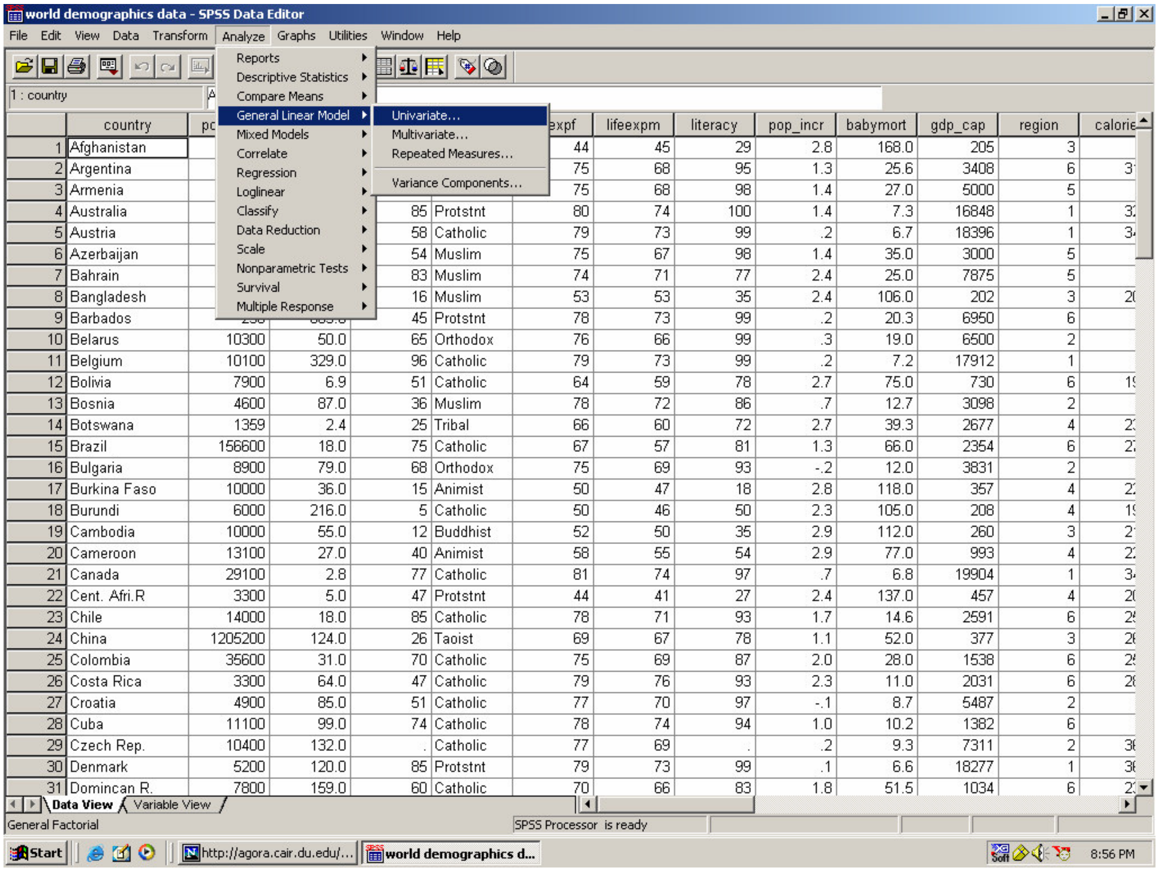
