Chapter 11B
Multi-Factor Repeated Measures ANOVA
Designs with one Between-Subjects and one Within-Subjects Factor

Multi-Factor Repeated Measures ANOVA
Designs with on Between-Subjects and one Within-Subjects Factors

1. Introduction

- Let's start with a simple example: one between subjects factor and one within-subjects factor

- Imagine that our previous data on the effects of a test prep class did not come from pre- and post-test scores from the same participants, but instead were scores from two different groups of people. In this case, we randomly assigned people to either take a test prep class, or to not take the test prep class.

No Training			Training		
Subscale1	Subscale2	Subscale3	Subscale1	Subscale2	Subscale3
42	42	48	48	60	78
42	48	48	36	48	60
48	48	54	66	78	78
42	54	54	48	78	90
54	66	54	48	66	72
36	42	36	36	48	54
48	48	60	54	72	84
48	60	66	54	72	90
54	60	54	48	72	78
48	42	54	54	66	78
46.2	51.0	52.8	49.2	66.0	76.2

Between-Subjects Comparison

Training	Subscale of test: Subscale 1	Subscale 2	Subscale 3	
No	$\overline{X}_{\cdot 11} = 46.2$	$\overline{X}_{\cdot 21} = 51.0$	$\overline{X}_{\cdot 31} = 52.8$	$\overline{X}_{\cdot\cdot 1} = 50.0$
Yes	$\overline{X}_{\cdot 12} = 49.2$	$\overline{X}_{\cdot 22} = 66.0$	$\overline{X}_{\cdot 32} = 76.2$	$\overline{X}_{\cdot\cdot 2} = 63.8$
	$\overline{X}_{\cdot 1\cdot} = 47.7$	$\overline{X}_{\cdot 2\cdot} = 58.5$	$\overline{X}_{\cdot 3\cdot} = 64.5$	

Within-Subjects Comparison

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

- To understand the structural model of a between and within design, let's start with the model of a design containing two within factors, and see what changes. We will consider the A factor the between subjects factor, and the B factor the within subjects factor

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

  - μ is the grand mean of all scores
  - α_j is the effect of the between subjects factor

No Training	Training
50.0	63.8

α_1 α_2

  - β_k is the effect of the within subjects factor

Subscale1	Subscale2	Subscale3
47.7	58.5	64.5

β_1 β_2 β_3

  - $\alpha\beta_{jk}$ is the interaction of the within and between subjects factors

$\alpha\beta_{11}$ $\alpha\beta_{12}$ $\alpha\beta_{13}$

	Subscale1	Subscale2	Subscale3
No Training	46.2	51.0	52.8
Training	49.2	66.0	76.2

$\alpha\beta_{21}$ $\alpha\beta_{22}$ $\alpha\beta_{23}$

  - These fixed effect parameters are computed exactly the same as for a all between- or all within-subjects design.

- π_σ is the subject effect, but we have a subject effect for each level of A, the between-subjects factor. We refer to this as the subject effect within (each level of) A, $\pi_{\sigma(j)}$

$\pi_{\sigma 1}$ $\pi_{\sigma 2}$

No Training	Training
44	62
46	48
50	74
50	72
58	62
38	46
52	70
58	72
56	66
48	66

- Note that $\pi_{\sigma(j)}$ measures how much the factor A effect varies by subject. We can think of the $\pi_{\sigma(j)}$ terms as a measure of the error in the α_j effect. For this to work, we will need $\pi_{\sigma(1)} = \pi_{\sigma(2)}$

- $(\alpha\pi)_\sigma$ is the interaction between subject and A. But subjects are not crossed with factor A. There are different subjects in each level of A. Thus, we cannot estimate this term.

  When a factor (Subjects) is not crossed with each level of another factor (A), but instead only appears within a single level of that factor (A), we say that subjects are <u>nested</u> within A

- $(\beta\pi)_{\sigma}$ is the interaction between subject and B. We will be able to estimate this term. Each subject gets each level of the within subjects factor.

    - Because we have two groups of subjects, we will have two estimates of $(\beta\pi)_{\sigma}$, one for each level. We refer to this as the subject by B effect within (each level of) A, $(\beta\pi)_{\sigma(j)}$

$\beta\pi_{\sigma 1}$ $\beta\pi_{\sigma 2}$

No Training			Training		
Subscale1	Subscale2	Subscale3	Subscale1	Subscale2	Subscale3
42	42	48	48	60	78
42	48	48	36	48	60
48	48	54	66	78	78
42	54	54	48	78	90
54	66	54	48	66	72
36	42	36	36	48	54
48	48	60	54	72	84
48	60	66	54	72	90
54	60	54	48	72	78
48	42	54	54	66	78

- Note that $(\beta\pi)_{\sigma(j)}$ measures how much the factor B effect varies by subject. $(\beta\pi)_{\sigma(j)}$ is also a measure of the extent to which the A*B interaction varies by subject
- Thus, we can think of the $(\beta\pi)_{\sigma(j)}$ terms as a measure of the error in the β_k and $\alpha\beta_{jk}$ effects. Again, for this to work nicely, we need $(\beta\pi)_{\sigma(1)} = (\beta\pi)_{\sigma(2)}$

- Finally, the $(\alpha\beta\pi)_{\sigma}$ effect is the three-way interaction between subject, A and B. But as we already noted, subjects are not crossed with factor A; subjects are nested within A. Thus, we cannot estimate how $(\beta\pi)_{\sigma(i)}$ varies across subjects.

- So we are left with the following model for a between (A) and within (B) factors design:

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

- Let's look at the expected mean squares for each of the terms in the model to see if our intuitions about the error terms are correct:

Source	E(MS)	F
Factor A	$\sigma_\varepsilon^2 + b\sigma_\pi^2 + \frac{nb\sum \alpha_j^2}{a-1}$	$\frac{MSA}{MS(S/A)}$
Subjects/A (Between Error)	$\sigma_\varepsilon^2 + b\sigma_\pi^2$	
Factor B	$\sigma_\varepsilon^2 + \sigma_{\beta\pi}^2 + \frac{na\sum \beta_k^2}{b-1}$	$\frac{MSB}{MS(B*S/A)}$
*A * B*	$\sigma_\varepsilon^2 + \sigma_{\beta\pi}^2 + \frac{n\sum \alpha\beta_{jk}^2}{(a-1)(b-1)}$	$\frac{MSAB}{MS(B*S/A)}$
*B*Subjects/A (Within Error)*	$\sigma_\varepsilon^2 + \sigma_{\beta\pi}^2$	

- ANOVA Table

Source	SS	df	MS	F
Factor A	*SSA*	*(a-1)*	$\frac{SSA}{a-1}$	$\frac{MSA}{MS(S/A)}$
Subjects/A (Between Error)	*SS(S/A)*	*N-a*	$\frac{SS(S/A)}{N-a}$	
Factor B	*SSB*	*(b-1)*	$\frac{SSB}{b-1}$	$\frac{MSB}{MS(B*S/A)}$
*A * B*	*SSAB*	*(a-1)(b-1)*	$\frac{SSAB}{(a-1)(b-1)}$	$\frac{MSAB}{MS(B*S/A)}$
*B*Subjects/A (Within Error)*	*SS(B*S/A)*	*(N-a)(b-1)*	$\frac{SS(B*S/A)}{(N-a)(b-1)}$	

3. Assumptions for between and within factor designs

- Assumptions for between-subjects tests: These assumptions are identical to the assumptions for a one-way between-subjects ANOVA.
  - To conduct the omnibus test for the between subjects effect, *assumptions are made on the marginal between-subjects means*.
    - Samples are independent and randomly drawn from the population
    - Each group is normally (symmetrically) distributed
    - All groups have a common variance
  - If you will perform simple effects tests on the between-subjects factor, then you need to make the following *assumptions on the between-subjects cell means at each level of the within-subjects factor*.
    - Each group is normally (symmetrically) distributed
    - All groups have a common variance

- Assumptions for within-subjects tests:
  - When examining the model parameters, we noted that we needed the error terms to be equal in the two samples: $\pi_{\sigma(1)} = \pi_{\sigma(2)}$ and $(\beta\pi)_{\sigma(1)} = (\beta\pi)_{\sigma(2)}$. To satisfy this assumption, we must have homogeneity of variance/covariance matrices for each sample/group

$$\overset{A_1}{\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}} = \overset{A_2}{\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}}$$

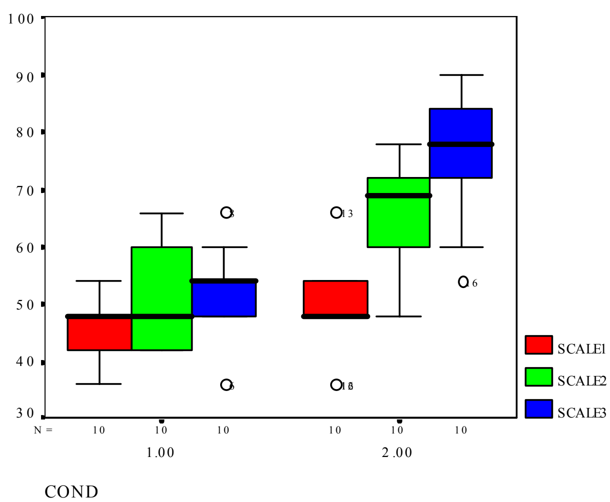
    - Homogeneity of variance/covariance matrices is required for any omnibus comparisons on the within-subjects marginal means or for omnibus interaction tests on between & within cell means.
    - SPSS provides Box's M test and Levine's test as a check of homogeneity of variance/covariance matrices.
    - If this assumption is violated, the omnibus tests may not be preformed for the main effect of the within-subjects effect or for the interaction between the within-subjects and between-subjects factor.

- *If homogeneity of variance/covariance matrices is satisfied,* then in order to conduct omnibus tests for the main effect of the within-subjects effect or for the interaction between the within-subjects and between-subjects factor we must have:
  - Sphericity of the pooled variance/covariance matrix.
  - Normality of repeated measures (but we already checked this)
  - Participants are independent and randomly selected from the population (but we already checked this)

- If we wish to conduct simple effects tests for the effect of the repeated measures factor at each level of the between-subject factor, then we must have sphericity of the variance/covariance matrix for each between subjects group.
  - Note that we do not need to have homogeneity of variance/covariance matrices in order to test this assumption.

- Testing assumptions: *Normality*
  - For all tests on the marginal within-subjects means and on the cell means, we need to check normality on a cell-by-cell basis.

```
EXAMINE VARIABLES=scale1 scale2 scale3 BY cond
  /PLOT BOXPLOT NPPLOT SPREADLEVEL
  /COMPARE VARIABLES.
```



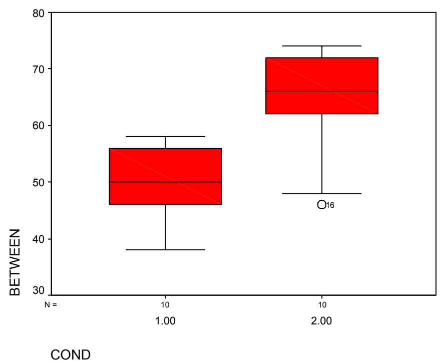
Tests of Normality

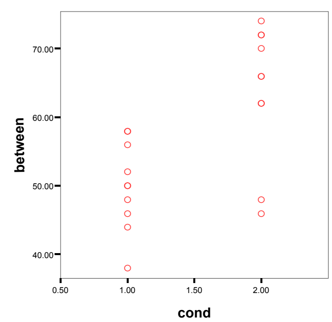
	cond	Shapiro-Wilk		
		Statistic	df	Sig.
scale1	1.00	.911	10	.287
	2.00	.897	10	.202
scale2	1.00	.886	10	.151
	2.00	.869	10	.097
scale3	1.00	.897	10	.203
	2.00	.892	10	.180

- No cell seems too asymmetrical. We appear to be OK for normality

- In order to conduct all tests on the between-subjects marginal means, we need the marginal means to be normally distributed. To test the marginal means, we must manually average across the repeated measures, compute the marginal effects, and conduct our usual tests for normality.

```
COMPUTE  between = (scale1+scale2+scale3)/3.
EXAMINE VARIABLES=between BY cond
  /PLOT BOXPLOT NPPLOT SPREADLEVEL.
```





Tests of Normality

	COND	Shapiro-Wilk Statistic	df	Sig.
BETWEEN	1.00	.952	10	.689
	2.00	.853	10	.064

- Normality (Symmetry) is satisfied.

- To check homogeneity/sphericity, we will adopt a three-step approach
  - Check the equality of the variance/covariance matrices across the different samples
  - Check the sphericity of the pooled variance/covariance matrix (Overall sphericity)
  - Check the sphericity of the variance/covariance matrix for each group separately (Multi-sample sphericity)

- To check the homogeneity of the variance/covariance matrices across the different samples, we use Box's test of equality of the variance/covariance matrices and Levene's test of variances.

```
GLM  scale1 scale2 scale3 by cond
 /WSFACTOR = scale 3
 /PRINT = DESC HOMO.
```

  - Box's test is an omnibus test of equivalence of variance/covariance matrices

    $H_0 : Var/Cov_1 = Var/Cov_2 = ... = Var/Cov_a$

    H_1 : At least 1 Var/Cov matrix differs from the others

$$\underset{A_1}{\begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}} = \underset{A_2}{\begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}}$$

    - Note that this test is not examining if the Var/Cov matrices are spherical, only if they are equal
    - If we reject the null hypothesis, we can not pool the matrices to test within-subject effects (and we will need to consider alternative approaches to omnibus analyses).

Box's Test of Equality of Covariance Matrices

Box's M	5.682
F	.774
df1	6
df2	2347.472
Sig.	.591

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

$F(6,2347.47) = 0.774, p = .59$

    - We fail to reject null hypothesis, so we have no evidence that the variance/covariance matrices are unequal.

- Levene's test is a more focused test of the equivalence of only the variances

$$\overset{A_1}{\begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}} = \overset{A_2}{\begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}}$$

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
SCALE1	.635	1	18	.436
SCALE2	.248	1	18	.624
SCALE3	1.204	1	18	.287

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

- This test only examines if the variances of the different groups are equal
- For subscale 1

  $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_a^2$

  $H_1 :$ At least 1 σ_i^2 differs from the others

  $F(1,18) = 0.64, p = .43$
- For subscale 2

  $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_a^2$

  $H_1 :$ At least 1 σ_i^2 differs from the others

  $F(1,18) = 0.25, p = .62$
- For subscale 3

  $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_a^2$

  $H_1 :$ At least 1 σ_i^2 differs from the others

  $F(1,18) = 1.20, p = .29$

- If either Box's test or any of Levene's Tests are significant, then we reject the assumption of homogeneity of the variance/covariance matrices.
- In this case, we have no evidence to conclude that the matrices are different, so we may pool them and test for sphericity.

- Next, we test for overall sphericity by averaging across the between subjects factor and examining the epsilon

Entire Sample

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SCALE	.961	1.000	.500

  - In this case, the overall sphericity assumption is satisfied
  - We may conduct unadjusted omnibus tests on the within-subjects factors
    - SCALE
    - SCALE*CONDITION

- If we plan on conducting simple effects tests (of the within-subjects factor at each level of the between-subjects factor), then we need to examine the epsilon for each condition (the multi-sample sphericity).

```
temporary.
select if cond=1.
GLM  scale1 scale2 scale3
  /WSFACTOR = scale 3.
```

CONDITION #1

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SCALE	.864	1.000	.500

```
temporary.
select if cond=2.
GLM  scale1 scale2 scale3
  /WSFACTOR = scale 3.
```

Condition #2

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SCALE	.776	.907	.500

  - We conclude that, separately, the var/cov matrix for each condition is not spherical, but the violation is fixable
    - If we want to conduct follow-up tests on each condition, we need to adjust all omnibus tests

- Overall we conclude that:
  - The var/cov matrix for condition 1 equals the var/cov matrix for condition 2
  - When we combine the 2 conditions, the overall var/cov matrix is spherical
  - BUT the neither the var/cov matrix for condition 1 nor the var/cov matrix for condition 2 is spherical!

- Remember that all this funny business of checking the var/cov matrix can be skipped if we avoid omnibus tests and stick to contrasts!

- If assumptions are violated: A recap

i. If normality/symmetry is not satisfied:
   - All F-tests may be biased.
   - Try advanced non-parametric/distribution-free tests

ii. If the variances are not equal between groups within each condition (Levene's test and boxplots suggest heterogeneity):
   - Then we cannot conduct between-subjects tests that require equal variances (omnibus tests and/or standard contrasts).
   - Test all between-subject contrasts with unequal variance contrasts.
   - Test all between-subject omnibus tests with the Brown-Forsyth F* Test.

iii. If variance/covariance matrices are not equal across all groups (Box's M is significant or Levene's test suggests heterogeneity):
   - Then we cannot pool var/cov matrices over the between-subjects groups.
   - The omnibus within-subject error term (used to test within-subject effects and between/within interactions) is not valid.
   - Use the MANOVA approach for omnibus tests of within-subject effects and between/within interactions OR use contrasts for between/within tests.

iv. If sphericity of the combined variance/covariance matrix (Overall sphericity) is violated:
   - Note: If assumption (iii.) is violated, then we cannot pool the var/cov matrices and this assumption is automatically violated.
   - The omnibus within-subject error term (used to test within-subject effects and between/within interactions) is not valid.
   - If violation is moderate, use epsilon-adjusted omnibus tests or contrasts.
   - If violation is extreme, use contrasts for between/within tests or the MANOVA approach for omnibus tests of within-subject effects and between/within interactions.

v. If the sphericity of the variance/covariance matrix for each group separately (Multi-sample sphericity) is violated:
   - Note: If assumption (iii.) is violated, then this assumption may still be satisfied.
   - The omnibus within-subject error term calculated at each level of the between subject factor (simple effects of the within-subjects factor at each level of the between subjects factor) is not valid.
   - If violation is moderate, use epsilon-adjusted simple effect omnibus tests or contrasts.
   - If violation is extreme, use the MANOVA approach for simple effect omnibus tests of within-subject effects or contrasts.

