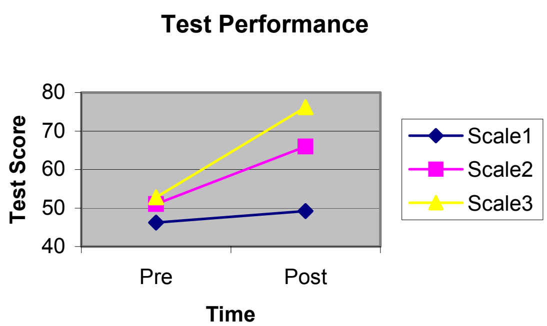
Chapter 11A
Multi-Factor Repeated Measures ANOVA
Repeated Measures on Both Factors

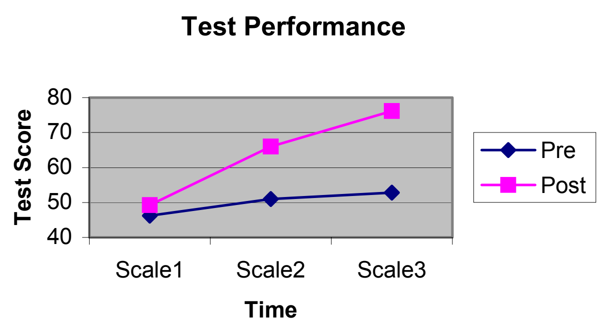
Repeated Measures ANOVA
Two-Factor Repeated Measures

1. Introduction

Participants take part in a training program to help them prepare for a standardized test. Before the training, they take the test and scores are recorded for all three sub-scales of the test. After the 12-week training program, participants retake the test.

	Pre-training			Post-training		
Participant	Subscale1	Subscale2	Subscale3	Subscale1	Subscale2	Subscale3
1	42	42	48	48	60	78
2	42	48	48	36	48	60
3	48	48	54	66	78	78
4	42	54	54	48	78	90
5	54	66	54	48	66	72
6	36	42	36	36	48	54
7	48	48	60	54	72	84
8	48	60	66	54	72	90
9	54	60	54	48	72	78
10	48	42	54	54	66	78
	46.2	51.0	52.8	49.2	66.0	76.2





- With this design, several questions come to mind:
  - Overall, does the training improve test scores?
    - Does training improve test scores for subscale 1?
    - Does training improve test scores for subscale 2?
    - Does training improve test scores for subscale 3?
  - Overall, is there a difference in performance on the three sub-scales?

- We have two repeated measures factors:
  - Pre-test and post-test scores
  - The three subscales of the test

- We can classify this design as a 2*3 repeated measures design, with repeated measures on both factors.

	Subscale of test (Factor A)			
Time (Factor B)	Subscale 1	Subscale 2	Subscale 3	
Pre-test	$\bar{X}_{\cdot 11} = 46.2$	$\bar{X}_{\cdot 21} = 51.0$	$\bar{X}_{\cdot 31} = 52.8$	$\bar{X}_{\cdot\cdot 1} = 50.0$
Post-test	$\bar{X}_{\cdot 12} = 49.2$	$\bar{X}_{\cdot 22} = 66.0$	$\bar{X}_{\cdot 32} = 76.2$	$\bar{X}_{\cdot\cdot 2} = 63.8$
$n = 10$	$\bar{X}_{\cdot 1\cdot} = 47.7$	$\bar{X}_{\cdot 2\cdot} = 58.5$	$\bar{X}_{\cdot 3\cdot} = 64.5$	

- Everything we learned about interpreting two-way between-subjects designs applies here. The only difference will be the assumptions of the test, and the construction of the error term.

2. Structural model, SS partitioning, and the ANOVA table

- We will only consider the case where the factors are fixed variables.

- Here is the structural model for a two-factor repeated measures design:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \pi_\sigma + (\alpha\beta)_{jk} + (\alpha\pi)_\sigma + (\beta\pi)_\sigma + (\alpha\beta\pi)_\sigma$$

  - Factor A (α_j) and Factor B (β_k) and the A*B interaction ($\alpha\beta_{jk}$) are fixed effects
  - The Subject effect (π_i) is a random effect. Thus, all interaction terms involving the subject effect are also random effects

  - Because we have one observation per participant, we do not have enough information to estimate both the $(\alpha\beta\pi)_\sigma$ interaction and the within cell residuals (ε_{ijk}).
    - In the randomized block design, we omitted the interaction term and retained the estimate of error
    - For factorial within-subjects resigns, we will omit the error term, and consider the information to be an estimate of the A*B*Subject interaction term.
    - This difference is a difference of terminology, not a sustentative difference.

- We can compute estimates for the fixed terms in the model, just as we have for factorial designs:

μ The overall mean of the scores

α_j The effect of being in level j of Factor A

$$\alpha_j = \mu._j. - \mu... \qquad \sum_{j=1}^{a} \alpha_j = 0$$

β_k The effect of being in level k of Factor B

$$\beta_k = \mu.._k - \mu... \qquad \sum_{k=1}^{b} \beta_k = 0$$

$(\alpha\beta)_{jk}$ The effect of being in level j of Factor A and level k of Factor B (the interaction of level j of Factor A and level k of Factor B)

$$(\alpha\beta)_{jk} = \mu._{jk} - \mu._j. - \mu.._k + \mu...$$

$$\sum_{j=1}^{a} (\alpha\beta)_{jk} = 0 \quad \textit{for each level of j}$$

$$\sum_{k=1}^{b} (\alpha\beta)_{jk} = 0 \quad \textit{for each level of k}$$

- The remaining terms are random effects.

- What do we do with all the random effect parameters? Let's take a closer look at the $(\beta\pi)_{\sigma}$ parameter. To look at the Factor B * Subject effect, we need to collapse across Factor A

Participant	Pre-Training	Post-Training	Difference
1	44	62	18
2	46	48	2
3	50	74	24
4	50	72	22
5	58	62	4
6	38	46	8
7	52	70	18
8	58	72	14
9	56	66	10
10	48	66	18
	50.0	63.8	13.8

- The B*Subject interaction examines if the effect of B (Pre vs. Post-training) is the same across all participants.

- In other words, the B*Subject interaction is a measure of the variability in the B effect or how much error we have in the measurement of the B effect

  (And so intuitively it makes sense that we can use the B*Subject term as an error term when we test the B effect)

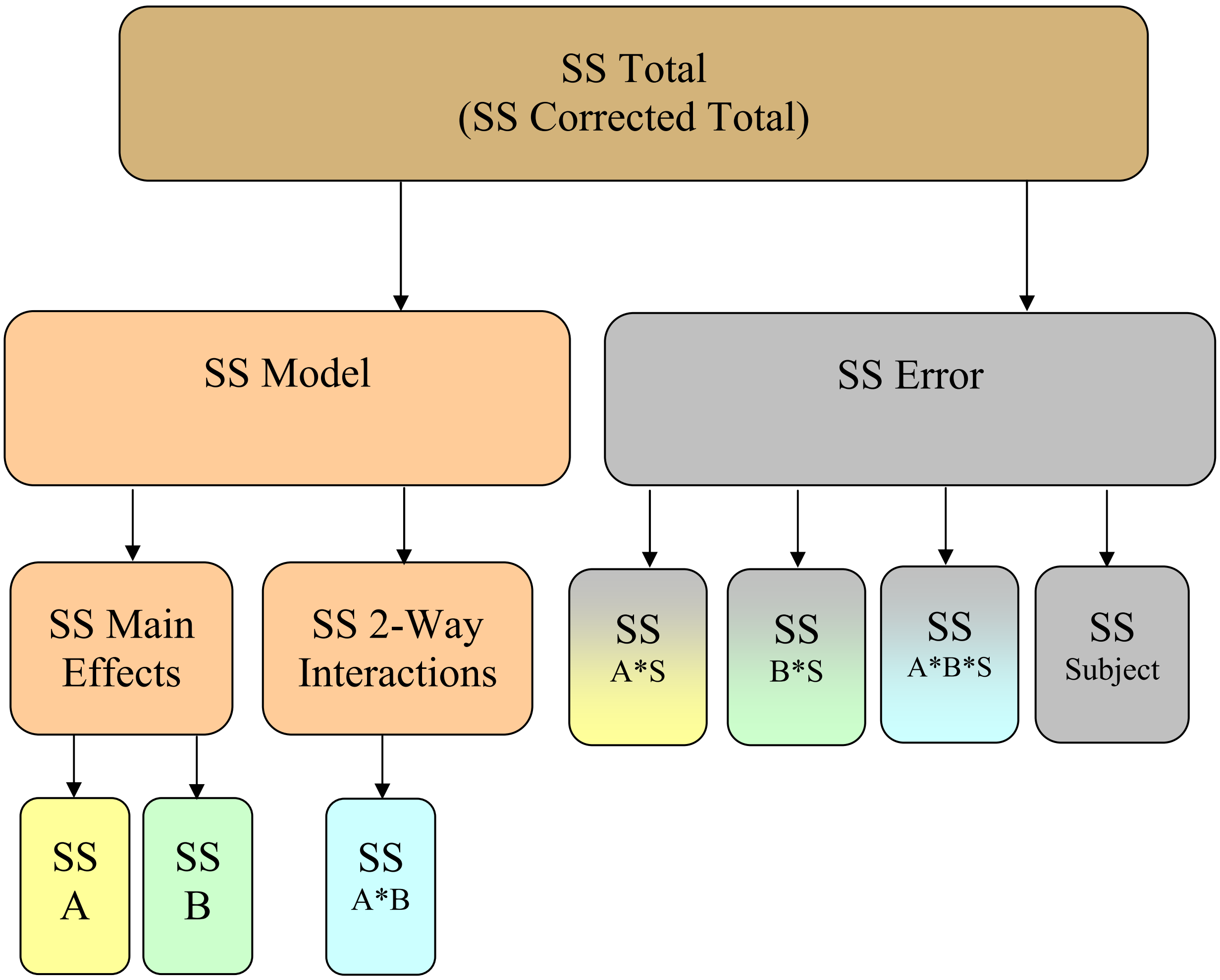
- This logic extends across each of the fixed effects

  ⇒ The A*Subject interaction measures the variability in the A effect

  ⇒ The B*Subject interaction measures the variability in the B effect

  ⇒ The A*B*Subject interaction measures the variability in the A*B interaction

- For a two-factor repeated measures design, we have the following SS decomposition.