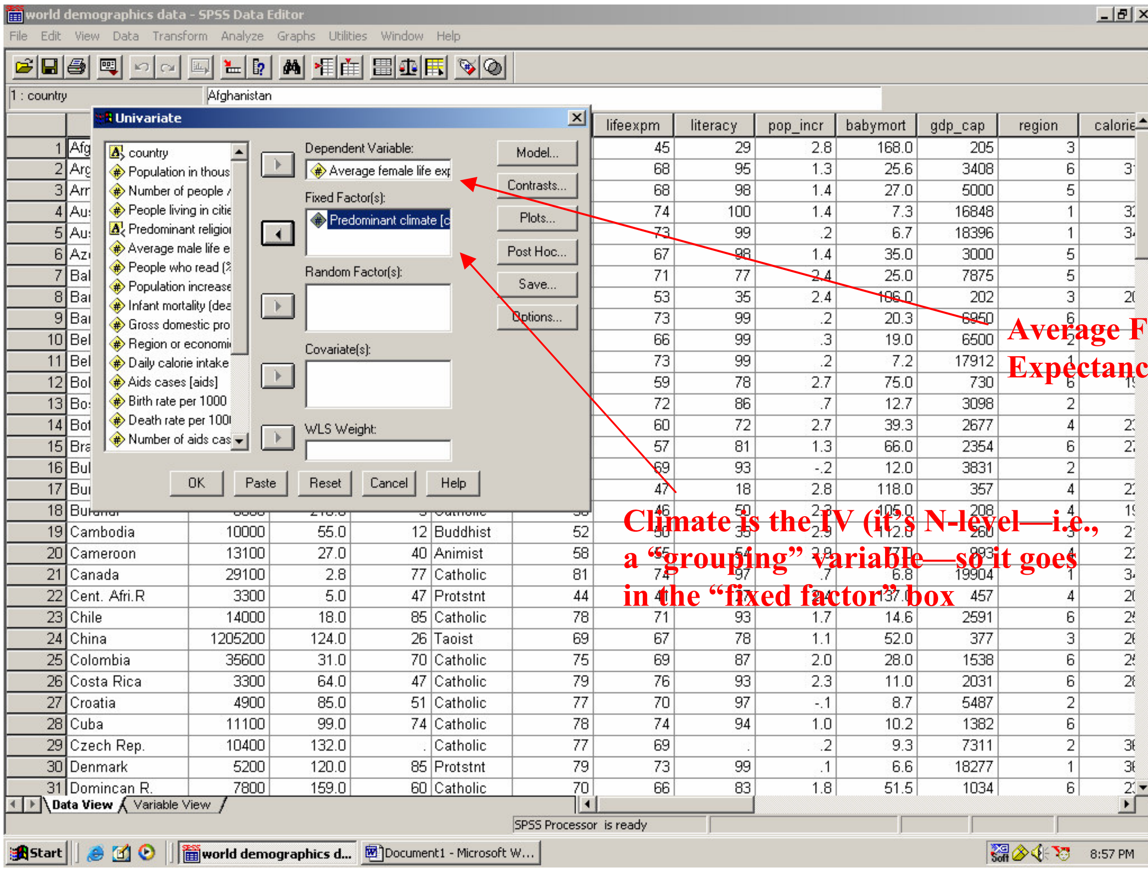
How to do ANCOVA Problems in SPSS:

For ANCOVA, use the same "General Linear Model" -> "Univariate" command that you use for a basic ANOVA. (Remember: "univariate" means "one dependent variable," regardless of how many independent variables there are. All we're doing here is to add more predictors—there's still just one criterion variable).



The same dialog box appears as is used in a one-way ANOVA. This data set gives you information about various demographic variables in different countries around the world. We're going to look at Average Female Life Expectancy as our criterion variable, and see whether we can predict this from the climate of the country where someone lives.



Hit "OK" to see the results (let's not do anything else fancy yet).

Here's the result of this initial test:

Tests of Between-Subjects Effects
Dependent Variable: Average female life expectancy

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2954.845	8	369.356	4.009	.000
Intercept	295057.782	1	295057.782	3202.530	.000
CLIMATE	2954.845	8	369.356	4.009	.000
Error	9029.005	98	92.133		
Total	536844.000	107			
Corrected Total	11983.850	106			

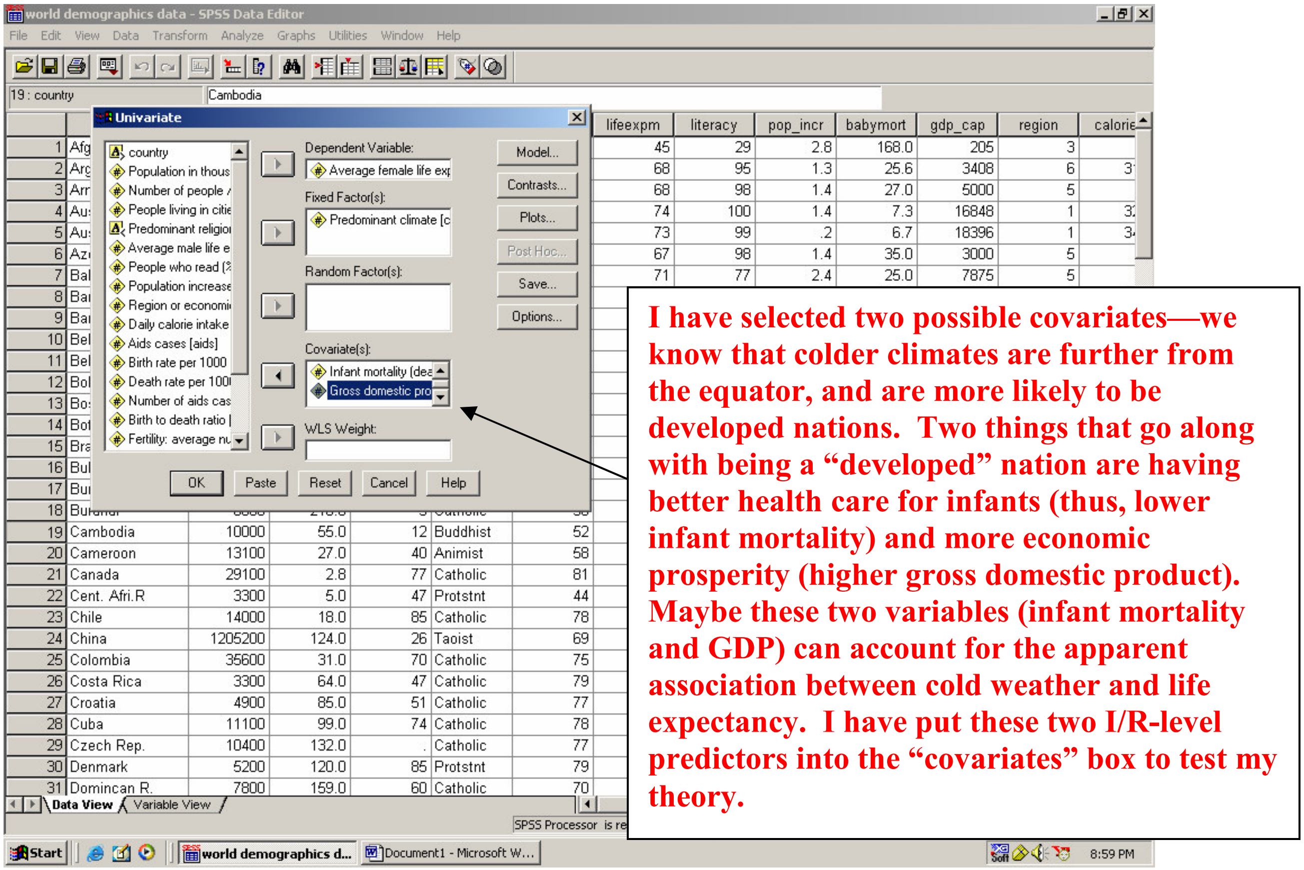
R Squared = .247 (Adjusted R Squared = .185)

This significant *p*-value says that the IV "climate" is a significant predictor of scores on the DV "average female life expectancy." (Remember to look at the *p*-value in the row that has the name of the predictor variable that you're interested in, or at the *p*-value in the row that says "corrected model" if you're interested in the effects of all the predictor variables together. In this case, there's only one predictor, so the *F*-tests for the single predictor and for the "model" are the same).

Now, if we were to look at the actual **direction of this effect**, we would find out that countries with colder climates tend to have longer life expectancies. (Try it: go back to the dialog box, click on the "options" button, and select "descriptive statistics," then re-run the test. See the website information on One-Way ANOVA if you need help finding the right commands).