4. Analysis of omnibus ANOVA effects

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

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

$$\hat{\eta}^2_{A} = \frac{SS_{A}}{SS_{A} + SS_{S/A}} \qquad \hat{\eta}^2_{B} = \frac{SS_{B}}{SS_{B} + SS_{B*S/A}} \qquad \hat{\eta}^2_{A*B} = \frac{SS_{A*B}}{SS_{A*B} + SS_{B*S/A}}$$

	Subscale of test			
Training	Subscale 1	Subscale 2	Subscale 3	
No	$\bar{X}_{\cdot 11} = 46.2$	$\bar{X}_{\cdot 21} = 51.0$	$\bar{X}_{\cdot 31} = 52.8$	$\bar{X}_{\cdot\cdot 1} = 50.0$
Yes	$\bar{X}_{\cdot 12} = 49.2$	$\bar{X}_{\cdot 22} = 66.0$	$\bar{X}_{\cdot 32} = 76.2$	$\bar{X}_{\cdot\cdot 2} = 63.8$
	$\bar{X}_{\cdot 1 \cdot} = 47.7$	$\bar{X}_{\cdot 2 \cdot} = 58.5$	$\bar{X}_{\cdot 3 \cdot} = 64.5$	

- In this case, we may conduct unadjusted within subjects tests (see p. 11B-12)

```
GLM scale1 scale2 scale3 by cond
  /WSFACTOR = scale 3
  /PRINT = DESC.
```

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SCALE	Sphericity Assumed	2899.200	2	1449.600	61.424	.000
	Greenhouse-Geisser	2899.200	1.921	1508.872	61.424	.000
	Huynh-Feldt	2899.200	2.000	1449.600	61.424	.000
	Lower-bound	2899.200	1.000	2899.200	61.424	.000
SCALE * COND	Sphericity Assumed	1051.200	2	525.600	22.271	.000
	Greenhouse-Geisser	1051.200	1.921	547.091	22.271	.000
	Huynh-Feldt	1051.200	2.000	525.600	22.271	.000
	Lower-bound	1051.200	1.000	1051.200	22.271	.000
Error(SCALE)	Sphericity Assumed	849.600	36	23.600		
	Greenhouse-Geisser	849.600	34.586	24.565		
	Huynh-Feldt	849.600	36.000	23.600		
	Lower-bound	849.600	18.000	47.200		

$$\hat{\eta}^2_{Scale} = \frac{2899.2}{2889.2 + 849.6} = .77 \qquad \hat{\eta}^2_{Time*Scale} = \frac{1051.2}{1051.2 + 849.6} = .55$$

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	194256.600	1	194256.600	949.446	.000
COND	2856.600	1	2856.600	13.962	.002
Error	3682.800	18	204.600		

$$\hat{\eta}^2_{Condition} = \frac{2856.6}{2856.6 + 3682.8} = .44$$

- Tests of the within subjects factors:
  - The main effect of scale: $F(2,36) = 61.42, p < .01, \eta^2 = .77$
    - ⇒ Collapsing across level of training, there are significant differences in the scores to the three sub-scales of the test
  - The scale by training interaction: $F(2,36) = 22.27, p < .01, \eta^2 = .55$
    - ⇒ The effect of training is not the same for each subscale of the test

- Tests of the between subjects factors:
  - The main effect of training: $F(1,18) = 13.96, p < .01, \eta^2 = .44$
    - ⇒ Averaging across subscales, those who received training performed better than those who do not receive training.

- To interpret effects, you use the same logic outlined for factorial ANOVA. Start with the highest order significant (or important) effect. Interpret lower order effects only if they are meaningful.
  - In this case, we have a significant scale by training interaction. We could follow-up this result with simple effect tests.

5. Contrasts
   - The contrast formulae remain the same:

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

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

   - For between-subjects tests on the marginal means:
     - If the homogeneity of variances assumption is satisfied, then MSE' will be the between-subjects error term, $MSE' = MS_{S/A}$ (with $df = N\text{-}a$).
     - If the homogeneity of variances assumption is not satisfied, then we can use the unequal variance test for contrast (Welsh's test).

   - For between-subjects tests within one level of the within-subject factor:
     - If the homogeneity of variances assumption is satisfied at that level of the within subjects factor, then MSE' will be the between-subjects error term, $MSE' = MS_{S/A\,j}$ (with $df = N\text{-}a$).
     - If the homogeneity of variances assumption is not satisfied, then we can use the unequal variance test for contrast (Welsh's test).

   - For within subjects tests (either on marginal within-subjects means or on the between*within cell means):
     - MSE' will be a contrast-specific error term (with $df = N\text{-}a$).
     - If the data are spherical, then we *could* use an omnibus error term. For contrasts on the marginal within-subjects means or on the between/within cell means, use the omnibus within-subjects error term, $MSE' = MS_{B*S/A}$ (with $df = (b\text{-}1)(N\text{-}a)$). However, I recommend that you <u>always</u> use the contrast-specific error term.

   - Note that *all* contrasts should have $df = N\text{-}a$.
     (Unless for some reason you decide to use the omnibus error term for within-subject or between*within contrasts.)

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

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

    - 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 between-subject contrasts, we can compute **d** directly from the *t*-statistic:

$$\hat{d} = \frac{2 * t}{\sqrt{df}}$$

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

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

- To perform contrasts on the *between subjects marginal means*, you need to compute an average across the within subjects factor.

Between-Subjects Marginal Means

	Subscale of test			
Training	Subscale 1	Subscale 2	Subscale 3	
No	$\overline{X}_{\cdot 11} = 46.2$	$\overline{X}_{\cdot 21} = 51.0$	$\overline{X}_{\cdot 31} = 52.8$	$\overline{X}_{\cdot\cdot 1} = 50.0$
Yes	$\overline{X}_{\cdot 12} = 49.2$	$\overline{X}_{\cdot 22} = 66.0$	$\overline{X}_{\cdot 32} = 76.2$	$\overline{X}_{\cdot\cdot 2} = 63.8$
	$\overline{X}_{\cdot 1 \cdot} = 47.7$	$\overline{X}_{\cdot 2 \cdot} = 58.5$	$\overline{X}_{\cdot 3 \cdot} = 64.5$	

Within-Subjects Marginal Means

- To run a test on the marginal between-subject means, we need to compute a new variable and then run an ANOVA (or t-test).

```
COMPUTE  between = (scale1+scale2+scale3)/3.
T-TEST GROUPS=cond(1 2)
 /VARIABLES=between .
```

Group Statistics

	cond	N	Mean	Std. Deviation	Std. Error Mean
between	1.00	10	50.0000	6.39444	2.02210
	2.00	10	63.8000	9.77298	3.09049

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
between	Equal variances assumed	1.414	.250	-3.737	18	.002	-13.80000	3.69324	-21.55920	-6.04080
	Equal variances not assumed			-3.737	15.512	.002	-13.80000	3.69324	-21.64936	-5.95064

$$d = \frac{2*t}{\sqrt{df}} = \frac{2*3.737}{\sqrt{18}} = 1.76$$

```
ONEWAY between by cond.
```

ANOVA

BETWEEN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	952.200	1	952.200	13.962	.002
Within Groups	1227.600	18	68.200		
Total	2179.800	19			

$$\eta^2 = \frac{SS_{Contrast}}{SS_{Contrast} + SS_{ErrorTermForContrast}} = \frac{952}{952 + 1227.6} = .44$$

  - Individuals in the training condition performed better than intervals without training: $t(18) = 3.73, p < .01, d = 1.76$
  - This test is identical to the main effect of training obtained from the repeated measures analysis, $F(1,18) = 13.96, p < .01, \eta^2 = .44$

- In this example, the between subjects factor has only two levels so follow-up tests are unnecessary.
  - If the between subjects factor had more than two levels, you could use the *CONTRAST* command to test the between subjects contrasts.
  - If the between-subjects variances are unequal, you can use unequal variance contrasts.
  - You may need to adjust the p-value of the tests, depending on whether the tests are planned or post-hoc.

- To conduct contrasts on the *within subjects marginal means*:

Training	Subscale of test: Subscale 1	Subscale 2	Subscale 3	
No	$\bar{X}_{\cdot 11} = 46.2$	$\bar{X}_{\cdot 21} = 51.0$	$\bar{X}_{\cdot 31} = 52.8$	$\bar{X}_{\cdot\cdot 1} = 50.0$
Yes	$\bar{X}_{\cdot 12} = 49.2$	$\bar{X}_{\cdot 22} = 66.0$	$\bar{X}_{\cdot 32} = 76.2$	$\bar{X}_{\cdot\cdot 2} = 63.8$
	$\bar{X}_{\cdot 1 \cdot} = 47.7$	$\bar{X}_{\cdot 2 \cdot} = 58.5$	$\bar{X}_{\cdot 3 \cdot} = 64.5$	

Within-Subjects Marginal Means

- The easiest approach to conducting contrasts on the within subjects marginal means is to use SPSS's built in contrasts:
  - Specify a type of contrast on the within-subject factor using the *WSFACTOR* subcommand:
  - To test if subscale 2 differs from subscale3:

```
GLM  scale1 scale2 scale3 by cond
 /WSFACTOR = scale 3 helmert.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SCALE	Type III Sum of Squares	df	Mean Square	F	Sig.
SCALE	Level 1 vs. Later	3808.800	1	3808.800	110.400	.000
	Level 2 vs. Level 3	720.000	1	720.000	14.876	.001
SCALE * COND	Level 1 vs. Later	1312.200	1	1312.200	38.035	.000
	Level 2 vs. Level 3	352.800	1	352.800	7.289	.015
Error(SCALE)	Level 1 vs. Later	621.000	18	34.500		
	Level 2 vs. Level 3	871.200	18	48.400		

$$\eta^2 = \frac{SS_{Contrast}}{SS_{Contrast} + SS_{ErrorTermForContrast}} = \frac{720}{720 + 871.2} = .45$$

  - We want to conduct tests on the marginal scale means (average across condition), so we need to read the line labeled "SCALE"
  - Averaging across level of training, we find that scores on scale 3 are higher than scores on scale 2, $F(1,18) = 14.87, p < .01, \eta^2 = .45$

- An alternative approach to conducting contrasts on the within subjects marginal means is to use the special command:

```
GLM  scale1 scale2 scale3 by cond
  /WSFACTOR = scale 3 special (  1  1  1
                                 0  -1 1
                                -1 0 1).
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SCALE	Type III Sum of Squares	df	Mean Square	F	Sig.
SCALE	L1	720.000	1	720.000	14.876	.001
	L2	5644.800	1	5644.800	103.007	.000
SCALE * COND	L1	352.800	1	352.800	7.289	.015
	L2	2080.800	1	2080.800	37.971	.000
Error(SCALE)	L1	871.200	18	48.400		
	L2	986.400	18	54.800		

  - The contrast labeled "SCALE L1" gives us the same results as the previous analysis

- If we try to create a new variable reflecting the contrast, and run a t-test, we get an incorrect result because the between-subjects factor is no longer included in the analysis (and we are ignoring the fact that we have different groups of participants). You should not use this method.

```
compute c1 = scale3 - scale2.
T-TEST /TESTVAL=0
  /VARIABLES=c1.
```

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
C1	3.343	19	.003	6.0000	2.2436	9.7564

  - When we convert this to an F-value, $F(1,19) = 11.18, p = .003$
  - The degrees of freedom are off by one, and this method uses a slightly different error term because this method of analysis completely drops between-subjects factor from the analysis.

- Again, depending on the nature of these tests, the p-values may need adjustment.

- To conduct *contrasts on the between subjects by within subjects cell means*, SPSS makes the task difficult.

  - To compute a between/within contrast in SPSS, we must be able to write the contrast as an interaction contrast (a difference of differences).
    - Suppose we want to examine if the difference between scores on subscale 2 and subscale 3 depends on training:

	Subscale1	Subscale2	Subscale3	
No Training		1	-1	
Training		-1	1	

  - <u>Method #1</u>: Use brand-name contrasts * condition tests.
    This contrast is a test of whether the (scale3 – scale2) contrast differs by condition.

    The effect of scale3 – scale2 for no training: $\mu_{NoTrain3} - \mu_{NoTrain2}$
    The effect of scale3 – scale2 for training: $\mu_{Train3} - \mu_{Train2}$

    Do these effects differ?

$$\psi : (\mu_{NoTrain3} - \mu_{NoTrain2}) - (\mu_{Train3} - \mu_{Train2})$$
$$\psi : \mu_{NoTrain3} - \mu_{NoTrain2} - \mu_{Train3} + \mu_{Train2}$$

    - I can obtain this contrast from SPSS by asking for an interaction between the (scale3 – scale2) contrast on the marginal scale means and a (Training – No Training) contrast on the marginal training condition means

	Subscale1	Subscale2	Subscale3	
No Training				-1
Training				1
		-1	1	

- This test result was printed when we asked for the Helmert contrasts:

```
GLM  scale1 scale2 scale3 by cond
 /WSFACTOR = scale 3 helmert.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SCALE	Type III Sum of Squares	df	Mean Square	F	Sig.
SCALE	Level 1 vs. Later	3808.800	1	3808.800	110.400	.000
	Level 2 vs. Level 3	720.000	1	720.000	14.876	.001
SCALE * COND	Level 1 vs. Later	1312.200	1	1312.200	38.035	.000
	Level 2 vs. Level 3	**352.800**	**1**	**352.800**	**7.289**	**.015**
Error(SCALE)	Level 1 vs. Later	621.000	18	34.500		
	Level 2 vs. Level 3	871.200	18	48.400		

$$\eta^2 = \frac{SS_{Contrast}}{SS_{Contrast} + SS_{ErrorTermForContrast}} = \frac{352.8}{352.8 + 871.2} = .29$$

$$F(1,18) = 7.29, p = .02, \eta^2 = .29$$

- Method #2: Use special contrasts * condition tests.

```
GLM  scale1 scale2 scale3 by cond
  /WSFACTOR = scale 3 special (  1  1  1
                                 0  -1 1
                                -1  0  1).
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SCALE	Type III Sum of Squares	df	Mean Square	F	Sig.
SCALE	L1	720.000	1	720.000	14.876	.001
	L2	5644.800	1	5644.800	103.007	.000
SCALE * COND	L1	352.800	1	352.800	7.289	.015
	L2	2080.800	1	2080.800	37.971	.000
Error(SCALE)	L1	871.200	18	48.400		
	L2	986.400	18	54.800		

$$F(1,18) = 7.29, p = .02, \eta^2 = .29$$

- Method #3: Manually compute the main effect contrast of interest, and run a t-test comparing that variable across levels of training:

```
compute c1 = scale3 - scale2.
UNIANOVA c1 by cond.
```

Tests of Between-Subjects Effects

Dependent Variable: C1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	352.800[a]	1	352.800	7.289	.015
Intercept	720.000	1	720.000	14.876	.001
COND	352.800	1	352.800	7.289	.015
Error	871.200	18	48.400		
Total	1944.000	20			
Corrected Total	1224.000	19			

a. R Squared = .288 (Adjusted R Squared = .249)

$$F(1,18) = 7.29, p = .02, \eta^2 = .29$$

```
T-TEST  GROUPS = cond(1 2)
 /VARIABLES = c1 .
```

Group Statistics

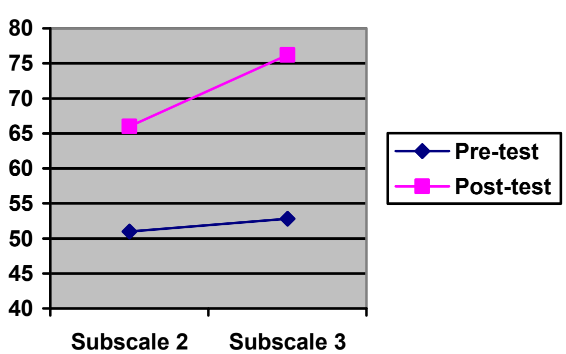
	cond	N	Mean	Std. Deviation	Std. Error Mean
c1	1.00	10	1.8000	8.02496	2.53772
	2.00	10	10.2000	5.69210	1.80000

Independent Samples Test

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
c1	-2.700	18	.015	-8.40000	3.11127	-14.93654	-1.86346

$$d = \frac{2*t}{\sqrt{df}} = \frac{2*2.7}{\sqrt{18}} = 1.28$$

$$t(18) = 2.70, p = .02, d = 1.28$$



- We conclude that the difference between scale 2 and scale 3 scores differs as a result of training. The difference between scores on scale 2 and scale 3 becomes larger after training

- For contrasts that are not differences across levels of the between-subjects factor, more advanced techniques are required (but it is not clear that you should be running these types of contrasts. How would you interpret this?).

	Subscale1	Subscale2	Subscale3
No Training	-1	1	
Training		-1	1

- If the between-subjects factor has more than two levels, then testing between/within contrasts is trickier (see example 1).

	Subscale1	Subscale2	Subscale3
No Training			
Type A Training		1	-1
Type B Training		-1	1

- If these contrasts are post-hoc and need adjustment, follow the adjustment procedures for factorial designs (using the appropriate error term and error degrees of freedom).

6. Simple Effects Tests

- To conduct *simple effects (of the between-subjects factor at each level of the within-subjects factor)*, we can run between-subject analyses on each scale.
  - We want to compute an error term based only on the within-subject information that is test-specific. Thus, it is acceptable to run separate tests on each subscale
  - The variances of the training conditions are equal for each subscale (recall the Levene's tests, p. 11B-11), so standard tests may be conducted.

	Subscale of test		
Training	Subscale 1	Subscale 2	Subscale 3
No	$\bar{X}_{\cdot 11} = 46.2$	$\bar{X}_{\cdot 21} = 51.0$	$\bar{X}_{\cdot 31} = 52.8$
Yes	$\bar{X}_{\cdot 12} = 49.2$	$\bar{X}_{\cdot 22} = 66.0$	$\bar{X}_{\cdot 32} = 76.2$

```
ONEWAY scale1 scale2 scale3 by cond.
```

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
SCALE1	Between Groups	45.000	1	45.000	.812	.379
	Within Groups	997.200	18	55.400		
	Total	1042.200	19			
SCALE2	Between Groups	1125.000	1	1125.000	11.598	.003
	Within Groups	1746.000	18	97.000		
	Total	2871.000	19			
SCALE3	Between Groups	2737.800	1	2737.800	27.543	.000
	Within Groups	1789.200	18	99.400		
	Total	4527.000	19			

$$\eta^2_{Scale1} = \frac{45}{45+997} = .04 \qquad \eta^2_{Scale2} = \frac{1125}{1125+1746} = .39 \qquad \eta^2_{Scale3} = \frac{2737.8}{2737.8+1789.2} = .60$$

$$p_{crit} = \frac{.05}{3} = 0.0167$$

- There is no effect of training on performance on subscale 1, $F(1,18) = 0.81$, *ns*. $F(1,18) = 0.82, ns, \eta^2 = .04$
- For subscales 2 and 3, training improves performance, $F(1,18) = 11.60, p < .05, \eta^2 = .39$, and $F(1,18) = 27.54, p < .05, \eta^2 = .60$, respectively.

