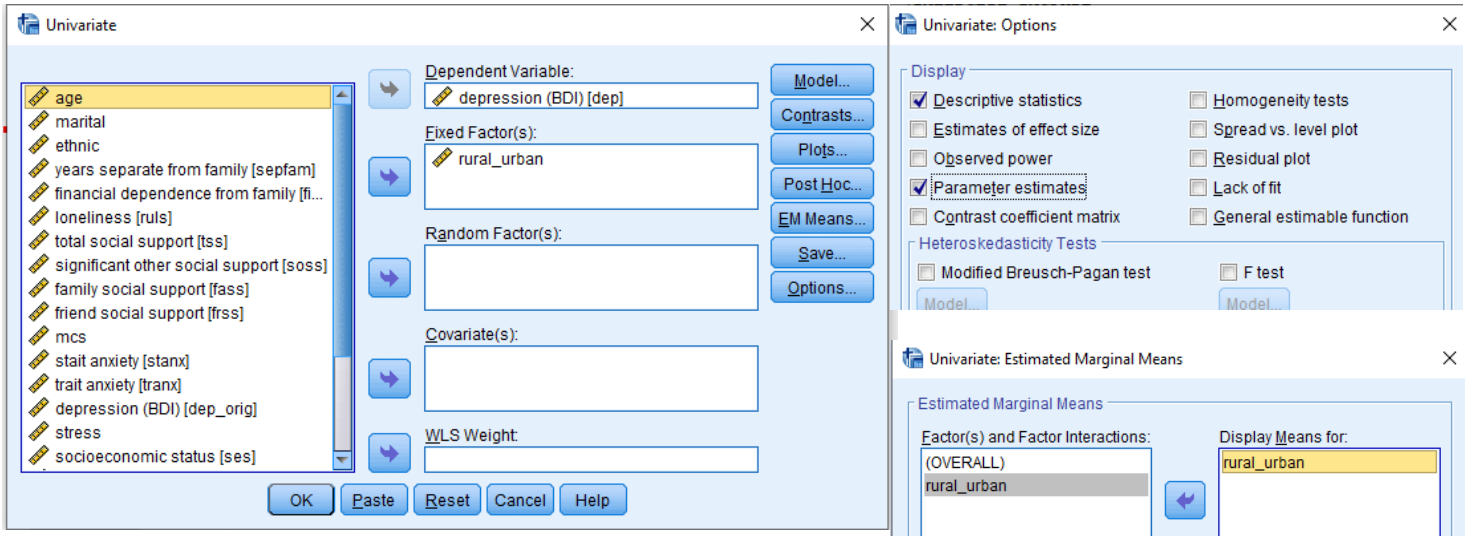
ANCOVA Example #2 – Multiple Covariates & Including a Categorical Covariate

The research question concerned depression differences between people living in rural and urban areas.

Initial ANOVA

Analyze → GLM → Univariate



click “options” & Select “Descriptive Statistics” & “Parameter estimates”
click “EM Means” & push the IV into the “Display means for” window

GLM output

Descriptive Statistics

Dependent Variable: depression (BDI)

rural_urban	Mean	Std. Deviation	N
rural	7.0500	5.99187	180
urban	8.7778	6.95036	225
Total	8.0099	6.59019	405

The mean difference is…

8.78 – 7.05 = 1.73

Tests of Between-Subjects Effects

Dependent Variable: depression (BDI)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	298.522[a]	1	298.522	6.975	.009
Intercept	25051.855	1	25051.855	585.356	.000
rural_urban	298.522	1	298.522	6.975	.009
Error	17247.439	403	42.798		
Total	43530.000	405			
Corrected Total	17545.960	404			

a. R Squared = .017 (Adjusted R Squared = .015)

Things to notice:

$F = t^2$ they both test group difference

GML uses a dummy code with the highest coded group as the control group, so …
a = mean of control group (urban)
b = group dif rural – urban = -1.728

Parameter Estimates

Dependent Variable: depression (BDI)