- The ANOVA table for a two-factor repeated measures design:

  - Remember that to construct a valid F-test for an effect, we need:
    - The numerator to contain exactly one more term than the denominator
    - The extra term must correspond to the effect being tested

  - When these conditions hold:
    - The F-ratio will equal 1 when the null hypothesis is true (because the numerator and denominator will be estimating the same effects)
    - The F-ratio will be greater than 1 when the null hypothesis is false

Source	SS	df	MS	E(MS)	F
Factor A	*SSA*	*a-1*	$\frac{SSA}{a-1}$	$\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2 + \frac{nb\sum\alpha_j^2}{a-1}$	$\frac{MSA}{MS(A*S)}$
*A*S* *(Factor A Error)*	*SS (A*S)*	*(a-1)(n-1)*	$\frac{SS(A*S)}{(a-1)(n-1)}$	$\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2$	
Factor B	*SSB*	*b-1*	$\frac{SSB}{b-1}$	$\sigma_\varepsilon^2 + a\sigma_{\beta\pi}^2 + \frac{na\sum\beta_k^2}{b-1}$	$\frac{MSB}{MS(B*S)}$
*B*S* *(Factor B Error)*	*SS (B*S)*	*(b-1)(n-1)*	$\frac{SS(B*S)}{(b-1)(n-1)}$	$\sigma_\varepsilon^2 + a\sigma_{\beta\pi}^2$	
*A * B*	*SSAB*	*(a-1)(b-1)*	$\frac{SSAB}{(a-1)(b-1)}$	$\sigma_\varepsilon^2 + \sigma_{\alpha\beta\pi}^2 + \frac{n\sum\alpha\beta_{jk}^2}{(a-1)(b-1)}$	$\frac{MSAB}{MS(A*B*S)}$
*A*B*S* *(A*B Error)*	*SS (A*B*S)*	*(a-1)(b-1) *(n-1)*	$\frac{SS(A*B*S)}{(a-1)(b-1)(n-1)}$	$\sigma_\varepsilon^2 + \sigma_{\alpha\beta\pi}^2$	
Subjects (S)	*SSS*	*(n-1)*	$\frac{SSS}{n-1}$	$\sigma_\varepsilon^2 + ab\sigma_\pi^2$	
Total	*SST*	*N-1*			

- For example, let's consider the test for Factor A

$$H_0 : \mu_{\cdot 1 \cdot} = \mu_{\cdot 2 \cdot} = \ldots = \mu_{\cdot a \cdot}$$
$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$$

$$F_A[(a-1),(a-1)(n-1)] = \frac{MSA}{MS(A*S)} = \frac{\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2 + \frac{nb\sum \alpha_j^2}{a-1}}{\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2}$$

If H_0 is true: $\sum \alpha_j^2 = 0$

Then $F_A = \frac{\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2}{\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2} = 1$

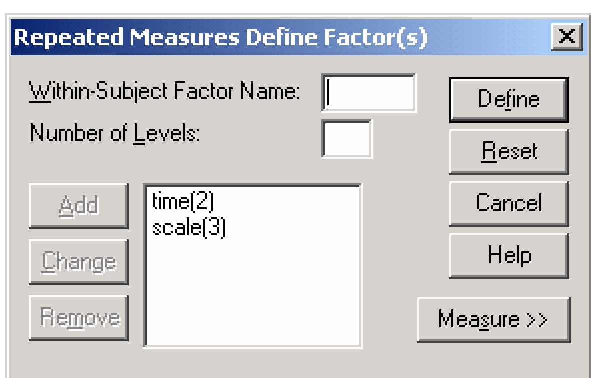
If H_0 is false: $\sum \alpha_j^2 > 0$

Then $F_A = \frac{\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2 + \frac{nb\sum \alpha_j^2}{a-1}}{\sigma_\varepsilon^2 + b\sigma_{\alpha\pi}^2} > 1$

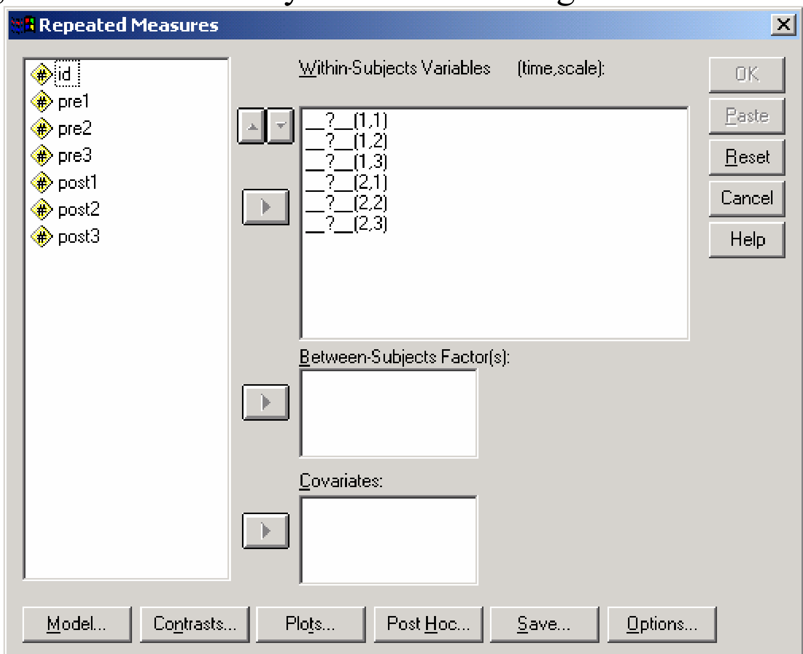
- Note that unlike the one-way within-subjects design, it is not possible to construct an F-test for the effect of subjects.

3. Two-Factor Repeated Measures ANOVA in SPSS

- Let's see how the ANOVA looks in SPSS.
  - We need to enter the within subjects factors correctly. First, we enter the name and number of levels of each repeated factor.

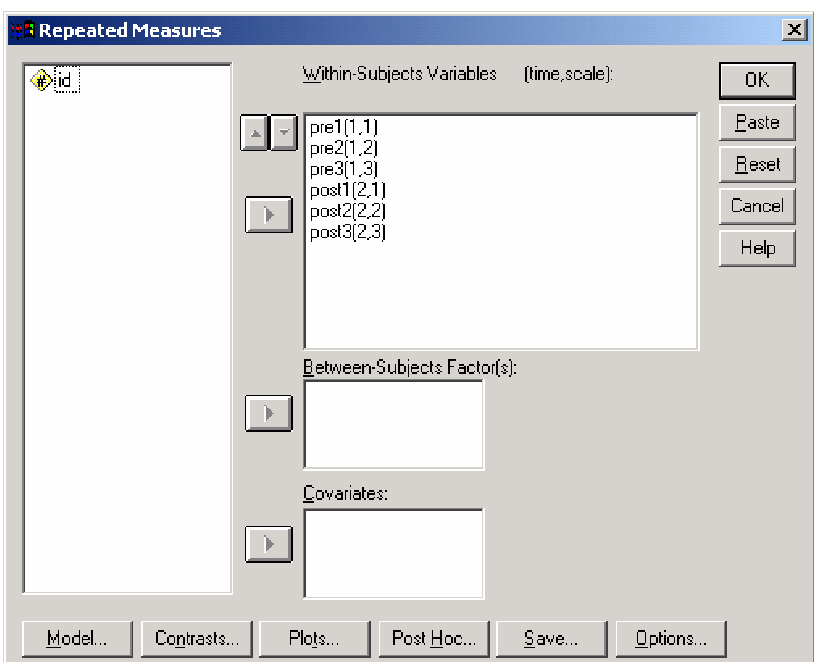


- Next, we need to identify which variables go with which factors:



(1,1) means time 1 and scale 1 $\Rightarrow$ pre1
(1,3) means time 1 and scale 3 $\Rightarrow$ pre3
(2,3) means time 2 and scale 3 $\Rightarrow$ post3

- If you do not identify the factors properly, you will misinterpret your results!



- Or you can enter the following syntax:

```
GLM  pre1 pre2 pre3 post1 post2 post3
  /WSFACTOR = time 2 scale 3
  /PRINT = DESC.
```

  - Time is the first repeated factor with 2 levels
  - Scale is the second repeated factor with 3 levels
  - The order of the variables needs to be

Time 1, Scale 1	pre1
Time 1, Scale 2	pre2
Time 1, Scale 3	pre3
Time 2, Scale 1	post1
Time 2, Scale 2	post2
Time 2, Scale 3	post3

- If we switched the order of the factors, we would need to also switch the order of the variables:

```
GLM  pre1 post1 pre2 post2 pre3 post3
  /WSFACTOR = scale 3 time 2
  /PRINT = DESC.
```

  - This syntax will give us exactly the same output as the syntax above

- Now, we can check the sphericity assumption (presumably, we already checked the normality assumption before starting to run the ANOVA)

Mauchly's Test of Sphericity

Measure: MEASURE_1

	Epsilon		
Within Subjects Effect	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	1.000	1.000	1.000
SCALE	.962	1.000	.500
TIME * SCALE	.904	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

  - We get an epsilon for each effect (main effect and interactions)
  - We can use our same rules of thumb for determining if we have compound symmetry. In this case, we are actually OK!

- Here is the SPSS ANOVA table with the epsilon-adjusted tests removed:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	2856.600	1	2856.600	33.766	.000
Error(TIME)	Sphericity Assumed	761.400	9	84.600		
SCALE	Sphericity Assumed	2899.200	2	1449.600	40.719	.000
Error(SCALE)	Sphericity Assumed	640.800	18	35.600		
TIME * SCALE	Sphericity Assumed	1051.200	2	525.600	45.310	.000
Error(TIME*SCALE)	Sphericity Assumed	208.800	18	11.600		

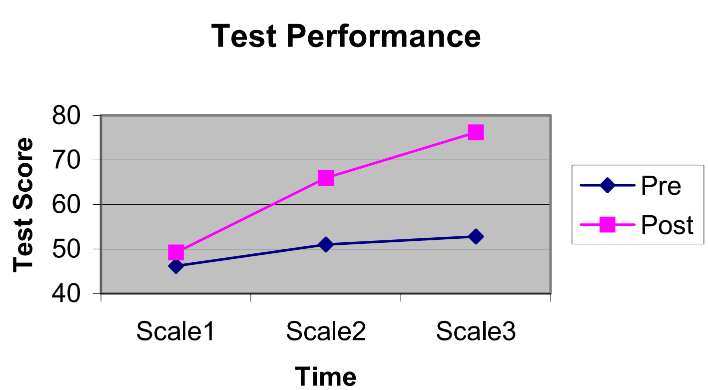
- Each within-subjects factor is immediately followed by its appropriate error term

Main effect of time: $F(1,9) = 33.77, p < .001$
Compares $\bar{X}_{..1} = 50.0$ vs. $\bar{X}_{..2} = 63.8$

Main effect of scale: $F(2,18) = 40.72, p < .001$
Compares $\bar{X}_{.1.} = 47.7$ vs. $\bar{X}_{.2.} = 58.5$ vs. $\bar{X}_{.3.} = 64.5$

Time by scale: $F(2,18) = 45.31, p < .001$
Examines if the time effect is the same for each scale
OR Examines if the scale effect is the same at each time

	Subscale of test (Factor A)			
Time (Factor B)	Subscale 1	Subscale 2	Subscale 3	
Pre-test	$\bar{X}_{.11} = 46.2$	$\bar{X}_{.21} = 51.0$	$\bar{X}_{.31} = 52.8$	$\bar{X}_{..1} = 50.0$
Post-test	$\bar{X}_{.12} = 49.2$	$\bar{X}_{.22} = 66.0$	$\bar{X}_{.32} = 76.2$	$\bar{X}_{..2} = 63.8$
$n = 10$	$\bar{X}_{.1.} = 47.7$	$\bar{X}_{.2.} = 58.5$	$\bar{X}_{.3.} = 64.5$	



- However, the main effect for scale and the time*scale interaction are omnibus tests. We previously stated that we wanted to avoid omnibus tests at all costs for repeated-measures designs

  Technically, in this case we are OK because we have spherical data but it is good practice to avoid omnibus tests for these designs.