  - These contrasts could also be run as t-tests.

```
T-TEST  GROUPS = cond(1 2)
 /VARIABLES = scale1 scale2 scale3.
```


- To conduct *simple effects (of the within-subjects factor at each level of the between-subjects factor)*, we can run within-subject analyses at each level of training.
  - These are omnibus within-subjects tests. An epsilon adjustment is required for each test.
  - The variance/covariance matrices for training and no training conditions are equal. Thus, we would like to pool information from both between-subject conditions to calculate the error term (in order to increase power and the precisions of the estimate of the error term).

	Subscale of test		
Training	Subscale 1	Subscale 2	Subscale 3
No	$\bar{X}_{\cdot 11} = 46.2$	$\bar{X}_{\cdot 21} = 51.0$	$\bar{X}_{\cdot 31} = 52.8$
Yes	$\bar{X}_{\cdot 12} = 49.2$	$\bar{X}_{\cdot 22} = 66.0$	$\bar{X}_{\cdot 32} = 76.2$

  - If we analyze the training and no training groups separately, the error terms will only contain information from the training and no training groups, respectively (Note that this procedure would be acceptable if the variances between the training and no training groups were unequal)
    - Thus, unless the between-subjects variances are unequal, we should avoid doing the following:

```
Temporary.
Select if cond = 1.
GLM  scale1 scale2 scale3
  /WSFACTOR = scale 3.
```

```
Temporary.
Select if cond = 2.
GLM  scale1 scale2 scale3
  /WSFACTOR = scale 3.
```

Each of these tests will only have *n-1* degrees of freedom (assuming equal *n* per group), rather than *N-a*. Thus, with this approach, we lose power and accuracy (assuming homogeneity of variance)

- However, we can select each group separately to obtain the sum of squares for the simple effects tests. We can then manually compute tests for the effect of time for training and no training groups separately using the omnibus within-subjects error term:

$$F[\hat{\varepsilon}(a-1), \hat{\varepsilon}(N-a)(b-1)] = \frac{MS_{Scale\,(No\,Training\,Only)}}{MS_{Scale*Subject\,/\,Training}}$$

From the full within-subjects omnibus tests we can obtain the appropriate *epsilon correction* and *error* mean squares.

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
scale	.961	1.000	.500

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
scale	Sphericity Assumed	2899.200	2	1449.600	61.424	.000
	Greenhouse-Geisser	2899.200	1.921	1508.872	61.424	.000
	Huynh-Feldt	2899.200	2.000	1449.600	61.424	.000
	Lower-bound	2899.200	1.000	2899.200	61.424	.000
scale * cond	Sphericity Assumed	1051.200	2	525.600	22.271	.000
	Greenhouse-Geisser	1051.200	1.921	547.091	22.271	.000
	Huynh-Feldt	1051.200	2.000	525.600	22.271	.000
	Lower-bound	1051.200	1.000	1051.200	22.271	.000
Error(scale)	Sphericity Assumed	849.600	36	23.600		
	Greenhouse-Geisser	849.600	34.586	24.565		
	Huynh-Feldt	849.600	36.000	23.600		
	Lower-bound	849.600	18.000	47.200		

$$F[(.725*2),(.725*36)] = \frac{MS_{Scale(No\,Training\,Only)}}{23.60}$$

From the within-subjects omnibus tests at each level of the between subjects factor, we can obtain the value of epsilon, and the appropriate *numerator* mean squares.

```
SORT CASES BY cond .
SPLIT FILE  LAYERED BY cond .
GLM  scale1 scale2 scale3
  /WSFACTOR = scale 3.
SPLIT FILE OFF.
```


Mauchly's Test of Sphericity

Measure: MEASURE_1

cond	Within Subjects Effect	Epsilon Greenhouse-Geisser	Huynh-Feldt	Lower-bound
1.00	scale	.864	1.000	.500
2.00	scale	.776	.907	.500

Tests of Within-Subjects Effects

Measure: MEASURE_1

cond	Source		Type III Sum of Squares	df	Mean Square	F	Sig.
1.00	scale	Sphericity Assumed	232.800	2	116.400	5.046	.018
		Greenhouse-Geisser	232.800	1.727	134.792	5.046	.024
	Error(scale)	Sphericity Assumed	415.200	18	23.067		
		Greenhouse-Geisser	415.200	15.544	26.711		
2.00	scale	Sphericity Assumed	3717.600	2	1858.800	77.022	.000
		Greenhouse-Geisser	3717.600	1.551	2396.735	77.022	.000
	Error(scale)	Sphericity Assumed	434.400	18	24.133		
		Greenhouse-Geisser	434.400	13.960	31.117		

$$MS_{Scale\,(No\ Training\ Only)} = 116.400 \qquad MS_{Scale\,(Training\ Only)} = 1858.80$$

$$\eta^2_{Scale(NoTrainingOnly)} = \frac{232.8}{232.8 + 849.6} = .22 \qquad \eta^2_{Scale(TrainingOnly)} = \frac{3717.6}{3717.6 + 849.6} = .81$$

$$F[2,36] = \frac{116.40}{23.60} = 4.93 \qquad F[2,36] = \frac{1858.8}{23.60} = 78.86$$

$$F(2,36) = 4.93, p = 0.01, \eta^2 = .22 \qquad F(2,36) = 78.86, p < .01, \eta^2 = .81$$

- We are conducting two simple effects tests, and thus, we need to apply a p-value correction.

$$p_{crit} = \frac{.05}{2} = 0.025$$

- Conclusions:
  - For the no training condition, we find a significant difference in performance over the three subscales, $F(2,36) = 4.93, p < .05, \eta^2 = .22$.
  - For the training condition, we find a significant difference in performance over the three subscales, $F(2,36) = 78.86, p < .05, \eta^2 = .81$.
  - Further pairwise tests must be conducted to understand these differences.

7. Final thoughts

- The approach to repeated measures that we have studied is known as the <u>univariate approach</u>. We assumed that all the differences of all the repeated measures were drawn from the same population. This assumption led us to a restrictive assumption on the covariance matrix and correlation matrix

$$\begin{bmatrix} \sigma^2 & \sigma_c & \sigma_c & \sigma_c \\ \sigma_c & \sigma^2 & \sigma_c & \sigma_c \\ \sigma_c & \sigma_c & \sigma^2 & \sigma_c \\ \sigma_c & \sigma_c & \sigma_c & \sigma^2 \end{bmatrix} \qquad \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

- Other approaches are possible, and if omnibus tests are called for are usually preferable. One approach is to assume that each difference of variables is drawn from a different population. This approach is known as the <u>multivariate approach</u> and leads to no assumptions on the covariance/correlation matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 \end{bmatrix} \qquad \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix}$$

- More recently, people have begun trying to model the structure of the variance covariance matrix:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

  - This approach is complicated, but it has much appeal if you
    - Have missing observations
    - Have unequal spacing in your repeated measurements
    - Are interested in the variance components

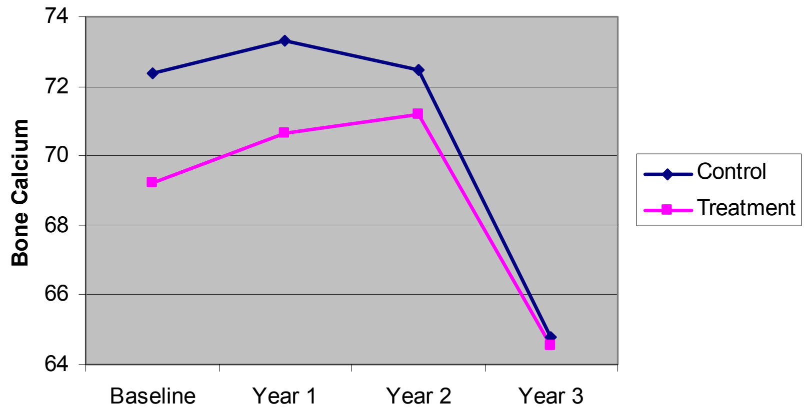
8. An example: Changes in bone calcium over time (2 * 4)

- A diet/exercise treatment was developed to stop bone calcium loss in women. A sample of older women was obtained and the women were placed in either a control group (n = 15) or a treatment group (n = 16). Bone calcium levels were obtained by photon absorptiometry readings of the dominant ulna bone at the time of enrollment in the study and at one year, two year, and three year follow-ups. Investigators were interested in:
  - Whether the treatment group had less bone loss than the control group.
  - Whether the rate of bone loss differs between the treatment group and the control group.

- The following data were obtained:

Control Group				Treatment Group			
Baseline	1 Year	2 Year	3Year	Baseline	1 Year	2 Year	3Year
87.3	86.9	86.7	75.5	83.3	85.5	86.2	81.2
59.0	60.2	60.0	53.6	65.3	66.9	67.0	60.6
76.7	76.5	75.7	69.5	81.2	79.5	84.5	75.2
70.6	76.1	72.1	65.3	75.4	76.7	74.3	66.7
54.9	55.1	57.2	49.0	55.3	58.3	59.1	54.2
78.2	75.3	69.1	67.6	70.3	72.3	70.6	68.6
73.7	70.8	71.8	74.6	76.5	79.9	80.4	71.6
61.8	68.7	68.2	57.4	66.0	70.9	70.3	64.1
85.3	84.4	79.2	67.0	76.7	79.0	76.9	70.3
82.3	86.9	79.4	77.4	77.2	74.0	77.8	67.9
68.6	65.4	72.3	60.8	67.3	70.7	68.9	65.9
67.8	69.2	66.3	57.9	50.3	51.4	53.6	48.0
66.2	67.0	67	56.2	57.3	57.0	57.5	51.5
81.0	82.3	86.8	73.9	74.3	77.7	72.6	68.0
72.3	74.6	75.3	66.1	74.0	74.7	74.5	65.7
				57.3	56.0	64.7	53.0

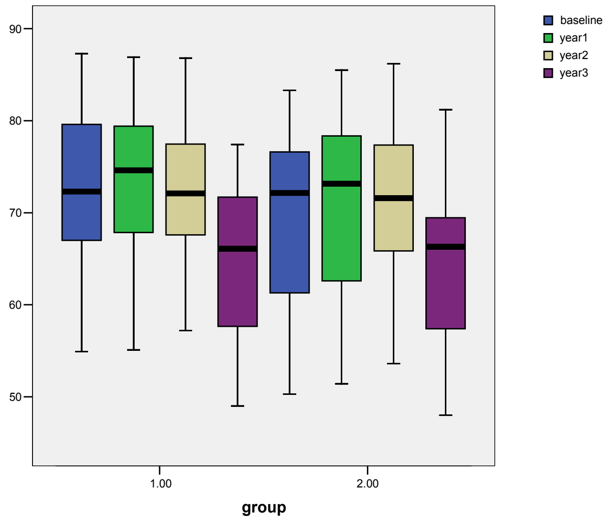
	Time				
Group	Baseline	Year 1	Year 2	Year 3	
Control	72.38	73.29	72.47	64.79	70.73
Treatment	69.23	70.66	71.18	64.53	68.90
	70.75	71.93	71.81	64.65	



- First, let's consider how we might test the hypotheses.
    - **Question #1**: Does the treatment group have less bone loss than the control group?
    - **Question #2**: Is the rate of bone loss different between the treatment group and the control group?

- Next, let's test all the assumptions for this model.
  - Normality

```
EXAMINE VARIABLES= baseline year1 year2 year3 BY group
 /PLOT BOXPLOT NPPLOT SPREADLEVEL
 /COMPARE VARIABLES.
```

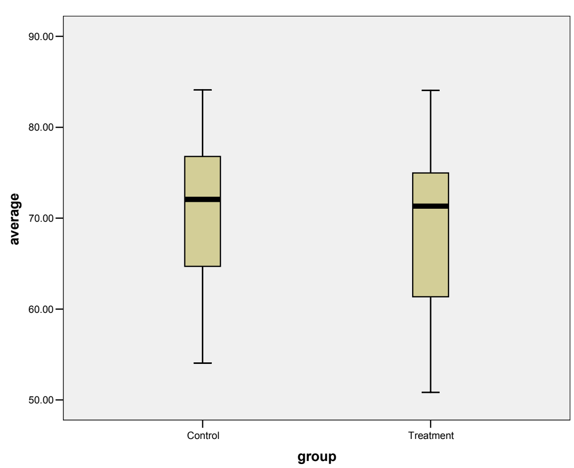


Tests of Normality

	group	Shapiro-Wilk Statistic	df	Sig.
baseline	Control	.979	15	.961
	Treatment	.935	16	.289
year1	Control	.964	15	.766
	Treatment	.918	16	.155
year2	Control	.968	15	.831
	Treatment	.976	16	.919
year3	Control	.957	15	.646
	Treatment	.952	16	.515

- If we want to perform tests on the marginal group means, then we should check normality on the marginal group means.
(Would tests on the marginal group means make sense?)

```
COMPUTE average = SUM(baseline,year1,year2,year3)/4.
EXAMINE VARIABLES= average  BY group
 /PLOT BOXPLOT  SPREADLEVEL.
```



Tests of Normality

	group	Shapiro-Wilk Statistic	df	Sig.
average	Control	.968	15	.827
	Treatment	.950	16	.497

- These data satisfy the normality assumption

- Homogeneity of variances / Sphericity

  - Homogeneity of the variance covariance matrices

```
GLM  baseline year1 year2 year3 BY group
  /WSFACTOR = time 4 Polynomial
  /PRINT = DESCRIPTIVE HOMOGENEITY.
```

Box's Test of Equality of Covariance Matrices

Box's M	17.926
F	1.522
df1	10
df2	3977.868
Sig.	.125

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
baseline	.076	1	29	.784
year1	.042	1	29	.839
year2	.163	1	29	.689
year3	.013	1	29	.911

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

We have no evidence that the variance/covariance matrices are different across the two treatment groups. This assumption is satisfied. We may average the data from the groups together to test within subject effects (marginal time means and time by group interaction effects)

  - Homogeneity of variances for between group tests.
    Necessary for equal variance tests of all between group effects.


```
GLM  baseline year1 year2 year3 BY group
  /WSFACTOR = time 4 Polynomial
  /PRINT = DESCRIPTIVE HOMOGENEITY.
```

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
baseline	.076	1	29	.784
year1	.042	1	29	.839
year2	.163	1	29	.689
year3	.013	1	29	.911

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.


```
COMPUTE average = SUM (baseline,year1,year2,year3)/4.
EXAMINE VARIABLES= average  BY group
  /PLOT BOXPLOT  SPREADLEVEL.
```

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
average	Based on Mean	.104	1	29	.749
	Based on Median	.066	1	29	.799
	Based on Median and with adjusted df	.066	1	28.388	.799
	Based on trimmed mean	.108	1	29	.745

We do not have any evidence that the variances are different across the two groups. We may conduct all between group tests under the assumption that the variances between groups are equal.

- Overall sphericity (averaging over the between subjects factor). Necessary for omnibus tests on the marginal time means and for omnibus time*group interaction tests (Are these tests meaningful?)

```
GLM  baseline year1 year2 year3 BY group
  /WSFACTOR = time 4 Polynomial
  /PRINT = DESCRIPTIVE HOMOGENEITY.
```

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.911	1.000	.333

The data are spherical. We can conduct omnibus tests for the within-subject effect (time) or for between/within subject interactions (group*time).

- Multi-sample sphericity: sphericity within each group/treatment level (the between subjects factor). Necessary for simple effect omnibus tests for the effect of time for the treatment group and the effect of time for the control group (Are these tests meaningful?)

```
SORT CASES BY group .
SPLIT FILE LAYERED BY group .
GLM  baseline year1 year2 year3
  /WSFACTOR = time 4 Polynomial
  /PRINT = DESCRIPTIVE HOMOGENEITY.
SPLIT FILE OFF.
```

Measure: MEASURE_1

group	Within Subjects Effect	Epsilon		
		Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Control	time	.879	1.000	.333
Treatment	time	.779	.932	.333

Within each treatment level, the data are not spherical, but the violation is fixable. We can conduct epsilon-adjusted simple effect omnibus tests for the within-subject effect (time) at each level of the between-subjects factor (group).

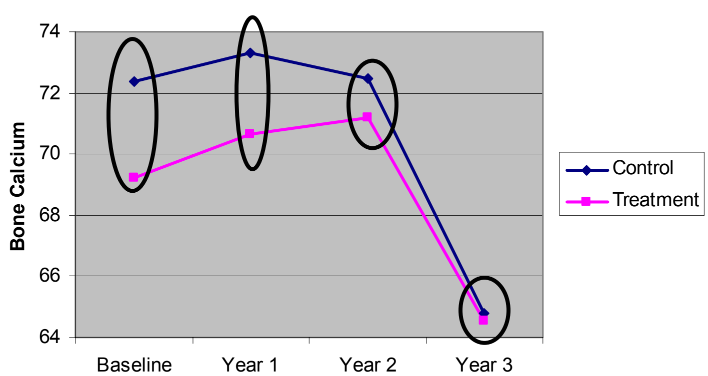
Group	Time: Baseline	Year 1	Year 2	Year 3	
Control	72.38	73.29	72.47	64.79	70.73
Treatment	69.23	70.66	71.18	64.53	68.90
	70.75	71.93	71.81	64.65	

- Conclusions from tests of assumptions:

  Tests of between subjects effects
  - We may perform an omnibus test (and/or standard contrasts) on the marginal between-subjects (group) means.
  - We may perform standard simple-effects tests (in this case, contrasts) for the effect of the between-subjects factor (group) at each level of the within-subjects factor (time).

