One-way ANCOVA

The one-way ANCOVA (**analysis of covariance**) can be thought of as an extension of the one-way ANOVA to incorporate a **covariate**. Like the one-way ANOVA, the one-way ANCOVA is used to determine whether there are any significant differences between two or more independent (unrelated) groups on a dependent variable. However, whereas the ANOVA looks for differences in the group means, the ANCOVA looks for differences in **adjusted means** (i.e., adjusted for the covariate). As such, compared to the one-way ANOVA, **the one-way ANCOVA has the additional benefit of allowing you to "statistically control" for a third variable (sometimes known as a "confounding variable"), which you believe will affect your results**. This third variable that could be confounding your results is called the **covariate** and you include it in your one-way ANCOVA analysis.

Note: You can have more than one covariate and although covariates are traditionally measured on a continuous scale, they can also be categorical. However, when the covariates are categorical, the analysis is not often called ANCOVA. In addition, the "one-way" part of one-way ANCOVA refers to the number of independent variables. If you have two independent variables rather than one, you could run a two-way ANCOVA.

Researchers wanted to investigate the effect of three **different types of exercise intervention on systolic blood pressure**. To do this, they recruited 60 participants to their study. They randomly allocated 20 participants to each of three interventions: a "low-intensity exercise intervention", a "moderate-intensity exercise intervention" and a "high-intensity exercise intervention". The exercise in all interventions burned the same number of calories. Each participant had their "systolic blood pressure" measured before the intervention and immediately after the intervention. The researcher wanted to know if the different exercise interventions had different effects on systolic blood pressure. To answer this question, the researchers wanted to determine whether there were any differences in mean systolic blood pressure after the exercise interventions (i.e., whether post-intervention mean systolic blood pressure different between the different interventions). **However, the researchers expected that the impact of the three different exercise interventions on mean systolic blood pressure would be affected by the participants' starting systolic blood pressure** (i.e., their systolic blood pressure before the interventions). To control the post-intervention systolic blood pressure for the differences in pre-intervention systolic blood pressure, you can run a one-way ANCOVA with pre-intervention systolic blood pressure as the covariate, intervention as the independent variable and post-intervention systolic blood pressure as the dependent variable. If you find a statistically significant difference between interventions, you can follow up a one-way ANCOVA with a post hoc test to determine which specific exercise interventions differed in terms of their effect on systolic blood pressure (e.g., whether the high-intensity exercise intervention had a greater effect on systolic blood pressure than the low-intensity exercise intervention).

Assumptions

When you choose to analyse your data using a one-way ANCOVA, part of the process involves checking to make sure that the data you want to analyse can actually be analysed using a one-way ANCOVA. You need to do this because it is only appropriate to use a one-way ANCOVA if your data "passes" nine assumptions that are required for a one-way ANCOVA to give you a valid result.

- **Assumption #1:** Your **dependent variable** and **covariate variable(s)** should be measured on a **continuous** scale (i.e., they are measured at the **interval** or **ratio** level).
- **Assumption #2:** Your **independent variable** should consist of **two or more categorical**, **independent groups**.
- **Assumption #3:** You should have **independence of observations**, which means that there is no relationship between the observations in each group or between the groups themselves. For example, there must be different participants in each group with no participant being in more than one group. This is more of a study design issue than something you can test for, but it is an important assumption of a one-way ANCOVA. If your study fails this assumption, you will need to use another statistical test instead of a one-way ANCOVA (e.g., a repeated measures design).
- **Assumption #4:** There should be **no significant outliers**. Outliers are simply data points within your data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores globally). The problem with outliers is that they can have a negative effect on the one-way ANCOVA, reducing the validity of your results.
- **Assumption #5:** Your **residuals** should be **approximately normally distributed for each category of the independent variable**. We talk about the ANCOVA only requiring **approximately** normal residuals because it is quite "robust" to violations of normality, meaning that the assumption can be violated to a degree and still provide valid results. You can test for normality using two Shapiro-Wilk tests of normality: one to test the within-group residuals and one to test the overall model fit. Both of these are easily tested for using SPSS Statistics.
- **Assumption #6:** There needs to be **homogeneity of variances**. You can test this assumption in SPSS Statistics using Levene's test for homogeneity of variances.
- **Assumption #7:** The **covariate should be linearly related to the dependent variable at each level of the independent variable**. You can test this assumption in SPSS Statistics by plotting a grouped scatterplot of the covariate, post-test scores of the dependent variable and independent variable. In our enhanced one-way ANCOVA guide, we show you how to (a) produce this grouped scatterplot in SPSS Statistics, (b) interpret the grouped scatterplot, and (c) present possible ways to continue with your analysis if your data fails to meet this assumption.
- **Assumption #8:** There needs to be **homoscedasticity**. You can test this assumption in SPSS Statistics by plotting a scatterplot of the standardized residuals against the predicted values.
- **Assumption #9:** There needs to be **homogeneity of regression slopes**, which means that there is no interaction between the covariate and the independent variable. By default, SPSS Statistics does not include an interaction term between a covariate and an independent in its GLM procedure so that you can test this.

