coefficients across three (or more) groups? | SPSS FAQ

Sometimes your research hypothesis may predict that the size of a regression coefficient may vary across groups. For example, you might believe that the regression coefficient of height predicting weight would differ across three age groups (young, middle age, senior citizen). Below, we have a data file with 10 fictional young people, 10 fictional middle age people, and 10 fictional senior citizens, along with their height in inches and their weight in pounds. The variable age indicates the age group and is coded 1 for young people, 2 for middle aged, and 3 for senior citizens. Below we show two ways that you can get this data file into SPSS. One way is to cut and paste the following code

into an SPSS syntax window and run it.

data list list / id age height weight.

begin data.

1 1 56 140

2 1 60 155

3 1 64 143

4 1 68 161

5 1 72 139

6 1 54 159

7 1 62 138

8 1 65 121

9 1 65 161

10 1 70 145

11 2 56 117

12 2 60 125

13 2 64 133

14 2 68 141

15 2 72 149

16 2 54 109

17 2 62 128

18 2 65 131

19 2 65 131

20 2 70 145

21 3 64 211

22 3 68 223

23 3 72 235

24 3 76 247

25 3 80 259

26 3 62 201

27 3 69 228

28 3 74 245

29 3 75 241

30 3 82 269

end data.

execute.

**Another way is to click on
<https://stats.idre.ucla.edu/wp-content/uploads/2016/02/compreg3.sav>**

and

then use the get file command (insert the proper drive letter if you did not place the file in your current directory):

**get file
'c:<https://stats.idre.ucla.edu/wp-content/uploads/2016/02/compreg3.sav>'.**

After first sorting by age, we analyze the data for each age group separately using the regression command.

In order to use just the data for a specific age group, we need to use a filter to "filter out" the other data. Remember that when you have completed the analysis, you need to turn the filter off.

sort cases by age.

split file by age.

regression

/dep weight

/method=enter height.

split file off.

exe.

The parameter estimates (coefficients) for the young, middle age, and senior citizens are shown below, and the results do seem to suggest that height is a stronger predictor of weight for seniors (3.18) than for the middle aged (2.09). The results also seem to suggest that height does not predict weight as strongly for the

young (-.37) as for the middle aged and seniors. However, we would need to perform specific significance tests to be able to make claims about the differences among these regression coefficients.

< some output omitted to save space >

Model Summary

age	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1.00	1	.170 ^a	.029	-.093	13.42080
2.00	1	.989 ^a	.978	.976	1.91504
3.00	1	.994 ^a	.988	.987	2.40738

a. Predictors: (Constant), height

ANOVA^b

age	Model		Sum of Squares	df	Mean Square	F	Sig.
1.00	1	Regression	42.657	1	42.657	.237	.640 ^a
		Residual	1440.943	8	180.118		
		Total	1483.600	9			
2.00	1	Regression	1319.561	1	1319.561	359.812	.000 ^a
		Residual	29.339	8	3.667		
		Total	1348.900	9			
3.00	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

age	Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
			B	Std. Error	Beta		
1.00	1	(Constant)	170.166	49.430		3.443	.009
		height	-.377	.774	-.170	-.487	.640
2.00	1	(Constant)	-2.397	7.053		-.340	.743
		height	2.096	.110	.989	18.969	.000
3.00	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 among these three age groups to test the null hypothesis

Ho: $B_1 = B_2 = B_3$

where B_1 is the regression for the young, B_2 is the regression for the middle aged, and B_3 is the regression for senior citizens. To do this analysis, we first make a dummy variable called age_1 that is coded 1 if young ($age=1$), 0 otherwise, and age_2 that is coded 1 if middle aged ($age=2$), 0 otherwise. We also create age_1ht that is age_1 times height, and age_2ht that is age_2 times height.

compute $age_1 = 0$.

```
compute age2 = 0.  
if age = 1 age1 = 1.  
if age = 2 age2 = 1.  
compute age1ht = age1*height.  
compute age2ht = age2*height.  
execute.
```

We can now use age1age2height, age1ht and age2ht as predictors in the regression equation in the regress command below. The regression command will be followed by

```
/method = test(age1 age2)
```

and

```
/method = test(age1ht age2ht)
```

The first one provides a 2 degree of freedom to determine if, taken together, the variable age is statistically significant. We have included this for the sake of completeness, because this is a

standard part of the analysis. The second subcommand tests the null hypothesis

Ho: B1 = B2 = B3

This test will also have 2 degrees of freedom because it compares among three regression coefficients.

regression

/dep weight

/method = enter height

/method=test(age1 age2)

/method = test(age1ht age2ht).

< some output omitted to save space >

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.714 ^a	.510	.492	35.27749
2	.974 ^b	.948	.942	11.93281
3	.989 ^c	.979	.974	7.94944

a. Predictors: (Constant), height

b. Predictors: (Constant), height, age2, age1

c. Predictors: (Constant), height, age2, age1, age2ht, age1ht

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.	R Square Change
1	Regression	36265.967	1	36265.967	29.141	.000 ^a	
	Residual	34846.033	28	1244.501			
	Total	71112.000	29				
2	Subset Tests	31143.844	2	15571.922	109.360	.000 ^b	.438
	Regression	67409.811	3	22469.937	157.803	.000 ^c	
	Residual	3702.189	26	142.392			
	Total	71112.000	29				
3	Subset Tests	2185.544	2	1092.772	17.292	.000 ^b	.031
	Regression	69595.355	5	13919.071	220.261	.000 ^d	
	Residual	1516.645	24	63.194			
	Total	71112.000	29				

a. Predictors: (Constant), height

b. Tested against the full model.

c. Predictors in the Full Model: (Constant), height, age2, age1.

d. Predictors in the Full Model: (Constant), height, age2, age1, age2ht, age1ht.

e. Dependent Variable: weight

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-158.525	61.382		-2.583	.015
	height	4.958	.918	.714	5.398	.000
2	(Constant)	108.492	27.745		3.910	.001
	height	1.765	.381	.254	4.635	.000
	age1	-74.524	6.261	-.722	-11.903	.000
	age2	-89.824	6.261	-.870	-14.347	.000
3	(Constant)	5.602	29.489		.190	.851
	height	3.190	.407	.459	7.838	.000
	age1	164.565	41.555	1.593	3.960	.001
	age2	-7.999	41.555	-.077	-.192	.849
	age1ht	-3.567	.613	-2.208	-5.817	.000
	age2ht	-1.094	.613	-.677	-1.784	.087

a. Dependent Variable: weight

The analysis below shows that the null hypothesis

Ho: $B_1 = B_2 = B_3$

can be rejected ($F=17.292$, $p = 0.000$). This means that the regression coefficients between height and weight do indeed significantly differ across the three age groups (young, middle age, senior citizen).

ARABPSYCHOLOGY.COM