**SAS Controlled-Trial Mixed Models**

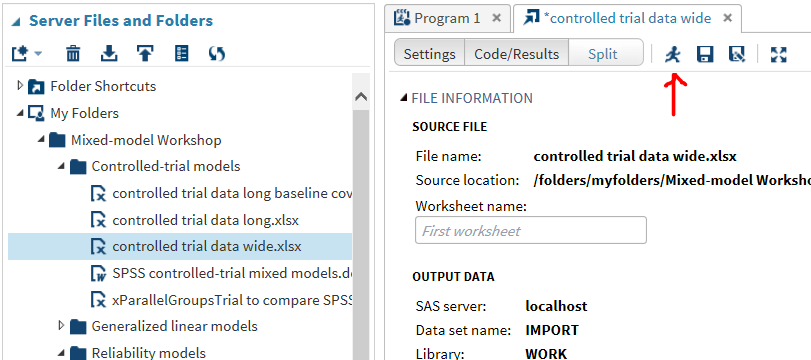
From now on, ignore instructions about finding the relevant folder and opening the relevant spreadsheet. Instead, create the relevant folder in SAS Studio and upload the spreadsheet from your laptop.

When you have pre and post measurements of a dependent variable in a treatment and control group, you have a classic controlled trial. You can analyze such data by converting the repeated measurements into a single change score for each subject, as shown in my controlled-trial spreadsheet. You are effectively keeping the data in "wide" format. Although there is no repeated measurement on subjects, you still use a mixed model to estimate the SD representing individual responses. You can also put the data into "long" format, in which each repeated measurement is a separate datum, row or "observation". This is the usual approach when you have lots of repeated measurements. We'll do both here, with the same data, and we'll compare what we get with the analyses in the controlled-trial spreadsheet.

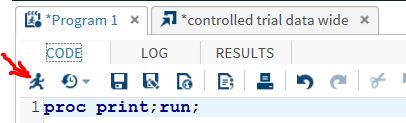
[Analyzing Change Scores, Data in Wide Format](#changescoresdatawide)  
[Analyzing Change Scores, Data in Wide Format, Plus a Covariate](#changescorespluscovariate)  
Analyzing Change Scores in a Crossover: see crossover two predictors SAS.docx  
[Analysis with Data in Long Format](#datalong)  
[Analysis with Data in Long Format and More Than One Repeated Measurement](#datalongmorethanonerep)  
[Adjusting for Baseline with Data in Long Format](#adjustbaseinedatalong)  
[Estimating the Modifying Effects of Sex](#effectsofsex)  
[Summary table](#summary)

**Analyzing Change Scores, Data in Wide Format**

1. Open SAS, open Server Files and Folders/My Folders/Mixed-model Workshop/Controlled-trial models, double-click controlled trial data wide.xlsx to open a data tab in SAS, and run it:



1. Click on the Program 1 tab, type in proc print; run; and run it:



Inspect the data in the RESULTS window:



There are 20 subjects in a control group followed by 20 subjects in an experimental group, with data for a pre test and a post test. The Pre and Post are Pre2 and Post1 in the spreadsheet **xParallelGroupsTrial to compare SAS.xlsx**. You could open that now for comparison, by right-clicking and downloading within SAS, not by double-clicking. In Row 3 of that spreadsheet it states that "Data (generated with Understanding Stats via Simulations) are for a training study in which the improvement in peak power in an incremental test depended on the initial weekly training (X) of the cyclists."

The only other important issue at this stage is that the variable Post\_Pre is the change score of the 100\*log-transformed peak power (LnPost-LnPre). So the numbers there represent approximate percent changes in peak power of each cyclist.

1. From now on, I am going to present you with code for Proc Mixed that you copy into Program 1 and run. The code is also available in the file **all controlled-trial programs SAS.docx**.

If you want to keep using Tasks/Statistics/Mixed Models, make sure you end up with the same code that I present here. SAS produces some extraneous bits of code, but if you run it and get different results, you've done something wrong.

1. Here is the code to copy and paste into Program 1:

\*set up options for proc mixed output;

ods noproctitle;

ods graphics / imagemap=on;

title "Analyzing change scores";

proc mixed data=import plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group;

model Post\_Pre=Group/residual;

random xVarExp/subject=Name;

\*random xVarExp\*Name; \*this does the same thing as the above line;

lsmeans Group/diff=control("Control") cl alpha=0.1;

run;

The plots in the proc mixed statement limits the plots to the only important ones.

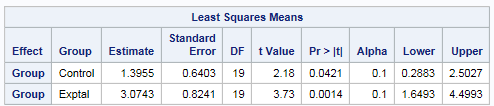
The model statement implies that we want means for the change scores in each of the levels of Group (Control and Exptal). residual makes the plots of standardized residuals.

The random statement adds a randomly chosen number to the change score for each athlete in the experimental group. Think back to the hat metaphor. There's a hat with the label xVarExp\*Name. The xVarExp ensures that a number comes out of the hat only for subjects in the experimental group. And Name implies that every time Name changes, another number has to come out of the hat. That makes each number an individual response for each athlete in that group. The variance of those numbers is the statistic summarizing the individual responses.

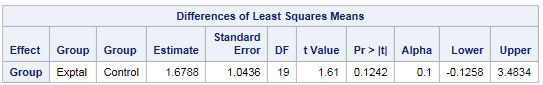
The lsmeans statement produces the means of the two groups. The diff=control("Control") forces SAS to use the Control group as the reference for the difference; otherwise it produces a difference or differences in alphabetical sequence, which here would be Control-Exptal. And of course alpha=0.1 cl provides confidence limits for the difference in the means.

Highlight the above code and run it.

1. Scroll down to the least-squares means and their differences:

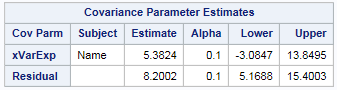


The control group experienced a 1.4% increase in power, while the experimental group experienced 3.1%.