Example

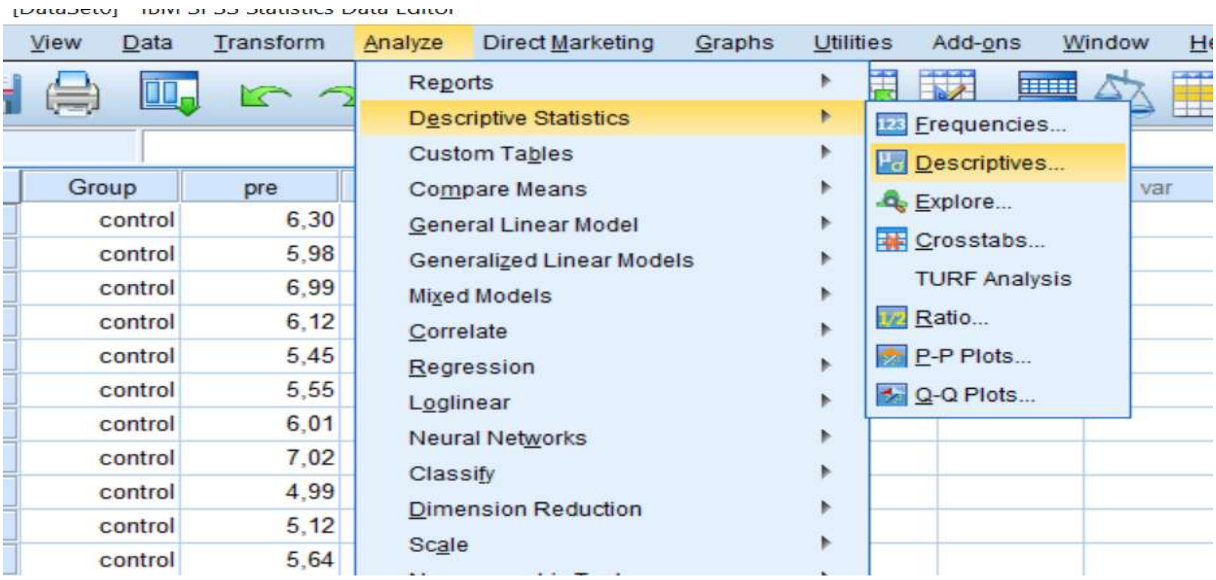
A researcher was interested in determining whether a six-week low- or high-intensity exercise-training programme was best at reducing blood cholesterol concentrations in middle-aged men. Both exercise programmes were designed so that the same number of calories was expended in the low- and high-intensity groups. As such, the duration of exercise differed between groups. The researcher expected that any reduction in cholesterol concentration elicited by the interventions would also depend on the participant's initial cholesterol concentration. As such, the researcher wanted to use pre-intervention cholesterol concentration as a covariate when comparing the post-intervention cholesterol concentrations between the interventions and a control group. Therefore, the researcher ran a one-way ANCOVA with: (a) post-intervention cholesterol concentration (post) as the dependent variable; (b) the control and two intervention groups as levels of the independent variable, group; and (c) the pre-intervention cholesterol concentrations as the covariate, pre.

In SPSS Statistics, we entered three variables: (1) the dependent variable, post, which is the post-intervention cholesterol concentration; (2) the independent variable, group, which has three categories: "control", "Int_1" (representing the low-intensity exercise intervention), and "Int_2" (representing the high-intensity exercise intervention); and (3) pre, which represents the pre-intervention cholesterol concentrations.

