The Pretest-Posttest x Groups Design: How to Analyze the Data

You could ignore the pretest scores and simply compare the groups on the posttest scores, but there is probably a good reason you collected the pretest scores in the first place (such as a desire to enhance power), so I'll dismiss that option.

To illustrate the analyses I shall use the AirportSearch data, available at http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Data.htm . Do see the Description of the data.

Mixed Factorial ANOVA

Treat the Pretest-Postest contrast as a within-subjects factor and the groups as a between-subjects factor. Since the within-subjects factor has only one degree of freedom, the multivariate-approach results will be identical to the univariate-approach results and sphericity will not be an issue.

Here is SPSS syntax and output.

```
GLM post pre BY race
 /WSFACTOR=PostPre 2 Polynomial
 /METHOD=SSTYPE(3)
 /CRITERIA=ALPHA(.05)
 /WSDESIGN=PostPre
 /DESIGN=race.
```

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	PostPre	Type III Sum of Squares	df	Mean Square	F	Sig.
PostPre	Linear	288.364	1	288.364	84.676	.000
PostPre * race	Linear	76.364	1	76.364	22.424	.000
Error(PostPre)	Linear	180.491	53	3.405		

Tests of Between-Subjects Effects

Measure:MEASURE_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1330.135	1	1330.135	307.006	.000
race	254.063	1	254.063	58.640	.000
Error	229.628	53	4.333		

It is the interaction term that is of most interest in this analysis, and it is significant. It indicates that the pre-post difference is not the same for Arab travelers as it is for Caucasian travelers. To further investigate the interaction, one can compare the groups on pretest only and posttest only and/or compare posttest with pretest separately for the two groups. I'll do the latter here.

```
SORT CASES  BY race.
SPLIT FILE SEPARATE BY race.
T-TEST PAIRS=post WITH pre (PAIRED)
  /CRITERIA=CI(.9500)
  /MISSING=ANALYSIS.
```

Arab Travelers

Paired Samples Statistics[a]

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Post-9-11	7.67	21	3.512	.766
	Pre-9-11	2.62	21	1.161	.253

a. race = Arab

Paired Samples Correlations[a]

		N	Correlation	Sig.
Pair 1	Post-9-11 & Pre-9-11	21	.065	.778

a. race = Arab

Paired Samples Test[a]

		Paired Differences				
		95% Confidence Interval of the Difference				
		Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Post-9-11 - Pre-9-11	3.397	6.698	6.379	20	.000

a. race = Arab

As you can see, the pre-post difference was significant for the Arab travelers.

Caucasian Travelers

Paired Samples Statistics[a]

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Post-9-11	2.82	34	1.290	.221
	Pre-9-11	1.21	34	1.572	.270

a. race = Caucasian

Paired Samples Correlations[a]

		N	Correlation	Sig.
Pair 1	Post-9-11 & Pre-9-11	34	.287	.099

a. race = Caucasian

Paired Samples Test[a]

		Paired Differences				
		95% Confidence Interval of the Difference				
		Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Post-9-11 - Pre-9-11	1.016	2.219	5.473	33	.000

a. race = Caucasian

As you can see, the (smaller) pre-post difference is significant for Caucasian travelers too. We conclude that both groups were searched more often after 9/11, but that the increase in searches was greater for Arab travelers than for Caucasian travelers. I should also note that the pre-post correlations are not very impressive here.

t Tests

Simple Comparison of the Difference Scores = The Interaction Term

We could simply compute Post minus Pre difference scores and then compare the two groups on those difference scores. Here is the output from exactly such a comparison.

```
COMPUTE Diff=post-pre.
VARIABLE LABELS  Diff 'Post Minus Pre'.
EXECUTE.
```

Group Statistics

	race	N	Mean	Std. Deviation	Std. Error Mean
Post Minus Pre	Arab	21	5.0476	3.62596	.79125
	Caucasian	34	1.6176	1.72354	.29558

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Post Minus Pre	Equal variances assumed	4.735	53	.000

The increase in number of searches is significantly greater for Arab travelers than for Caucasian travelers. Notice that the value of *t* is 4.735 on 53 *df.* An ANOVA on the same contrast would yield an *F* with one *df* in the numerator and the same error *df*. The value of *F* would be the square of the value of *t*. When you square our *t* here you get $F(1, 53) = 22.42$. The one-tailed *p* for this *F* is identical to the two-tailed *p* for the *t*. Yes, with ANOVA one properly employs a one-tailed *p* to evaluate a nondirectional hypothesis. Look back at the source table for the mixed factorial ANOVA. Notice that the $F(1, 53)$ for the interaction term is 22.42. Is this mere coincidence? No, it is a demonstration that **a *t* test on difference scores is absolutely equivalent to the test of the interaction term in a 2 x 2 mixed factorial ANOVA**. Many folks find this hard to believe, but it is easy to demonstrate, as I have done above. Try it with any other Pre-Post x Two Groups design if you are not yet convinced.

The logical next step here would be to test, for each group, whether or not the mean difference score differs significantly from zero. Here are such tests:

```
SORT CASES  BY race.
SPLIT FILE SEPARATE BY race.
T-TEST
  /TESTVAL=0
  /MISSING=ANALYSIS
  /VARIABLES=Diff
  /CRITERIA=CI(.95).
```

One-Sample Test[a]

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Post Minus Pre	6.379	20	.000	5.04762	3.3971	6.6981

a. race = Arab

One-Sample Test[a]

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Post Minus Pre	5.473	33	.000	1.61765	1.0163	2.2190

a. race = Caucasian

Please do notice that the values of *t*, *df*, and *p* here are identical to those obtained earlier with pre-post correlated samples *t* tests.