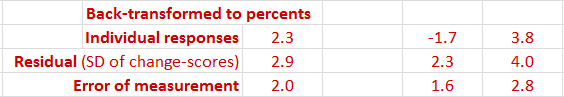


The differences is 1.7%. If these were competitive cyclists, the smallest important change in mean power in a competition time trial would be ~1.0%. Assuming a change in performance in the peak-power test translates into a similar change in mean power in a time trial, the effect of the treatment here would be clear.

1. Scroll back to the covariance parameters:



Let's back-transform these properly with a spreadsheet. Open **controlled trial process SAS output.xlsx** (not in SAS!) and click on the tab **change scores**, where you will see this:



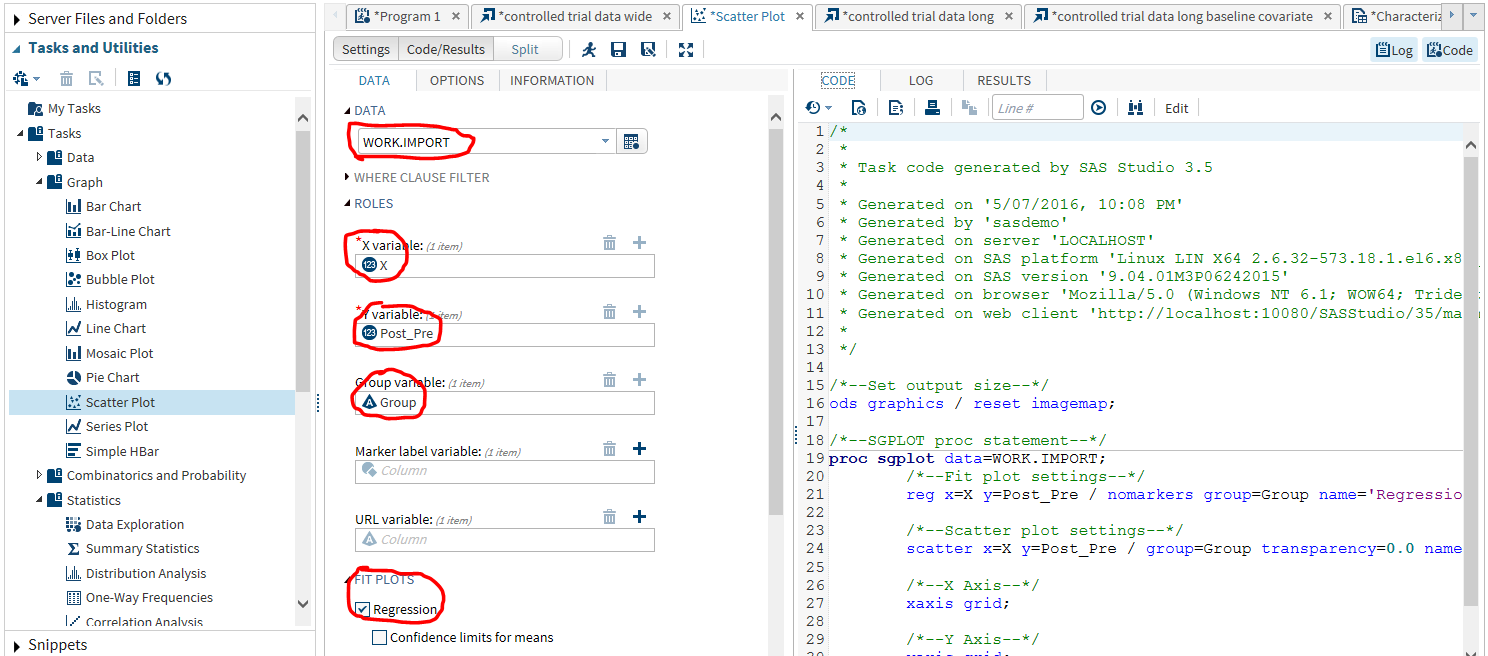
Check the formulae to see what's going on. The variance for change scores is twice that for the error variance on the pre and post scores (or equivalently, the error of measurement is the SD of change scores divided by √2). Compare the outcomes in **xParallelGroupsTrial to compare SAS.xlsx**, where I have highlighted relevant cells in pale green.

Don't forget that you have to double SDs (or halve thresholds) to interpret their magnitudes. If the smallest important change in power is 1%, the thresholds for moderate, large and very large are 3.0%, 5.5% and 8.7% (see the formulae in the spreadsheet), which means the individual responses are moderate, and they could even be large, but they are unclear, because of the moderate negative lower confidence limit.

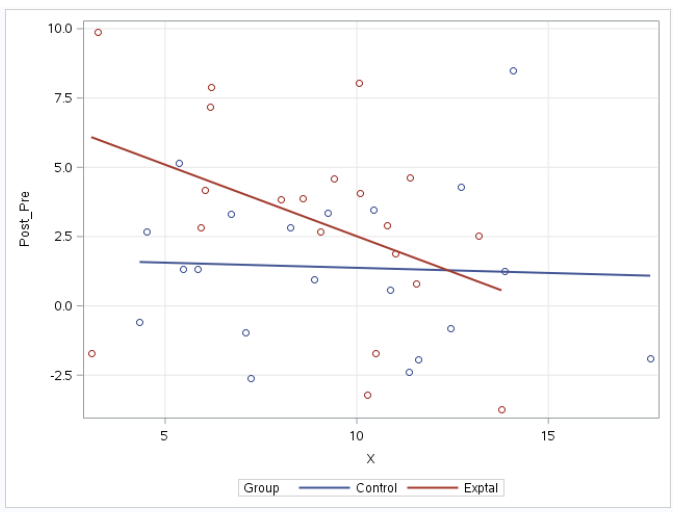
The error of measurement isn't too bad for a test of maximal power output. The best you can hope for is ~1%. But these are simulated data anyway!