  Tests of within subjects effects
  - We may perform standard omnibus tests on the marginal within-subjects (time) effect and on the between/within (group by time) interaction.
  - Within each group level, the data are not spherical, but the violation is fixable. We can conduct epsilon-adjusted simple effect omnibus tests for the within-subject effect (time) at each level of the between-subjects factor (group).

- **Question#1:** Does the treatment group have less bone loss than the control group?
  - We can perform tests of the effect of group at each year. The hypotheses about the rate of bone loss are more important – those will be our planned tests. Thus, we will consider these four pairwise comparisons to be post-hoc tests.



- The easiest way to run these tests is as 4 separate independent samples t-tests.

```
T-TEST GROUPS = group(1 2)
 /VARIABLES = baseline year1 year2 year3.
```

Group Statistics

	group	N	Mean	Std. Deviation	Std. Error Mean
baseline	Control	15	72.3800	9.59786	2.47816
	Treatment	16	69.2313	9.89186	2.47297
year1	Control	15	73.2933	9.43803	2.43689
	Treatment	16	70.6563	10.02975	2.50744
year2	Control	15	72.4733	8.47884	2.18923
	Treatment	16	71.1813	9.29245	2.32311
year3	Control	15	64.7867	8.68586	2.24268
	Treatment	16	64.5313	9.02306	2.25577

Independent Samples Test

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
baseline	.898	29	.376	3.14875	3.50450	-4.01876	10.31626
year1	.753	29	.458	2.63708	3.50362	-4.52862	9.80278
year2	.404	29	.690	1.29208	3.20186	-5.25645	7.84062
year3	.080	29	.937	.25542	3.18494	-6.25851	6.76934

$$d_{baseline} = \frac{2*0.898}{\sqrt{29}}.33 \quad d_{Year1} = \frac{2*0.753}{\sqrt{29}}.29 \quad d_{Year2} = \frac{2*0.404}{\sqrt{29}} = .15 \quad d_{Year3} = \frac{2*0.080}{\sqrt{29}}.03$$

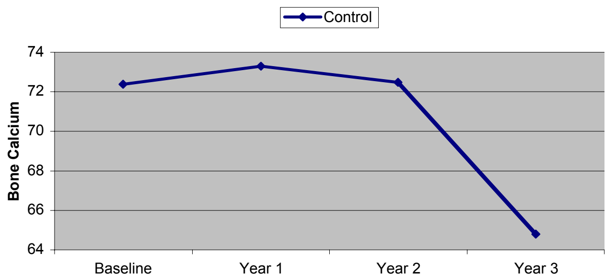
We should apply a Tukey HSD post-hoc correction to these tests. Because none of these tests are significant, it is not necessary to do the calculations, we can report the tests are non-significant with the Tukey HSD procedure. However, for completeness, here is the correction:

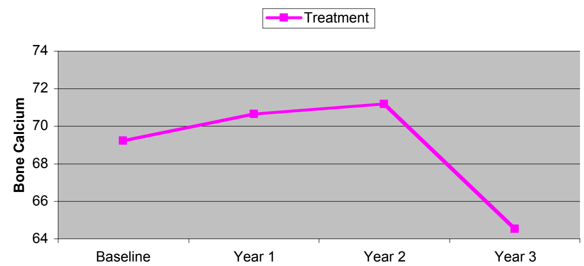
$$t_{crit} = \frac{q(1-\alpha,8,29)}{\sqrt{2}} = \frac{4.613}{\sqrt{2}} = 3.26$$

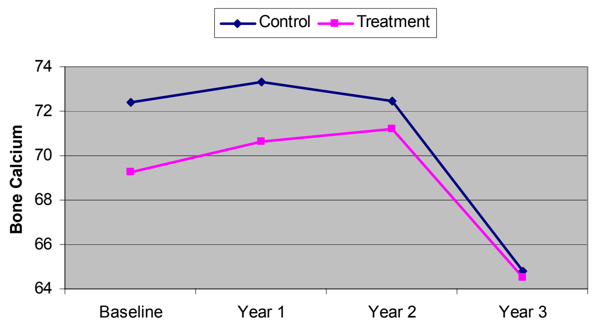
Applying a Tukey HSD correction to these pairwise tests, we find:

- No evidence that the treatment and the control group differed in their calcium bone density at baseline, $t(29) = 0.90, ns, d = .33$.
- No evidence that the treatment and the control group differed in their calcium bone density at the one year follow-up, $t(29) = 0.75, ns, d = .29$.
- No evidence that the treatment and the control group differed in their calcium bone density at the two year follow-up, $t(29) = 0.40, ns, d = .15$.
- No evidence that the treatment and the control group differed in their calcium bone density at the two year follow-up, $t(29) = 0.08, ns, d = .03$.

- **Question #2**: Is the rate of bone loss different between the treatment group and the control group?
    - We can test for:
        - (Downward) polynomial trends in the control condition,
        - (Downward) polynomial trends in the treatment condition,
        - And for differences in the polynomial trends between the groups.







- Although there are many tests here (nine), they are the key tests of the hypotheses and we have a strong theory supporting these hypotheses. Thus, were these my own data, I would not apply a p-value correction to them.
- If you were to apply a correction:
    - These are complex contrasts, so you could use a Scheffé correction:

$$3 * F(\alpha = .05, 3, 29) = 3 * 2.934 = 8.802$$

    - Alternatively, you will be conducting 9 planned contrasts, so a Bonferroni correction could also be appropriate:

$$p_{crit} = \frac{.05}{9} = .0056$$

    - You can select whichever of these two methods is less conservative. In this case, the Bonferroni correction is less conservative by a hair, so were we to apply a correct we should use the Bonferroni correction.

- **Question #2 A and B**: Are there polynomial trends in the control condition? Are there polynomial trends in the treatment condition?
  - These are contrast tests within one level of the between-subjects variable.
  - Method #1: Select the level of the between-subjects variable of interest and conduct polynomial trends on that level.

```
SORT CASES BY group .
SPLIT FILE LAYERED BY group .
GLM  baseline year1 year2 year3
  /WSFACTOR = time 4 Polynomial.
SPLIT FILE OFF.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

group	Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
Control	time	Linear	417.720	1	417.720	40.180	.000
		Quadratic	277.350	1	277.350	32.683	.000
		Cubic	19.763	1	19.763	3.125	.099
	Error(time)	Linear	145.547	14	10.396		
		Quadratic	118.805	14	8.486		
		Cubic	88.540	14	6.324		
Treatment	time	Linear	147.425	1	147.425	38.266	.000
		Quadratic	260.823	1	260.823	241.391	.000
		Cubic	31.500	1	31.500	6.050	.027
	Error(time)	Linear	57.790	15	3.853		
		Quadratic	16.208	15	1.081		
		Cubic	78.106	15	5.207		

$$\eta^2_{Linear-Control} = \frac{417.72}{417.72 + 145.547} = .74 \quad \eta^2_{Quadratic-Control} = \frac{277.35}{277.35 + 118.805} = .70$$

$$\hat{\eta}^2_{Cubic-Control} = \frac{19.763}{19.763 + 88.50} = .18$$

$$\eta^2_{Linear-Treatment} = \frac{147.425}{147.425 + 57.790} = .72 \quad \eta^2_{Quadratic-Treatment} = \frac{260.823}{260.823 + 16.208} = .94$$

$$\hat{\eta}^2_{Cubic-Treatment} = \frac{31.5}{31.5 + 78.106} = .29$$

$F_{Linear-Control}(1,14) = 40.18, p < .01, \eta^2 = .74$

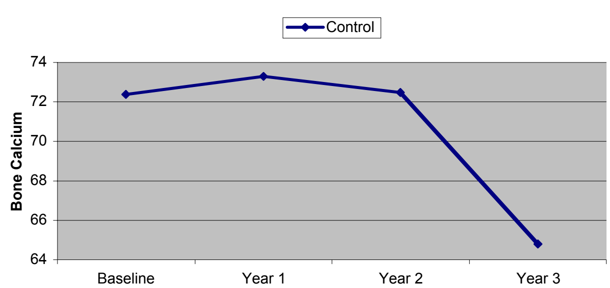
$F_{Quadratic-Control}(1,14) = 32.68, p < .01, \eta^2 = .70$

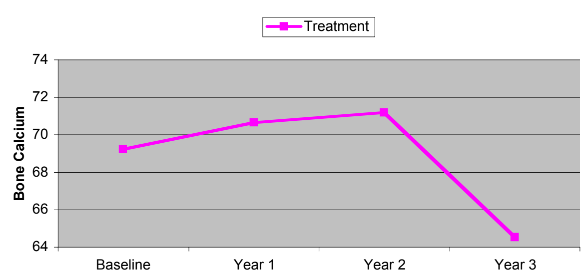
$F_{Cubic-Control}(1,14) = 3.13, p = .10, \eta^2 = .18$

$F_{Linear-Treatment}(1,15) = 38.27, p < .01, \eta^2 = .72$

$F_{Quadratic-Treatment}(1,15) = 241.39, p < .01, \eta^2 = .94$

$F_{Cubic-Treatment}(1,15) = 6.50, p = .03, \eta^2 = .29$





- Advantages of method #1
  - It is easy to run
- Disadvantages of method #1
  - Each test has fewer than *(N-a)* degrees of freedom. If the variances between groups are homogeneous, then we are (voluntarily) sacrificing accuracy and power.

- Method #2: Compute the contrast of interest. Trick SPSS into testing it within each group separately using an error term with information from all between-subjects groups. (This method is only appropriate if you have equal variances between groups).

```
compute linear = -3*baseline + -1*year1 + 1*year2 + 3*year3.
compute quad = 1*baseline + -1*year1 + -1*year2 + 1*year3.
compute cubic = -1*baseline + 3*year1 + -3*year2 + 1*year3.

ONEWAY  linear quad cubic BY group
 /CONTRAST= 1 0
 /CONTRAST= 0 1
 /STATISTICS DESCRIPTIVES HOMOGENEITY.
```

  - What is being tested by the contrast commands?

$$\psi_{Linear:Control} = -3*\mu_{Control:Baseline} + -1*\mu_{Control:Year1} + 1*\mu_{Control:Year2} + 3*\mu_{Control:Year3}$$
$$\psi_{Linear:Treatment} = -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3}$$

Contrast #1

$$H_0: 1*\psi_{Linear:Control} + 0*\psi_{Linear:Treatment} = 0$$
$$H_0: \psi_{Linear:Control} = 0$$
$$H_0: -3*\mu_{Control:Baseline} + -1*\mu_{Control:Year1} + 1*\mu_{Control:Year2} + 3*\mu_{Control:Year3} = 0$$

Contrast #2

$$H_0: 0*\psi_{Linear:Control} + 1*\psi_{Linear:Treatment} = 0$$
$$H_0: \psi_{Linear:Treatment} = 0$$
$$H_0: -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3} = 0$$

  - By using the contrast subcommand, we obtain an error term that uses information from both groups (and has *N-a* dfs).

- Note that the *assume equal variances* tests have *N-a* dfs and that the *does not assume equal variances* tests are identical to Method 1 where we ran the contrast only on the (between-subjects) group of interest.

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
linear	Assume equal variances	1	-23.6000[a]	3.05758	-7.719	29	.000
		2	-13.5750[a]	2.96049	-4.585	29	.000
	Does not assume equal variances	1	-23.6000[a]	3.72312	-6.339	14.000	.000
		2	-13.5750[a]	2.19449	-6.186	15.000	.000
quad	Assume equal variances	1	-8.6000[a]	1.11422	-7.718	29	.000
		2	-8.0750[a]	1.07884	-7.485	29	.000
	Does not assume equal variances	1	-8.6000[a]	1.50431	-5.717	14.000	.000
		2	-8.0750[a]	.51974	-15.537	15.000	.000
cubic	Assume equal variances	1	-5.1333[a]	2.76800	-1.855	29	.074
		2	-6.2750[a]	2.68011	-2.341	29	.026
	Does not assume equal variances	1	-5.1333[a]	2.90385	-1.768	14.000	.099
		2	-6.2750[a]	2.55123	-2.460	15.000	.027

a. The sum of the contrast coefficients is not zero.

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

- Assume Equal Variance Tests
  N-a degrees of freedom

$$t_{Linear-Control}(29) = -7.72, p < .01, r = .82$$
$$t_{Quadratic-Control}(29) = -7.72, p < .01, r = .82$$
$$t_{Cubic-Control}(29) = -1.86, p = .07, r = .33$$

$$t_{Linear-Treatment}(29) = -4.59, p < .01, r = .65$$
$$t_{Quadratic-Treatment}(29) = -7.49, p < .01, r = .81$$
$$t_{Cubic-Treatment}(29) = -2.34, p = .03, r = .40$$

- Does Not Assume Equal Variance Tests
  Matches Method #1 output
  n_j-*1* degrees of freedom

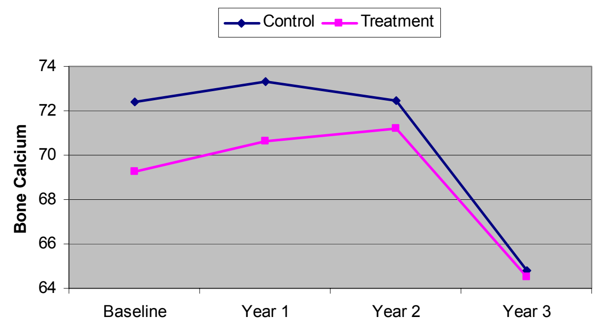
$$t_{Linear-Control}(14) = -6.34, p < .01, r = .86$$
$$t_{Quadratic-Control}(14) = -5.72, p < .01, r = .84$$
$$t_{Cubic-Control}(14) = -1.77, p = .10, r = .42$$

$$t_{Linear-Treatment}(15) = -6.19, p < .01, r = .85$$
$$t_{Quadratic-Treatment}(15) = -15.54, p < .01, r = .97$$
$$t_{Cubic-Treatment}(15) = -2.46, p = .03, r = .53$$

- Which method is right? It depends!
  If the variances between groups are equal, then we should pool the error term to include information from both groups. This procedure results in more accurate error estimates and tests with greater power.
  If the variances between groups are unequal, then we should not pool the error term and we should only use information from the group of interest to calculate the error term.

- Advantages of method #2
  - It gives us both equal variance and unequal variance output
- Disadvantages of method #2
  - More time consuming to run than method #1

- **Question #2 C**: Are there differences in the polynomial trends between the groups?



Linear: Treatment

Group	Time Baseline	Year 1	Year 2	Year 3
Control				
Treatment	-3	-1	1	3

Linear: Control

Group	Time Baseline	Year 1	Year 2	Year 3
Control	-3	-1	1	3
Treatment				

Linear: Treatment – Linear: Control

Group	Time Baseline	Year 1	Year 2	Year 3
Control	3	1	-1	-3
Treatment	-3	-1	1	3

$$\psi_{Linear:Treatment} = -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3}$$

$$\psi_{Linear-Control} = -3*\mu_{Control:Baseline} + -1*\mu_{Control:Year1} + 1*\mu_{Control:Year2} + 3*\mu_{Control:Year3}$$

$$\psi_{Linear:Treatment-Control} = \psi_{Linear:Treatment} - \psi_{Linear:Control}$$

$$= -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3}$$
$$-1(-3*\mu_{Control:Baseline} + -1*\mu_{Control:Year1} + 1*\mu_{Control:Year2} + 3*\mu_{Control:Year3})$$

$$= -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3}$$
$$3*\mu_{Control:Baseline} + 1*\mu_{Control:Year1} - 1*\mu_{Control:Year2} - 3*\mu_{Control:Year3}$$

- We can repeat this procedure for differences in the quadratic and cubic trends

- Method #1: Examine the interaction between the polynomial trends on time (the repeated measures factor) and condition.

```
GLM  baseline year1 year2 year3 BY group
  /WSFACTOR = time 4 Polynomial.
```

Time * Group (Linear)

Group	Baseline	Year 1	Year 2	Year 3	
Control					-1
Treatment					1
	-3	-1	1	3	

Group	Baseline	Year 1	Year 2	Year 3
Control	3	1	-1	-3
Treatment	-3	-1	1	3

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	534.960	1	534.960	76.296	.000
	Quadratic	538.172	1	538.172	115.597	.000
	Cubic	50.381	1	50.381	8.767	.006
time * group	Linear	38.903	1	38.903	5.548	.025
	Quadratic	.533	1	.533	.115	.737
	Cubic	.505	1	.505	.088	.769
Error(time)	Linear	203.337	29	7.012		
	Quadratic	135.013	29	4.656		
	Cubic	166.645	29	5.746		

$$\eta^2_{DiffernceInLinear} = \frac{38.903}{38.903 + 203.337} = .16 \quad \eta^2_{DiffernceInQuadratic} = \frac{0.533}{0.533 + 135.013} < .01$$

$$\eta^2_{DiffernceInCubic} = \frac{0.505}{0.505 + 166.645} < .01$$

Difference in linear trends: $F(1,29) = 5.55, p = .03, \eta^2 = .16$
Difference in quadratic trends: $F(1,29) = 0.12, p = .74, \eta^2 < .01$
Difference in cubic trends: $F(1,29) = 0.09, p = .77, \eta^2 < .01$

- Advantages of method #1
  - Easy to run
- Disadvantages of method #1
  - Only works (provides a *1 df* contrast test of difference between polynomial trends) when *a=2*.

- Method #2: Compute the contrast of interest and (manually) ask for a comparison between the treatment group and the control group.