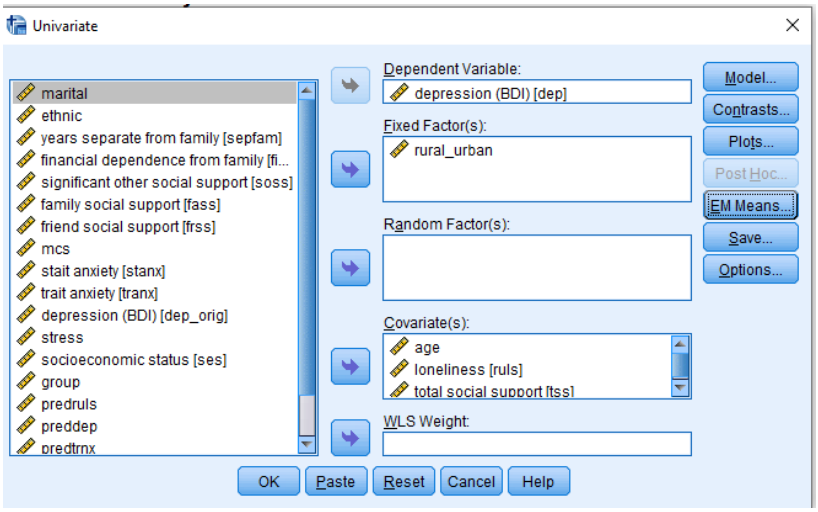
Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	8.778	.436	20.126	.000	7.920	9.635
[rural_urban=1]	-1.728	.654	-2.641	.009	-3.014	-.442
[rural_urban=2]	0[a]	.	.	.	.	.

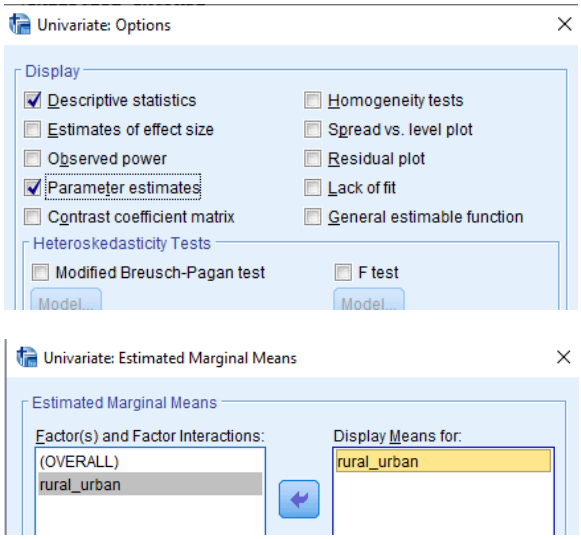
a. This parameter is set to zero because it is redundant.

mean of rural = a + b = 8.778 + (-1.728) = 7.050

ANCOVA with Multiple Covariates

Analyze → GLM → Univariate





"Covariates" can be any quantitative, binary, or coded variables.

Adding variables to the "Covariates" window will create a ANCOVA.

Request the same output as the last analysis

GLM output

Descriptive Statistics

Dependent Variable: depression (BDI)

rural_urban	Mean	Std. Deviation	N
rural	7.0500	5.99187	180
urban	8.7778	6.95036	225
Total	8.0099	6.59019	405

Tests of Between-Subjects Effects

Dependent Variable: depression (BDI)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6281.245[a]	4	1570.311	55.760	.000
Intercept	40.821	1	40.821	1.450	.229
rural_urban	580.881	1	580.881	20.627	.000
tss	74.807	1	74.807	2.656	.104
ruls	2982.872	1	2982.872	105.919	.000
age	930.640	1	930.640	33.046	.000
Error	11264.715	400	28.162		
Total	43530.000	405			
Corrected Total	17545.960	404			

a. R Squared = .358 (Adjusted R Squared = .352)

rural_urban

Dependent Variable: depression (BDI)

rural_urban	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
rural	6.642[a]	.400	5.855	7.429
urban	9.104[a]	.357	8.402	9.806

a. Covariates appearing in the model are evaluated at the following values: age = 28.48, loneliness = 37.21, total social support = 5.6233.