**Analyzing Change Scores, Data in Wide Format, Plus a Covariate**

1. Now we adjust for and estimate the modifying effect of a subject characteristic, X, a measure of baseline training. First, let's visualize the relationship between the characteristic and the change scores by producing the same graph as in the xParallelGroupsTrial spreadsheet. In the navigation window, open  
   Tasks and Utilities/Tasks/Graph, and double-click Scatter Plot. In the DATA tab choose these…



…and click Run. Maximize the window so you can see this graph…



The Y axis is the individual change scores of log-transformed peak power, so they are approximate percent changes. The X axis the hours per week of training at baseline. It should be obvious that there is a substantial negative effect of baseline training (the more training, the less the effect) in the experimental group, and no substantial effect in the control group.

1. Here's the code to run the mixed model, with the changes to the previous program highlighted:

title "Analyzing change scores, plus a modifying covariate";

proc mixed data=import plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group;

model Post\_Pre=Group Group\*X/noint solution ddfm=sat residual alpha=0.1;

random xVarExp/subject=Name;

\*random xVarExp\*Name; \*this does the same thing as the above line;

lsmeans Group/diff=control("Control") cl alpha=0.1;

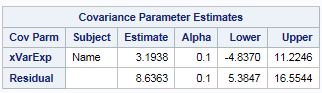
run;

**Group\*X** produces different slopes for the two groups. You could include X in the model, which would then produce the slope for one of the groups, while Group\*X would give the difference in the slopes. Try it, if you like. We will want the difference in the slopes, but we'll get it another way, shortly.

**noint** stands for no intercept, which suppresses the usual constant, which means Group provides the intercept for both groups. **solution** outputs the intercepts and the two slopes (the estimates of the parameters of the fixed effects). **ddfm=sat** should be the default, because without it you get wrong degrees of freedom; sat stands for Satterthwaite, a mid-20th Century statistician.

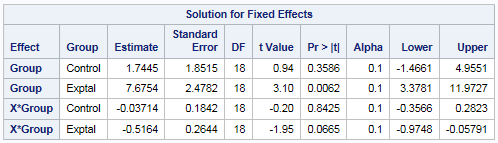
Copy the above code into the bottom of Program 1 and run it.

1. First interesting thing in the RESULTS window is the covparms:



The individual responses are the square root of 3.19, which comes to 1.8: smaller than before, because the covariate X is accounting for some individual responses. I haven't converted these to percent SDs in a spreadsheet, but you should make sure you can.

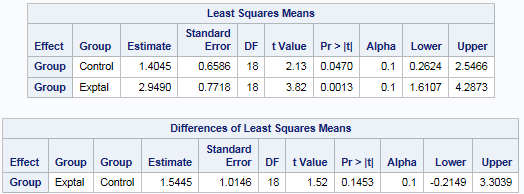
1. Next is the solution for fixed effects:



The two Group estimates are the Y-axis intercepts of the two straight lines in the scatterplot. Check it out. You'll notice that the X axis doesn't extend down to zero, but if it did, the intercepts of the straight lines would be the values here.

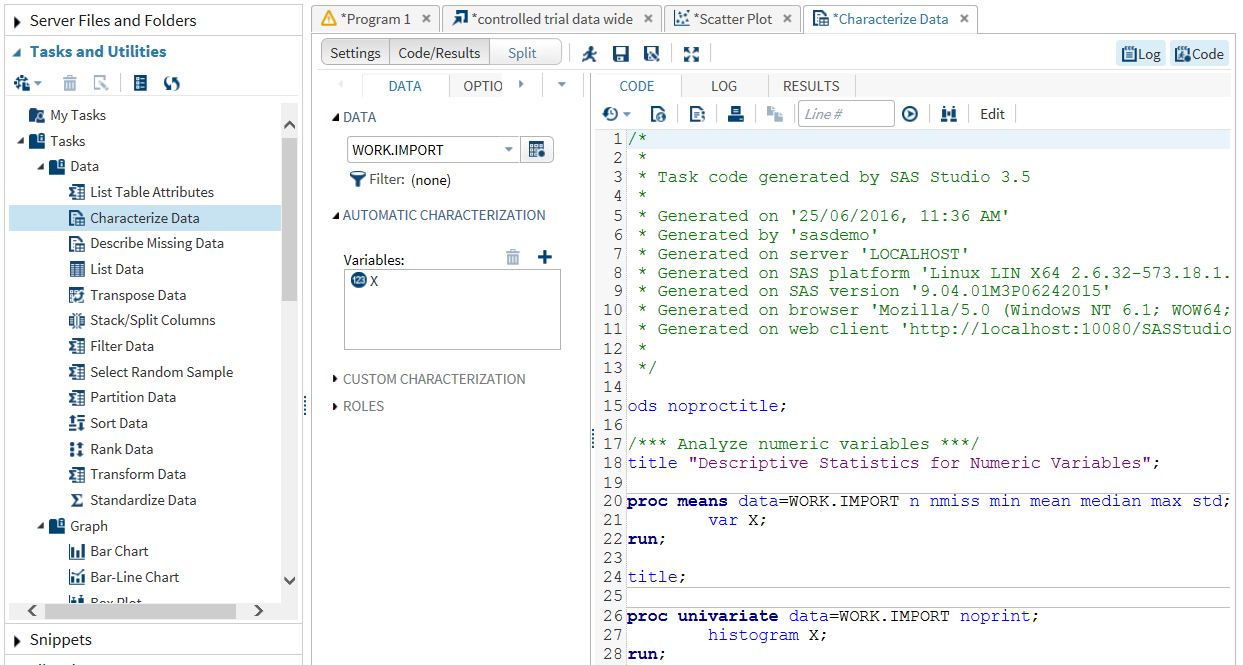
The two X\*Group estimates are the slopes of the straight lines.

1. As before, the treatment effect comes from the least-squares means:

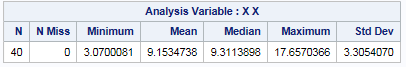


These are now evaluated at the mean value of X. The spreadsheet shows how by putting a vertical dashed line through the mean in the scatterplot, when you enter the mean in the appropriate cell, AA35. Where the line intersects the two straight lines is the two values of the least-squares mean change, and the spreadsheet evaluates the difference. The spreadsheet also evaluates the difference at any chosen value of X entered into AA35. How do we do that in SAS? First, let's get the mean and SD of X.