```
ONEWAY  linear quad cubic BY group
 /CONTRAST= -1 1.
```

  - What is being tested by the contrast command?

$$\psi_{Linear:Control} = -3*\mu_{Control:Baseline} + -1*\mu_{Control:Year1} + 1*\mu_{Control:Year2} + 3*\mu_{Control:Year3}$$
$$\psi_{Linear:Treatment} = -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3}$$

$$H_0 : -1*\psi_{Linear:Control} + 1*\psi_{Linear:Treatment} = 0$$

$$H_0 : -1*(-3*\mu_{Control:Baseline} + -1*\mu_{Control:Year1} + 1*\mu_{Control:Year2} + 3*\mu_{Control:Year3}) + 1*(-3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3}) = 0$$

$$H_0 : 3*\mu_{Control:Baseline} + 1*\mu_{Control:Year1} + -1*\mu_{Control:Year2} + -3*\mu_{Control:Year3} + -3*\mu_{Treatment:Baseline} + -1*\mu_{Treatment:Year1} + 1*\mu_{Treatment:Year2} + 3*\mu_{Treatment:Year3} = 0$$

	Time			
Group	Baseline	Year 1	Year 2	Year 3
Control	3	1	-1	-3
Treatment	-3	-1	1	3

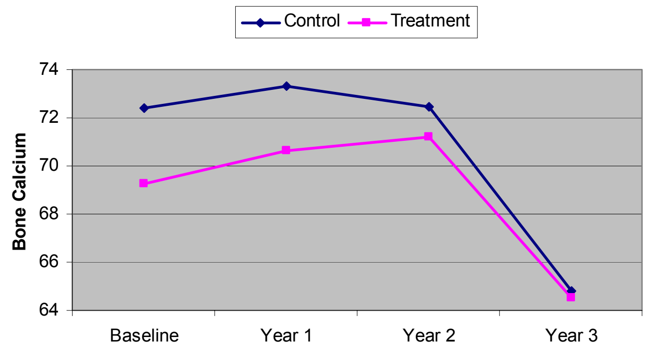
  - Thus, the contrast command tests for a difference in linear, quadratic, and cubic trends between the control and treatment groups (exactly the same as Method #1).

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
linear	1	10.0250	4.25597	2.356	29	.025
quad	1	.5250	1.55093	.339	29	.737
cubic	1	-1.1417	3.85290	-.296	29	.769

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

$$r_{LinearDiff} = \sqrt{\frac{2.356^2}{2.356^2 + 29}} = .40 \quad r_{QuadDiff} = \sqrt{\frac{0.339^2}{0.339^2 + 29}} = .06 \quad r_{CubicDiff} = \sqrt{\frac{0.296^2}{0.296^2 + 29}} = .05$$



Difference in linear trends: $t(29) = 2.36, p = .03, r = .40$
Difference in quadratic trends: $t(29) = 0.34, p = .74, r = .06$
Difference in cubic trends: $t(29) = 0.30, p = .77, r = .05$

- Advantages of method #2
  - Can be used to test for differences in trends when there are more than 2 between-subject groups in the factor ($a>2$).
  - Also provides output to test the contrasts when the variance between groups is not homogeneous
- Disadvantages of method #2
  - More time consuming to run than method #1.

- Conclusions
  - **Question #2**: Is the rate of calcium loss different between the treatment group and the control group?
    - Yes. There are significant linear, quadratic, and (significant or marginally significant) cubic trends in calcium bone loss for both the treatment and control group. These trends indicate that over time, participants in both groups are losing calcium in their bones. However, the *linear rate of calcium bone loss* is stronger in the control group than in the treatment group. Thus, there is some evidence that the treatment is associated with less bone loss.
  - **Question #1**: Does the treatment group have less calcium loss than the control group?
    - No. At the same time, there were no differences in bone calcium levels at any of the follow-up assessments.

- This example is an illustration of growth curve analysis. In growth curve analysis, the rate/pattern of change over time is modeled and usually compared between 2 or more groups.

Appendix
Two Additional Between/Within Examples

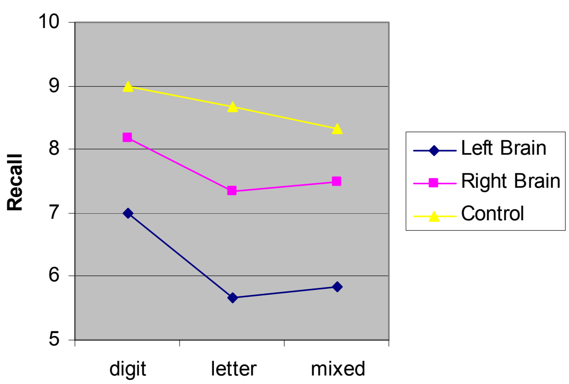
9. : Effects of brain damage on memory (3 * 3)

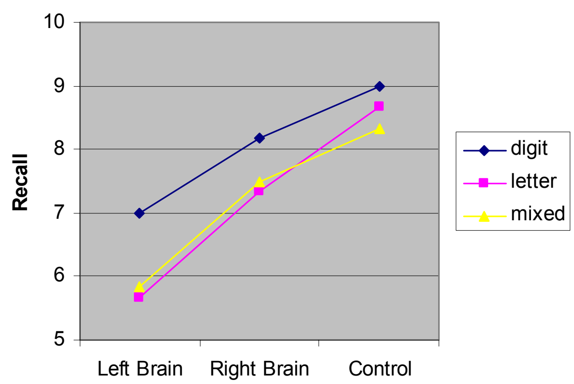
- A neuropsychologist is exploring short-term memory deficits in brain-damaged individuals. Patients were classified as either having left-hemisphere damage, right-hemisphere damage, or no damage (control).
 Participants viewed stimuli consisting of string of all digits, all letter, and mixed letters and digits. The longest string that each participant could remember in each condition is listed below:

	Stimuli					
Damage	Digits		Letters		Mixed	
Left Brain	6	8	5	5	6	8
	8	6	7	4	5	7
	7	7	7	6	4	5
Right Brain	9	7	8	8	6	8
	8	7	8	6	7	7
	9	9	7	8	8	7
Control	8	9	8	7	7	9
	10	8	9	8	9	8
	9	10	10	10	8	9

- The researcher would like to know:
  - Does recall vary by type of stimuli?
  - Does this difference vary by type of brain damage?

  - Does recall vary by type of brain damage?
  - Does this difference vary by type stimuli?





- Tests of assumptions
  - Normality: Cell means

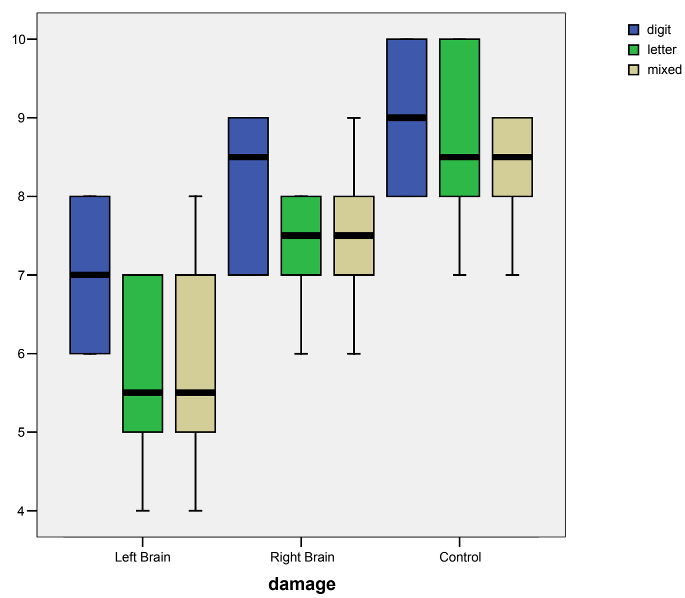
```
EXAMINE VARIABLES=digit letter mixed BY damage
  /PLOT BOXPLOT STEMLEAF NPPLOT
  /COMPARE GROUP.
```

Descriptives

	damage		Statistic	Std. Error
digit	Left Brain	Skewness	.000	.845
		Kurtosis	-1.875	1.741
	Right Brain	Skewness	-.456	.845
		Kurtosis	-2.390	1.741
	Control	Skewness	.000	.845
		Kurtosis	-1.875	1.741
letter	Left Brain	Skewness	-.075	.845
		Kurtosis	-1.550	1.741
	Right Brain	Skewness	-.857	.845
		Kurtosis	-.300	1.741
	Control	Skewness	-.075	.845
		Kurtosis	-1.550	1.741
mixed	Left Brain	Skewness	.418	.845
		Kurtosis	-.859	1.741
	Right Brain	Skewness	.000	.845
		Kurtosis	-.248	1.741
	Control	Skewness	-.857	.845
		Kurtosis	-.300	1.741

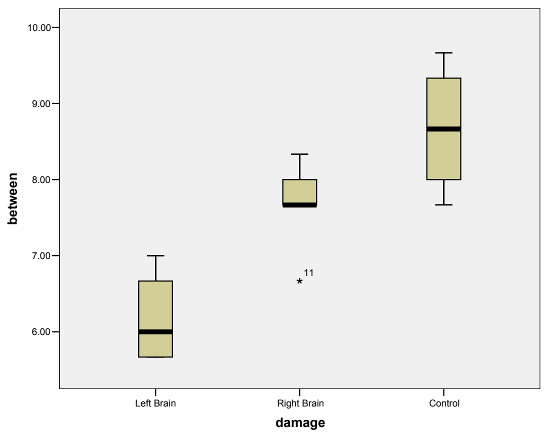
Tests of Normality

		Shapiro-Wilk		
	DAMAGE	Statistic	df	Sig.
DIGIT	Left Brain	.853	6	.167
	Right Brain	.775	6	.035
	Control	.853	6	.167
LETTER	Left Brain	.907	6	.415
	Right Brain	.822	6	.091
	Control	.907	6	.415
MIXED	Left Brain	.958	6	.804
	Right Brain	.960	6	.820
	Control	.822	6	.091



- Normality: Marginal between-subjects means

```
COMPUTE between = (digit + letter + mixed)/3.
EXAMINE VARIABLES=between BY damage
  /PLOT BOXPLOT  SPREADLEVEL.
```



Tests of Normality

	damage	Shapiro-Wilk Statistic	df	Sig.
between	Left Brain	.863	6	.201
	Right Brain	.873	6	.238
	Control	.950	6	.739

Descriptives

	damage		Statistic	Std. Error
between	Left Brain	Skewness	.811	.845
		Kurtosis	-1.029	1.741
	Right Brain	Skewness	-1.153	.845
		Kurtosis	2.500	1.741
	Control	Skewness	.000	.845
		Kurtosis	-1.875	1.741

  - The data look relatively symmetrical

- Homogeneity of variances / Sphericity
  - Homogeneity of variances for between group tests:

```
GLM digit letter mixed BY damage
  /WSFACTOR = recall 3
  /PRINT = DESC HOMO.
```

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
DIGIT	.250	2	15	.782
LETTER	1.000	2	15	.391
MIXED	1.250	2	15	.315

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

```
COMPUTE between = (digit + letter + mixed)/3.
EXAMINE VARIABLES=between BY damage
  /PLOT BOXPLOT  SPREADLEVEL.
```

Test of Homogeneity of Variance

	Levene Statistic	df1	df2	Sig.
BETWEEN	1.573	2	15	.240

We do not have any evidence that the variances are different across the between-subjects groups. This assumption is satisfied.

- Homogeneity of the variance covariance matrices

```
GLM digit letter mixed BY damage
  /WSFACTOR = recall 3
  /PRINT = DESC HOMO.
```

Box's Test of Equality of Covariance Matrices

Box's M	24.372
F	1.422
df1	12
df2	1090.385
Sig.	.149

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
DIGIT	.250	2	15	.782
LETTER	1.000	2	15	.391
MIXED	1.250	2	15	.315

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

We do not have any evidence that the variance/covariance matrices are different across the three groups. This assumption is satisfied.

- Overall sphericity (averaging over the between subjects factor):

```
GLM digit letter mixed BY damage
  /WSFACTOR = recall 3
  /PRINT = HOMOGENIETY.
```

Mauchly's Test of Sphericity

Measure: MEASURE_1

	Epsilon		
Within Subjects Effect	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
RECALL	.689	.837	.500

The data are not spherical and the violation is severe. We cannot conduct omnibus tests for the within-subject effect (recall) or for between/within subject interactions (recall*stimuli).

- Multi-Sample Sphericity: Sphericity at each level of the between subjects factor:

```
SORT CASES BY damage .
SPLIT FILE LAYERED BY damage .
GLM digit letter mixed
  /WSFACTOR = recall 3.
SPLIT FILE OFF.
```

Mauchly's Test of Sphericity

Measure: MEASURE_1

		Epsilon		
damage	Within Subjects Effect	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Left Brain	recall	.545	.581	.500
Right Brain	recall	.937	1.000	.500
Control	recall	.571	.630	.500

The data are only spherical for patients with right brain damage. For the other two groups, the data are not spherical and the violation is severe and unfixable. If we want to use the same methods to test effects at each level, then we cannot conduct simple effect omnibus tests for the within-subject effect (recall) within each level of the between-subjects factor (stimuli).

- Conclusions from tests of assumptions:
  - We may perform an omnibus test and/or standard contrasts on the marginal between-subjects (damage) means.
  - We may not perform any omnibus tests involving within-subjects effects. Tests on the marginal within-subjects (stimuli) means or on the between/within interaction (damage by stimuli) must use a contrast-specific error term.

	Stimuli			
Damage	Digits	Letters	Mixed	
Left Brain	$\bar{X} = 7.00$	$\bar{X} = 5.67$	$\bar{X} = 5.83$	$\bar{X} = 6.17$
Right Brain	$\bar{X} = 8.17$	$\bar{X} = 7.33$	$\bar{X} = 7.50$	$\bar{X} = 7.67$
Control	$\bar{X} = 9.00$	$\bar{X} = 8.67$	$\bar{X} = 8.33$	$\bar{X} = 8.67$
	$\bar{X} = 8.06$	$\bar{X} = 7.22$	$\bar{X} = 7.22$	

- Hypothesis testing:
  - The researcher is basically asking for all possible tests of interest to be conducted. We will consider all tests to be exploratory (post-hoc).

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\bar{X} = 7.00$	$\bar{X} = 5.67$	$\bar{X} = 5.83$	$\bar{X} = 6.17$
Right Brain	$\bar{X} = 8.17$	$\bar{X} = 7.33$	$\bar{X} = 7.50$	$\bar{X} = 7.67$
Control	$\bar{X} = 9.00$	$\bar{X} = 8.67$	$\bar{X} = 8.33$	$\bar{X} = 8.67$
	$\bar{X} = 8.06$	$\bar{X} = 7.22$	$\bar{X} = 7.22$	

  - Does recall vary by type of stimuli?
    - Main effect for stimuli (Within subject effect)
      We cannot conduct a standard omnibus test
      We will conduct pairwise tests on marginal (within-subject) stimuli means.

  - Does this difference vary by type of brain damage?
    - Interaction between damage and stimuli (between by within effect)
      We cannot conduct a standard interaction omnibus test
      We will conduct pairwise tests on the effect of stimuli within each level of brain damage.

  - Does recall vary by type of brain damage?
    - Main effect for brain damage (between subject effect)
      We can conduct a standard omnibus test
      We will follow this test with pairwise tests on marginal (between-subject) brain damage means.

  - Does this difference vary by type of stimuli?
    - Interaction between damage and stimuli (between by within effect)
      We cannot conduct a standard omnibus interaction test
      We can examine the simple effect of brain damage within each level of stimuli and follow each test with pairwise comparisons to identify differences.

- Does recall vary by type of stimuli?
  We will conduct pairwise tests on marginal stimuli means.

Damage	Stimuli: Digits	Letters	Mixed	
Left Brain	$\overline{X} = 7.00$	$\overline{X} = 5.67$	$\overline{X} = 5.83$	$\overline{X} = 6.17$
Right Brain	$\overline{X} = 8.17$	$\overline{X} = 7.33$	$\overline{X} = 7.50$	$\overline{X} = 7.67$
Control	$\overline{X} = 9.00$	$\overline{X} = 8.67$	$\overline{X} = 8.33$	$\overline{X} = 8.67$
	$\overline{X} = 8.06$	$\overline{X} = 7.22$	$\overline{X} = 7.22$	

```
GLM digit letter mixed BY damage
  /WSFACTOR = recall 3 simple (1).
GLM digit letter mixed BY damage
  /WSFACTOR = recall 3 simple (2).
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	recall	Type III Sum of Squares	df	Mean Square	F	Sig.
recall	Level 2 vs. Level 1	12.500	1	12.500	10.714	.005
	Level 3 vs. Level 1	12.500	1	12.500	7.353	.016
Error(recall)	Level 2 vs. Level 1	17.500	15	1.167		
	Level 3 vs. Level 1	25.500	15	1.700		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	recall	Type III Sum of Squares	df	Mean Square	F	Sig.
recall	Level 3 vs. Level 2	.000	1	.000	.000	1.000
Error(recall)	Level 3 vs. Level 2	53.000	15	3.533		

$$\eta^2_{DigitsV.Letters} = \frac{12.5}{12.5+17.5} = .42 \quad \eta^2_{DigitsV.Mixed} = \frac{12.5}{12.5+25.5} = .33 \quad \eta^2_{LettersV.Mixed} = \frac{0}{0+535} = .0$$

Tukey HSD critical value: $F_{crit} = \frac{(q(.05,3,15))^2}{2} = \frac{(3.67)^2}{2} = 6.747$

Digits vs. Letters: $F(1,15) = 10.71, p < .05, \eta^2 = .42$
Digits vs. Mixed: $F(1,15) = 7.35, p < .05, \eta^2 = .33$
Letters vs. Mixed: $F(1,15) = 0.00, ns, \eta^2 = 0.00$

- Does this difference in recall of types of stimuli vary by type of brain damage?
  - We want to repeat the three contrasts we just ran, but we want to look within each level of brain damage (rather than averaging over the types of damage).