So, our basic conclusion from this *F*-test is that "cold weather is good for you." Intuitively, this doesn't make much sense. So, let's see if we can find a **covariate** that can **account for the apparent association between cold weather and health**. If such a covariate exists, then the *F*-test for climate will become *nonsignificant* after the covariate is included in the model (assuming that we're using the Type III sums of squares, where variables don't get credit for any "shared variability" that they have with any of the other predictors).

Going back to the "univariate" dialog box, we can put in some possible confounding variables as "covariates." *Remember that* *<u>covariates</u>* *have to be I/R-level variables*. We can do the same process with other N-level predictors, but we would have to put them in as additional "fixed factors," test for interaction effects, etc. That would make it a 2-way ANOVA, instead of an ANCOVA.



Again, hit "OK" to go on.

Here are the new test results, including the covariates:

Tests of Between-Subjects Effects
Dependent Variable: Average female life expectancy

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11195.472	10	1119.547	136.326	.000
Intercept	89766.619	1	89766.619	10930.778	.000
BABYMORT	5650.664	1	5650.664	688.075	.000
GDP_CAP	7.100	1	7.100	.865	.355
CLIMATE	92.036	8	11.504	1.401	.206
Error	788.379	96	8.212		
Total	536844.000	107			
Corrected Total	11983.850	106			

R Squared = .934 (Adjusted R Squared = .927)

Notice that the *F*-test value for "climate" is no longer significant. This tells us that once we take into account the effects of infant mortality ("babymort") and economic prosperity ("gdp_cap"), there is no longer a significant effect of climate on average female life expectancy.

Note that the "model" as a whole was still significant—it *is* possible to predict female life expectancy from these three variables (babymort, gdp_cap, and climate). It's just that the effects of climate *alone* are no longer significant, after controlling for the correlated effects of the other variables on the DV.

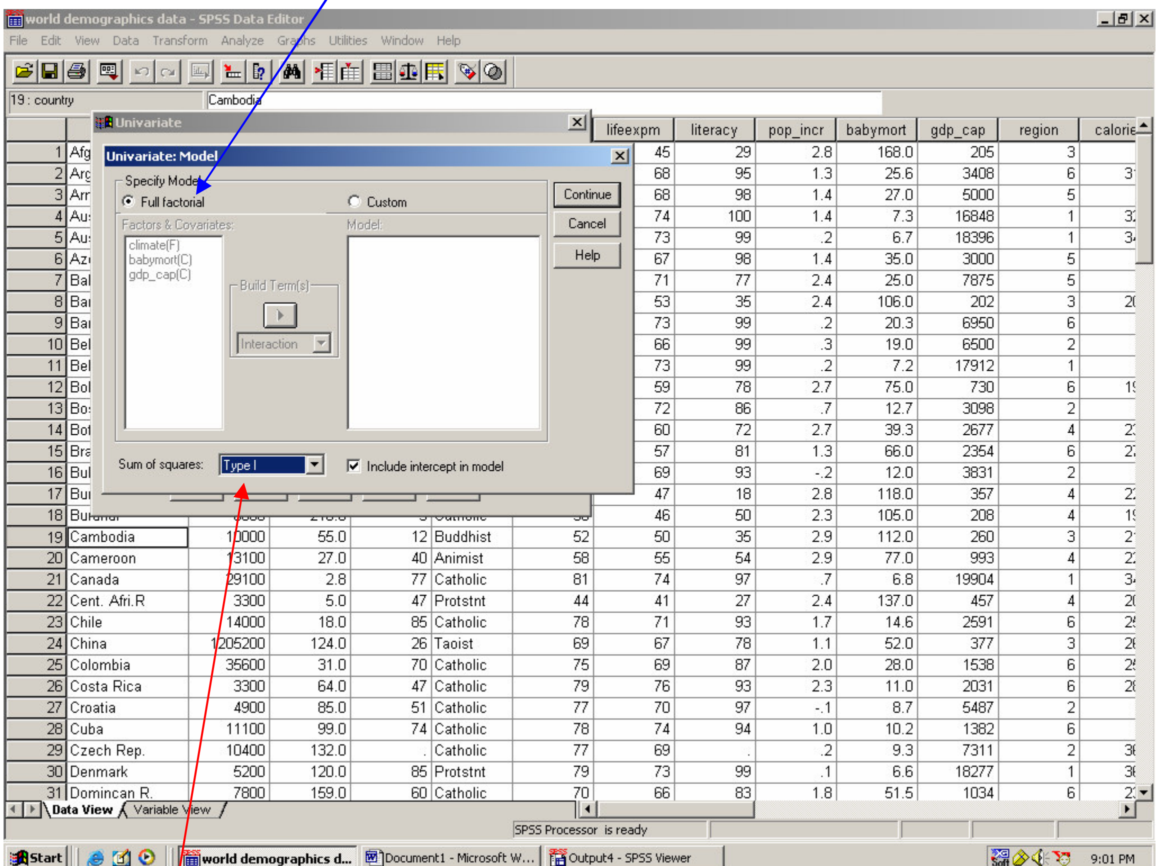
If we wanted to get a "semipartial R-square" for just the effects of climate, after controlling for these other two variables, we would look at the Type III SS for climate, and compare it to the "corrected total" Type III SS:

$$\text{semipartial R2 for "climate"} = \frac{\text{SS climate}}{\text{SS corrected total}} = \frac{92.036}{11983.95} = .00768$$

$$= \underline{0.77\%}$$

Let's go back to the dialog box, and see **what happens when we select "Type I" sums of squares, instead of "Type III."**

To do this, you will need to hit the "Model" button (again, see last week's class example for details). Under "specify model," leave the selection on its default, "full factorial." This automatically includes all covariates, all fixed factors, and all possible interactions between fixed factors in your model.



Under "sums of squares," this time you should use the drop-down menu to select "Type I" instead of its default value, which is "Type III."

Hit "continue" to get back to the main dialog box, and then hit "OK" to run the results.

Here's the revised output:

Tests of Between-Subjects Effects
Dependent Variable: Average female life expectancy

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11195.472	10	1119.547	136.326	.000
Intercept	524860.150	1	524860.150	63911.620	.000
BABYMORT	11089.966	1	11089.966	1350.412	.000
GDP_CAP	13.470	1	13.470	1.640	.203
CLIMATE	92.036	8	11.504	1.401	.206
Error	788.379	96	8.212		
Total	536844.000	107			
Corrected Total	11983.850	106			

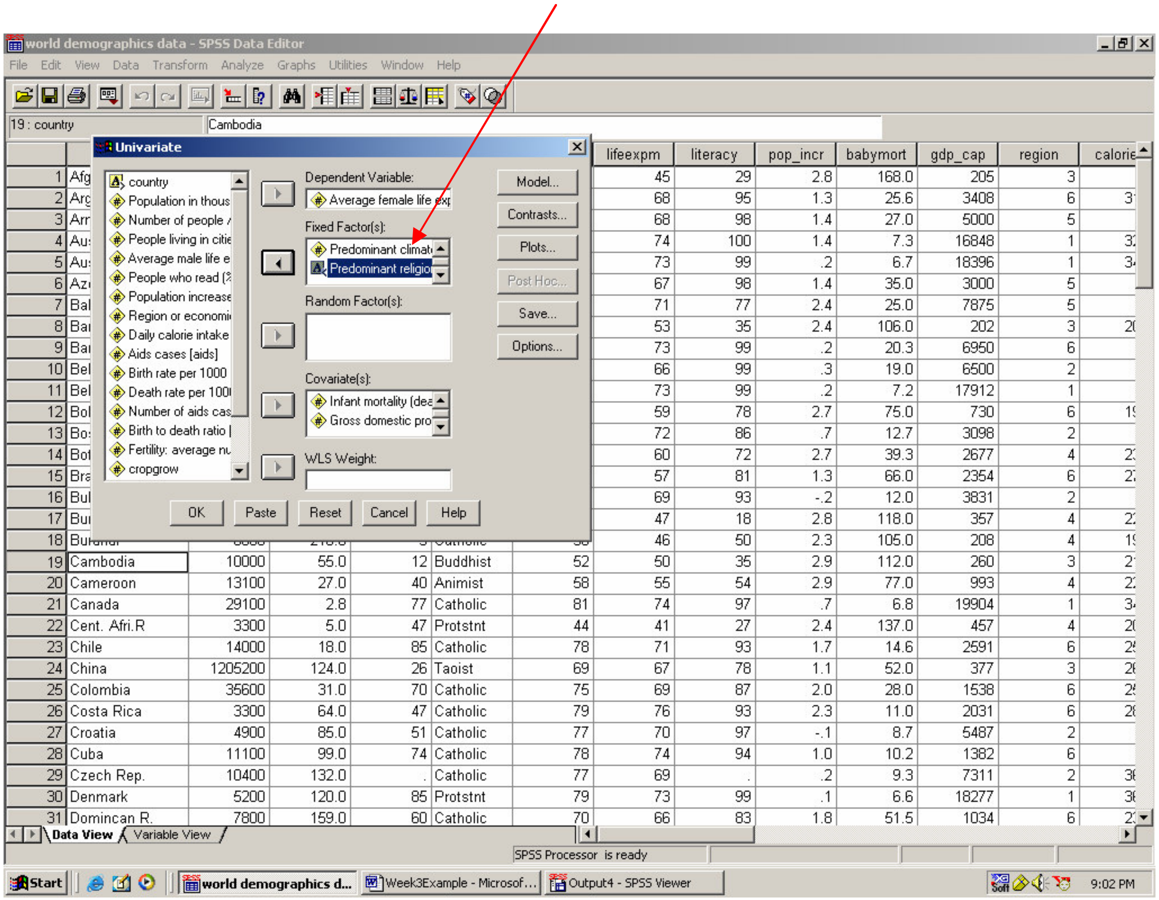
R Squared = .934 (Adjusted R Squared = .927)

Notice the change in sums of squares for these two variables in particular—for each one, the Type I SS was **greater than** the Type III SS. This is the usual pattern that you will see—when you look at Type I SS, it's generally larger, because there is often some shared variability with other variables (which doesn't get counted when you look at Type III SS, but does get counted when you look at Type I SS).

Also, remember that the shared variability is now getting counted *multiple times*. That's why the sum of SS for "babymort," "gdp_cap," and "climate" now adds up to **more than the total for the "corrected model" SS**.