1. Go to Tasks and Utilities/Tasks/Data and double-click Characterize Data, and choose the variable X:



You can see SAS is about to invoke two basic procs, means and univariate. You should learn the code for both these procs. Click the Run icon. Here's the output from proc means:



So now let's see how to evaluate the effects of the treatment at something other than what the least-squares means provide, the effects at the mean of X. First, we'll do it *at* the mean of X to make sure we get the same answer as the least-squares means. It's achieved with an ESTIMATE statement, which is not available in the Mixed Model task menu–we have to add it ourselves to the code. (The same thing happens in SPSS, where it's done with a TEST statement.) Here's the extra lines of code:

estimate "Means @ X=9.15:";

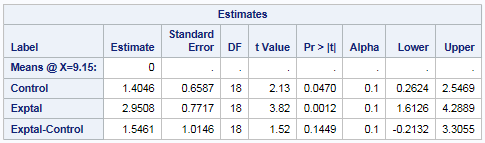
estimate " Control" Group 1 0 Group\*X 9.15 0/cl alpha=0.1;

estimate " Exptal" Group 0 1 Group\*X 0 9.15/cl alpha=0.1;

estimate " Exptal-Control" Group -1 1 Group\*X -9.15 9.15/cl alpha=0.1;

The first estimate is cosmetic. The other three estimates estimate what it says in the quote marks. And what follows after the quote marks sort-of makes sense: the effect in the control group is the intercept for that group plus the slope for that group times the mean value of X, 9.15. The values of the control group come before those of the experimental group alphabetically, which is why the zeros come where they do.

Copy and paste these into Program 1 above the last run; statement, highlight whole proc mixed step, click the Run icon, and this is what you get:



Check these against the least-squares means above and you will see they are almost exactly the same, the only slight difference arising from the fact that the mean of X isn't exactly 9.15.

Now, the SD for X is 3.31, so if you wanted to evaluate the effects of the treatment at 1 SD below the mean, you would use 5.84 (9.15-3.31) instead of 9.15 in the above estimate statements. And if you wanted to get the effect for cyclists who do no baseline training, you would drop Group\*X altogether. What you would get is the same values for Group as shown previously in the solution for fixed effects.

1. Now, what about the effect of X itself? We've already seen the two slopes in that solution panel. We need to reproduce those, and derive the difference between them. The magnitude of the effect of a linear covariate should be evaluated for 2 SD of the covariate: you effectively compare the effect for subjects 1 SD above the mean with the effect for subjects 1 SD below the mean. (It's the only way to make the scale of magnitudes for correlations fit with the scale of magnitudes for standardized differences in means.) Two SD of X is 2\*3.305 = 6.61, so here are the appropriate estimate statements:

estimate "Effect of X:";

estimate " Control x1" Group\*X 1 0/cl alpha=0.1;

estimate " Exptal x1" Group\*X 0 1/cl alpha=0.1;

estimate " Exp-Con x1" Group\*X -1 1/cl alpha=0.1;

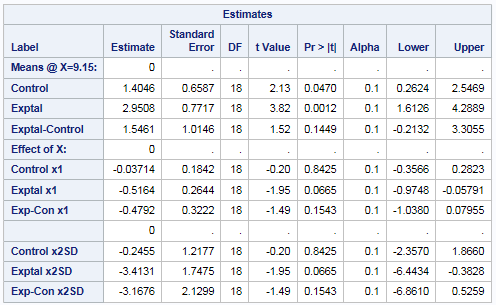
estimate "";

estimate " Control x2SD" Group\*X 6.61 0/cl alpha=0.1;

estimate " Exptal x2SD" Group\*X 0 6.61/cl alpha=0.1;

estimate " Exp-Con x2SD" Group\*X -6.61 6.61/cl alpha=0.1;

Copy and paste these into Program 1 above the last run; statement, highlight whole proc mixed step, click the Run icon, and this is what you get:

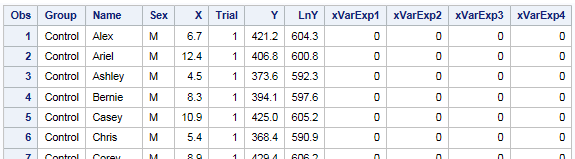


Check that the estimates for x1 here are almost exactly the same as in the previous solution for fixed effects. You can also check the values in the xParallelGroupsTrial spreadsheet: put 1 in Cell AA38 (making sure you have something in Cell AA35) and you will get the difference in slopes as an effect; put 6.61 into Cell AA38 and you will get the above x2SD values.

What about **an extra covariate**? The file all controlled-trial programs SAS.docx shows the code first for inclusion of a nominal covariate **sex** then a numeric covariate, **LnPre**, the baseline of the dependent (log-transformed, because the dependent is log transformed before analysis). At this point you can also check out crossover two predictors SAS.docx for programs to analyze change scores from a crossover with two covariates, using the data from the post-only crossover spreadsheet at Sportscience.

**Analysis with Data in Long Format**

1. The data are the same as above, a control and experimental group each with two pre-intervention and two post-intervention trials. Find the Excel file controlled trial data long.xlsx in the navigation window, double-click to open a data window, run it, then do a proc print in Program 1. Inspect the data:



The dependent variable is called Y (and LnY), and Trial has values 1 to 4. I have added four dummy variables with values of 1 for each of the four trials in the experimental group. We won't use all these.

1. To compare this approach with the change-scores approach, we have to limit the analyses to Trials 2 and 3. This new dataset is IMPORT1. Create a new dataset with the following code copied into Program 1:

data long;

set import1;

if 1<Trial<4; \*this makes a subset of the data;

run;

You could achieve the same thing with if 1<Trial and Trial<4;   
or if Trial>1 and Trial<4; or if Trial=2 or Trial=3; Run it and check the LOG.