4. Contrasts and Effect Sizes
   - The formulae for tests of contrasts are the same formulae we used for one-factor within-subjects designs.

$$t_{observed} = \frac{\hat{\psi}}{\text{standard error}'(\hat{\psi})} = \frac{\sum c_j \overline{X}_{\cdot j}}{\sqrt{MSE' \sum \frac{c_j^2}{n}}}$$

$$SS\hat{\psi} = \frac{\hat{\psi}^2}{\sum \frac{c_j^2}{n}} \qquad F(1, df') = \frac{SS\hat{\psi}}{MSE'}$$

   - The strongly recommended (and the SPSS) approach
     - MSE' will be the contrast-specific error term (with $df = n\text{-}1$).
   - The alternative, use at your own risk approach relies on the data being spherical. If the data are spherical, then we can use the appropriate omnibus error term:
     - For contrasts on the marginal Factor A means, use the omnibus Factor A error term, $MSE' = MS_{A*S}$ (with $df = (a\text{-}1)(n\text{-}1)$).
     - For contrasts on the marginal Factor B means, use the omnibus Factor B error term, $MSE' = MS_{B*S}$ (with $df = (b\text{-}1)(n\text{-}1)$).
     - For contrasts on the A*B cell means, use the omnibus A*B interaction error term, $MSE' = MS_{A*B*S}$ (with $df = (a\text{-}1)(b\text{-}1)(n\text{-}1)$).
     - I recommend that you always use the contrast-specific error term.

- Just as for one-factor within-subjects designs, we have a number of options for effect sizes

  - Partial eta-squared is a measure of percentage of the variance accounted for (in the sample) that can be used for omnibus tests or contrasts:

$$\hat{\eta}^2_{(Effect)} = \frac{SS_{effect}}{SS_{effect} + SS_{ErrorTermForEffect}}$$

$$\hat{\eta}^2_A = \frac{SS_A}{SS_A + SS_{A*S}} \qquad \hat{\eta}^2_B = \frac{SS_B}{SS_B + SS_{B*S}} \qquad \hat{\eta}^2_{A*B} = \frac{SS_{S*B}}{SS_{A*B} + SS_{A*B*S}}$$

$$\hat{\eta}^2_{Contrast} = \frac{SS_{Contrast}}{SS_{Contrast} + SS_{ErrorTermForContrast}}$$

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	2856.600	1	2856.600	33.766	.000
Error(TIME)	Sphericity Assumed	761.400	9	84.600		
SCALE	Sphericity Assumed	2899.200	2	1449.600	40.719	.000
Error(SCALE)	Sphericity Assumed	640.800	18	35.600		
TIME * SCALE	Sphericity Assumed	1051.200	2	525.600	45.310	.000
Error(TIME*SCALE)	Sphericity Assumed	208.800	18	11.600		

$$\hat{\eta}^2_{Time} = \frac{2856.6}{2856.6 + 761.4} = .80 \qquad \hat{\eta}^2_{Scale} = \frac{2899.2}{2899.2 + 640.8} = .82$$

$$\hat{\eta}^2_{Time*Scale} = \frac{1051.2}{1051.2 + 208.8} = .83$$

This formula can be used for omnibus tests and for contrasts.

- For contrasts (except maybe polynomial trends), we can also compute a *d* as a measure of the effect size, just as we did for the paired t-test.

$$\hat{d} = \frac{\overline{\psi}}{\hat{\sigma}_\psi} \qquad \text{but if and only if } \sum |c_i|$$

Where: $\overline{\psi}$ is the average value of the contrast of interest
$\hat{\sigma}_\psi$ is the standard deviation of the contrast values

- For all contrasts, we can also compute an *r* as a measure of the effect size.

$$\hat{r} = \sqrt{\frac{t^2_{Contrast}}{t^2_{Contrast} + df_{contrast}}} = \sqrt{\frac{F_{Contrast}}{F_{Contrast} + df_{contrast}}}$$

- There are four methods we can use in SPSS to test contrasts:
  - Create a new variable reflecting the value of the contrast and conduct a one-sample t-test on this new variable
  - Selecting only the groups of interest and running a contrast or paired t-test on those groups
  - SPSS's brand-name contrasts
  - SPSS's *special* subcommand

- <u>Method 1</u>: Compute a new variable for each contrast, and test if the value of the contrast differs from zero.
  - Let's start by testing three of our hypotheses

    i. Does training improve test scores for subscale 1?

```
compute diff1 = post1-pre1.
T-TEST /TESTVAL=0
  /VARIABLES=diff1.
```

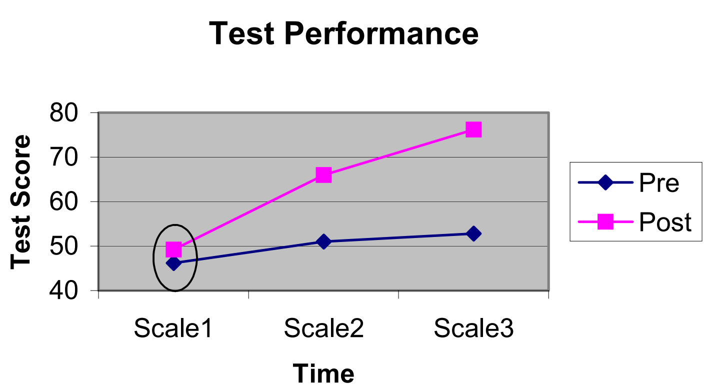
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
DIFF1	10	3.0000	7.61577	2.40832

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
DIFF1	1.246	9	.244	3.0000	-2.4480	8.4480

$$\hat{d} = \frac{\overline{\psi}}{\hat{\sigma}_{\psi}} = \frac{3.0}{7.61577} = 0.39$$



No. The scores on sub-scale 1 do not change significantly between pre- and post-test, $t(9) = 1.25$, $p = .24$, $d = .39$

ii. Does training improve test scores for subscale 2?

```
compute diff2 = post2-pre2.
T-TEST /TESTVAL=0
 /VARIABLES=diff2.
```

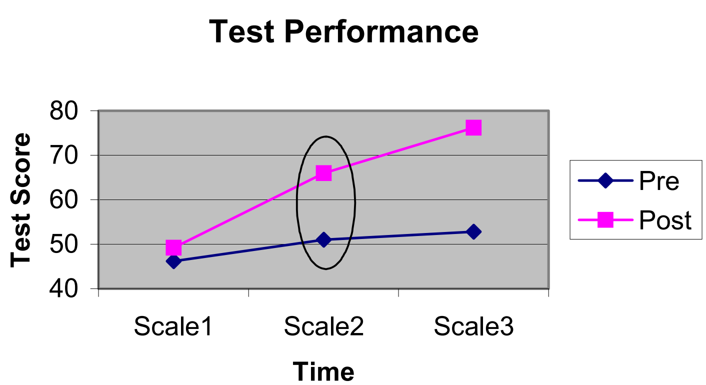
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
diff2	10	15.0000	10.67708	3.37639

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
diff2	4.443	9	.002	15.00000	7.3621	22.6379

$$\hat{d} = \frac{\overline{\psi}}{\hat{\sigma}_{\psi}} = \frac{15}{10.67708} = 1.40$$



Yes. The scores on sub-scale 2 significantly improve between pre- and post-test, $t(9) = 4.44, p < .01, d = 1.40$

iii. Does training improve test scores for subscale 3?

```
compute diff3 = post3-pre3.
T-TEST /TESTVAL=0
 /VARIABLES=diff3.
```

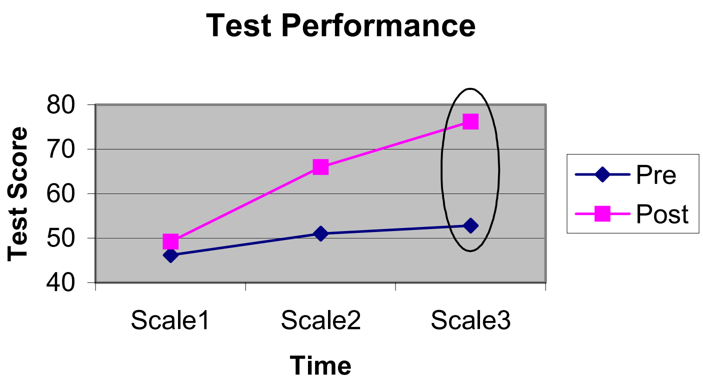
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
diff3	10	23.4000	6.60303	2.08806

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
diff3	11.207	9	.000	23.40000	18.6765	28.1235

$$\hat{d} = \frac{\overline{\psi}}{\hat{\sigma}_{\psi}} = \frac{22.4}{6.60303} = 3.54$$



Yes. The scores on sub-scale 3 significantly improve between pre- and post-test, $t(9) = 11.21$, $p < .01$, $d = 3.54$

- Method 2: Selecting the groups of interest and running paired-test on those groups
    - o Note that this approach would not be recommended for between subjects designs. However, for within-subjects designs, the default is to use a contrast-specific error term. That is, we only use information from the groups that are involved in the contrast to construct the error term. Thus, for within-subjects designs, it is acceptable to select the groups of interest and run a test only on those groups.

```
T-TEST PAIRS = pre1 pre2 pre3  WITH post1 post2 post3 (PAIRED).
```

