Chapter 10
One-Factor Repeated Measures ANOVA

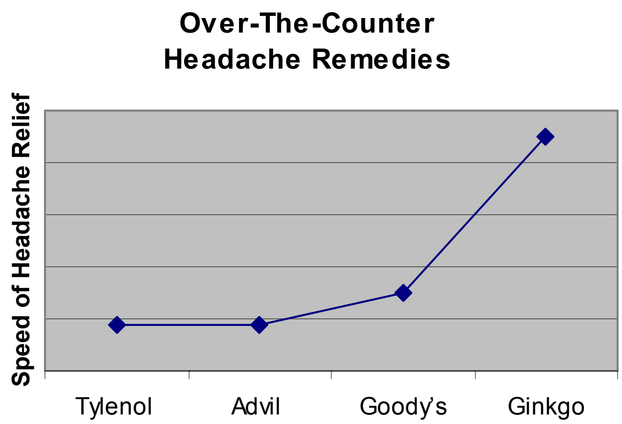
Repeated Measures ANOVA
One-Factor Repeated Measures

1. Types of repeated measures designs

   i. Each participant/unit is observed in *a* different treatment conditions

      - Example: Testing over the counter headache remedies

        Each participant is given four over-the-counter headache medicines:
        - Tylenol (Acetaminophen)
        - Advil (Ibuprofen)
        - Goody's Headache Powder (??)
        - An herbal remedy (Ginkgo)



      - If each participant takes the treatments in random order, then this design can be treated as a randomized block design with participant as the block

		Treatment Order			
		1	2	3	4
Block	1	A	T	GI	GO
	2	GO	A	GI	T
	3	T	GO	A	GI
	4	GI	T	GO	A
	5	GO	A	GI	T

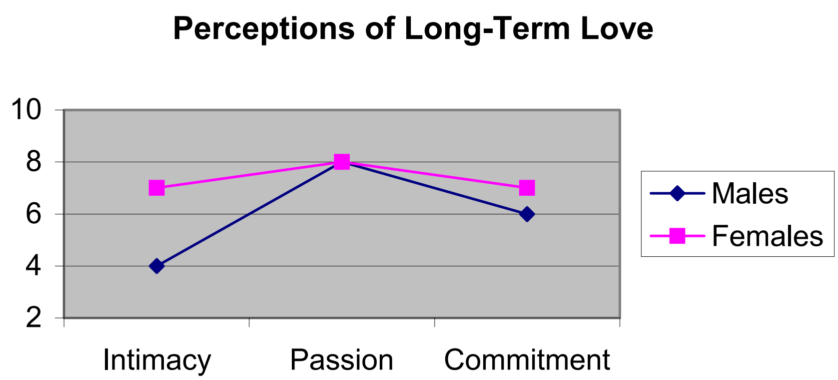
ii. Profile analysis
Scores on *a* different tests are compared for each participant

- Each participant completes several different scales, and the profile of responses to those scores is examined.

- Example: Perceptions of a loving relationship

  Single male and female participants complete three scales, each designed to measure a different aspect of Sternberg's (1988) love triangle:
  - Intimacy
  - Passion
  - Commitment

  Participants rate the importance of each component for a long-term relationship.



- This example has a repeated measures factor and a between-subjects factor.