1. Here is the code for the mixed model, with new stuff highlighted:

title "Analyzing actual scores for Trials 2 and 3";

proc mixed data=long plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group Trial;

model LnY=Group\*Trial/residual ddfm=sat;

random Intercept xVarExp3/subject=Name;

\*random Name xVarExp3\*Name; \*this does the same thing as the above line;

\*repeated Trial/subject=Name type=cs; \*this line plus the next…;

\*random xVarExp3\*Name; \*…does the same thing as the above;

lsmeans Group\*Trial;

lsmestimate Group\*Trial

"Control Post-Pre" -1 1 0 0,

"Exptal Post-Pre" 0 0 -1 1,

"Exp-Con Post-Pre" 1 -1 -1 1/cl alpha=0.1;

run;

Obviously **Trial** now has to be included, and **Group\*Trial** estimates the four means of interest, in the order Control Trial2, Control Trial3, Exptal Trial2, Exptal Trial3.

Now that we aren't modeling change scores, we can estimate a value for each subject, hence the **Intercept** term in the random statement, and of course **xVarExp3** specifies individual responses in the experimental group in Trial 3.

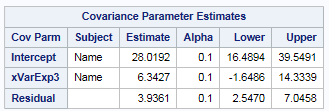
The next **random** statement is just a trivially different way to state the same two random effects.

The **repeated** statement is an alternative and sometimes useful way to specify random effects involving repeated measurements on subjects, where there are complicated relationships between the repeated measurements. The option **type=cs** stands for "compound symmetry", which means the same independent residual error and the same value for the subject random effect on every repeated measurement. Since we have extra error representing individual responses on one of the measurements, we have to specify that error with an extra random statement.

The **lsmestimate** statement is a very cool way of getting combination of the levels of nominal (class) fixed effects without having to specify mean values of covariates with estimate statements. We didn't have to worry about this issue when modeling change scores, because lsmeans Group/diff provided the difference we wanted, but all /diff would do here is produce only all the pairwise differences. We want a combination of differences. The syntax of the lsmestimate statement should be obvious.

Copy it into Program 1 and Run it!

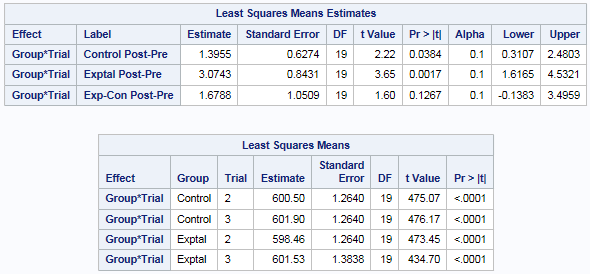
1. Here are the random effects:



xVarExp3 is the individual responses in the experimental group on the post test. You will see it's a bit different from the value we got with the analysis of change scores (5.38). The residual (3.94) at first glance is quite different from the residual in the previous analysis (8.20). In fact it is about half the previous value, which is right, because the residual in the previous analysis represents the overall SD of change scores, which are √2 times the typical error. These estimates here are variances, which makes the variance of the change scores twice the variance for the typical error. Only it's not quite twice! Why?

The estimates differ slightly, because the mixed model here is also trying to estimate a random effect for subjects; that is, an individual mean value for each subject. With the analysis of change scores, the differences between subjects disappear. The optimization of the fit of the data to the model therefore has to contend with something extra here, so it converges on a slightly different answer. The values for each subject are summarized by the estimate for Intercept as a variance, 28.0. The square root is 5.3, which is the approximate difference between subjects expressed as a percent SD or coefficient of variation.

1. The least-squares mean estimates come out ahead of the least-squares means. The effect Exp-Con Post-Pre is identical to the effect estimated with change scores:



1. Now let's include the covariate X. Here's the code, with new bits highlighted:

title "Analyzing actual scores for Trials 2 and 3, plus a covariate";

proc mixed data=long plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group Trial;

model LnY=Group\*Trial Group\*Trial\*X/residual ddfm=sat;

\*random Intercept xVarExp3/subject=Name;

\*random Name xVarExp3\*Name; \*this does the same thing as the above line;

repeated Trial/subject=Name type=cs; \*this line plus the next…;

random xVarExp3\*Name; \*does the same thing as the above two lines;

lsmeans Group\*Trial;

lsmestimate Group\*Trial

"Control Post-Pre" -1 1 0 0,

"Exptal Post-Pre" 0 0 -1 1,

"Exp-Con Post-Pre" 1 -1 -1 1/cl alpha=0.1;

estimate "Effect of X:";

\*estimate " Control x1" Group\*Trial\*X -1 1 0 0/cl alpha=0.1;

\*estimate " Exptal x1" Group\*Trial\*X 0 0 1 -1/cl alpha=0.1;

\*estimate " Exp-Con x1" Group\*Trial\*X 1 -1 -1 1/cl alpha=0.1;

\*estimate "";

estimate " Control x2SD" Group\*Trial\*X -6.61 6.61 0 0/cl alpha=0.1;

estimate " Exptal x2SD" Group\*Trial\*X 0 0 -6.61 6.61/cl alpha=0.1;

estimate " Exp-Con x2SD" Group\*Trial\*X 6.61 -6.61 -6.61 6.61/cl alpha=0.1;

run;

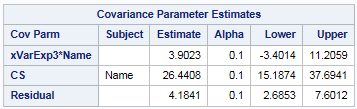
**Group\*Trial\*X** estimates a different slope at each of the four levels of Group\*Trial. We only ever estimate changes and differences in changes in these slopes, as shown by the **estimate** statements. I've starred off the x1 estimates.

The **repeated** statement is an alternative to our usual random intercept/subject=Name; CS stands for compound symmetry. There's probably a good explanation!

I've starred off the first random statement and unstarred the repeated and last random to show you the output with this equivalent specification of the random effects.

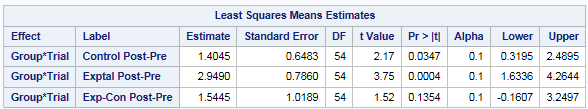
Copy all this code into the end of Program 1 and run it.