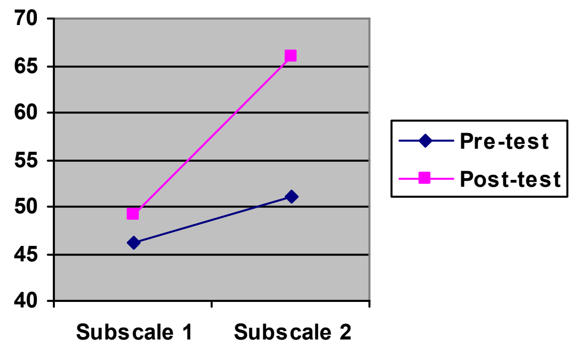
Paired Samples Test

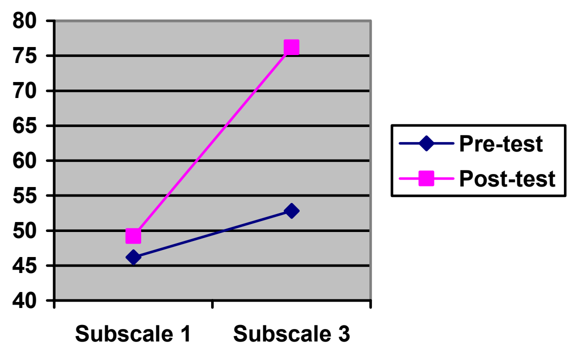
		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	pre1 - post1	-3.00000	7.61577	2.40832	-8.44800	2.44800	-1.246	9	.244
Pair 2	pre2 - post2	-15.00000	10.67708	3.37639	-22.63792	-7.36208	-4.443	9	.002
Pair 3	pre3 - post3	-23.40000	6.60303	2.08806	-28.12352	-18.67648	-11.207	9	.000

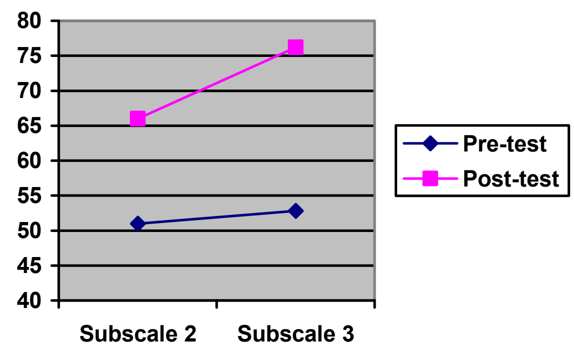
- Scale 1, pre vs post: $t(9) = 1.25, p = .24, d = .39$
- Scale 2, pre vs post: $t(9) = 4.44, p < .01, d = 1.40$
- Scale 3, pre vs post: $t(9) = 11.21, p < .01, d = 3.54$

- These analyses are identical to the previously conducted analyses on the difference scores we computed.

- <u>Method 3: SPSS's brand-name contrasts</u>. SPSS conducts contrasts on the marginal main effect means of the repeated measures factor, using contrast specific error estimates
  - Contrasts can only be specified on the marginal means. Tests on the cell means are obtained by multiplying together main effect contrasts.
  - So far, we have examined the effect of training for each subscale. Now, we would like to test whether:
    - The effect of training on subscale 1 is the same as the effect of training on subscale 2
    - The effect of training on subscale 1 is the same as the effect of training on subscale 3
    - The effect of training on subscale 2 is the same as the effect of training on subscale 3








```
GLM  pre1 pre2 pre3 post1 post2 post3
  /WSFACTOR = time 2 Simple (1) scale 3 simple (1)
  /PRINT = DESC.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	scale	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Level 2 vs. Level 1		1904.400	1	1904.400	33.766	.000
Error(time)	Level 2 vs. Level 1		507.600	9	56.400		
scale		Level 2 vs. Level 1	1166.400	1	1166.400	38.368	.000
		Level 3 vs. Level 1	2822.400	1	2822.400	66.566	.000
Error(scale)		Level 2 vs. Level 1	273.600	9	30.400		
		Level 3 vs. Level 1	381.600	9	42.400		
time * scale	Level 2 vs. Level 1	Level 2 vs. Level 1	1440.000	1	1440.000	45.000	.000
		Level 3 vs. Level 1	4161.600	1	4161.600	83.903	.000
Error(time*scale)	Level 2 vs. Level 1	Level 2 vs. Level 1	288.000	9	32.000		
		Level 3 vs. Level 1	446.400	9	49.600		

- There are only two groups in the time factor, so no matter what we ask SPSS to do, it will give us a pairwise contrast

	Pre-test	Post test
	$\bar{X}_{\cdot\cdot 1} = 50.0$	$\bar{X}_{\cdot\cdot 2} = 63.8$
Coefficients	-1	+1

$$\hat{\eta}^2_{Contrast} = \frac{SS_{Contrast}}{SS_{Contrast} + SS_{ErrorTermForContrast}} = \frac{1904.4}{1904.4 + 507.6} = .79$$

$$F(1,9) = 33.77, p < .01, \eta^2 = .79$$

- For the scale effect there are three groups, so we can ask SPSS to conduct two main effect contrasts:

	Subscale 1	Subscale 2	Subscale 3
	$\bar{X}_{\cdot 1 \cdot} = 47.7$	$\bar{X}_{\cdot 2 \cdot} = 58.5$	$\bar{X}_{\cdot 3 \cdot} = 64.5$
Level 2 vs Level 1	-1	1	0
Level 3 vs Level 1	-1	0	1

$$\hat{\eta}^2_{2vs1} = \frac{1166.4}{1166.4 + 273.6} = .81 \qquad \hat{\eta}^2_{3vs1} = \frac{2822.4}{2822.4 + 381.6} = .88$$

Level 2 vs Level 1: $F(1,9) = 38.37, p < .01, \eta^2 = .81$

Level 3 vs Level 1: $F(1,9) = 66.57, p < .01, \eta^2 = .88$

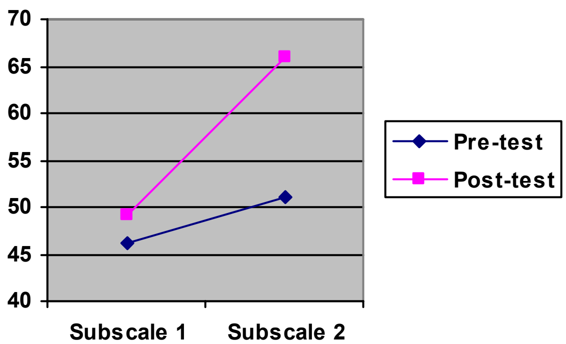
- For the time by scale interaction there are 2 dfs, so SPSS will provide two follow-up tests. SPSS *multiplies* each of the main effect contrasts together to obtain interaction contrasts.
    - These are the tests we are interested in!

Level 2 vs Level 1* Level 2 vs Level 1	Subscale 1	Subscale 2	Subscale 3	
Pre-test				-1
Post-test				+1
	-1	1	0	

Level 2 vs Level 1* Level 2 vs Level 1	Subscale 1	Subscale 2	Subscale 3
Pre-test	1	-1	0
Post-test	-1	1	0

- This contrast tests whether the effect of training on Subscale 1 is the same as the effect of training for Subscale 2 (an interaction!)

$$H_0 : \mu_{Pre1} - \mu_{Post1} = \mu_{Pre2} - \mu_{Post2}$$
$$H_1 : \mu_{Pre1} - \mu_{Post1} \neq \mu_{Pre2} - \mu_{Post2}$$



$$\hat{\eta}^2_{Contrast} = \frac{1440}{1440 + 288} = .83$$

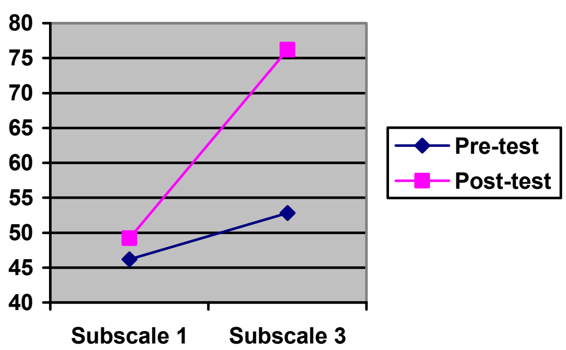
$$F(1,9) = 45.00, p < .01, \eta^2 = .98$$

Level 2 vs Level 1* Level 3 vs Level 1	Subscale 1	Subscale 2	Subscale 3	
Pre-test				-1
Post-test				+1
	-1	0	1	

Level 2 vs Level 1* Level 3 vs Level 1	Subscale 1	Subscale 2	Subscale 3
Pre-test	-1	0	1
Post-test	1	0	-1

- This contrast tests whether the effect of training on Subscale 1 is the same as the effect of training for Subscale 3

$$H_0 : \mu_{Pre1} - \mu_{Post1} = \mu_{Pre3} - \mu_{Post3}$$
$$H_1 : \mu_{Pre1} - \mu_{Post1} \neq \mu_{Pre3} - \mu_{Post3}$$



$$\hat{\eta}^2_{Contrast} = \frac{4161.6}{4161.6 + 446} = .90$$

$$F(1,9) = 83.90, p < .01, \eta^2 = .90$$

- To test whether the effect of training on Subscale 2 is the same as the effect of training for Subscale 3, we need to run a new command.

```
GLM  pre1 pre2 pre3 post1 post2 post3
  /WSFACTOR = time 2 Simple (1) scale 3 simple (2)
  /PRINT = DESC.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

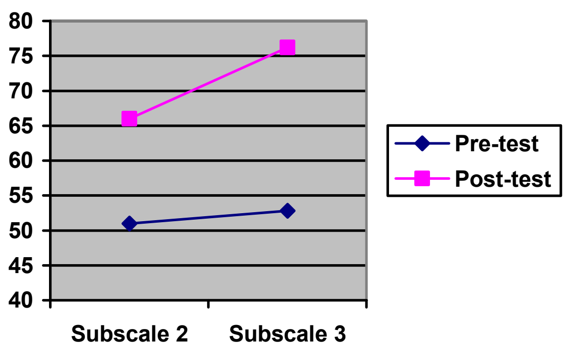
Source	time	scale	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Level 2 vs. Level 1		1904.400	1	1904.400	33.766	.000
Error(time)	Level 2 vs. Level 1		507.600	9	56.400		
scale		Level 1 vs. Level 2	1166.400	1	1166.400	38.368	.000
		Level 3 vs. Level 2	360.000	1	360.000	10.588	.010
Error(scale)		Level 1 vs. Level 2	273.600	9	30.400		
		Level 3 vs. Level 2	306.000	9	34.000		
time * scale	Level 2 vs. Level 1	Level 1 vs. Level 2	1440.000	1	1440.000	45.000	.000
		Level 3 vs. Level 2	705.600	1	705.600	12.250	.007
Error(time*scale)	Level 2 vs. Level 1	Level 1 vs. Level 2	288.000	9	32.000		
		Level 3 vs. Level 2	518.400	9	57.600		

Level 2 vs Level 1* Level 3 vs Level 2	Subscale 1	Subscale 2	Subscale 3	
Pre-test				-1
Post-test				+1
	0	-1	1	

Level 2 vs Level 1* Level 2 vs Level 1	Subscale 1	Subscale 2	Subscale 3
Pre-test	0	1	-1
Post-test	0	-1	1

$$H_0 : \mu_{Pre1} - \mu_{Post1} = \mu_{Pre2} - \mu_{Post2}$$

$$H_1 : \mu_{Pre1} - \mu_{Post1} \neq \mu_{Pre2} - \mu_{Post2}$$



$$\hat{\eta}^2_{Contrast} = \frac{705.6}{705.6 + 518.4} = .58$$

$$F(1,9) = 12.25, p < .01, \eta^2 = .58$$

- Most cell mean contrasts that you would find of interest can be obtained by multiplying two main effect contrasts (with the exception of simple effect contrasts).