You will not generally use the Type I SS for anything, in routine practice. You would only look at them if there was some specific reason to. *For most routine analyses that use the general linear model—2-way ANOVA, ANCOVA, and others—rely on the Type III sums of squares*.

Finally, let's go back to the main dialog box, and add a second "grouping" variable. Let's say that we're interested in the effect of predominant *religious group* on life expectancy, and whether religion has any effect above and beyond those of the other predictors. Because "religion" is N-level (groups), we don't call it a "covariate." Instead, we put it in as another fixed factor, just like we did with "climate."



Hit "OK" to continue.

Here is one last set of results:

Tests of Between-Subjects Effects
Dependent Variable: Average female life expectancy

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11382.949	33	344.938	41.621	.000
Intercept	520240.340	1	520240.340	62772.890	.000
BABYMORT	11086.572	1	11086.572	1337.720	.000
GDP_CAP	13.209	1	13.209	1.594	.211
CLIMATE	98.148	8	12.269	1.480	.180
RELIGION	53.840	9	5.982	.722	.687
CLIMATE * RELIGION	131.179	14	9.370	1.131	.347
Error	596.711	72	8.288		
Total	532220.000	106			
Corrected Total	11979.660	105			

R Squared = .950 (Adjusted R Squared = .927)

If you left the "model" set on "full factorial," then you'll see an interaction effect here, as well as the effects of each individual predictor. This gives you a test for each predictor variable, including the interaction between the two N-level variables.

Congratulations! You have just completed an advanced statistical analysis!
This analysis had …

- 2 N-level predictors (fixed factors)
- 2 I/R-level predictors (covariates)
- and 1 I/R-level criterion variable (DV)

Therefore, you have just completed a 2-way univariate analysis of covariance (ANCOVA).