Digits vs. Letters

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\overline{X} = 7.00$	$\overline{X} = 5.67$	$\overline{X} = 5.83$	$\overline{X} = 6.17$
Right Brain	$\overline{X} = 8.17$	$\overline{X} = 7.33$	$\overline{X} = 7.50$	$\overline{X} = 7.67$
Control	$\overline{X} = 9.00$	$\overline{X} = 8.67$	$\overline{X} = 8.33$	$\overline{X} = 8.67$
	$\overline{X} = 8.06$	$\overline{X} = 7.22$	$\overline{X} = 7.22$	

Damage	Digit-Letters
Left Brain	1.33
Right Brain	0.84
Control	0.33

To examine these differences at each level of brain damage, we can compute the difference of interest and use the *ONEWAY* command and the *CONTRAST* subcommand:

```
Compute dig_let = digit - letter.
ONEWAY dig_let by damage
  /STAT = DESC
  /CONT = 1 0 0
  /CONT = 0 1 0
  /CONT = 0 0 1.
```

ANOVA

dig_let

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.000	2	1.500	1.286	.305
Within Groups	17.500	15	1.167		
Total	20.500	17			

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
dig_let	1	1.3333	.44096	3.024	15	.009
	2	.8333	.44096	1.890	15	.078
	3	.3333	.44096	.756	15	.461

$$t_{crit} = \frac{q(.05/3,3,15)}{\sqrt{2}} = \frac{4.473}{\sqrt{2}} = 3.163 \quad t_{crit} = \frac{q(.10/3,3,15)}{\sqrt{2}} = \frac{3.973}{\sqrt{2}} = 2.809$$

$$d_{LeftBrain} = \frac{1.33}{1.29} = 1.03 \quad d_{RightBrain} = \frac{0.833}{1.29} = 0.64 \quad d_{Control} = \frac{.3333}{1.29} = 0.26$$

Digits vs. Letters

Left Brain: $t(15) = 3.02, p < .10, d = 1.03$

Right Brain: $t(15) = 1.89, ns, d = 0.64$

No Damage: $t(15) = 0.76, ns, d = 0.26$

Digits vs. Mixed

Damage	Stimuli: Digits	Letters	Mixed	
Left Brain	$\bar{X} = 7.00$	$\bar{X} = 5.67$	$\bar{X} = 5.83$	$\bar{X} = 6.17$
Right Brain	$\bar{X} = 8.17$	$\bar{X} = 7.33$	$\bar{X} = 7.50$	$\bar{X} = 7.67$
Control	$\bar{X} = 9.00$	$\bar{X} = 8.67$	$\bar{X} = 8.33$	$\bar{X} = 8.67$
	$\bar{X} = 8.06$	$\bar{X} = 7.22$	$\bar{X} = 7.22$	

Damage	Digit-Mixed
Left Brain	1.167
Right Brain	0.667
Control	0.667

```
Compute dig_mix = digit - mixed.
ONEWAY dig_mix by damage
 /STAT = DESC
 /CONT = 1 0 0
 /CONT = 0 1 0
 /CONT = 0 0 1.
```

ANOVA

dig_mix

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.000	2	.500	.294	.749
Within Groups	25.500	15	1.700		
Total	26.500	17			

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
dig_mix	1	1.1667	.53229	2.192	15	.045
	2	.6667	.53229	1.252	15	.230
	3	.6667	.53229	1.252	15	.230

$$t_{crit} = \frac{q(.05/3,3,15)}{\sqrt{2}} = \frac{4.473}{\sqrt{2}} = 3.163 \qquad t_{crit} = \frac{q(.10/3,3,15)}{\sqrt{2}} = \frac{3.973}{\sqrt{2}} = 2.809$$

$$d_{LeftBrain} = \frac{1.167}{1.303} = 0.90 \qquad d_{RightBrain} = \frac{0.667}{1.303} = 0.51 \qquad d_{Control} = \frac{.667}{1.303} = 0.52$$

Digits vs. Mixed

Left Brain: $t(15) = 2.19, ns, d = .90$

Right Brain: $t(15) = 1.25, ns, d = .52$

No Damage: $t(15) = 1.25, ns, d = .52$

Letter vs. Mixed

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\bar{X} = 7.00$	$\bar{X} = 5.67$	$\bar{X} = 5.83$	$\bar{X} = 6.17$
Right Brain	$\bar{X} = 8.17$	$\bar{X} = 7.33$	$\bar{X} = 7.50$	$\bar{X} = 7.67$
Control	$\bar{X} = 9.00$	$\bar{X} = 8.67$	$\bar{X} = 8.33$	$\bar{X} = 8.67$
	$\bar{X} = 8.06$	$\bar{X} = 7.22$	$\bar{X} = 7.22$	

Damage	Letter-Mixed
Left Brain	0.167
Right Brain	0.167
Control	0.333

```
Compute let_mix = letter - mixed.
ONEWAY let_mix by damage
 /STAT = DESC
 /CONT = 1 0 0
 /CONT = 0 1 0
 /CONT = 0 0 1.
```

ANOVA

let_mix

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.000	2	.500	.142	.869
Within Groups	53.000	15	3.533		
Total	54.000	17			

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
let_mix	1	-.1667	.76739	-.217	15	.831
	2	-.1667	.76739	-.217	15	.831
	3	.3333	.76739	.434	15	.670

$$t_{crit} = \frac{q(.05/3,3,15)}{\sqrt{2}} = \frac{4.473}{\sqrt{2}} = 3.163 \qquad t_{crit} = \frac{q(.10/3,3,15)}{\sqrt{2}} = \frac{3.973}{\sqrt{2}} = 2.809$$

$$d_{LeftBrain} = \frac{0.1667}{1.880} = 0.09 \qquad d_{RightBrain} = \frac{0.1667}{1.880} = 0.09 \qquad d_{Control} = \frac{.333}{1.880} = 0.18$$

Letters vs. Mixed

Left Brain: $t(15) = -0.22, ns, d = .09$

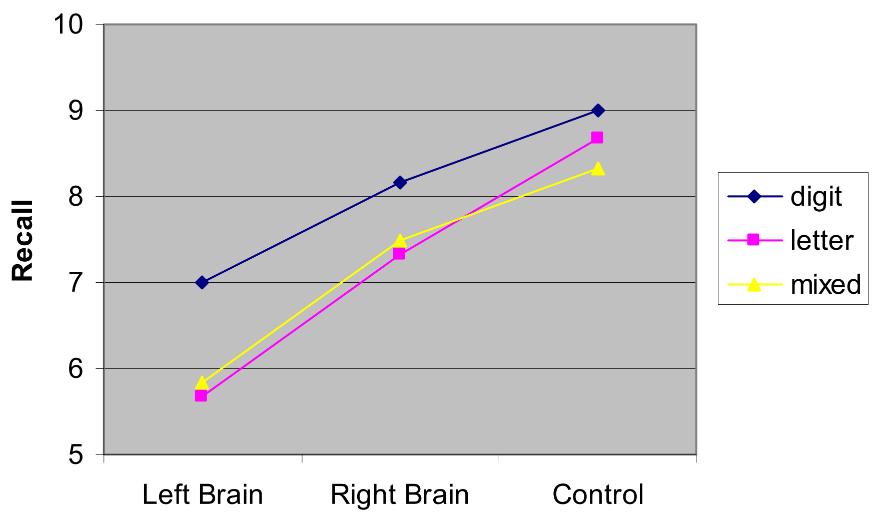
Right Brain: $t(15) = -0.22, ns, d = .09$

No Damage: $t(15) = 0.43, ns, d = 18$

Overall, the effects are relatively consistent within each level of brain damage, although there is some (marginal) evidence that the advantage of digits over letters is stronger in left-brain damaged individuals than in control or right-brain damaged participants.

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\overline{X} = 7.00$	$\overline{X} = 5.67$	$\overline{X} = 5.83$	$\overline{X} = 6.17$
Right Brain	$\overline{X} = 8.17$	$\overline{X} = 7.33$	$\overline{X} = 7.50$	$\overline{X} = 7.67$
Control	$\overline{X} = 9.00$	$\overline{X} = 8.67$	$\overline{X} = 8.33$	$\overline{X} = 8.67$
	$\overline{X} = 8.06^{bc}$	$\overline{X} = 7.22^{b}$	$\overline{X} = 7.22^{c}$	

Note: Within each row, means with a common subscript are significantly different from each other.



- Does recall vary by type of brain damage?
  - Main effect for brain damage (between subject effect)
    We will follow-up this test with pairwise tests on marginal brain damage means.

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\overline{X} = 7.00$	$\overline{X} = 5.67$	$\overline{X} = 5.83$	$\overline{X} = 6.17$
Right Brain	$\overline{X} = 8.17$	$\overline{X} = 7.33$	$\overline{X} = 7.50$	$\overline{X} = 7.67$
Control	$\overline{X} = 9.00$	$\overline{X} = 8.67$	$\overline{X} = 8.33$	$\overline{X} = 8.67$
	$\overline{X} = 8.06$	$\overline{X} = 7.22$	$\overline{X} = 7.22$	

```
GLM digit letter mixed BY damage
  /WSFACTOR = recall 3
  /POSTHOC = damage (TUKEY)
```

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	3037.500	1	3037.500	2462.838	.000
DAMAGE	57.000	2	28.500	23.108	.000
Error	18.500	15	1.233		

$$\eta^2 = \frac{57}{57 + 18.5} = .75$$

There is a significant main effect for brain damage, $F(2,15) = 23.10, p < .01, \eta^2 = .76$. Overall, recall varies by type of brain damage.

Multiple Comparisons

Dependent Variable: BETWEEN
Tukey HSD

(I) DAMAGE	(J) DAMAGE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Left Brain	Right Brain	-1.5000*	.37019	.003	-2.4615	-.5385
	Control	-2.5000*	.37019	.000	-3.4615	-1.5385
Right Brain	Left Brain	1.5000*	.37019	.003	.5385	2.4615
	Control	-1.0000*	.37019	.041	-1.9615	-.0385
Control	Left Brain	2.5000*	.37019	.000	1.5385	3.4615
	Right Brain	1.0000*	.37019	.041	.0385	1.9615

*. The mean difference is significant at the .050 level.

SPSS computes appropriate Tukey-HSD adjusted p-values.

$$d_{LeftVRight} = \frac{1.5}{1.110} = 1.35 \quad d_{LeftVControl} = \frac{2.5}{1.110} = 2.25 \quad d_{RightVControl} = \frac{1}{1.110} = 0.9$$

Left-brain vs. right brain: $t(15) = 4.05, p < .01, d = 1.35$
Left-brain vs. control: $t(15) = 6.75, p < .01, d = 2.25$
Right-brain vs. control: $t(15) = 2.70, p = .04, d = 0.90$

- These exact same tests can be conducted by manually averaging over the within-subjects factor and conducting an ANOVA on this average variable.

```
COMPUTE between = (digit + letter + mixed)/3.
ONEWAY between by damage
  /POSTHOC = TUKEY.
```

- Does this difference vary by type of stimuli?
  - We can examine the simple effect of brain damage within each level of stimuli and follow each test with pairwise comparisons to identify differences.

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\bar{X} = 7.00$	$\bar{X} = 5.67$	$\bar{X} = 5.83$	$\bar{X} = 6.17$
Right Brain	$\bar{X} = 8.17$	$\bar{X} = 7.33$	$\bar{X} = 7.50$	$\bar{X} = 7.67$
Control	$\bar{X} = 9.00$	$\bar{X} = 8.67$	$\bar{X} = 8.33$	$\bar{X} = 8.67$
	$\bar{X} = 8.06$	$\bar{X} = 7.22$	$\bar{X} = 7.22$	

  - To conduct simple effects within each level of stimuli, we can select the appropriate level and run an omnibus test comparing the levels of damage. (Note that we can select a level of stimuli because stimuli is a within-subjects factor. For within subjects factors, we compute error estimates based only on information involved in the comparison).

  - Simple effect of damage for digits only:

```
ONEWAY digit by damage
 /STAT = DESC
 /CONTRAST = -1 1 0
 /CONTRAST = -1 0 1
 /CONTRAST =  0 -1 1.
```

ANOVA

DIGIT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12.111	2	6.056	7.078	.007
Within Groups	12.833	15	.856		
Total	24.944	17			

$$p_{crit} = \frac{.05}{3} = .0167 \qquad \eta^2 = \frac{12.111}{12.111 + 12.833} = .49$$

There is a significant simple effect for brain damage on recall of digits only, $F(2,15) = 7.08, p < .05, \eta^2 = .49$. Overall, recall of digits only varies by type of brain damage.

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
digit	1	1.1667	.53403	2.185	15	.045
	2	2.0000	.53403	3.745	15	.002
	3	.8333	.53403	1.560	15	.139

$$t_{crit} = \frac{q(.05/3,3,15)}{\sqrt{2}} = \frac{4.473}{\sqrt{2}} = 3.163$$

$$d_{LeftVRight} = \frac{1.1667}{.925} = 1.26 \quad d_{LeftVControl} = \frac{2}{.9252} = 2.16 \quad d_{RightVControl} = \frac{.8333}{.9252} = 0.90$$

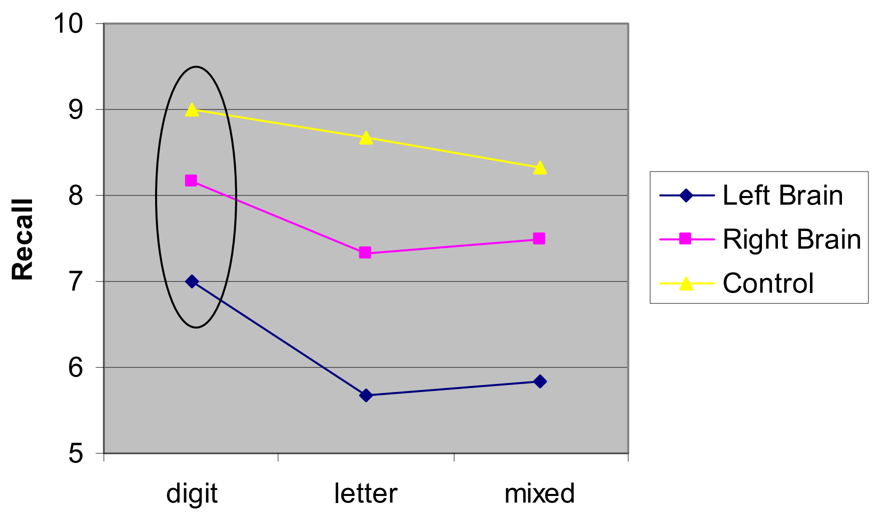
Recall of digits:

Left-brain vs. right brain: $t(15) = 2.18, ns, d = 1.26$

Left-brain vs. control: $t(15) = 3.75, p < .05, d = 2.16$

Right-brain vs. control: $t(15) = 1.56, ns, d = 0.90$

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\bar{X} = 7.00$	$\bar{X} = 5.67$	$\bar{X} = 5.83$	$\bar{X} = 6.17$
Right Brain	$\bar{X} = 8.17$	$\bar{X} = 7.33$	$\bar{X} = 7.50$	$\bar{X} = 7.67$
Control	$\bar{X} = 9.00$	$\bar{X} = 8.67$	$\bar{X} = 8.33$	$\bar{X} = 8.67$
	$\bar{X} = 8.06$	$\bar{X} = 7.22$	$\bar{X} = 7.22$	



- Simple effect of damage for letters only:

```
ONEWAY letter by damage
 /STAT = DESC
 /CONTRAST = -1 1 0
 /CONTRAST = -1 0 1
 /CONTRAST =  0 -1 1.
```

ANOVA

LETTER

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	27.111	2	13.556	11.296	.001
Within Groups	18.000	15	1.200		
Total	45.111	17			

$$p_{crit} = \frac{.05}{3} = .0167 \qquad \eta^2 = \frac{27.111}{27.111 + 18.000} = .60$$

There is a significant simple effect for brain damage on recall of letters only, $F(2,15) = 11.30, p < .05, \eta^2 = .60$. Overall, recall of letters only varies by type of brain damage.

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
letter	1	1.6667	.63246	2.635	15	.019
	2	3.0000	.63246	4.743	15	.000
	3	1.3333	.63246	2.108	15	.052

$$t_{crit} = \frac{q(.05/3,3,15)}{\sqrt{2}} = \frac{4.473}{\sqrt{2}} = 3.163$$

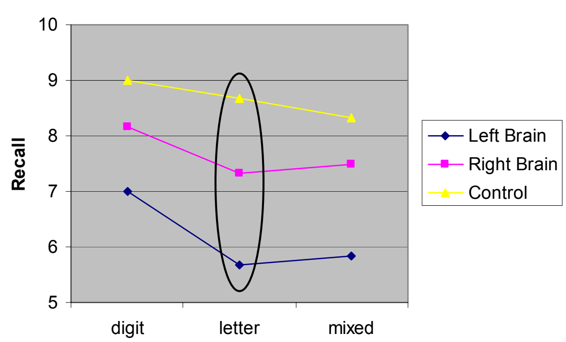
$$d_{LeftVRight} = \frac{1.6667}{1.095} = 1.52 \quad d_{LeftVControl} = \frac{3}{1.095} = 2.74 \quad d_{RightVControl} = \frac{1.333}{1.095} = 1.21$$

Recall of letters:

Left-brain vs. right brain: $t(15) = 2.64, ns, d = 1.52$

Left-brain vs. control: $t(15) = 4.74, p < .05, d = 2.74$

Right-brain vs. control: $t(15) = 2.10, ns, d = 1.21$



- Simple effect of damage for mixed stimuli:

```
ONEWAY mixed by damage
 /STAT = DESC
 /CONTRAST = -1 1 0
 /CONTRAST = -1 0 1
 /CONTRAST =  0 -1 1.
```

ANOVA

mixed

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19.444	2	9.722	7.415	.006
Within Groups	19.667	15	1.311		
Total	39.111	17			

$$p_{crit} = \frac{.05}{3} = .0167 \qquad \eta^2 = \frac{19.444}{19.444 + 19.667} = .50$$

There is a significant simple effect for brain damage on recall of letters and numbers, $F(2,15) = 7.41, p < .05, \eta^2 = .50$. Overall, recall of letters only varies by type of brain damage.