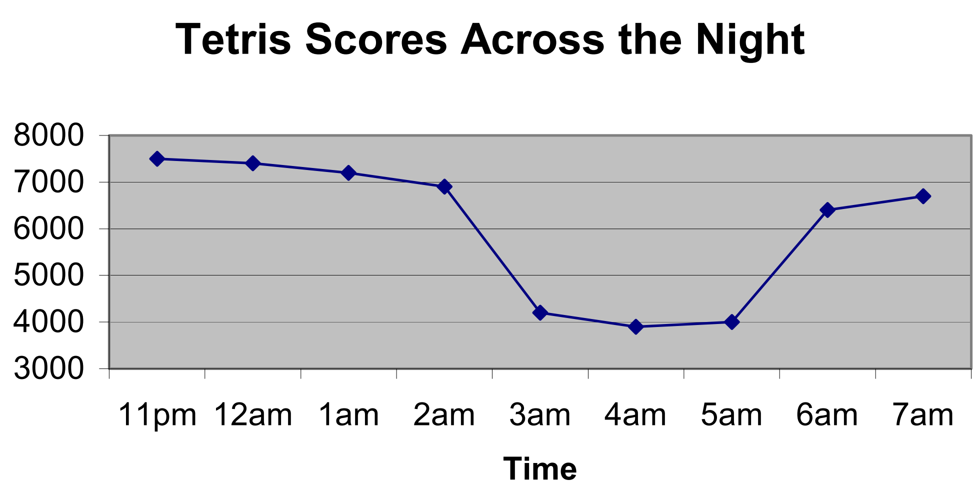
- For a profile analysis, it is highly desirable that each questionnaire is scaled similarly (with the same mean and standard deviation) so that the profile comparison is meaningful.

- A common use of profile analysis is to examine the profile of MMPI subscale scores.

iii. Some aspect of each participant/unit is measured at *a* different times.

- This use of repeated measures designs is the most common. (In fact, some authors reserve the term repeated measures for this type of design, and refer to all three types of designs as within-subjects designs.)
- When the research question involves modeling the trend of the change over time (rather than just the presence or absence of differences over time), this type of design is called <u>growth-curve analysis</u>.

- <u>Example</u>: Participants play Tetris from 11pm to 8am. Scores for each hour are averaged.



2. Advantages and Disadvantages of Repeated Measures Designs

- Advantages
    - Each participant serves as his or her own control. (You achieve perfect equivalence of all groups / perfect blocking)

    - Fewer participants are needed for a repeated measures design to achieve the same power as a between-subjects design. (Individual differences in the DV are measured and removed from the error term.)

    - The most powerful designs for examining change and/or trends

- Disadvantages

    - Practice effects

    - Differential carry-over effects

    - Demand characteristics

3. The Paired t-test

- The simplest example of a repeated measures design is a paired t-test.
    - Each subject is measured twice (time 1 and time 2) on the same variable
    - Or each pair of matched participants are assigned to one of two treatment levels

- The analysis of a paired t-test is exactly equivalent to a one sample t-test conducted on the difference of the time 1 and time 2 scores (or on the difference of the matched pair)

- Understanding the paired t-test
  - Recall that for an independent samples t-test, we used the following formula (see p. 2-39)

$$t_{obs} = \frac{estimate}{std\ error\ of\ estimate} = \frac{\overline{X}_1 - \overline{X}_2}{s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$s_{pooled} = \sqrt{\frac{SS_1 + SS_2}{(n_1 + n_2 - 2)}}$$

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

$df = n_1 + n_2 - 2$

  - For a paired t-test, we will conduct the analysis on the <u>difference</u> of the time 1 and time 2 observations.

Subject	Pre-test	Post-Test	Difference
1	6	9	-3
2	4	6	-2
3	6	5	1
4	7	10	-3
5	4	10	-6
6	5	8	-3
7	5	7	-2
8	12	10	2
9	6	6	0
10	1	5	-4
Average	5.4	7.4	-2