Parameter Estimates

Dependent Variable: depression (BDI)

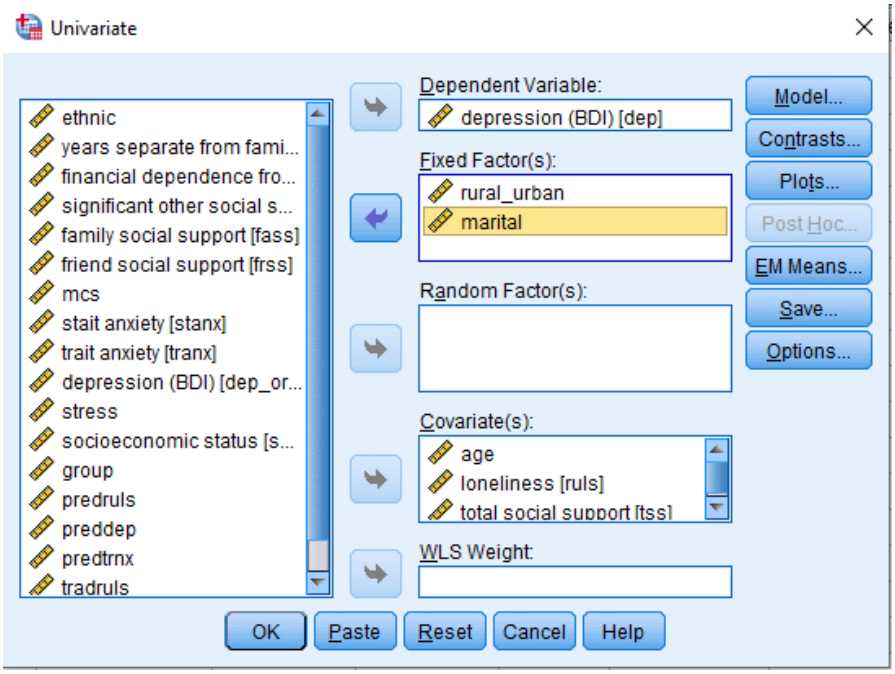
Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.340	2.633	1.649	.100	-.835	9.515
[rural_urban=1]	-2.462	.542	-4.542	.000	-3.528	-1.396
[rural_urban=2]	0[a]	.	.	.	.	.
tss	-.473	.290	-1.630	.104	-1.043	.097
ruls	.311	.030	10.292	.000	.251	.370
age	-.145	.025	-5.749	.000	-.195	-.096

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

Again, the regression weight for rural_urban is the same as the corrected mean difference (9.104-6.642 = 2.462), but now corrected for multiple covariates.

The regression weights tell and test the slope of each quantitative covariate for both groups, correcting for the other variables in the model.

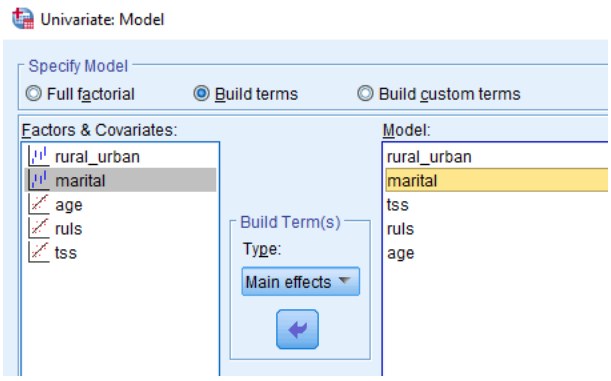
ANCOVA with Multiple Covariates Including a Categorical Covariate

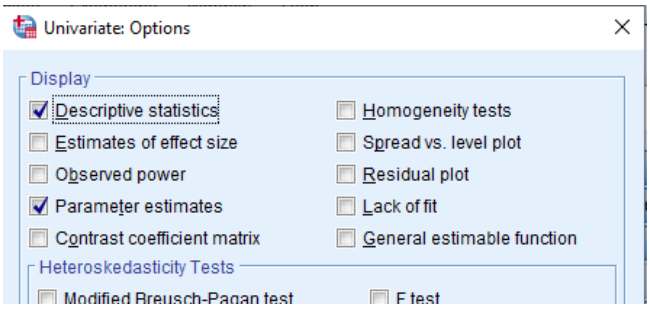


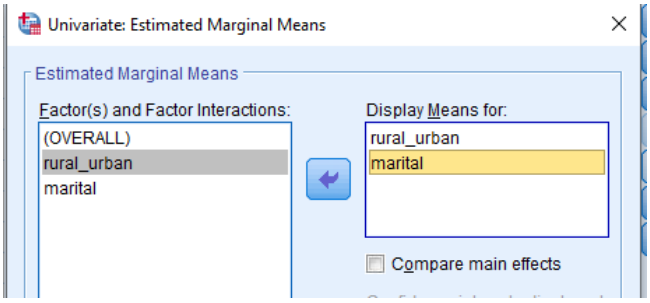
If we put more than one variable into the “Fixed Factors” window, we will obtain a factorial analysis.