Independent *t* on Mean(Pre, Post) = The Main Effect of Groups

I computed, for each subject, the mean of e's pretest and posttest scores. I then conducted an independent samples t test comparing the groups on MeanPrePost. Here are the results:

Group Statistics

	race	N	Mean	Std. Deviation	Std. Error Mean
MeanPrePost	Arab	21	5.1429	1.88509	.41136
	Caucasian	34	2.0147	1.15132	.19745

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
MeanPrePost	Equal variances assumed	7.658	53	.000

Notice that if you take the value of *t* and square it, you get the value of the F for the main effect of groups in the ANOVA, with the same degrees of freedom and the same *p* value. This will always be the case for this design. **An independent samples *t* test comparing groups on the mean of pre/post is mathematically equivalent to the ANOVA *F* test on the main effect of groups**.

Correlated *t* Comparing Pre to Post ≠ The ANOVA Pre-Post Comparison

The correlated t will have one more error degree of freedom than will the ANOVA F, and the t will not be the square root of the F. The F test will typically have more power, having removed from what otherwise would be error variance that variance due to any interaction between groups and pre/post.

Analysis of Covariance

Here we treat the pretest scores as a covariate and the posttest scores as the outcome variables. Please note that this involves the assumption that the relationship between pretest and posttest is linear and that the slope is identical in both groups. These assumptions are easily evaluated. For example, the slope for predicting posttest scores from pretest scores in the Arab group is PostTest = 7.148 + .198∗PreTest. For the Caucasian group it is PostTest = 2.539 + .236∗PreTest.

```
UNIANOVA post BY race WITH pre
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(race) WITH(pre=MEAN)
  /CRITERIA=ALPHA(.05)
  /DESIGN=pre race.
```

Tests of Between-Subjects Effects

Dependent Variable:Post-9-11

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	310.064[a]	2	155.032	27.231	.000
Intercept	437.205	1	437.205	76.795	.000
pre	5.563	1	5.563	.977	.327
race	214.378	1	214.378	37.655	.000
Error	296.045	52	5.693		
Total	1807.000	55			
Corrected Total	606.109	54			

a. R Squared = .512 (Adjusted R Squared = .493)

Estimated Marginal Means

race

Dependent Variable:Post-9-11

race	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Arab	7.469[a]	.558	6.350	8.588
Caucasian	2.946[a]	.427	2.088	3.803

a. Covariates appearing in the model are evaluated at the following values: Pre-9-11 = 1.75.

We conclude that the two groups differ significantly on posttest scores after adjusting for the pretest scores. Notice that for these data the *F* for the effect of interest is larger with the ANCOV in the mixed factorial ANOVA. In other words, ANCOV had more power.

Which Analysis Should I Use?

If the difference scores are intrinsically meaningful (generally this will involve prettest and posttest having both been measured on the same metric), the simple comparison of the groups on mean differences scores is appealing and, as I have shown earlier, is mathematically identical to the mixed factorial. The ANCOV, however, generally has more power.

Huck and McLean (1975) addressed the issue of which type of analysis to use for the pretest-postest control group design. They did assume that assignment to groups was random. They explained that it is the interaction term that is of interest if the mixed factorial ANOVA is employed and that a simple *t* test comparing the groups on pre-post difference scores is absolutely equivalent to such an ANOVA. They pointed out that the *t* test and the mixed factorial ANOVA are equivalent to an ANCOV (with pretest as covariate) if the linear correlation between pretest and posttest is positive and perfect. They argued that the ANCOV is preferred over the others due to fact that it generally will have more power and can easily be adopted to resolve problems such as heterogeneity of regression (the groups differ with respect to slope for predicting the posttest from the pretest) and nonlinearity of the relationship between pretest and posttest.

Maxwell, Delaney, and Dill (1984) demonstrated that:

- The mixed factorial ANOVA may employ data from a randomized blocks design (where subjects have been matched/blocked up on one or more variables thought to be well associated with the outcome variable) and then, within blocks, randomly assigned to treatment groups, or it may employ data where no random assignment was employed (as in my example, where subjects were not randomly assigned a race/ethnicity). This matters. The randomized blocks design equates the groups on the blocking variables.
- If you can obtain scores on the concomitant variable (here the pretest) prior to assigning subjects to groups, matching/blocking the subjects on that concomitant variable and then randomly assigning subjects to treatment groups will enhance power relative to ignoring the concomitant variable when assigning subjects to treatment groups. Even with the randomized blocks design, one can use ANCOV rather than mixed ANOVA for the analysis and doing so will increase power.
- If the relationship between the concomitant variable and the outcome variable is not linear, the ANCOV is problematic. You may want to consider transformations to straighten up the (monotonic) nonlinear relationship or polynomial regression analysis.

References and Recommended Readings

- Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin*, *82*, 511-518.
- Maxwell, S. E., Delaney, H. D., & Dill, C. A. (1984). Another look at ANCOVA versus Blocking. *Psychological Bulletin*, *95*, 136-147.
- Rausch, J. R., Maxwell, S. E., & and Kelley, K. (2003). Analytic methods for questions pertaining to a randomized pretest, posttest, follow-up design. *Journal of Clinical Child and Adolescent Psychology*, *32*, 467-486.

Links

- Wuensch’s Stats Lessons
  - Least Squares Analyses of Variance and Covariance

Karl L. Wuensch
August, 2015