Quiz #2 -- ANCOVA to increase statistical power (& reduce error variance)

Here are the data from a two group design:

```
data list free / grp dv cov.
begin data.
1  12  21        2   15  20
1  15  20        2   20  21
1  18  26        2   21  27
1  21  35        2   26  34
1  24  38        2   27  39
1  27  43        2   32  42
end data.
```

Here are the correlations among the variables:

```
Corrs:    DV       GRP      COV
DV      1.0000    .3522    .9015*
GRP      .3522  1.0000     .0000
COV      .9015*  .0000   1.0000
```

Note: non-zero but not sig corr of grp and dv
zero corr of grp and cov
strong corr cov and dv

Here's the "regular ANOVA" for these data -- both DV and COV.

```
GRP   DV                     Source of Variation  Sum of         Mean           Signif
                                                  Squares   DF   Square     F     of F
       1          2          GRP                   48.000    1   48.000   1.416   .262
   19.50      23.50          Residual             339.000   10   33.900
 (     6)   (     6)         Total                387.000   11   35.182
```

```
               COV                                Sum of         Mean          Signif
GRP  COV                     Source of Variation  Squares   DF   Square     F     of F
       1          2          GRP                     .000    1     .000    .000  1.000
   30.50      30.50          Residual             883.000   10   88.300
 (     6)   (     6)         Total                883.000   11   80.273
```

Notice (about same information as the corrs above):
-- nonsig relationship of GRP and DV & -- no relationship of GRP and COV (i.e., nothing to correct for!!)
Also notice the very large amount of w/in group variation on both DV and COV -- we'll come back to this.

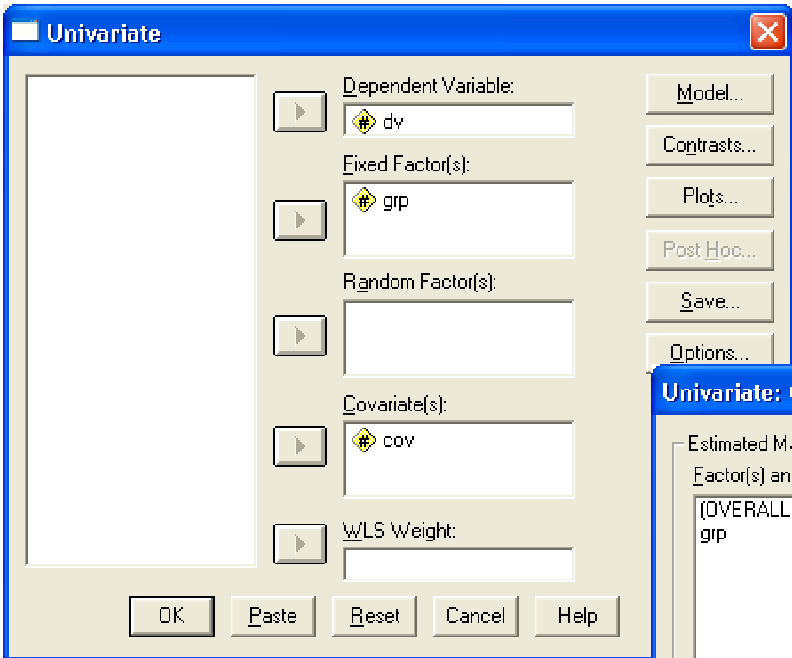
- The "problem" in this case is that there is a clear numeric DV difference between the groups.
- Also, the covariate is not "disrupting" the DV mean difference between the groups -- the groups have the same COV mean
- But because there is substantial variation **within each condition** on the DV, the DV mean difference is not significant.
- So, is the statistical result "proper" or a Type II error, produced by having an "too-heterogeneous" sample (i.e., large SS wg) ??

This is the situation for which ANCOVA was originally designed:
- A between group mean difference on the DV is "masked" by within-group DV variability.
- The covariate is not correlated with the IV (i.e., no between group mean difference on the covariate),
- The covariate is correlated with the DV.