If we want an ANCOVA instead of a factorial, we can specify that we want a “main effects model” -- as shown below on the left.

We would also want to get **both** the *corrected group means* for each of the categorical variables (rural_urban and marital status) that go with the ANCOVA F-tests for these variables **and** the *regression* parameters that tell about the corrected effects of the quantitative variables (age, ruls,TSS) that go with the ANCOVA F-tests for these variables – as shown below on the right.







GLM output

Descriptive Statistics

Dependent Variable: depression (BDI)

rural_urban	marital	Mean	Std. Deviation	N
rural	single	7.2459	5.67336	122
	married	6.1064	6.69923	47
	divorced	8.9091	6.18796	11
	Total	7.0500	5.99187	180
urban	single	10.0417	8.10995	120
	married	6.8919	4.57081	74
	divorced	8.3871	5.80044	31
	Total	8.7778	6.95036	225
Total	single	8.6322	7.11327	242
	married	6.5868	5.48281	121
	divorced	8.5238	5.83195	42
	Total	8.0099	6.59019	405

1. rural_urban

Dependent Variable: depression (BDI)

			95% Confidence Interval	
rural_urban	Mean	Std. Error	Lower Bound	Upper Bound
rural	6.547[a]	.514	5.535	7.558
urban	8.991[a]	.455	8.097	9.885

a. Covariates appearing in the model are evaluated at the following values: age = 28.48, loneliness = 37.21, total social support = 5.6233.

2. marital

Dependent Variable: depression (BDI)

			95% Confidence Interval	
marital	Mean	Std. Error	Lower Bound	Upper Bound
single	8.222[a]	.452	7.333	9.111
married	7.160[a]	.636	5.910	8.411
divorced	7.924[a]	.981	5.995	9.854

a. Covariates appearing in the model are evaluated at the following values: age = 28.48, loneliness = 37.21, total social support = 5.6233.

The "raw" means are above and the "corrected" means are on the right.

Tests of Between-Subjects Effects

Dependent Variable: depression (BDI)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6328.474[a]	6	1054.746	37.423	.000
Intercept	11.239	1	11.239	.399	.528
rural_urban	568.049	1	568.049	20.155	.000
marital	47.229	2	23.614	.838	.433
tss	57.999	1	57.999	2.058	.152
ruls	2981.050	1	2981.050	105.769	.000
age	208.332	1	208.332	7.392	.007
Error	11217.487	398	28.185		
Total	43530.000	405			
Corrected Total	17545.960	404			

a. R Squared = .361 (Adjusted R Squared = .351)

Notice again that the F-tests and t-tests tell the same story, except for Marital status

The F-test is a test of the "marital effect", while the t-tests of the individual dummy codes test specific pairwise comparisons.

If either of the dummy codes (pairwise comparisons) are significant, then the F-test of that k-group effect will be significant.

Parameter Estimates

Dependent Variable: depression (BDI)

					95% Confidence Interval	
Parameter	B	Std. Error	t	Sig.	Lower Bound	Upper Bound
Intercept	3.207	3.150	1.018	.309	-2.985	9.399
[rural_urban=1]	-2.445	.545	-4.489	.000	-3.515	-1.374
[rural_urban=2]	0[a]	.	.	.	.	.
[marital=1]	.298	1.213	.246	.806	-2.087	2.683
[marital=2]	-.764	.969	-.789	.431	-2.669	1.141
[marital=3]	0[a]	.	.	.	.	.
tss	-.421	.293	-1.435	.152	-.997	.156
ruls	.311	.030	10.284	.000	.251	.370
age	-.114	.042	-2.719	.007	-.197	-.032

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

Note: It is possible to have a significant F, without either of the dummy-code t-tests to be significant → if the comparison group happens to be the middle-value mean, then it might be different from neither the group with the higher nor the lower mean, while those higher and lower means are different from each other

Because there are no interactions (i.e.,making the regression homogeneity assumption) the regression weights tell and test the slope of each quantitative covariate for both groups, correcting for the other variables in the model.