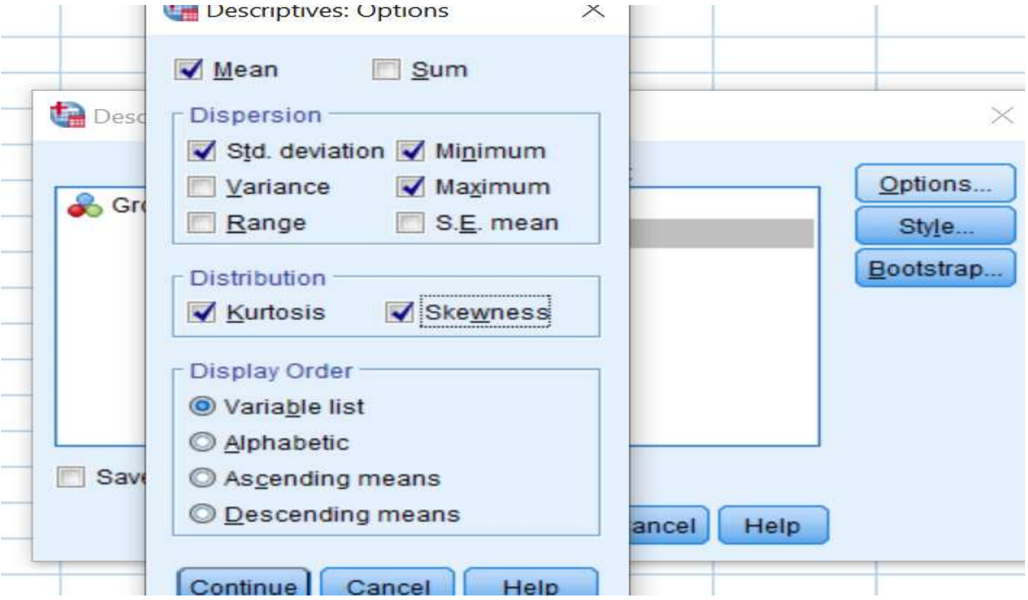
control	PRE	Int_1	PRE	Int_2	PRE
6,30	6,17	5,53	5,65	5,83	5,85
5,98	5,76	6,12	6,18	5,26	5,86
6,99	6,24	6,14	6,16	5,32	5,76
6,12	5,57	6,66	5,49	5,13	5,75
5,45	5,60	5,14	5,47	5,67	5,59
5,55	5,73	5,51	6,36	6,05	6,21
6,01	6,01	5,49	6,65	5,51	6,11
7,02	7,06	6,00	6,54	4,55	6,00
4,99	5,92	5,94	6,13	5,48	5,53
5,12	6,82	6,10	5,80	5,31	5,38
5,64	6,90	6,50	6,54	5,78	6,04
6,00	6,05	6,00	6,48	5,16	5,89
6,11	6,97	5,20	5,26	4,26	5,74
6,31	6,69	5,30	4,43	5,84	6,61
6,85	6,90	5,90	6,39	5,06	5,57

So first we should check the assumptions;

- Normality of the DV and covariate



Choose pre and post ; Ia m interested to find out whether the distribution of these two variable is normal. So I choose kurtosis and skewness.



[DataSet0]

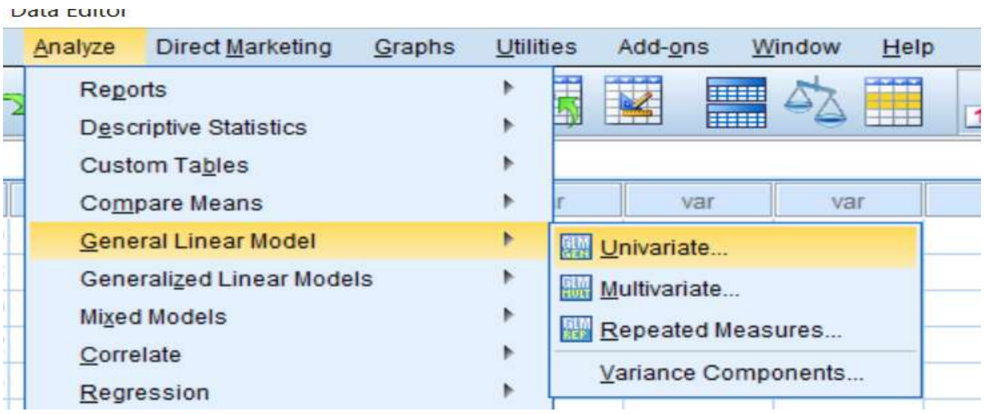
Descriptive Statistics

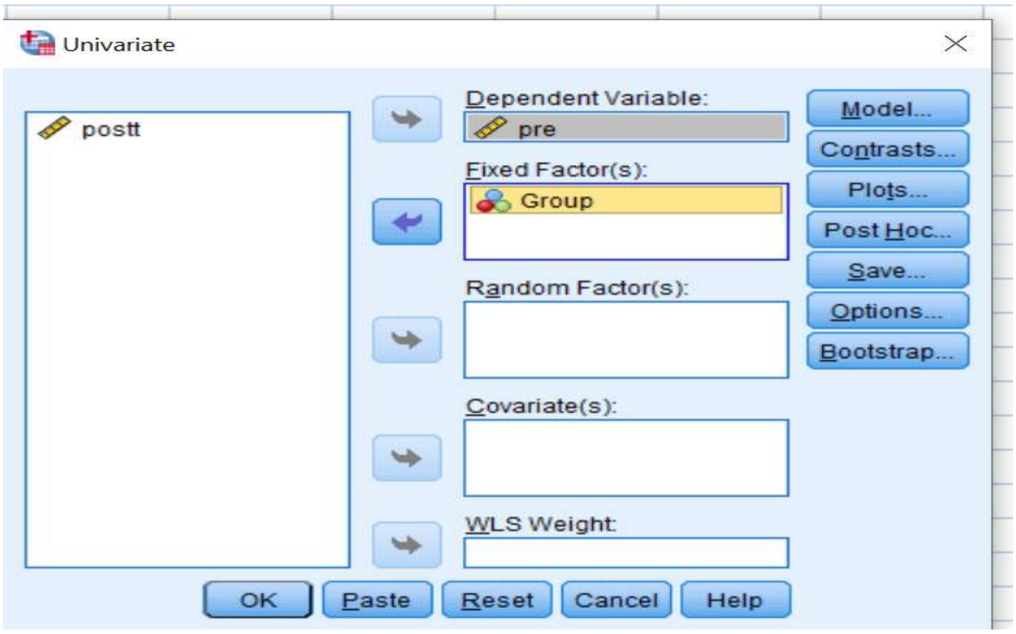
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
postt	45	4,26	7,02	5,7662	,59652	-,058	,354	,157	,695
pre	45	4,43	7,57	6,1267	,56365	-,200	,354	,958	,695
Valid N (listwise)	45								