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
mixed	1	1.6667	.66109	2.521	15	.024
	2	2.5000	.66109	3.782	15	.002
	3	.8333	.66109	1.261	15	.227

$$t_{crit} = \frac{q(.05/3,3,15)}{\sqrt{2}} = \frac{4.473}{\sqrt{2}} = 3.163$$

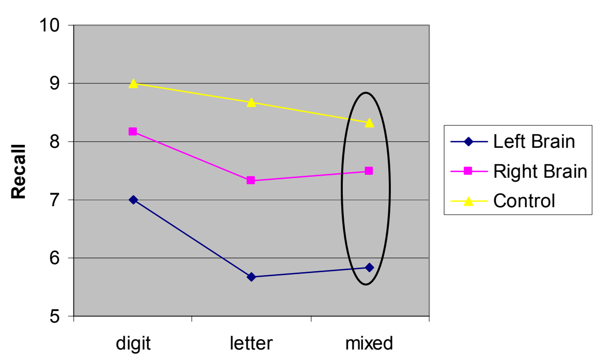
$$d_{LeftVRight} = \frac{1.6667}{1.145} = 1.46 \quad d_{LeftVControl} = \frac{2.5}{1.145} = 2.18 \quad d_{RightVControl} = \frac{0.8333}{1.145} = 0.73$$

Recall of mixed stimuli (digits and letters):

Left-brain vs. right brain: $t(15) = 2.52, ns, d = 1.46$

Left-brain vs. control: $t(15) = 3.78, p < .05, d = 2.18$

Right-brain vs. control: $t(15) = 1.26, ns, d = 0.73$



- Conclusions from simple effects of the effect of brain damage on recall for each type of stimulus

Damage	Stimuli Digits	Letters	Mixed	
Left Brain	$\overline{X} = 7.00^{a}$	$\overline{X} = 5.67^{b}$	$\overline{X} = 5.83^{c}$	$\overline{X} = 6.17^{d}$
Right Brain	$\overline{X} = 8.17$	$\overline{X} = 7.33$	$\overline{X} = 7.50$	$\overline{X} = 7.67^{d}$
Control	$\overline{X} = 9.00^{a}$	$\overline{X} = 8.67^{b}$	$\overline{X} = 8.33^{c}$	$\overline{X} = 8.67^{d}$
	$\overline{X} = 8.06$	$\overline{X} = 7.22$	$\overline{X} = 7.22$	

Note: Within each column, means with a common subscript are significantly different from each other.

- The simple effects and pairwise tests allow us to indirectly test the stimuli by damage interaction. However, we never actually tested any interaction contracts (all of our contrasts on cell means were within a level of a factor). When the between-subjects factor has more than two levels, testing interaction contrasts is not straightforward.

10. Relationship between time of year and cholesterol (2 * 4)

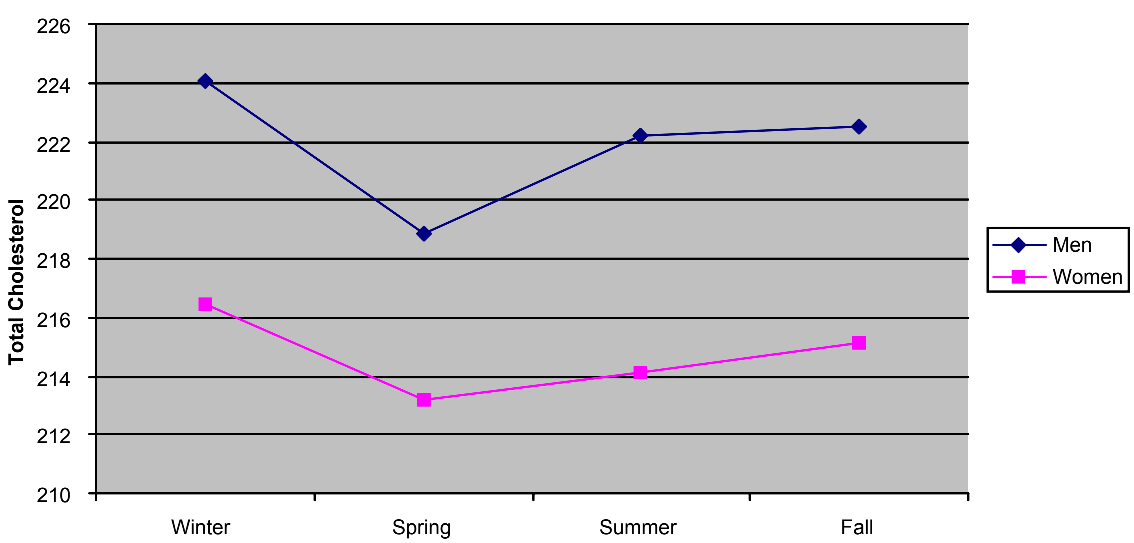
- Example #2: The Seasons data come from a longitudinal study recently conducted by the UMass Medical School (Merriam et al., 1999). Subjects were volunteers recruited from the membership of a large HMO in central Massachusetts. For some of the variables, subjects provided data during each season of the year. The number at the end of the variable name indicates the season: 1=winter; 2=spring; 3=summer; and 4=fall.

  Participants' total cholesterol (TC) level was measured in each of the four seasons. The researcher would like to know if total cholesterol levels varied season, and if this variation differed for men and women.

Descriptive Statistics

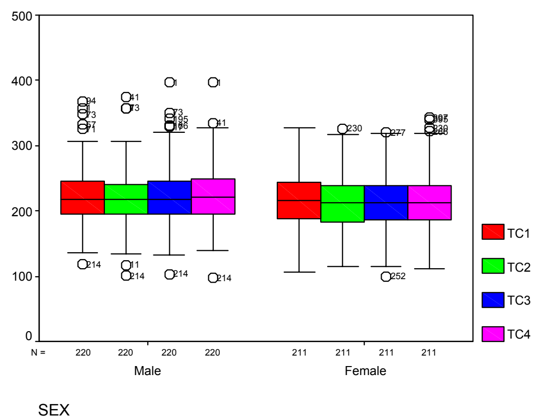
	SEX	Mean	Std. Deviation	N
TC1	Male	224.0591	40.79346	220
	Female	216.4171	42.84937	211
	Total	220.3179	41.93859	431
TC2	Male	218.8182	40.11304	220
	Female	213.2204	40.43307	211
	Total	216.0777	40.32061	431
TC3	Male	222.1636	41.60071	220
	Female	214.0924	41.07910	211
	Total	218.2123	41.49518	431
TC4	Male	222.5182	39.90822	220
	Female	215.0948	42.98048	211
	Total	218.8840	41.55878	431

Total Cholesterol by Season



- Tests of assumptions
  - Normality

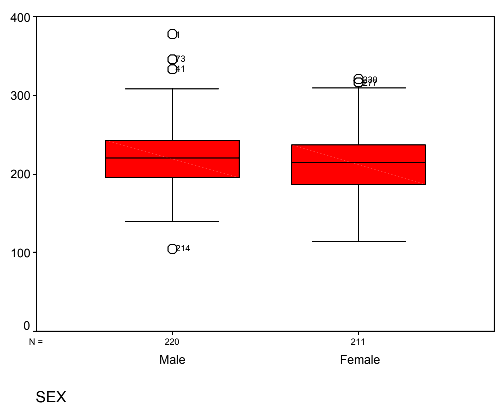
```
EXAMINE VARIABLES=tc1 tc2 tc3 tc4 BY sex
  /PLOT BOXPLOT NPPLOT SPREADLEVEL
  /COMPARE VARIABLES.
```



Tests of Normality

	SEX	Shapiro-Wilk Statistic	df	Sig.
TC1	Male	.978	220	.002
	Female	.995	211	.759
TC2	Male	.979	220	.003
	Female	.991	211	.205
TC3	Male	.971	220	.000
	Female	.994	211	.554
TC4	Male	.982	220	.007
	Female	.988	211	.088

```
EXAMINE VARIABLES=tc_mean BY sex
  /PLOT BOXPLOT NPPLOT SPREADLEVEL
  /COMPARE VARIABLES.
```



Tests of Normality

	SEX	Shapiro-Wilk Statistic	df	Sig.
TC_MEAN	Male	.981	220	.005
	Female	.994	211	.634

- The data look relatively symmetrical, but there are a number of outliers. A sensitivity analysis would be in order.

- Homogeneity of variances / Sphericity

  - Homogeneity of the variance covariance matrices

```
GLM  tc1 tc2 tc3 tc4 BY sex
  /WSFACTOR = time 4
  /PRINT = DESC HOMO.
```

Box's Test of Equality of Covariance Matrices

Box's M	14.276
F	1.413
df1	10
df2	876387.4
Sig.	.167

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

We do not have any evidence that the variance/covariance matrices are different across the three groups. This assumption is satisfied.

  - Homogeneity of variances for between group tests:

```
GLM  tc1 tc2 tc3 tc4 BY sex
  /WSFACTOR = time 4
  /PRINT = DESC HOMO.
```

Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
TC1	.927	1	429	.336
TC2	.565	1	429	.453
TC3	.364	1	429	.546
TC4	.620	1	429	.431

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

```
COMPUTE mean_tc = (tc1 + tc2 + tc3 + tc4)/4.
EXAMINE VARIABLES= mean_tc BY sex
  /PLOT BOXPLOT  SPREADLEVEL.
```

Test of Homogeneity of Variance

	Levene Statistic	df1	df2	Sig.
TC_MEAN	.693	1	429	.406

We do not have any evidence that the variances are different across the two groups. This assumption is satisfied.

- Overall sphericity (averaging over the between subjects factor):

```
GLM  tc1 tc2 tc3 tc4 BY sex
  /WSFACTOR = time 4
  /PRINT = DESC HOMO.
```

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.980	.989	.333

The data are spherical. We can conduct omnibus tests for the within-subject effect (time) or for between/within subject interactions (sex*time).

- Sphericity within each level of the between subjects factor:

```
Temporary.
select if sex = 0.
GLM  tc1 tc2 tc3 tc4
  /WSFACTOR = time 4.
```

```
Temporary.
select if sex = 1.
GLM  tc1 tc2 tc3 tc4
  /WSFACTOR = time 4.
```

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.979	.994	.333

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Epsilon		
	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.952	.966	.333

Within each level of stimuli the data are spherical. We can conduct simple effect omnibus tests for the within-subject effect (time) within each level of the between-subjects factor (sex).

- Conclusions from tests of assumptions:
  - We may perform an omnibus test and/or standard contrasts on the marginal between-subjects (sex) means.
  - We may perform standard omnibus tests on the marginal within-subjects (time) effect and on the between/within (sex by time) interaction.
  - We may perform standard simple-effect omnibus tests for the effect of the within-subjects factor (time) within each level of the between-subjects factor (sex).
  - Contrasts on the marginal within-subjects (time) means or on the between/within (sex by time) cell means may use the omnibus error term. However, I recommend always using a contrast-specific error term, so all tests will use these contrast-specific error terms.
  - There are a number of outliers; a sensitivity analysis should be conducted.

- General ANOVA omnibus tests:

```
GLM  tc1 tc2 tc3 tc4 BY sex
 /WSFACTOR = time 4
 /PRINT = DESC HOMO.
```

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	3982.386	3	1327.462	5.382	.001
	Greenhouse-Geisser	3982.386	2.939	1355.223	5.382	.001
	Huynh-Feldt	3982.386	2.968	1341.813	5.382	.001
	Lower-bound	3982.386	1.000	3982.386	5.382	.021
TIME * SEX	Sphericity Assumed	384.530	3	128.177	.520	.669
	Greenhouse-Geisser	384.530	2.939	130.857	.520	.665
	Huynh-Feldt	384.530	2.968	129.562	.520	.667
	Lower-bound	384.530	1.000	384.530	.520	.471
Error(TIME)	Sphericity Assumed	317410.663	1287	246.628		
	Greenhouse-Geisser	317410.663	1260.637	251.786		
	Huynh-Feldt	317410.663	1273.235	249.295		
	Lower-bound	317410.663	429.000	739.885		

- Omnibus tests using the within-subjects error term $MS_{Time*Sub/Sex}$:
  - Main effect of time: $F(3,1287) = 5.38, p = .001$
  - Time by gender interaction: $F(3,1287) = 0.52, p = .67$

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	82119676.7	1	82119676.65	13558.029	.000
SEX	22231.744	1	22231.744	3.670	.056
Error	2598411.686	429	6056.904		

- Omnibus tests using the between-subjects error term $MS_{Sub/Sex}$:
  - Main effect of gender: $F(1,429) = 3.67, p = .056$

- Simple effects of season within each gender:
  - There are two simple effects tests (for men and for women). We need to use an adjusted critical p-value to maintain $\alpha_{FW} = .05$

$$p_{crit} = \frac{.05}{2} = .025$$

  - We want our test of season to be based on an error term containing information from both men and women (because overall sphericity is satisfied, we should use the omnibus within-subjects error term).
    - If we select men and women separately, the error terms will only contain information from the male and female participants, respectively.
    - However, we can select each group separately to obtain the sum of squares for the simple effects tests. We can then manually compute tests for the effect of time for men and women separately using the omnibus error term:

$$F(a-1,(N-a)(b-1)) = \frac{MS_{Time(Men\ Only)}}{MS_{Time*Sub/Sex}} \qquad F(a-1,(N-a)(b-1)) = \frac{MS_{Time(Women\ Only)}}{MS_{Time*Sub/Sex}}$$

- Simple effect of season for men:

```
Temporary.
select if sex = 0.
GLM  tc1 tc2 tc3 tc4
  /WSFACTOR = time 4.
```

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	3214.312	3	1071.437	4.356	.005
	Greenhouse-Geisser	3214.312	2.938	1093.868	4.356	.005
	Huynh-Feldt	3214.312	2.983	1077.619	4.356	.005
	Lower-bound	3214.312	1.000	3214.312	4.356	.038
Error(TIME)	Sphericity Assumed	161583.937	657	245.942		
	Greenhouse-Geisser	161583.937	643.528	251.091		
	Huynh-Feldt	161583.937	653.231	247.361		
	Lower-bound	161583.937	219.000	737.826		

- We can use the *SS* and *MS* for the effect of time within men, but we should not use the F-test. Because overall sphericity is satisfied, we should use the omnibus within-subject error term for this simple effect test

$$F(a-1,(N-a)(b-1)) = \frac{MS_{Time(Men\ Only)}}{MS_{Time*Sub/Sex}} \qquad F(3,1287) = \frac{1071.437}{246.648} = 4.344, p = .0048$$

$$p_{crit} = \frac{.05}{2} = .025$$

$$F(3,1287) = \frac{1071.437}{246.648} = 4.344, p < .05$$

- There is a significant simple effect of time on total cholesterol levels for men, $F(3,1287) = 4.34, p < .05$.

- Simple effect of season for women:

```
Temporary.
select if sex = 1.
GLM  tc1 tc2 tc3 tc4
  /WSFACTOR = time 4.
```

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	1194.775	3	398.258	1.610	.186
	Greenhouse-Geisser	1194.775	2.855	418.543	1.610	.188
	Huynh-Feldt	1194.775	2.898	412.266	1.610	.187
	Lower-bound	1194.775	1.000	1194.775	1.610	.206
Error(TIME)	Sphericity Assumed	155826.725	630	247.344		
	Greenhouse-Geisser	155826.725	599.467	259.942		
	Huynh-Feldt	155826.725	608.594	256.044		
	Lower-bound	155826.725	210.000	742.032		

$$F(a-1,(N-a)(b-1)) = \frac{MS_{Time(Women\ Only)}}{MS_{Time*Sub/Sex}} \qquad F(3,1287) = \frac{398.258}{246.648} = 1.6147, p = .1841$$

$$p_{crit} = \frac{.05}{2} = .025$$

$$F(3,1287) = \frac{398.258}{246.648} = 1.61, ns$$

- There is no significant simple effect of time on total cholesterol levels for women, $F(3,1287) = 1.62, p < .05$.

- Simple effects of gender within each time:
  - There are four simple effects tests (one for each season). We need to use an adjusted critical p-value to maintain $\alpha_{FW} = .05$

$$p_{crit} = \frac{.05}{4} = .0125$$

  - Simple effect of gender in winter (time 1):

```
GLM tc1 by sex
 /PRINT = DESC.
```

Tests of Between-Subjects Effects

Dependent Variable: TC1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6289.922[a]	1	6289.922	3.598	.059
Intercept	20896457.5	1	20896457.46	11952.558	.000
SEX	6289.922	1	6289.922	3.598	.059
Error	750013.530	429	1748.283		
Total	21677027.0	431			
Corrected Total	756303.452	430			

a. R Squared = .008 (Adjusted R Squared = .006)

- We find no evidence for a significant simple effect of gender at time 1(winter):

$$F(1,429) = 3.60, ns$$

  - Simple effect of gender in spring (time 2):

```
GLM tc2 by sex
 /PRINT = DESC.
```

Tests of Between-Subjects Effects

Dependent Variable: TC2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3374.917[a]	1	3374.917	2.081	.150
Intercept	20103556.2	1	20103556.17	12396.791	.000
SEX	3374.917	1	3374.917	2.081	.150
Error	695698.230	429	1621.674		
Total	20822283.8	431			
Corrected Total	699073.146	430			

a. R Squared = .005 (Adjusted R Squared = .003)

- We find no evidence for a significant simple effect of gender at time 2(spring):

$$F(1,429) = 2.08, ns$$

- Simple effect of gender in summer (time 3):

```
GLM tc3 by sex
 /PRINT = DESC.
```

Tests of Between-Subjects Effects

Dependent Variable: TC3

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7016.268[a]	1	7016.268	4.104	.043
Intercept	20497967.4	1	20497967.41	11990.563	.000
SEX	7016.268	1	7016.268	4.104	.043
Error	733379.057	429	1709.508		
Total	21263152.8	431			
Corrected Total	740395.325	430			

a. R Squared = .009 (Adjusted R Squared = .007)

  - We find no evidence for a significant simple effect of gender at time 3(summer):

$$F(1,429) = 4.10, ns$$

- Simple effect of gender in fall (time 4):

```
GLM tc4 by sex
 /PRINT = DESC.
```

Tests of Between-Subjects Effects

Dependent Variable: TC4

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5935.168[a]	1	5935.168	3.456	.064
Intercept	20625678.0	1	20625678.00	12010.367	.000
SEX	5935.168	1	5935.168	3.456	.064
Error	736731.532	429	1717.323		
Total	21391963.5	431			
Corrected Total	742666.700	430			

a. R Squared = .008 (Adjusted R Squared = .006)

  - We find no evidence for a significant simple effect of gender at time 4(fall):

$$F(1,429) = 3.46, ns$$

- Tests of polynomial trends over time:

Gender	Season Winter	Spring	Summer	Fall
Male	224.06	218.82	222.16	222.51
Female	216.42	213.22	214.09	216.09
	220.32	216.07	218.21	218.88