Given these conditions, ANCOVA allows us to examine whether the GRP effect "emerges" when the SSwg has been statistically reduced by "removing" the within-group variation that is associated with the COV. Put differently, using ANCOVA we ask if there would be a significant GRP effect (mean difference on the DV) if there were less within-group variation on the COV.

ANCOVA using SPSS ANOVA

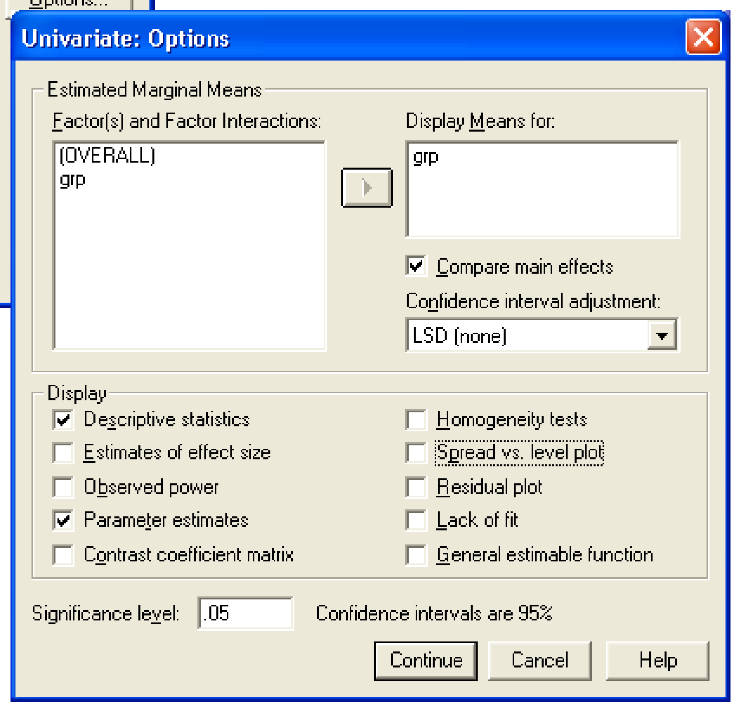
Analyze → General Linear Model → Univariate



In the main window…

Move the DV, Grouping variable and covariate into their respective windows.

Click “Options”



“Descriptive statistics” gets the raw means

“Parameter estimates” gives the regression weights for the model (the group with the highest code “2” → comparison group)

“Estimated marginal means” gets the group means after correction for the covariate(s)

Descriptive Statistics

Dependent Variable: DV

GRP	Mean	Std. Deviation	N
1.00	19.5000	5.61249	6
2.00	23.5000	6.02495	6
Total	21.5000	5.93143	12

The uncorrected or raw means are the same as in the ANOVA

Tests of Between-Subjects Effects

Dependent Variable: DV

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	362.529[a]	2	181.264	66.666	.000
Intercept	9.560	1	9.560	3.516	.094
COV	314.529	1	314.529	115.678	.000
GRP	48.000	1	48.000	17.653	.002
Error	24.471	9	2.719		
Total	5934.000	12			
Corrected Total	387.000	11			

a. R Squared = .937 (Adjusted R Squared = .923)

Notice:

- The SSgrp is 48.00, as in the ANOVA because there is no between group difference on the covariate to “correct for”
- But the error variance is smaller. And so,
- the Grp F is larger and the GRP effect is now significant

How's that work ????

ANOVA Model

$$\underset{SS_{total}}{387.00} = \underset{SS_{IV}}{48.00} + \underset{SS_{error}}{339.00}$$

ANCOVA model (when no grp dif on covariate)

$$\underset{387.00}{SS_{total}} = \underset{48.00}{SS_{IV}} + \underset{314.53}{SS_{cov}} + \underset{24.47}{SS_{error}}$$

- Adding the covariate to the model, "takes up" part of the variation that was error variation -- so, the error variation is smaller & test of the IV effect is more powerful (smaller error term).