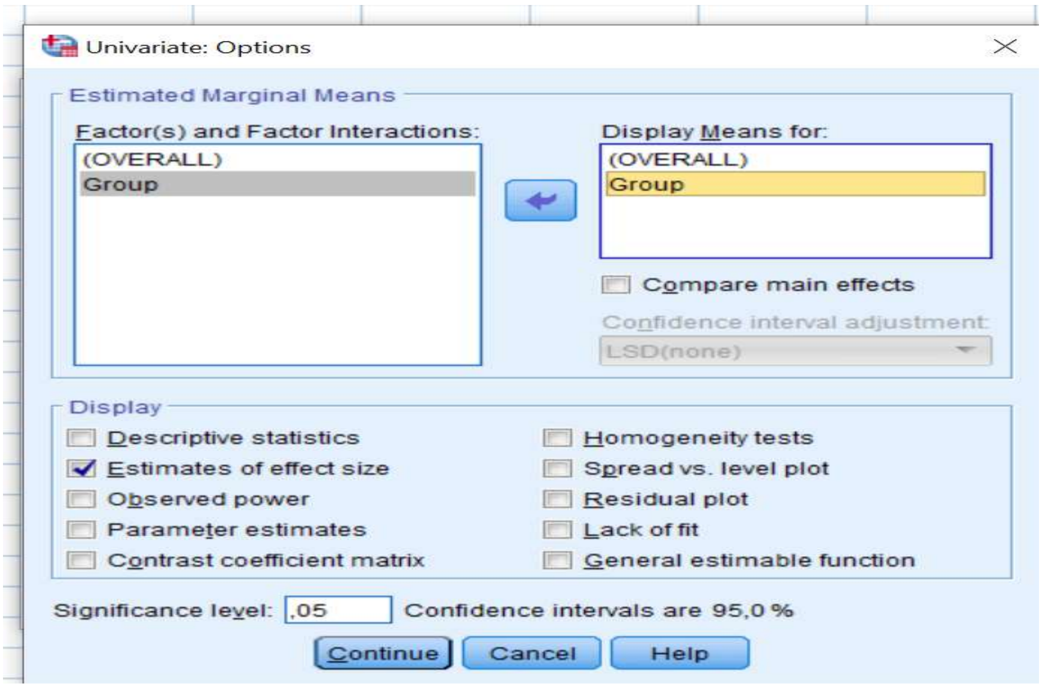
When we check the obtained statistics belong to skewness and kurtosis both of them for both variables are between -2 to +2 .Normality assumption is met.

- Other assumption is The IV does not have a significant effect on the covariate.

So we need to look into relationship between the group and pre sores. To check that


Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Custom Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
var
var
Univariate...
Multivariate...
Repeated Measures...
Variance Components...


Univariate
postt
Dependent Variable:
pre
Fixed Factor(s):
Group
Random Factor(s):
Covariate(s):
WLS Weight:
Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...
Bootstrap...
OK
Paste
Reset
Cancel
Help


Univariate: Options
Estimated Marginal Means
Factor(s) and Factor Interactions:
(OVERALL)
Group
Display Means for:
(OVERALL)
Group
Compare main effects
Confidence interval adjustment:
LSD(none)
Display
Descriptive statistics
Estimates of effect size
Observed power
Parameter estimates
Contrast coefficient matrix
Homogeneity tests
Spread vs. level plot
Residual plot
Lack of fit
General estimable function
Significance level: ,05
Confidence intervals are 95,0 %
Continue
Cancel
Help

By doing this we will se whether groups has an impact on pre _ cholesterol concentration