- Method 4: SPSS's *special* subcommand.

  - Let's return to our original questions:
    i. Overall, does the training improve test scores?
    ii. Does training improve test scores for subscale 1?
    iii. Does training improve test scores for subscale 2?
    iv. Does training improve test scores for subscale 3?
    v. Overall, is there a difference in performance on the three sub-scales?
    This final hypothesis is an omnibus hypothesis, so we will only consider the first four

  - To use the *special* subcommand:
    - Treat your design as a one-factor repeated-measures design
    - Enter the appropriate coefficients in the special command

```
GLM  pre1 pre2 pre3 post1 post2 post3
  /WSFACTOR = factor 6 special (  1   1  1  1  1  1
                                 -1 -1 -1  1  1  1
                                 -1  0  0  1  0  0
                                  0 -1  0  0  1  0
                                  0  0 -1  0  0  1
                                 -1 -1 2   1  1 -2).
```

    - Note that contrast 6 is of no interest to us, but we must enter five contrasts after the row of ones.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	FACTOR	Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR	L1	17139.600	1	17139.600	33.766	.000
	L2	90.000	1	90.000	1.552	.244
	L3	2250.000	1	2250.000	19.737	.002
	L4	5475.600	1	5475.600	125.587	.000
	L5	8294.400	1	8294.400	45.474	.000
Error(FACTOR)	L1	4568.400	9	507.600		
	L2	522.000	9	58.000		
	L3	1026.000	9	114.000		
	L4	392.400	9	43.600		
	L5	1641.600	9	182.400		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	FACTOR	Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR	L1	17139.600	1	17139.600	33.766	.000
	L2	90.000	1	90.000	1.552	.244
	L3	2250.000	1	2250.000	19.737	.002
	L4	5475.600	1	5475.600	125.587	.000
	L5	8294.400	1	8294.400	45.474	.000
Error(FACTOR)	L1	4568.400	9	507.600		
	L2	522.000	9	58.000		
	L3	1026.000	9	114.000		
	L4	392.400	9	43.600		
	L5	1641.600	9	182.400		

i. Overall, does the training improve test scores? (L1)

Yes. Averaging across the subscales, post-test scores are higher than pre-test scores, $F(1,9) = 33.77, p < .01$

ii. Does training improve test scores for subscale 1? (L2)

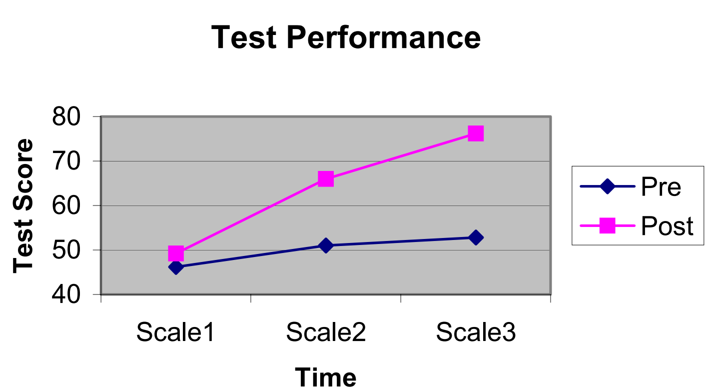
No. The scores on sub-scale 1 do not change significantly between pre- and post-test, $F(1,9) = 1.55, p = .24$

iii. Does training improve test scores for subscale 2? (L3)

Yes. The scores on sub-scale 2 are higher at post-test than at pre-test, $F(1,9) = 19.74, p < .01$

iv. Does training improve test scores for subscale 3? (L4)

Yes. The scores on sub-scale 3 are higher at post-test than at pre-test, $F(1,9) = 125.59, p < .01$



- Depending on how these contrasts are conducted, you may need to adjust their p-values

  - The recommended approach is to forgo the omnibus tests, and to conduct three of fewer planned contrasts. If you take this approach, no correction is necessary
    - If you have a large number of planned tests, you may need to apply a Bonferroni correction.

  - The alternative is to conduct the tests for main effects and interactions, and then conduct the contrasts as follow-up tests. Now, the contrasts are post-hoc tests. If they are pairwise, then you need to use the Tukey procedure; if they are complex, you need to use the Scheffé procedure to adjust the p-values.

    - To use Tukey's HSD, compute $q(1\text{-}\alpha, a, \nu)$

      Where α = Familywise error rate
      a = Number of repeated-measures in the family
      ν = *df*(error)

      ⇒ For single-*df* tests, *df*(error) should be $(n-1)$, the *df* associated with the contrast-specific error estimate.

      ⇒ To determine significance at the $(1\text{-}\alpha)$ level,

      $$\text{Compare } t_{observed} \text{ to } \frac{q_{crit}}{\sqrt{2}} \qquad \text{or} \qquad F_{observed} \text{ to } \frac{(q_{crit})^2}{2}$$

    - To use the Scheffé correction, compute $F_{Crit} = (r)F_{\alpha=.05;r,\nu}$

      Where α = Familywise error rate
      r = Degrees of freedom associated with the family
      ν = *df*(error), $(n-1)$ for a contrast-specific error estimate.

      Compare $F_{observed}$ to F_{crit}

- As an example, let's consider the following question as a post-hoc test
Does training improve test scores for subscale 2?

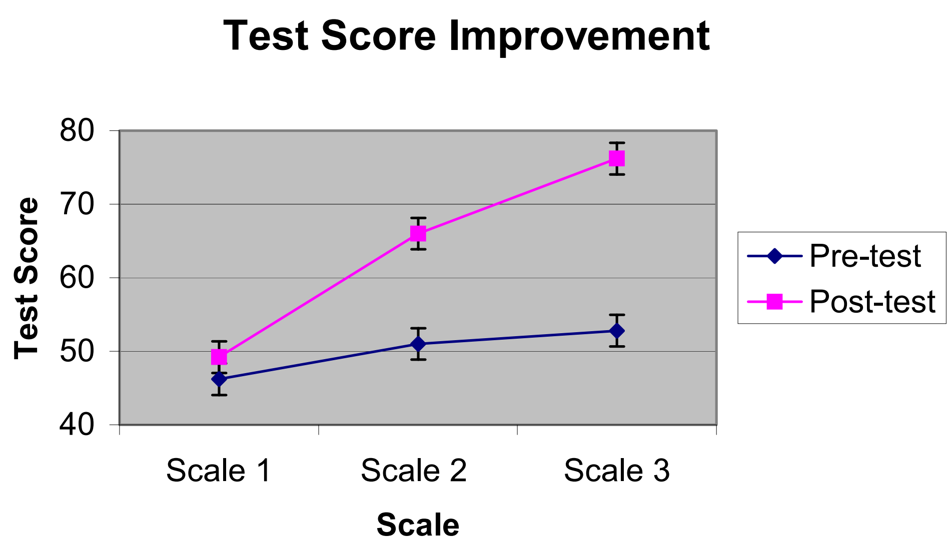
  When we tested this contrast, we found the test statistic to be:
  $F(1,9) = 19.74$

  We need to discard the p-value and compute a Tukey adjusted critical value.

  $q(1-\alpha,a,\nu)$ with $\alpha = .05$ $a = 6$ $\nu = 9$

  $q(.95,6,9) = 5.02$ $F_{crit} = \frac{(5.02)^2}{2} = 12.60$

  And so we can report the test to be significant at the $\alpha = .05$ level:
  $F(1,9) = 19.74, p < .05$



- We should have a separate variance estimate of the error bars for each cell

$$StdError = \sqrt{\frac{s_{jk}^2}{n}}$$

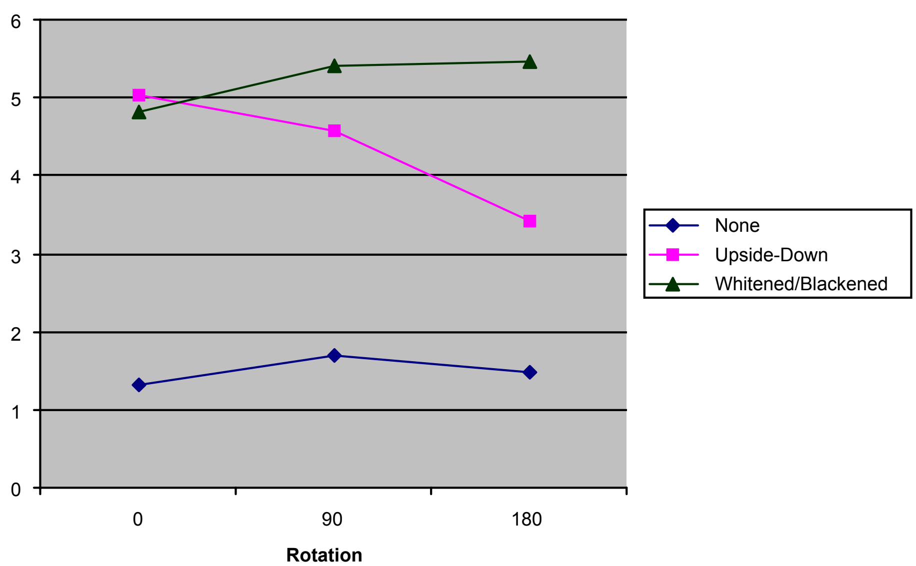
- SPSS and EXCEL are not good for plotting separate error bars for each cell. The best you can do is to compute a common standard error based on the error term for the highest order interaction. This error bar is misleading (because you did not actually use it in your analyses). If you plan to publish using repeated measures data, get better graphical software.