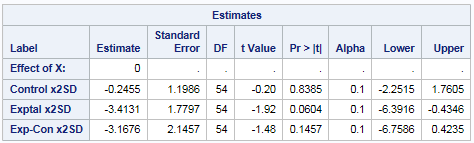
1. Here are the random effects…



Once again inclusion of the covariate X has explained some of the individual responses, because the variance of 3.9 is less than the 6.3 in the previous analysis. The random effect for Name is now called CS.

1. The least-squares mean estimates and the fixed-effect estimates are almost exactly what we've already seen for the analysis of change scores:





**Analysis with Data in Long Format and More Than One Repeated Measurement**

1. Let's add an extra trial, the second post test, Trial 4. Copy this code into the bottom of Program 1, highlight it and run it:

data long1;

set import1;

if 1<Trial;

run;

Check the log to make sure it worked OK.

1. Now here's the code to take into account the extra trial:

title "Analyzing actual scores for Trials 2, 3 & 4, plus a covariate";

proc mixed data=long1 plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group Trial;

model LnY=Group\*Trial Group\*Trial\*X/residual ddfm=sat;

random Intercept/subject=Name;

random xVarExp3 xVarExp4/subject=Name type=un;

lsmeans Group\*Trial;

lsmestimate Group\*Trial

"Control Post1-Pre" -1 1 0 0 0 0,

"Exptal Post1-Pre" 0 0 0 -1 1 0,

"Exp-Con Post1-Pre" 1 -1 0 -1 1 0,

"Exp-Con Post2-Pre" 1 0 -1 -1 0 1/cl alpha=0.1;

estimate "Effect of X:";

estimate " Control Post1-Pre x2SD" Group\*Trial\*X -6.61 6.61 0 0 0 0/cl alpha=0.1;

estimate " Exptal Post1-Pre x2SD" Group\*Trial\*X 0 0 0 -6.61 6.61 0/cl alpha=0.1;

estimate " Exp-Con Post1-Pre x2SD" Group\*Trial\*X 6.61 -6.61 0 -6.61 6.61 0/cl alpha=0.1;

estimate " Exp-Con Post2-Pre x2SD" Group\*Trial\*X 6.61 0 -6.61 -6.61 0 6.61/cl alpha=0.1;

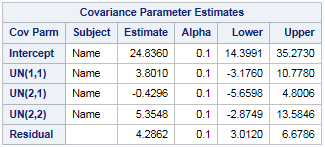
run;

I've made a random effect for Name and separate random effects for individual responses in Trials 3 and 4. The crucial point here is the type=un, which allows the individual responses to be correlated: you would expect subjects who responded well in the first post test to still respond well in the second post test, assuming the experimental treatment was still at least partially effective in the second post test.

The changes to the lsmestimate and estimate statements take into account the extra trial. I have provided only a few of the possible statements here.

Highlight the code and run it.

1. Only the random effects need explaining:



We've met UN(1,1) and so on already. UN(1,1) and UN(2,2) correspond to xVarExp3 and xVarExp4, individual responses in the first and second post tests. Neither would be clear. Interestingly their covariance UN(2,1) is slightly negative rather than positive. Well, the confidence limits allow for positive covariance, and in any case, I may not have simulated correlated individual responses!

**Adjusting for Baseline with Data in Long Format**

1. The baseline or pre-test value of the dependent variable can be an important modifier of a treatment effect (e.g., subjects with high initial values have less headroom for improvement). It's easy to include the pre-test as a modifier of change scores, but when modeling in long format, it's trickier. **Warning**: this gets complex.

When using the pre test as a covariate, it simply doesn't make sense to include the pre test as a repeated measurement, because you'd be predicting that measurement with itself! Instead, you have to delete all the observations for that trial, but copy it alongside all the other trials like any other subject characteristic. It's as if you start the repeated measurements with the first post test. In the analysis, you include baseline as a modifying covariate, which means you are effectively analyzing change scores.

I have modified the dataset we have been using as an example. I've made Trial 2 the baseline, to make it easy to develop the codes from the previous programs, and I have dropped Trial 1 from the dataset altogether, for simplicity.

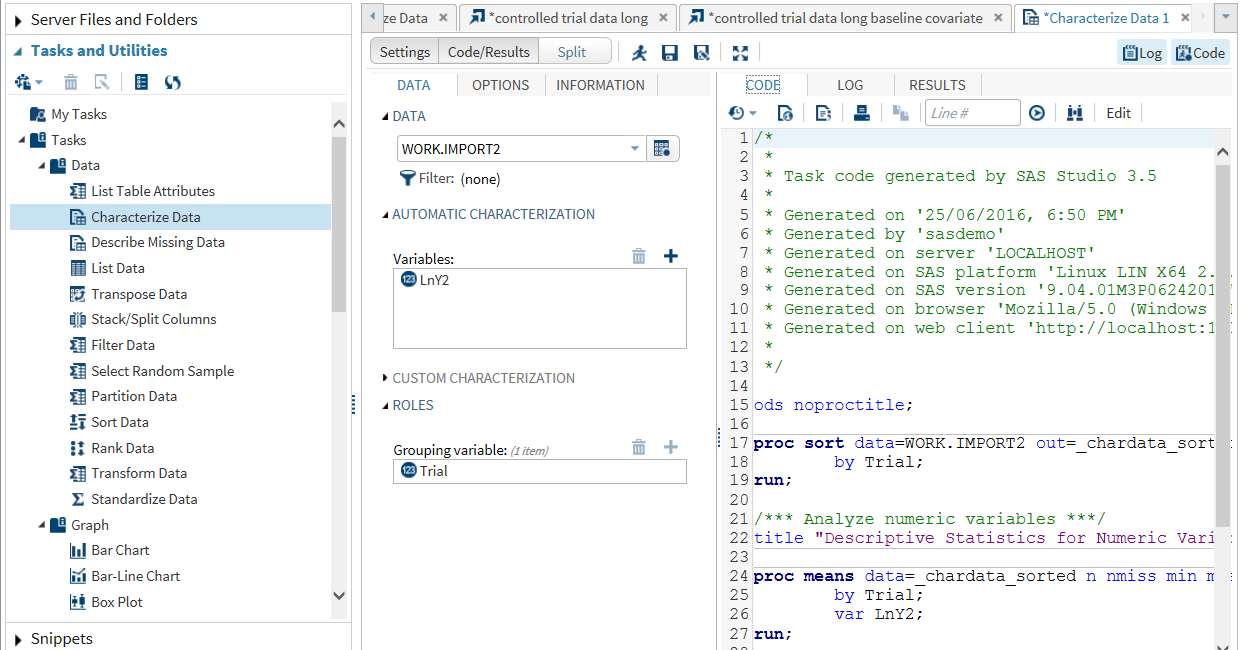
1. Find the file controlled trial data long baseline covariate.xlsx, double-click it and run the resulting data window. It comes in as IMPORT2.