  - Polynomial trends on the marginal season means:

```
GLM  tc1 tc2 tc3 tc4 BY sex
  /WSFACTOR = time 4
  /PRINT = DESC HOMO.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Linear	102.937	1	102.937	.362	.548
	Quadratic	2583.051	1	2583.051	10.575	.001
	Cubic	1296.398	1	1296.398	6.140	.014
TIME * SEX	Linear	17.789	1	17.789	.063	.803
	Quadratic	52.504	1	52.504	.215	.643
	Cubic	314.237	1	314.237	1.488	.223
Error(TIME)	Linear	122043.560	429	284.484		
	Quadratic	104784.659	429	244.253		
	Cubic	90582.444	429	211.148		

  - These are complex post-hoc tests and require a Scheffe correction:

$$F_{crit} = df_{season} * F(.05, df_{season}, df_{error}) = 3 * F(.05,3,429) = 3 * 3.01 = 9.05$$

    - Linear trend in total cholesterol over seasons: $F(1,429) = 0.36, ns$
    - Quadratic trend in total cholesterol over seasons: $F(1,429) = 10.58, p < .05$
    - Cubic trend in total cholesterol over seasons: $F(1,429) = 6.14, ns$

- Next, we test if these polynomial trends differ by gender:
  - These tests (linear*sex, quadratic*sex, and cubic*sex) were printed in the previous analysis.
  - These are complex post-hoc tests and require a Scheffe correction:
    $F_{crit} = df_{season*sex} * F(.05, df_{season*sex}, df_{error}) = 3 * F(.05,3,429) = 3*3.01 = 9.05$

	Season			
Linear	Winter	Spring	Summer	Fall
Male	-3	-1	1	3
Female	3	1	-1	-3

There is no difference in linear trends in total cholesterol over seasons between men and women: $F(1,429) = 0.06, ns$

	Season			
Quadratic	Winter	Spring	Summer	Fall
Male	1	-1	-1	1
Female	-1	1	1	-1

There is no difference in quadratic trends in total cholesterol over seasons between men and women: $F(1,429) = 0.22, ns$

	Season			
Cubic	Winter	Spring	Summer	Fall
Male	-3	1	-1	3
Female	3	-1	1	-3

There is no difference in cubic trends in total cholesterol over seasons between men and women: $F(1,429) = 1.49, ns$

- Next, we conduct repeated contrasts *on the marginal time means* (comparing each level to the previous level):

```
GLM  tc1 tc2 tc3 tc4 BY sex
  /WSFACTOR = time 4 repeated
  /PRINT = DESC HOMO.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Level 1 vs. Level 2	7667.696	1	7667.696	17.697	.000
	Level 2 vs. Level 3	1915.740	1	1915.740	4.288	.039
	Level 3 vs. Level 4	198.305	1	198.305	.402	.527
TIME * SEX	Level 1 vs. Level 2	450.076	1	450.076	1.039	.309
	Level 2 vs. Level 3	658.904	1	658.904	1.475	.225
	Level 3 vs. Level 4	45.200	1	45.200	.092	.762
Error(TIME)	Level 1 vs. Level 2	185873.819	429	433.272		
	Level 2 vs. Level 3	191652.790	429	446.743		
	Level 3 vs. Level 4	211850.594	429	493.824		

  - These are post-hoc pairwise comparisons and require a Tukey HSD correction.

$$F_{crit} = \frac{(q(.05,4,429))^2}{2} = \frac{(3.633)^2}{2} = 6.60$$

    - Winter vs. Spring: $F(1,429) = 17.70, p < .05$
    - Spring vs. Summer: $F(1,429) = 4.29, ns$
    - Summer vs. Fall: $F(1,429) = 0.40, ns$

- We also want to test if these repeated contrasts differ for men and women.
  - Again, tests of these contrasts were provided as interaction contrasts when we asked for the repeated contrasts.
  - These are complex, interaction post-hoc tests and require a Scheffe correction:

$$F_{crit} = df_{season*sex} * F(.05, df_{season*sex}, df_{error}) = 3 * F(.05,3,429) = 3 * 3.01 = 9.05$$

Winter vs. Spring	Season			
	Winter	Spring	Summer	Fall
Male	1	-1	0	0
Female	-1	1	0	0

- Difference in winter vs. spring total cholesterol levels between men and women: $F(1,429) = 1.04, ns$

Spring vs. Summer	Season			
	Winter	Spring	Summer	Fall
Male	0	1	-1	0
Female	0	-1	1	0

- Difference in spring vs. summer total cholesterol levels between men and women: $F(1,429) = 1.48, ns$

Summer vs. Fall	Season			
	Winter	Spring	Summer	Fall
Male	0	0	1	-1
Female	0	0	-1	1

- Difference in summer vs. fall total cholesterol levels between men and women: $F(1,429) = 0.09, ns$

- These differences in repeated contrasts between men and women can also be conducted by computing repeated contrasts on the marginal season means, and then testing if these contrasts differ by gender

  - Step 1: Compute a contrast comparing winter to spring.

$$\psi_{Winter-Spring} = \mu_{Winter} - \mu_{Spring}$$

Winter vs. Spring	Season: Winter	Spring	Summer	Fall
Male				
Female				
	1	-1		

  - Step 2: Test if this contrast differs by gender.

$$\begin{aligned}\psi_{interaction} &= \psi_{Winter-Spring}(men) - \psi_{Winter-Spring}(women) \\ &= \big(\mu_{Winter}(men) - \mu_{Spring}(men)\big) - \big(\mu_{Winter}(women) - \mu_{Spring}(women)\big) \\ &= \mu_{Winter}(men) - \mu_{Spring}(men) - \mu_{Winter}(women) + \mu_{Spring}(women)\end{aligned}$$

Winter vs. Spring	Season: Winter	Spring	Summer	Fall
Male	1	-1		
Female	-1	1		

  - A test of whether the difference in winter vs. spring total cholesterol levels are equal for men and women is equivalent to a test of the interaction contrast, $H_0 : \psi_{interaction} = 0$

$$\begin{aligned}&H_0 : \psi_{interaction} = 0 \\ &H_0 : \psi_{Winter-Spring}(men) - \psi_{Winter-Spring}(women) = 0 \\ &H_0 : \psi_{Winter-Spring}(men) = \psi_{Winter-Spring}(women)\end{aligned}$$

- In SPSS:

```
Compute t1vst2 = tc1  - tc2.
Compute t2vst3 = tc2  - tc3.
Compute t3vst4 = tc3  - tc4.

T-TEST GROUPS=sex(0 1)
  /VARIABLES=t1vst2 t2vst3 t3vst4.
```

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
T1VST2	Equal variances assumed	.491	.484	1.019	429	.309	2.0442	2.00570
	Equal variances not assumed			1.021	427.693	.308	2.0442	2.00164
T2VST3	Equal variances assumed	.856	.355	-1.214	429	.225	-2.4734	2.03664
	Equal variances not assumed			-1.217	426.741	.224	-2.4734	2.03177
T3VST4	Equal variances assumed	.656	.418	.303	429	.762	.6478	2.14127
	Equal variances not assumed			.302	416.015	.763	.6478	2.14734

$$t_{crit} = \sqrt{df_{season*sex} * F(.05, df_{season*sex}, df_{error})} = \sqrt{3 * F(.05,3,429)} = \sqrt{3 * 3.01} = 3.01$$

- Difference in winter vs. spring total cholesterol levels between men and women: $t(429) = 1.02, ns$
- Difference in spring vs. summer total cholesterol levels between men and women: $t(429) = 1.21, ns$
- Difference in summer vs. fall total cholesterol levels between men and women: $t(429) = 0.30, ns$

- These results exactly match the results we obtain by asking for repeated contrasts (and repeated*gender interaction contrasts) in the repeated-measures ANOVA. Both analyses test the same hypothesis and include gender as a between-subjects factor in the design.
- An advantage of this method is that it can be used when the variance between the male and female (contrast) scores are not equal.

- If we wish to conduct simple contrasts (comparing cholesterol levels at each time point to the cholesterol levels in winter):

```
GLM  tc1 tc2 tc3 tc4 BY sex
  /WSFACTOR = tc 4 simple(1)
  /PRINT = DESC HOMO.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Level 2 vs. Level 1	7667.696	1	7667.696	17.697	.000
	Level 3 vs. Level 1	1918.108	1	1918.108	3.468	.063
	Level 4 vs. Level 1	882.930	1	882.930	1.621	.204
TIME * SEX	Level 2 vs. Level 1	450.076	1	450.076	1.039	.309
	Level 3 vs. Level 1	19.839	1	19.839	.036	.850
	Level 4 vs. Level 1	5.148	1	5.148	.009	.923
Error(TIME)	Level 2 vs. Level 1	185873.819	429	433.272		
	Level 3 vs. Level 1	237265.107	429	553.066		
	Level 4 vs. Level 1	233599.217	429	544.520		

  - These are post-hoc pair-wise comparisons and require a Tukey HSD correction.

$$F_{crit} = \frac{(q(.05,4,429))^2}{2} = \frac{(3.633)^2}{2} = 6.60$$

    - Winter vs. Spring: $F(1,429) = 17.70, p < .05$
    - Winter vs. Summer: $F(1,429) = 3.47, ns$
    - Winter vs. Fall: $F(1,429) = 1.62, ns$

- We can also test if these simple contrasts differ for men and women.
  - These are complex, interaction post-hoc tests and require a Scheffe correction:

$$F_{crit} = df_{season*sex} * F(.05, df_{season*sex}, df_{error}) = 3 * F(.05,3,429) = 3 * 3.01 = 9.05$$

- The tests of these contrasts were provided as interaction contrasts when we asked for the repeated contrasts:

Winter vs. Spring	Season: Winter	Spring	Summer	Fall
Male	1	-1	0	0
Female	-1	1	0	0

  - Difference in winter vs. spring total cholesterol levels between men and women: $F(1,429) = 1.04, ns$

Spring vs. Summer	Season: Winter	Spring	Summer	Fall
Male	1	0	-1	0
Female	-1	0	1	0

  - Difference in winter vs. summer total cholesterol levels between men and women: $F(1,429) = 0.04, ns$

Summer vs. Fall	Season: Winter	Spring	Summer	Fall
Male	1	0	0	-1
Female	-1	0	0	1

  - Difference in winter vs. fall total cholesterol levels between men and women: $F(1,429) = 0.01, ns$

- Again, differences in simple contrasts between men and women can also be conducted by computing simple contrasts on the marginal season means, and then testing if these contrasts differ by gender

- In SPSS:

```
Compute t1vst2 = tc1  - tc2.
Compute t1vst3 = tc1  - tc3.
Compute t1vst4 = tc1  - tc4.

T-TEST  GROUPS=sex(0 1)
 /VARIABLES=t1vst2 t1vst3 t1vst4.
```

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
T1VST2	Equal variances assumed	.491	.484	1.019	429	.309	2.0442	2.00570
	Equal variances not assumed			1.021	427.693	.308	2.0442	2.00164
T1VST3	Equal variances assumed	.153	.696	-.189	429	.850	-.4292	2.26608
	Equal variances not assumed			-.189	428.071	.850	-.4292	2.26630
T1VST4	Equal variances assumed	.018	.895	.097	429	.923	.2186	2.24850
	Equal variances not assumed			.097	427.401	.923	.2186	2.24941

$$t_{crit} = \sqrt{df_{season*sex} * F(.05, df_{season*sex}, df_{error})} = \sqrt{3 * F(.05,3,429)} = \sqrt{3 * 3.01} = 3.01$$

- Difference in winter vs. spring total cholesterol levels between men and women: $t(429) = 1.02, ns$
- Difference in winter vs. summer total cholesterol levels between men and women: $t(429) = 0.19, ns$
- Difference in winter vs. fall total cholesterol levels between men and women: $t(429) = 0.01, ns$

* Finally, when we look at the data, we may decide to examine some complex contrasts on the marginal season means:

	Season			
Gender	Winter	Spring	Summer	Fall
Male	224.06	218.82	222.16	222.51
Female	216.42	213.22	214.09	216.09
	220.32	216.07	218.21	218.88

i. Do cholesterol levels in winter differ from average cholesterol levels in summer and fall?

Season			
Winter	Spring	Summer	Fall
2	0	-1	-1

ii. Do cholesterol levels in spring differ from average cholesterol levels in summer and fall?

Season			
Winter	Spring	Summer	Fall
0	2	-1	-1

iii. Do average cholesterol levels in the winter and fall differ from average cholesterol levels in summer and fall?

Season			
Winter	Spring	Summer	Fall
-1	1	1	-1

iv.

- We cannot test these hypotheses on the marginal means in SPSS by computing a value reflecting this contrast (because we need to keep gender in the analysis).
- We must enter these contrasts in the special subcommand as contrasts coefficients on the marginal season means.

- These are complex post-hoc tests and require a Scheffe correction:

$$F_{crit} = df_{season} * F(.05, df_{season}, df_{error}) = 3 * F(.05,3,429) = 3 * 3.01 = 9.05$$

```
GLM tc1 tc2 tc3 tc4 by sex
  /WSFACTOR = time 4 special (1 1 1 1   2 0 -1 -1  0 -2 1 1  -1 1 1 -1)
```

.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	L1	5403.773	1	5403.773	3.176	.075
	L2	10326.706	1	10326.706	7.505	.006
	L3	10332.205	1	10332.205	10.575	.001
TIME * SEX	L1	4.775	1	4.775	.003	.958
	L2	1990.511	1	1990.511	1.447	.230
	L3	210.014	1	210.014	.215	.643
Error(TIME)	L1	729878.055	429	1701.347		
	L2	590257.230	429	1375.891		
	L3	419138.635	429	977.013		

- Winter vs. (Summer and Fall): $F(1,429) = 3.18, ns$
- Spring vs. (Summer and Fall): $F(1,429) = 7.50, ns$
- (Winter and Fall) vs. (Spring and Summer): $F(1,429) = 10.58, p < .05$

- We also should check to test if these comparisons differ for men and women.
  - Again, these are complex, interaction post-hoc tests and require a Scheffe correction:
    $F_{crit} = df_{season*sex} * F(.05, df_{season*sex}, df_{error}) = 3 * F(.05,3,429) = 3*3.01 = 9.05$

- The tests of these contrasts were provided as interaction contrasts when we asked for the special contrasts:

L1* Sex	Season: Winter	Spring	Summer	Fall
Male	2	0	-1	-1
Female	-2	0	1	1

- Difference in winter vs. (summer and fall) total cholesterol levels between men and women: $F(1,429) = 0.01, ns$

L2* Sex	Season: Winter	Spring	Summer	Fall
Male	0	2	-1	-1
Female	0	-2	1	1

- Difference in spring vs. (summer and fall) total cholesterol levels between men and women: $F(1,429) = 1.45, ns$

L3* Sex	Season: Winter	Spring	Summer	Fall
Male	-1	1	1	-1
Female	1	-1	-1	1

- Difference in (winter and fall) vs. (spring and summer) total cholesterol levels between men and women: $F(1,429) = .22, ns$

- These differences in complex interaction contrasts can also be conducted by computing the complex contrasts on the marginal season means, and then testing if these contrasts differ by gender

  - For example, first compute a contrast comparing winter to (summer and fall):

$$\psi_{Win-SumFall} = 2\mu_{Win} - (\mu_{Sum} + \mu_{Fall})$$

Winter vs. Spring	Season			
	Winter	Spring	Summer	Fall
Male				
Female				
	2		-1	-1

  - Next, test if this contrast differs by gender

$$\begin{aligned}
\psi_{interaction} &= \psi_{Win-SumFall}(men) - \psi_{Win-SumFall}(women) \\
&= \left(2\mu_{Win}(men) - (\mu_{Sum}(men) + \mu_{Fall}(men))\right) - \\
&\quad \left(2\mu_{Win}(women) - (\mu_{Sum}(women) + \mu_{Fall}(women))\right) \\
&= 2\mu_{Win}(men) - \mu_{Sum}(men) - \mu_{Fall}(men) + \\
&\quad -2\mu_{Win}(women) + \mu_{Sum}(women) + \mu_{Fall}(women)
\end{aligned}$$

Winter vs. Spring	Season			
	Winter	Spring	Summer	Fall
Male	2	0	-1	-1
Female	-2	0	1	1

  - A test of whether the difference in winter vs. (summer and fall) total cholesterol levels is equal for men and women is equivalent to testing if the interaction contrast differs from zero ($H_0 : \psi_{interaction} = 0$).

$$H_0 : \psi_{interaction} = 0$$
$$H_0 : \psi_{Win-SumFall}(men) - \psi_{Win-SumFall}(women) = 0$$
$$H_0 : \psi_{Win-SumFall}(men) = \psi_{Win-SumFall}(women)$$

- In SPSS:

```
compute t1vst34 = tc1  - (tc3 + tc4)/2.
compute t2vst34 = tc2  - (tc3 + tc4)/2.
compute t14vst23 = (tc1 + tc4)/2  - (tc2 + tc3)/2.

T-TEST GROUPS=sex(0 1)
  /VARIABLES= t1vst34  t2vst34 t14vst23.
```

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
T1VST34	Equal variances assumed	.007	.934	-.053	429	.958	-.1053	1.98725
	Equal variances not assumed			-.053	428.862	.958	-.1053	1.98625
T2VST34	Equal variances assumed	.712	.399	-1.203	429	.230	-2.1495	1.78710
	Equal variances not assumed			-1.205	427.915	.229	-2.1495	1.78366
T14VST23	Equal variances assumed	.743	.389	.464	429	.643	.6982	1.50594
	Equal variances not assumed			.463	421.550	.644	.6982	1.50881

$$t_{crit} = \sqrt{df_{season*sex} * F(.05, df_{season*sex}, df_{error})} = \sqrt{3 * F(.05,3,429)} = \sqrt{3 * 3.01} = 3.01$$

- Difference in winter vs. (summer and fall) total cholesterol levels between men and women: $t(429) = 0.05, ns$
- Difference in spring vs. (summer and fall) total cholesterol levels between men and women: $t(429) = 1.20, ns$
- Difference in (winter and fall) vs. (spring and summer) total cholesterol levels between men and women: $t(429) = 0.46, ns$

- These results exactly match the results we obtain by asking for special contrasts (and special*gender interaction contrasts) in the repeated-measures ANOVA. Both analyses test the same hypotheses and include gender as a between-subjects factor in the design.

- Remember that our check of assumptions revealed a number of outliers. We should conduct a sensitivity analysis to see of the outliers affected any of our conclusions.