5. An Example

- Consider an experiment on facial perception. Faces vary on two dimensions: orientation (upright, 90° rotation, and 180° rotation) and distortion (none, eyes & mouth upside down, eyes whitened and teeth blackened). Participants rate each of the six resulting faces on how “bizarre” each face looks on a 7-point scale, with higher numbers indicating more bizarreness. The following data were obtained:

	No Distortion			Upside-Down			Whitened and Blackened		
Subject	0°	90°	180°	0°	90°	180°	0°	90°	180°
1	1.18	2.40	2.48	4.76	4.93	3.13	5.56	4.93	5.21
2	1.14	1.55	1.25	4.81	4.73	3.89	4.85	5.43	4.89
3	1.02	1.25	1.30	4.98	3.85	3.05	4.28	5.64	6.49
4	1.05	1.63	1.84	4.91	5.21	2.95	5.13	5.52	5.69
5	1.81	1.65	1.01	5.01	4.18	3.51	4.90	5.18	5.52
6	1.69	1.67	1.04	5.65	4.56	3.94	4.12	5.76	4.99

Distortion	0°	90°	180°	
None	1.32	1.69	1.49	1.50
Upside-Down	5.02	4.58	3.41	4.34
Whitened/Blackened	4.81	5.41	5.47	5.23
	3.72	3.89	3.46	3.69



- We can solve for the (fixed) model parameters

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \pi_\sigma + (\alpha\beta)_{jk} + (\alpha\pi)_\sigma + (\beta\pi)_\sigma + (\alpha\beta\pi)_\sigma$$

μ The overall mean of the scores

$$\hat{\mu} = 3.69$$

α_j The effect of being in level *j* of Orientation

$$\hat{\alpha}_j = \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}$$

$$\hat{\alpha}_1 = 3.72 - 3.69 = 0.03$$
$$\hat{\alpha}_2 = 3.89 - 3.69 = 0.20$$
$$\hat{\alpha}_3 = 3.46 - 3.69 = -0.23$$

β_k The effect of being in level *k* of Distortion

$$\hat{\beta}_k = \bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdots}$$

$$\hat{\beta}_1 = 1.50 - 3.69 = -2.19$$
$$\hat{\beta}_2 = 4.34 - 3.69 = 0.65$$
$$\hat{\beta}_3 = 5.23 - 3.69 = 1.54$$

$(\alpha\beta)_{jk}$ The effect of being in level *j* of Orientation and level *k* of Distortion

$$(\hat{\alpha\beta})_{jk} = \bar{Y}_{\cdot jk} - \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{\cdots}$$

$$(\hat{\alpha\beta})_{11} = 1.32 - 3.72 - 1.50 + 3.69 = -0.21$$
$$(\hat{\alpha\beta})_{12} = 5.02 - 3.72 - 4.34 + 3.69 = 0.65$$
$$(\hat{\alpha\beta})_{13} = 4.81 - 3.72 - 5.23 + 3.69 = -0.45$$

.

.

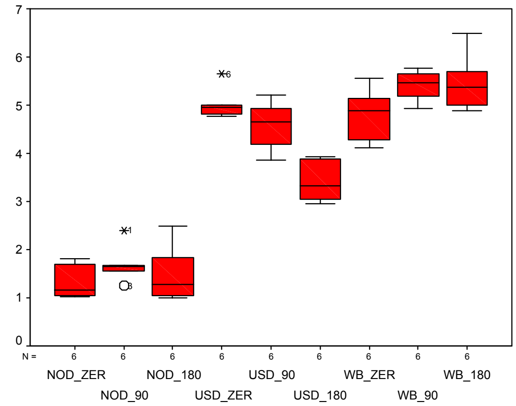
$$(\hat{\alpha\beta})_{33} = 5.47 - 3.46 - 5.23 + 3.69 = 0.47$$

- First, we need to check assumptions
  - This design is a two-factor repeated measures design
  - Participants must be independent and randomly selected from the population
  - Normality/ symmetry of difference scores (but in practice normality within each condition)


```
EXAMINE
  VARIABLES=nod_zer nod_90 nod_180 usd_zer usd_90 usd_180 wb_zer wb_90 wb_180
  /PLOT BOXPLOT STEMLEAF NPPLOT
  /COMPARE VARIABLES.
```

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
NOD_ZER	.807	6	.068
NOD_90	.831	6	.109
NOD_180	.846	6	.147
USD_ZER	.771	6	.032
USD_90	.980	6	.950
USD_180	.872	6	.235
WB_ZER	.954	6	.776
WB_90	.955	6	.779
WB_180	.909	6	.432



- Sphericity

Mauchly's Test of Sphericity

Measure: MEASURE_1

	Epsilon		
Within Subjects Effect	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
ORIENTAT	.963	1.000	.500
DISTORT	.932	1.000	.500
ORIENTAT * DISTORT	.462	.720	.250

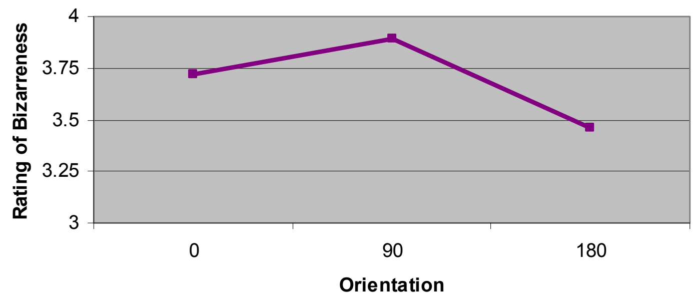
  - Sphericity is not satisfied.
  - We must either conduct only non-parametric tests or contrasts

- Contrasts of interest (all post-hoc):
  - Are there linear (and quadratic) trends in the marginal orientation means?
  - Are there linear (and quadratic) trends in the orientation means within each level of distortion?
  - Are the linear (and quadratic) trends in the orientation means within each level of distortion different from each other?

- First, let's test for linear and quadratic trends in orientation.

Distortion		0°	90°	180°
None				
Upside-Down				
Whitened/Blackened				
	Linear	-1	0	+1
	Quadratic	+1	-2	+1

Marginal Orientation Means



- Method 1: Use SPSS's built-in contrasts to test main effect contrasts

```
GLM nod_zer usd_zer wb_zer  nod_90 usd_90 wb_90
        nod_180 usd_180 wb_180
  /WSFACTOR = orientat 3 polynomial distort 3
  /PRINT = DESCRIPTIVE.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	orientat	distort	Type III Sum of Squares	df	Mean Square	F	Sig.
orientat	Linear		.606	1	.606	5.772	.061
	Quadratic		1.143	1	1.143	13.537	.014
Error(orientat)	Linear		.525	5	.105		
	Quadratic		.422	5	.084		

$$\eta^2_{Linear} = \frac{SS_{Linear}}{SS_{Linear} + SS_{ErrorTermForLinear}} = \frac{.606}{.606 + .525} = .54$$

$$\eta^2_{Quadratic} = \frac{1.143}{1.143 + .422} = .73$$

These are complex, post-hoc tests, so a Scheffé correction is required.

$$F_{Crit} = 2 * F(.05,2,5) = 2 * 5.78 = 11.57$$

$$F_{Crit} = 2 * F(.10,2,5) = 2 * 3.78 = 7.56$$

Linear trend: $F(1,5) = 5.78, ns, \eta^2 = .54$

Quadratic trend: $F(1,5) = 13.54, p < .05, \eta^2 = .73$

- Method 2: Compute and test the contrasts manually

```
compute lin_ori = -1*nod_zer + 0*nod_90 +1*nod_180
                  - 1*usd_zer + 0*usd_90+ 1* usd_180
                  - 1*wb_zer + 0*wb_90+1* wb_180.
compute quad_ori = 1*nod_zer -2*nod_90 +1*nod_180
                   + 1*usd_zer -2*usd_90+1* usd_180
                   + 1*wb_zer -2*wb_90+1* wb_180.

T-TEST  /TESTVAL=0
  /VARIABLES=lin_ori quad_ori.
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
LIN_ORI	6	-.7783	.79356	.32397
QUAD_ORI	6	-1.8517	1.23274	.50327

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
LIN_ORI	-2.402	5	.061	-.7783	-1.6111	.0545
QUAD_ORI	-3.679	5	.014	-1.8517	-3.1454	-.5580

These results exactly match the results obtained from using SPSS's built-in main effect contrasts.

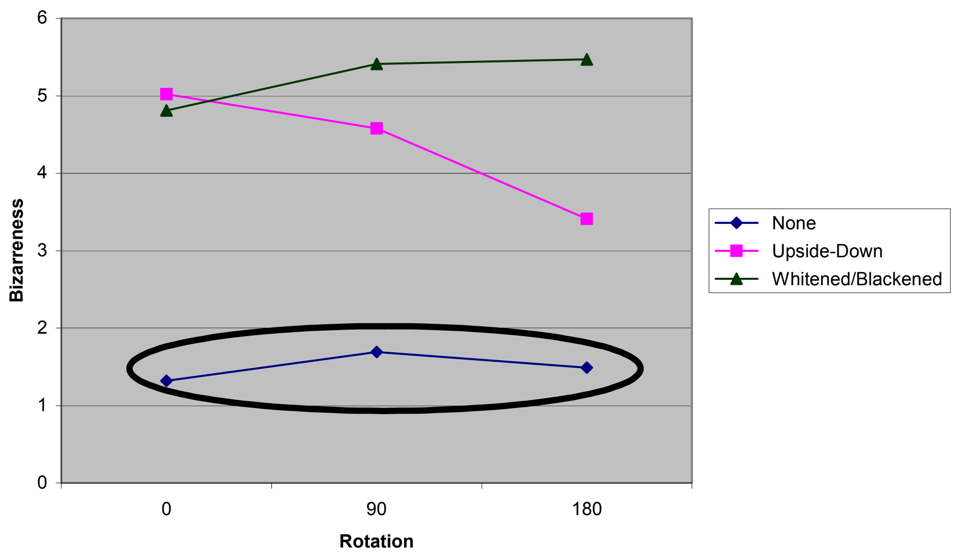
$$r_{linear} = \sqrt{\frac{t^2_{Contrast}}{t^2_{Contrast} + df_{contrast}}} = \sqrt{\frac{2.402^2}{2.404^2 + 5}} = .73 \qquad r_{quad} = \sqrt{\frac{3.679^2}{3.679^2 + 5}} = .85$$

Linear trend: $F(1,5) = 5.78, ns, r = .73$

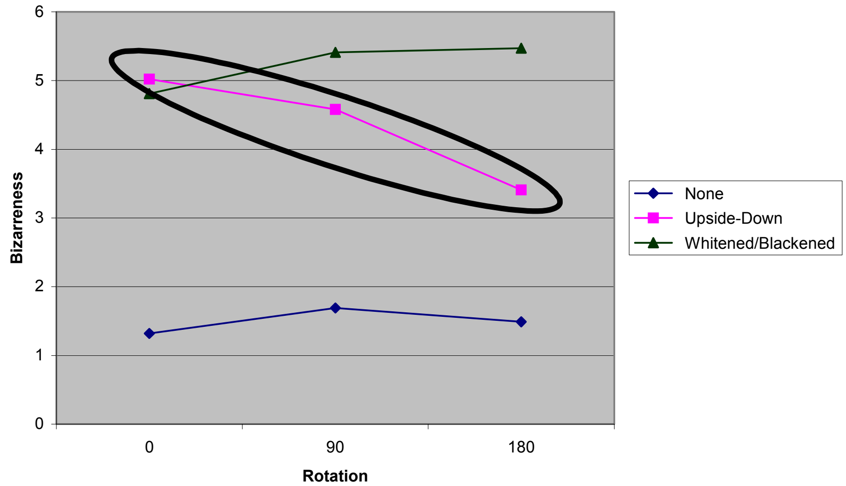
Quadratic trend: $F(1,5) = 13.54, p < .05, r = .85$

- Second, let's test for linear and quadratic trends in orientation *within each level of distortion*.