Tests of Between-Subjects Effects

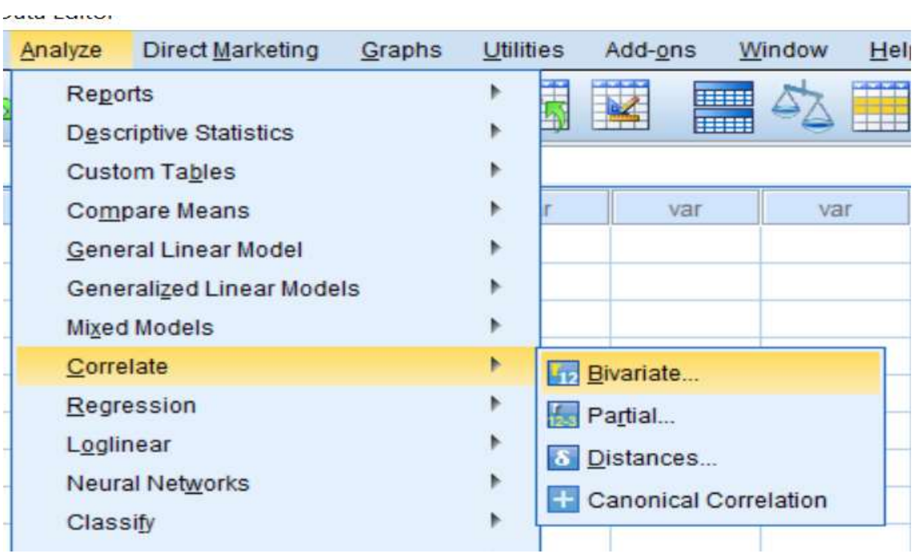
Dependent Variable: pre

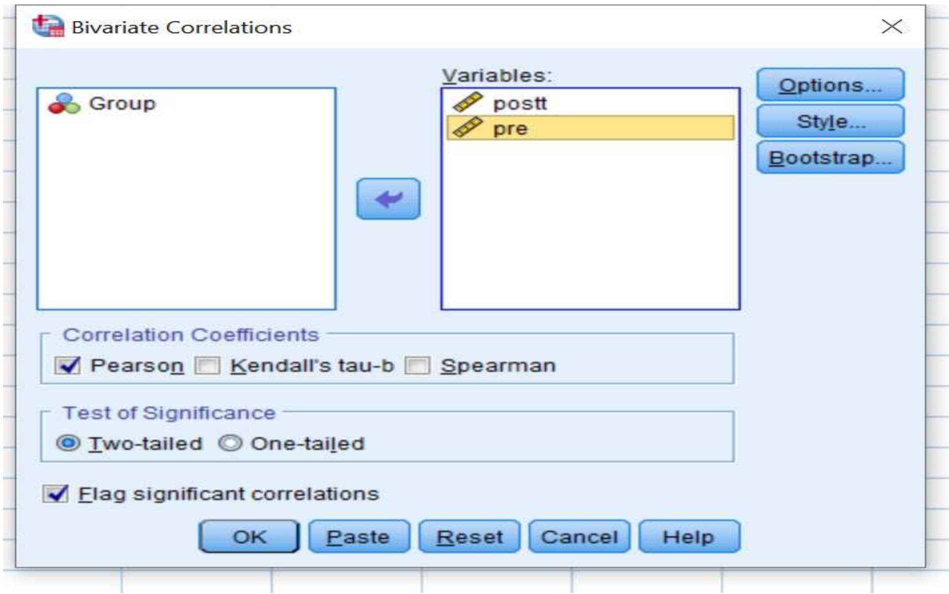
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1,204[a]	2	,602	1,979	,151	,086
Intercept	1689,122	1	1689,122	5553,405	,000	,992
Group	1,204	2	,602	1,979	,151	,086
Error	12,775	42	,304			
Total	1703,101	45				
Corrected Total	13,979	44				

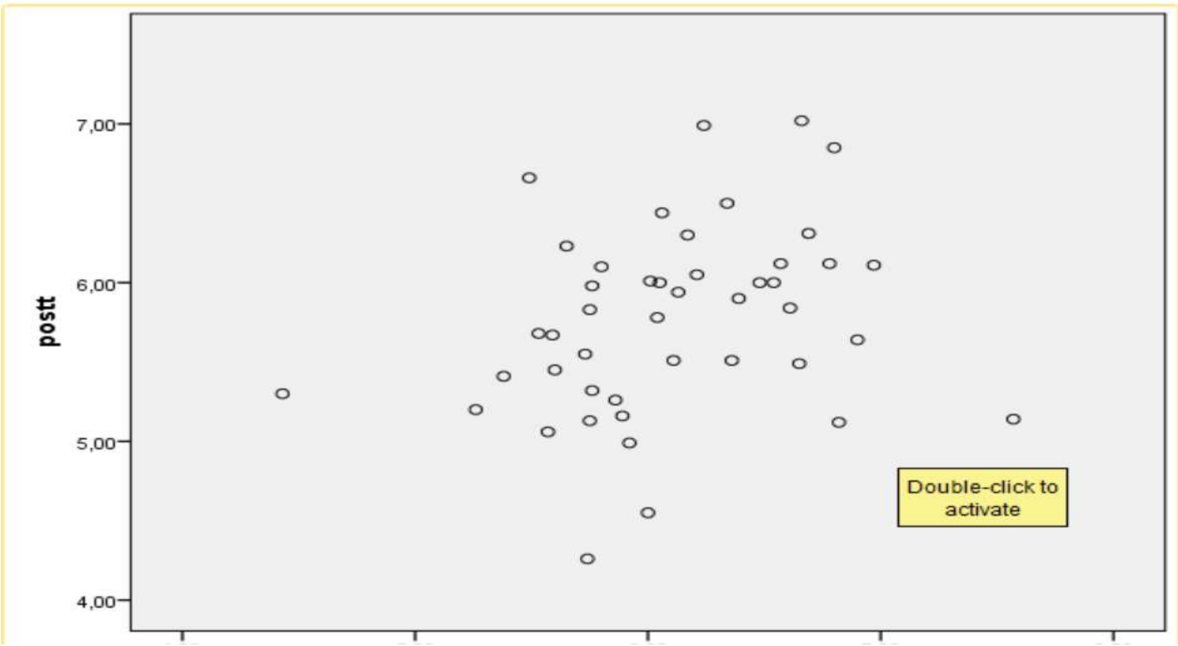
a. R Squared = ,086 (Adjusted R Squared = ,043)