Do a proc print at the bottom of Program 1 to view the data. Here are the first few observations:



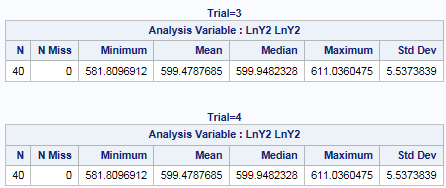
There are now only two trials, Trial 3 and 4, and Trial 2 has been copied into each trial as the new variables Y2 and LnY2. And there are now only two dummy variables, xVarExp3 and xVarExp4.

1. To fully evaluate the effect of baseline LnY2, we will need the mean and SD of LnY2. Get these via Tasks/Data and double-click Characterize Data…
2. Choose these values in the DATA window:



I've chosen Trial as a grouping variable. It's a minor point, because LnY2 has the same values for Trial 3 and Trial 4, but if you get means and SD for both trials combined, the SD is slightly different.

Click the Run icon. Here are the means and SD…



…so we'll use 600 for the mean and 11.1 (2\*5.54) for 2 SD.

1. Here's the code to analyze this dataset, with changes highlighted and a sample of estimate statements:

title "Analyzing actual scores for Trials 3 & 4, with X and a baseline covariate";

proc mixed data=import2 plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group Trial;

model LnY=Group\*Trial Group\*Trial\*X Group\*Trial\*LnY2/residual ddfm=sat;

random Intercept/subject=Name;

random xVarExp3 xVarExp4/subject=Name type=un;

lsmeans Group\*Trial/cl alpha=0.1;

lsmestimate Group\*Trial

"Exp-Con Post1-Pre" -1 0 1 0,

"Exp-Con Post2-Pre" 0 -1 0 1/cl alpha=0.1;

estimate "Effect of X x2SD:";

estimate " Control Post1-Pre" Group\*Trial\*X 6.61 0 0 0/cl alpha=0.1;

estimate " Exptal Post1-Pre" Group\*Trial\*X 0 0 6.61 0/cl alpha=0.1;

estimate " Exp-Con Post1-Pre" Group\*Trial\*X -6.61 0 6.61 0/cl alpha=0.1;

estimate "Effect of LnY2 x2SD:";

estimate " Control Post1-Pre" Group\*Trial\*LnY2 11.1 0 0 0/cl alpha=0.1;

estimate " Exptal Post1-Pre" Group\*Trial\*LnY2 0 0 11.1 0/cl alpha=0.1;

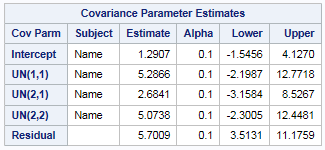
estimate " Exp-Con Post1-Pre" Group\*Trial\*LnY2 -11.1 0 11.1 0/cl alpha=0.1;

run;

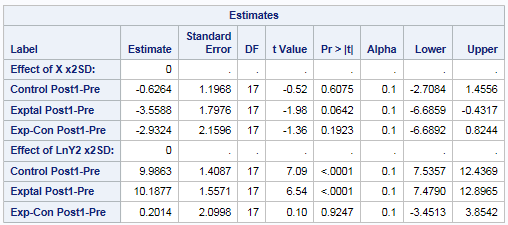
These are sort-of obvious. One thing that's missing is estimates of the Post1-Pre and Post2-Pre changes in the control and experimental groups. These have to be derived "manually", by subtracting the mean for LnPre2 from the least-squares mean estimates for Post1 and Post2 in each group.

Copy the above code into the end of Program 1, highlight it, and run it.

1. First up are the random effects:

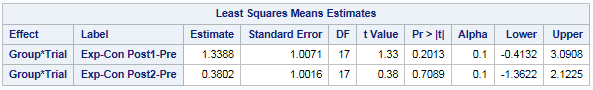


Check back to the values for the previous model, without the baseline. The individual responses in the first post test, UN(1,1), have actually increased a bit: 5.3 vs 3.8. I suspect that the baseline is not having any modifying effect on the effect! Skip down to the estimates to see:

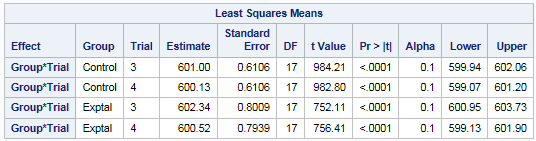


The modifying effect of 2 SD of LnY2 in each group appears to be really large (10.0 and 10.2, approximate percents), but these are just the slopes of LnY3 vs LnY2 and LnY4 vs LnY2 multiplied by 11.1. It's their difference that matters, and it's a trivial 0.2%.

1. The least-squares mean estimates provide the net treatment effect Post1-Pre and Post2-Pre:



1. And if you subtract 600 off each of the following least-squares means, you will get the Post1-Pre (Trial3 – Trial2) and Post2-Pre (Trial4-Trial2) changes in the control and experimental groups:



1. At this stage you could use comments to annotate Program 1 and save it. Otherwise all the programs, including the following with Sex, are in the file **all controlled-trial programs SAS.docx**.