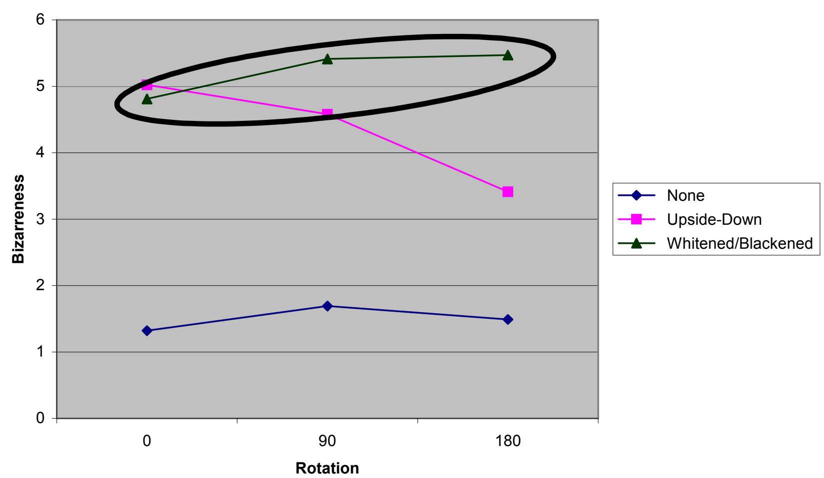
Distortion		0°	90°	180°
None	Linear	-1	0	+1
	Quadratic	+1	-2	+1
Upside-Down				
Whitened/Blackened				



Distortion		0°	90°	180°
None				
Upside-Down	Linear	-1	0	+1
	Quadratic	+1	-2	+1
Whitened/Blackened				



Distortion		0°	90°	180°
None				
Upside-Down				
Whitened/Blackened	Linear	-1	0	+1
	Quadratic	+1	-2	+1



- Method 1: Compute and test the contrasts manually

```
Compute lin_nod = -nod_zer  + 0*nod_90 + nod_180.
Compute quad_nod = nod_zer  - 2*nod_90 + nod_180.

Compute lin_usd = -usd_zer  + 0*usd_90 + usd_180.
Compute quad_usd = usd_zer  - 2*usd_90 + usd_180.

Compute lin_wb = -wb_zer  + 0*wb_90 + wb_180.
Compute quad_wb = wb_zer  - 2*wb_90 + wb_180.

T-TEST  /TESTVAL=0
  /VARIABLES=lin_nod quad_nod lin_usd quad_usd lin_wb quad_wb .
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
LIN_NOD	6	.1717	.81121	.33117
QUAD_NOD	6	-.5817	.33030	.13484
LIN_USD	6	-1.6083	.38039	.15529
QUAD_USD	6	-.7217	1.28395	.52417
LIN_WB	6	.6583	.87894	.35883
QUAD_WB	6	-.5483	1.13125	.46183

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
LIN_NOD	.518	5	.626	.1717	-.6796	1.0230
QUAD_NOD	-4.314	5	.008	-.5817	-.9283	-.2350
LIN_USD	-10.357	5	.000	-1.6083	-2.0075	-1.2091
QUAD_USD	-1.377	5	.227	-.7217	-2.0691	.6258
LIN_WB	1.835	5	.126	.6583	-.2641	1.5807
QUAD_WB	-1.187	5	.288	-.5483	-1.7355	.6388

$$r_{LinearNoDistort} = \sqrt{\frac{t^2_{Contrast}}{t^2_{Contrast} + df_{contrast}}} = \sqrt{\frac{0.518^2}{0.518^2 + 5}} = .23 \qquad r_{QuadNoDistort} = \sqrt{\frac{4.314^2}{4.314^2 + 5}} = .89$$

$$r_{LinearUpsideDown} = \sqrt{\frac{10.357^2}{10.357^2 + 5}} = .98 \qquad r_{QuadUpsideDown} = \sqrt{\frac{1.377^2}{1.377^2 + 5}} = .28$$

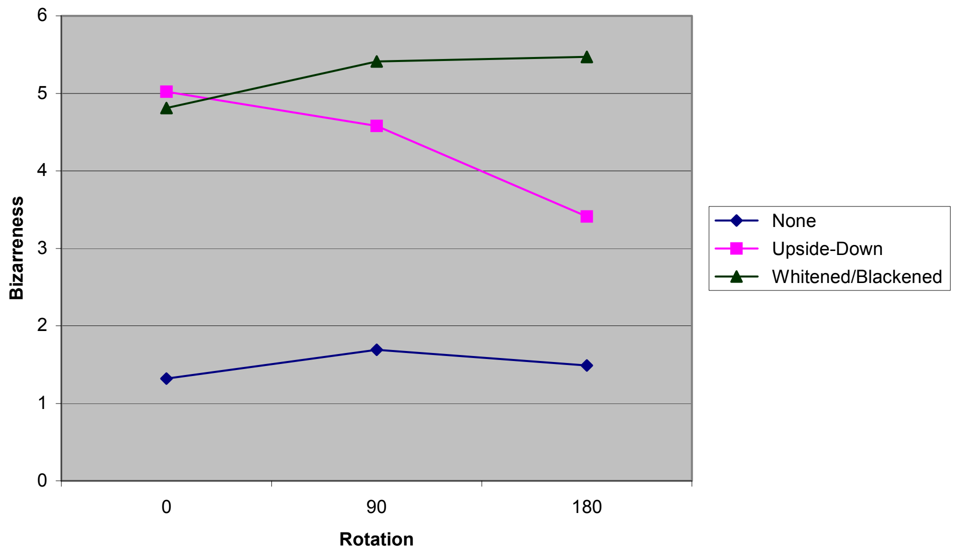
$$r_{LinearWhitenedBlacked} = \sqrt{\frac{1.835^2}{1.835^2 + 5}} = .63 \qquad r_{QuadWhitenedBlackened} = \sqrt{\frac{1.187^2}{1.187^2 + 5}} = .47$$

- These are complex, post-hoc tests, so a Scheffé correction is required.

$$F_{Crit} = 4 * F(.05,4,5) = 4 * 5.19 = 20.76 \qquad F_{Crit} = 4 * F(.10,4,5) = 4 * 3.52 = 14.08$$

$$t_{crit} = \sqrt{F_{crit}} = \sqrt{20.76} = 4.56 \qquad t_{crit} = \sqrt{F_{crit}} = \sqrt{14.08} = 3.75$$

- For faces that were not distorted: There is a marginally significant quadratic trend such that sideways faces are rated to be most bizarre and deviations from 90° are less bizarre, $t(5) = -4.31, p < .10, r = .89$.
- For faces with upside-down mouths and faces: There is a linear trend in ratings of bizarreness such that as orientation increases, bizarreness decreases, $t(5) = -10.36, p < .05, r = .98$
- For faces with whitened eyes and blacked teeth: Ratings of bizarreness are unaffected by orientation, $rs \leq .63$.



- Method 2: Selecting only the groups of interest and running a contrast on those groups

  - No Distortion: Linear and quadratic trends

```
GLM nod_zer nod_90 nod_180
  /WSFACTOR = orientat 3 Polynomial.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	orientat	Type III Sum of Squares	df	Mean Square	F	Sig.
orientat	Linear	.088	1	.088	.269	.626
	Quadratic	.338	1	.338	18.608	.008
Error(orientat)	Linear	1.645	5	.329		
	Quadratic	.091	5	.018		

$$\eta^2_{LinearNoDistortion} = \frac{.08841}{.08841 + 1.645} = .05 \qquad \eta^2_{QuadraticNoDistortion} = \frac{.338}{.338 + .09091} = .79$$

  - Upside-down eyes and mouths: Linear and quadratic trends

```
GLM usd_zer usd_90 usd_180
  /WSFACTOR = orientat 3 Polynomial.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	orientat	Type III Sum of Squares	df	Mean Square	F	Sig.
orientat	Linear	7.760	1	7.760	107.262	.000
	Quadratic	.521	1	.521	1.896	.227
Error(orientat)	Linear	.362	5	.072		
	Quadratic	1.374	5	.275		

$$\eta^2_{LinearUpsideDown} = \frac{7.76}{7.76 + .362} = .96 \qquad \eta^2_{QuadraticUpsideDown} = \frac{.521}{.521 + 1.374} = .27$$

  - Whitened eyes and blackened mouths: Linear and quadratic trends

```
GLM wb_zer wb_90 wb_180
  /WSFACTOR = orientat 3 Polynomial.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	orientat	Type III Sum of Squares	df	Mean Square	F	Sig.
orientat	Linear	1.300	1	1.300	3.366	.126
	Quadratic	.301	1	.301	1.410	.288
Error(orientat)	Linear	1.931	5	.386		
	Quadratic	1.066	5	.213		

$$\eta^2_{LinearWhitenedBlackened} = \frac{1.300}{1.300 + 1.931} = .40 \qquad \eta^2_{QuadraticWhitenedBlackened} = \frac{.301}{.301 + 1.066} = .22$$

- Finally, we'd like to test for differences between these trends
  - For example, does the linear trend for no distortion differ from the linear trend for upside-down eyes and mouth?