So group does not have an impact on covariate. This assumption is also met.

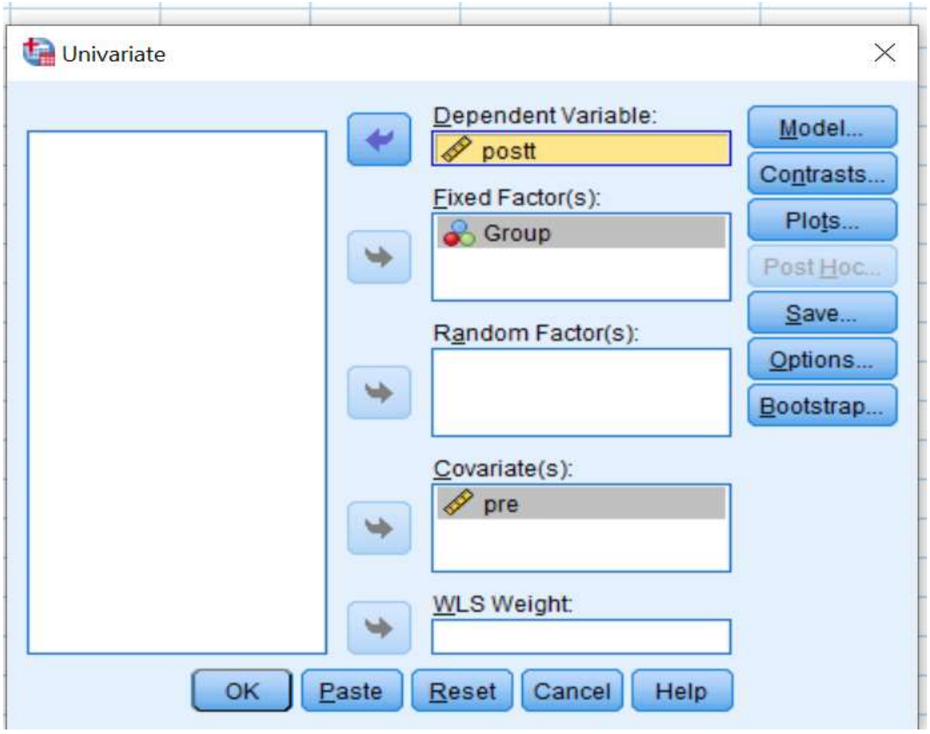
- Another assumption is DV and Covariates are linearly related >>>> correlation