Parameter Estimates

Dependent Variable: DV

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Partial Eta Squared
Intercept	5.297	1.821	2.908	.017	1.176	9.417	.484
COV	.597	.055	10.755	.000	.471	.722	.928
[GRP=1.00]	-4.000	.952	-4.202	.002	-6.154	-1.846	.662
[GRP=2.00]	0[a]	.	.	.	.	.	.

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

You can get the same parameter estimates that you are used to getting from a regression analysis (though you don't get the β weights, which can be helpful for comparing the relative contributions among the predictors).

SPSS uses the group with the highest code (group = 2, here) as the comparison group in a dummy code. So the regression weight for GRP=1.00 (with Group 1 as the target group) shows us that Group 1 has a mean 4 less than Group 2, after controlling for (correcting for, covarying out) the covariate.

Estimated Marginal Means

Estimates

Dependent Variable: DV

GRP	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.00	19.500[a]	.673	17.977	21.023
2.00	23.500[a]	.673	21.977	25.023

a. Evaluated at covariates appeared in the model: COV = 30.5000.

Finally, you get the corrected mean for each group.

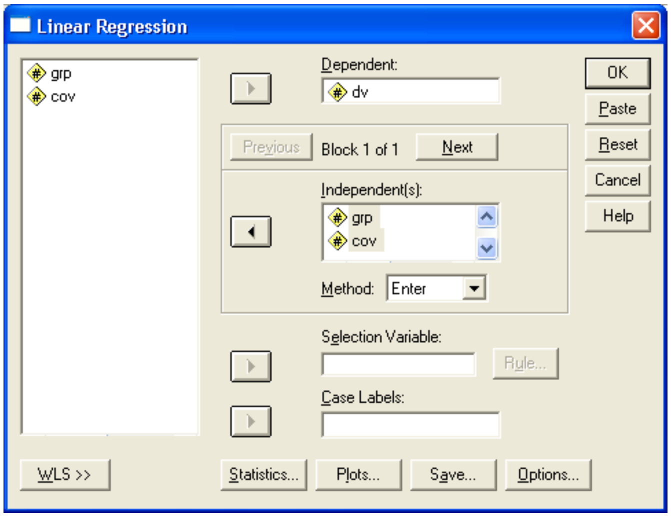
For this analysis the corrected and uncorrected means are the same, because there was no difference between groups on the covariate.

You are also given the mean of the covariate (30.5, here), reminding you that these are the group means after holding the score on the covariate constant at its mean.

ANCOVA via Regression

Start by creating a dummy code using Group=2 as the comparison

recode grp (1 = 1) (2 = 0).



Put grp (now recoded) and the covariate into a regression model.

Model Summary

Model	R	R Square
1	.968[a]	.937

a. Predictors: (Constant), GRP, COV

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	362.529	2	181.264	66.666	.000[a]
	Residual	24.471	9	2.719		
	Total	387.000	11			

a. Predictors: (Constant), GRP, COV

b. Dependent Variable: DV

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.297	1.821		2.908	.017
	COV	.597	.055	.902	10.755	.000
	GRP	-4.000	.952	-.352	-4.202	.002

a. Dependent Variable: DV

Everything matches with the ANCOVA output…

SS, F, b, etc.

Things to notice:

- the F-test for "explained variance" is the same for this regression model as the F-test in the ANCOVA model
- COV regression weight is same here as in simple regression model using just COV -- because there is no correlation between COV and GRP (no collinearity)
- GRP regression weight is same here as in the simple regression model using just GRP -- because there is no correlation between COV and GRP (no collinearity)
- **BUT**, the test of the regression weight for GRP is significant (with $t^2 = F$ and the same p-value as in the ANCOVA above) -- because there is less error (residual) variation
- Also notice that the β for the covariate is much higher than for the group/treatment variable – suggesting which of these is a “more important” descriptor/predictor of the criterion behavior (remember to interpret covariates, not just to “correct” for them!!!)