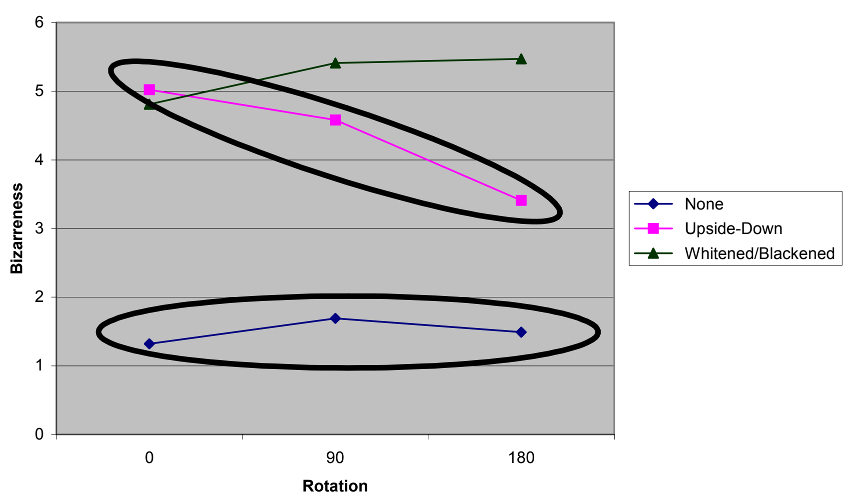
  - Method 1: Compute and test the contrasts manually
    - Linear (No distortion) vs. Linear (Up-side down)

Distortion		0°	90°	180°
None	Linear	-1	0	+1
Upside-Down	Opposite Linear	+1	0	-1
Whitened/Blackened				

- Quadratic (No distortion) vs. Quadratic (Up-side down)

Distortion		0°	90°	180°
None	Quadratic	+1	-2	+1
Upside-Down	Opposite Quadratic	-1	+2	-1
Whitened/Blackened				

- The syntax and output for this method is not included here.



- Approach 2: Use SPSS's built-in contrasts
  - Notice that differences in trends can be obtained by examining the interaction between polynomial contrasts on orientation and simple contrasts on distortion

  - Linear (orientation) by None vs.Upside-down (distortion)

Distortion	0°	90°	180°	
None	-1	0	+1	+1
Upside-Down	+1	0	-1	-1
Whitened/Blackened				
	-1	0	+1	

  - Quadratic (orientation) by None vs.Upside-down (distortion)

Distortion	0°	90°	180°	
None	+1	-2	+1	+1
Upside-Down	-1	+2	-1	-1
Whitened/Blackened				
	+1	-2	+1	

```
GLM nod_zer usd_zer wb_zer nod_90 usd_90 wb_90 nod_180 usd_180 wb_180
  /WSFACTOR = orientat 3 Polynomial distort 3 Simple(1)
  /PRINT = DESCRIPTIVE.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	ORIENTAT	DISTORT	Type III Sum of Squares	df	Mean Square	F	Sig.
ORIENTAT	Linear		.202	1	.202	5.772	.061
	Quadratic		.381	1	.381	13.537	.014
Error(ORIENTAT)	Linear		.175	5	3.499E-02		
	Quadratic		.141	5	2.814E-02		
DISTORT		Level 2 vs. Level 1	145.010	1	145.010	382.927	.000
		Level 3 vs. Level 1	250.358	1	250.358	621.185	.000
Error(DISTORT)		Level 2 vs. Level 1	1.893	5	.379		
		Level 3 vs. Level 1	2.015	5	.403		
ORIENTAT * DISTORT	Linear	Level 2 vs. Level 1	9.505	1	9.505	20.243	.006
		Level 3 vs. Level 1	.711	1	.711	.755	.425
	Quadratic	Level 2 vs. Level 1	1.960E-02	1	1.960E-02	.082	.787
		Level 3 vs. Level 1	1.111E-03	1	1.111E-03	.004	.951
Error(ORIENTAT*DISTORT)	Linear	Level 2 vs. Level 1	2.348	5	.470		
		Level 3 vs. Level 1	4.705	5	.941		
	Quadratic	Level 2 vs. Level 1	1.201	5	.240		
		Level 3 vs. Level 1	1.338	5	.268		

$F_{Crit} = 4*F(.05,4,5) = 4*5.19 = 20.76$ $\quad F_{Crit} = 4*F(.10,4,5) = 4*3.52 = 14.08$

$t_{crit} = \sqrt{F_{crit}} = \sqrt{20.76} = 4.56$ $\quad t_{crit} = \sqrt{F_{crit}} = \sqrt{14.08} = 3.72$

- Difference in trends between no distortion and upside-down eyes and mouth:

$$\eta^2_{LinearDiff} = \frac{9.505}{9.505 + 2.348} = .80 \qquad \eta^2_{QuadraticDiff} = \frac{.0196}{.0196 + 1.201} = .02$$

Difference in linear trends: $F(1,5) = 20.24, p < .10, \eta^2 = .80$

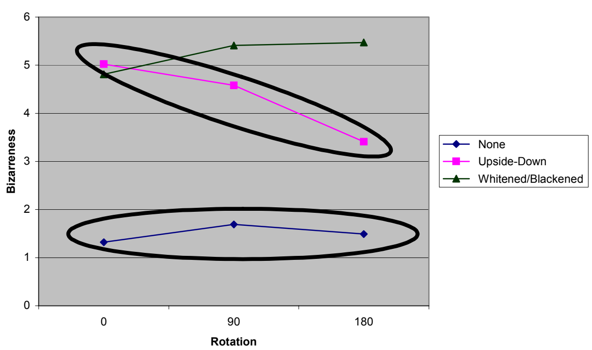
Difference in quadratic trends: $F(1,5) = 0.08, ns, \eta^2 = .02$

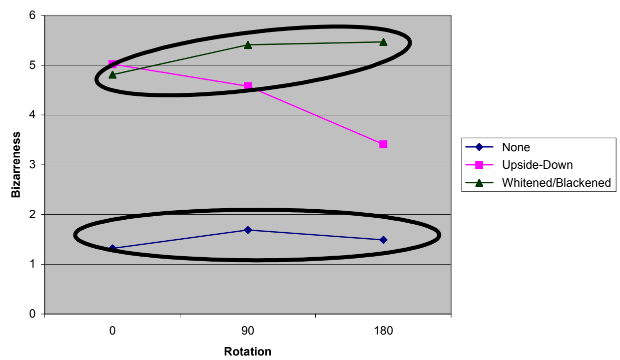
- Difference in trends between no distortion and whitened eyes and black teeth:

$$\eta^2_{LinearDiff} = \frac{.711}{.711 + 4.705} = .13 \qquad \eta^2_{QuadraticDiff} = \frac{.00111}{.00111 + 1.338} = .0008$$

Difference in linear trends: $F(1,5) = 0.77, ns, \eta^2 = .13$

Difference in quadratic trends: $F(1,5) = 0.01, ns, \eta^2 < .01$



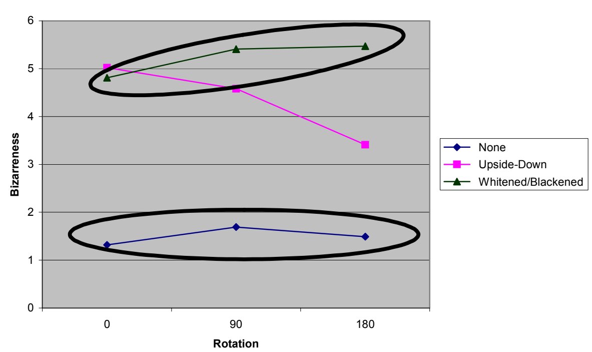


- To obtain differences in trends between upside-down eyes and mouth and whitened eyes and black teeth, we need to run another analysis:

```
GLM nod_zer usd_zer wb_zer nod_90 usd_90 wb_90 nod_180 usd_180 wb_180
  /WSFACTOR = orientat 3 Polynomial distort 3 Simple(2)
  /PRINT = DESCRIPTIVE.
```

Difference in linear trends: $F(1,5) = 24.18, p < .05, \eta^2 = .83$

Difference in quadratic trends: $F(1,5) = 0.04, ns, \eta^2 < .01$



- All of the previous analysis examined the effect of orientation or the effect of orientation within each level of distortion.
- Alternatively, we may be interested in the effect of distortion or the effect of distortion within each level of orientation. The following analysis are a few examples of these types of contrasts

Distortion	0°	90°	180°	
None	1.32	1.69	1.49	1.50
Upside-Down	5.02	4.58	3.41	4.34
Whitened/Blackened	4.81	5.41	5.47	5.23
	3.72	3.89	3.46	3.69

- Within each level of orientation, let's compare the distorted faces to the non-distorted control.

```
compute  comp1 = usd_zer - nod_zer.
compute  comp2 = wb_zer - nod_zer.

compute  comp3 = usd_90 - nod_90.
compute  comp4 = wb_90 - nod_90.

compute  comp5 = usd_180 - nod_180.
compute  comp6 = wb_180 - nod_180.

T-TEST  /TESTVAL=0
 /VARIABLES=comp1 to comp6.
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
COMP1	6	3.7050	.29187	.11916
COMP2	6	3.4917	.71065	.29012
COMP3	6	2.8850	.42505	.17353
COMP4	6	3.7183	.64691	.26410
COMP5	6	1.9250	.90697	.37027
COMP6	6	3.9783	.82956	.33867

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
COMP1	31.093	5	.000	3.7050	3.3987	4.0113
COMP2	12.035	5	.000	3.4917	2.7459	4.2374
COMP3	16.626	5	.000	2.8850	2.4389	3.3311
COMP4	14.079	5	.000	3.7183	3.0394	4.3972
COMP5	5.199	5	.003	1.9250	.9732	2.8768
COMP6	11.747	5	.000	3.9783	3.1078	4.8489

- These are pair-wise posthoc comparisons, so a Tukey HSD correction is required.

$$t_{crit} = \frac{q_{crit}(.05,9,5)}{\sqrt{2}} = \frac{6.80}{\sqrt{2}} = 4.81$$

- Within each level of orientation, all distorted faces are rated as more bizarre than the control, non-distorted faces, all $ps < .05$, $ds > 2.12$.

o We decide to follow these tests up with pair wise comparisons between the two distorted faces at each level of orientation.

```
compute comp7 = usd_zer - wb_zer.
compute comp8 = usd_90 - wb_90.
compute comp9 = usd_180 - wb_180.
T-TEST /TESTVAL=0
 /VARIABLES=comp7 to comp9.
```

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
COMP7	6	.2133	.80746	.32964
COMP8	6	-.8333	.64242	.26227
COMP9	6	-2.0533	.95007	.38786

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
COMP7	.647	5	.546	.2133	-.6340	1.0607
COMP8	-3.177	5	.025	-.8333	-1.5075	-.1592
COMP9	-5.294	5	.003	-2.0533	-3.0504	-1.0563

- Again, a Tukey HSD correction is required.

$$t_{crit} = \frac{q_{crit}(.05,9,5)}{\sqrt{2}} = \frac{6.80}{\sqrt{2}} = 4.81$$

- When faces are presented upside-down, then the faces with eyes whitened and teeth blacked are rated as more bizarre than faces with up-side down eyes and mouth, $t(5) = 5.29, p < .05, d = 2.16$
- In other orientations (upright and 90°), there are no significant differences in ratings of the two distorted faces, $ds < 1.29$.

Distortion	0°	90°	180°	
None	1.32	1.69	1.49	1.50
Upside-Down	5.02	4.58	3.41	4.34
Whitened/Blackened	4.81	5.41	5.47	5.23
	3.72	3.89	3.46	3.69