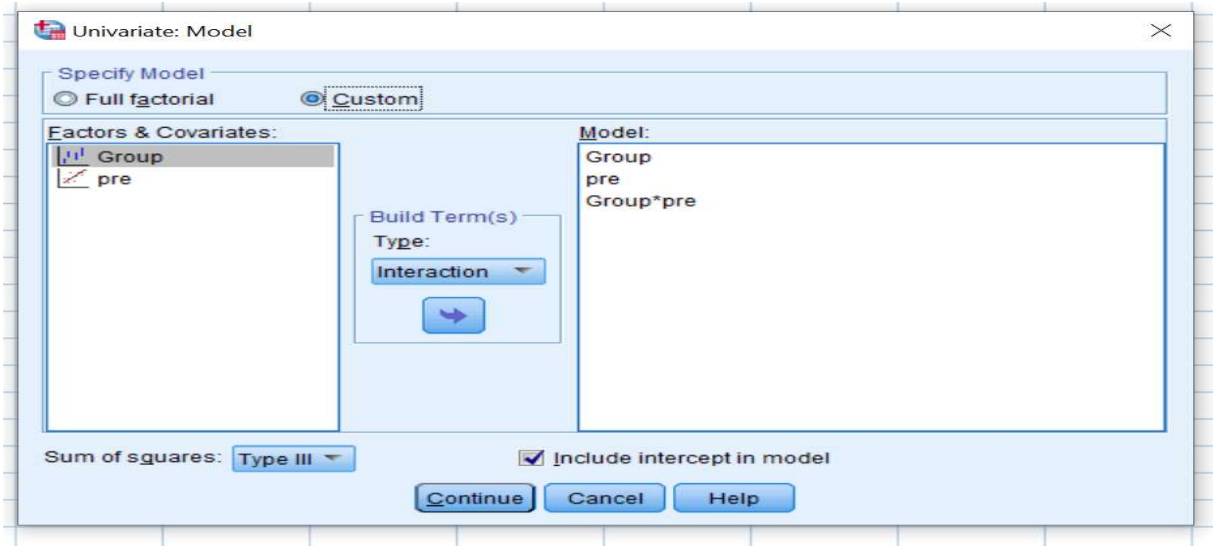






- Homogenty of regression <<<< non _significant interactions between the DV and covariate





Here to be able to see the interaction effect whether its important or not we create our model by ourself.

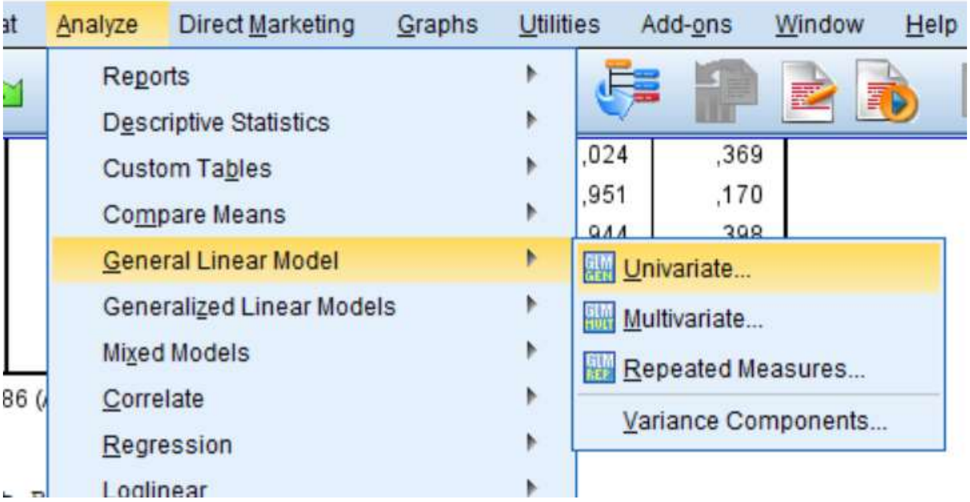
Dependent Variable: postt

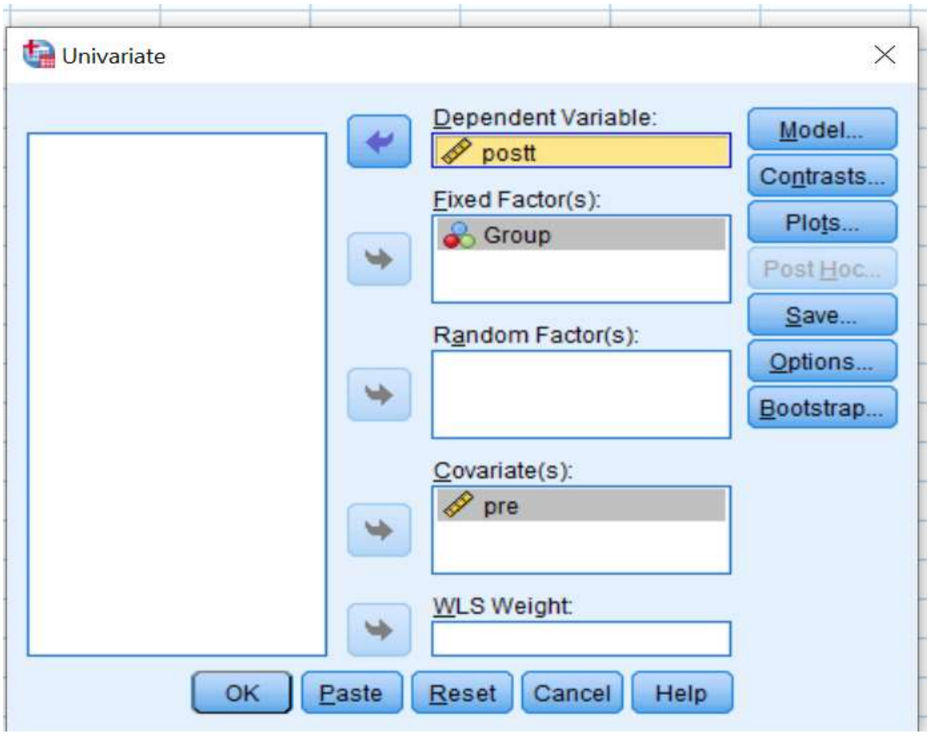
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4,476a	5	,895	3,122	,018
Intercept	3,485	1	3,485	12,155	,001
Group	,587	2	,294	1,024	,369
pre	,559	1	,559	1,951	,170
Group * pre	,541	2	,271	,944	,398
Error	11,181	39	,287		
Total	1511,876	45			
Corrected Total	15,657	44			

a. R Squared = ,286 (Adjusted R Squared = ,194)

As it is expected interaction effect is not significant.