**Estimating the Modifying Effects of Sex**

1. When you have two or more distinct subgroups of subjects, such as females and males, in a controlled trial, you should investigate the effects of the treatment in each subgroup, the difference between the subgroups, and possibly the effect averaged over the two subgroups (which is similar to the effect with equal numbers of females and males and without sex in the model). You do it either by doing separate analyses with "by group" processing (here, by Sex) or by doing a single analyses and extending the lsmestimate or estimate statements. As an example, let's return to the data in long format with a pre and post trial…
2. Here is how to do it with "by group" processing. Additions to the original code are highlighted:

proc sort data=long;

by Sex;

title "Analyzing actual scores for Trials 2 and 3";

proc mixed data=long plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Group Trial;

model LnY=Group\*Trial/residual ddfm=sat;

random Intercept xVarExp3/subject=Name;

lsmeans Group\*Trial;

lsmestimate Group\*Trial

"Control Post-Pre" -1 1 0 0,

"Exptal Post-Pre" 0 0 -1 1,

"Exp-Con Post-Pre" 1 -1 -1 1/cl alpha=0.1;

by Sex;

run;

If you run this code, you will see substantial differences in the random and fixed effects between females and males. To compare and combine them inferentially, use the [Combine/compare effects](http://www.sportsci.org/resource/stats/xCombineGroups.xls) spreadsheet.

1. Here is the code to account for Sex in a single analysis, with additions highlighted:

title "Analyzing actual scores for Trials 2 and 3";

proc mixed data=long plots(only)=StudentPanel(conditional) alpha=0.1 nobound;

class Name Sex Group Trial;

model LnY=Sex\*Group\*Trial/residual ddfm=sat;

random Intercept xVarExp3/subject=Name group=Sex;

repeated/group=Sex;

lsmeans Sex\*Group\*Trial;

lsmestimate Sex\*Group\*Trial

"F Control Post-Pre" -1 1 0 0 0 0 0 0,

"F Exptal Post-Pre" 0 0 -1 1 0 0 0 0,

"F Exp-Con Post-Pre" 1 -1 -1 1 0 0 0 0,

" " 0,

"M Control Post-Pre" 0 0 0 0 -1 1 0 0,

"M Exptal Post-Pre" 0 0 0 0 0 0 -1 1,

"M Exp-Con Post-Pre" 0 0 0 0 1 -1 -1 1,

" " 0,

"F-M Exp-Con Post-Pre" 1 -1 -1 1 -1 1 1 -1,

"(F+M)/2 Exp-Con Post-Pre" 1 -1 -1 1 1 -1 -1 1 divisor=2

/cl alpha=0.1;

run;

**IMPORTANT**: Make the order of the variables in the class statement (Sex Group Trial) the same as the order of the terms in the fixed effect (Sex\*Group\*Trial). SAS will otherwise scramble the order of the terms in the fixed effect, and the resulting estimate and lsmestimate statements produce wrong answers.

**group=Sex** makes SAS estimate separate random effects. Unfortunately, to get separate residuals (the females and males could differ in their reliability), we need the **repeated** statement as well. The use of **divisor=2** to average the females and males should be obvious.

When I ran the above, the analysis failed and the LOG showed   
WARNING: Stopped because of infinite likelihood.   
This sort of thing happens occasionally. My fix was first to remove the nobound, run the program to see the values of the covariance parameters, then use their approximate values to make the following statement…  
parms 30 30 0 10 10 2;  
…which I inserted below the repeated statement (but it can go anywhere before the run;). parms stands for starting values of the covariance parameters. SAS sometimes needs to be told reasonable values or it can't get started on the iterative process that finds the values of the fixed and random-effect parameters.

I then put nobound back into the proc mixed statement and ran the program. It gave the same answers for females and males as the by Sex program.

Doing separate analyses with a by group variable is easier than doing a single analysis, but beware: if the groups are not independent, then the spreadsheet for combining and comparing the effects doesn't work properly. In that case you have to do a single analysis, and the random effects have to include provision for shared variance (interdependence).

**Summary Table**

|  |  |  |
| --- | --- | --- |
| **Various methods for analyzing reliability, controlled trials and other repeated measures.** Anything other than my spreadsheets has the disadvantage that you have to do lots of processing to deal with log transformation and magnitude-based inferences. SPSS also has the disadvantage of not allowing negative variances. R has the further disadvantage of not providing confidence limits for random effects. | | |
| Method | Advantages | Disadvantages |
| **Reliability Analyses** | | |
| Sportscience consecutive pairwise spreadsheet | Best for performance or fitness tests.  Logs and graphs all done for you. | No good for clustering (multiple sources of variability). |
| Sportscience one- and two-way spreadsheets | Good for simple block of repeated measurements, if discard familiarization trial(s).  Logs and graphs all done for you. | No good for clustering (multiple sources of variability).  Can’t handle missing data in 2-way analyses. |
| SPSS Scale/Reliability | Good for simple block of repeated measurements in wide format. Don't include any familiarization trial(s). | No good for clustering (multiple sources of variability).  Can’t handle missing data in 2-way analyses. |
| Mixed modeling | Best for clustering (multiple sources of variability).  Add covariates for studies of monitoring. | Can't create random-effect model for consecutive pairwise analysis (autoregressive models don't work). |
| **Controlled-trial Analyses** | | |
| Spreadsheet for analysis of change scores | Good for only a few trials. Easy to choose and/or combine trials.  Logs, graphs and MBIs all done for you.  Modifying effect of covariates is easy to understand from the graphs.  Modifying effect of baseline is easy.  Mediator analysis with change scores for a covariate is easy. | Only one covariate at a time.  Only two groups at a time.  (These are what stop the spreadsheet from being best in this category.) |
| Mixed modeling of change scores,  data in wide format | Best for only a few trials.  Simple fixed and random effects.  Modifying effect of baseline is easy. | Only one "pre" trial at a time. |
| Mixed modeling,  data in long format | Classic mixed model with random effect for subjects–especially with lots of repeated measurements and clustering.  Several trial effects and >2 groups can be estimated in one analysis. | Fixed effects are a bit complex with the group\*covariate interactions. |
| Mixed modeling,  data in long format, adjusting for baseline | As above. | Have to modify the long-format data to turn the baseline trial into a subject characteristic.  Less intuitive specification of model and estimates. |