Now it is time to perform ancova;

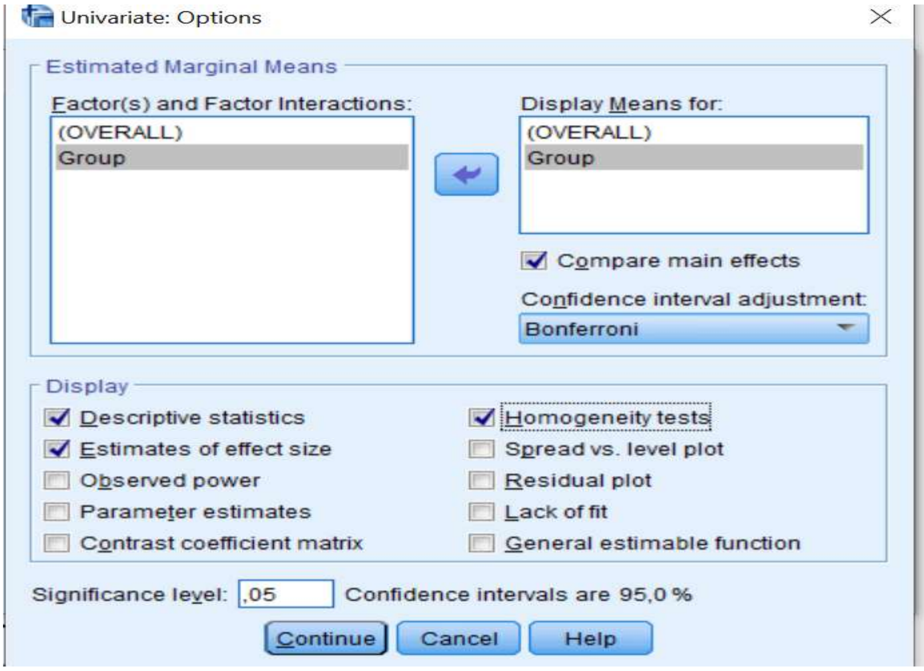




Model needs to be choosen full factorial

We should also look at the relationships between group and the dependent variable by making a bonferronni coreection, if you are conserned as to whether you may commit Type 1 error.

If you are interested to look at the distribution of the standardized values for your dependent variable or covariate.



Between-Subjects Factors

		Value Label	N
Group	1,00	control	15
	2,00	Int_1	15
	3,00	Int_2	15

Descriptive Statistics

Dependent Variable: postt

Group	Mean	Std. Deviation	N
control	6,0293	,61946	15
Int_1	5,9020	,47778	15
Int_2	5,3673	,49187	15
Total	5,7662	,59652	45

Levene's Test of Equality of Error Variancesa

Dependent Variable: postt

F	df1	df2	Sig.
,207	2	42	,814

Tests the null hypothesis that the error variance of the dependent variable is equal

Levene test tells us that the assumption is met.

Tests of Between-Subjects Effects

Dependent Variable: postt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4,475[a]	3	1,492	5,749	,002	,296
Intercept	3,343	1	3,343	12,884	,001	,239
pre	1,139	1	1,139	4,389	,042	,097
Group	1,791	2	,895	3,451	,041	,144
Error	10,638	41	,259			
Total	1488,356	45				
Corrected Total	15,113	44				

a. R Squared = ,296 (Adjusted R Squared = ,245)

So group has significant impact on our dependent variable.(post)So groups has an effect on post cholostrol levels.

Pairwise Comparisons

Dependent Variable: postt

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
control	Int_1	,280	,186	,422	-,185	,745
	Int_2	,516*	,199	,039	,021	1,012
Int_1	control	-,280	,186	,422	-,745	,185
	Int_2	,237	,202	,744	-,267	,741
Int_2	control	-,516*	,199	,039	-1,012	-,021
	Int_1	-,237	,202	,744	-,741	,267

Based on estimated marginal means

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

b. Adjustment for multiple comparisons: Bonferroni.

WE MAY CHECK THE pairwise comparison between the groups.

Estimates

Dependent Variable: postt

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
control	5,987[a]	,133	5,718	6,256
Int_1	5,707[a]	,135	5,435	5,979
Int_2	5,471[a]	,140	5,187	5,754

a. Covariates appearing in the model are evaluated at the following values: pre = 6,1691.

This analyses results tells us that groups have a significant impact on post cholesterol levels, if we control for the contribution of pre chost level on post cholesterol scores . Pre cost level has an ipact on post cholesterol level as well.