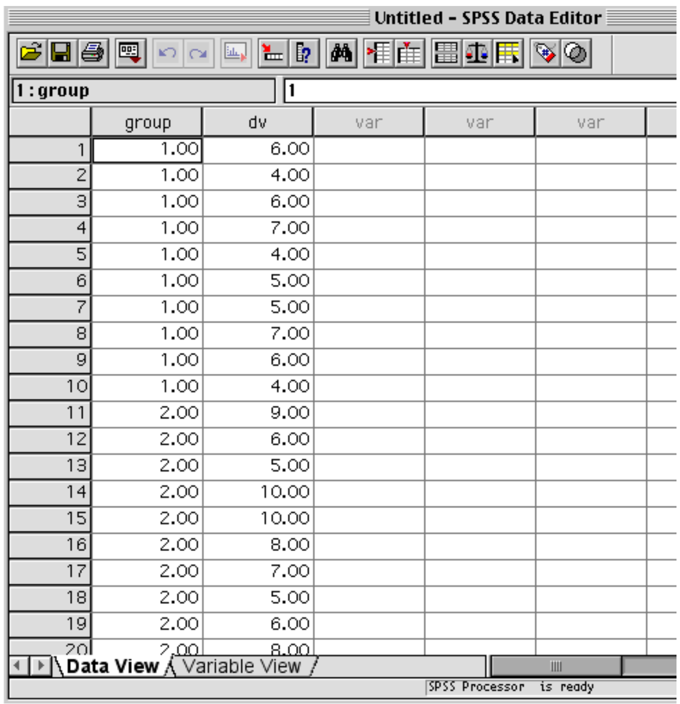
$$t_{obs} = \frac{estimate}{std\ error\ of\ estimate} = \frac{\overline{D}}{s_D\sqrt{\frac{1}{n}}}$$

where $s_D = \sqrt{\frac{SS_D}{n-1}}$

$df = n - 1$

- First, let's see how this analysis would differ if we "forgot" we had a repeated measures design, and treated it as an independent-samples design

  - We treat each of the 20 observations as independent observations
  - To analyze in SPSS, we need 20 lines of data, ostensibly one line for each participant




```
T-TEST GROUPS=group(1 2)
 /VARIABLES=dv.
```

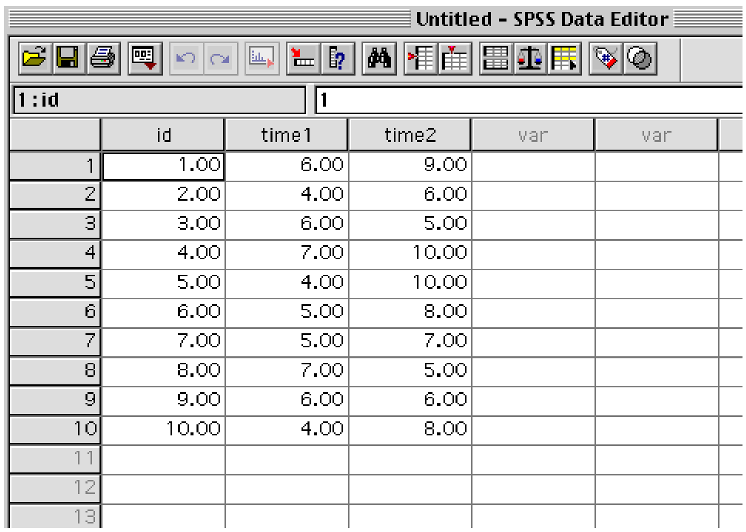
Independent Samples Test

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DV	Equal variances assumed	-1.819	18	.086	-2.0000	1.09949	-4.30995	.30995

$t(18) = -1.82, p = .086$

  - But this analysis is erroneous! We do not have 20 independent data points (18 df).

- Now, let's conduct the proper paired t-test in SPSS:
  - We properly treat the data as 10 pairs of observations.
  - To conduct this analysis in SPSS, we need to enter our data differently than for a between-subjects design. The data from each participant are entered on one line



	id	time1	time2	var	var
1	1.00	6.00	9.00		
2	2.00	4.00	6.00		
3	3.00	6.00	5.00		
4	4.00	7.00	10.00		
5	5.00	4.00	10.00		
6	6.00	5.00	8.00		
7	7.00	5.00	7.00		
8	8.00	7.00	5.00		
9	9.00	6.00	6.00		
10	10.00	4.00	8.00		
11					
12					
13					

```
T-TEST PAIRS time1 time2.
```

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	TIME1 & TIME2	10	.546	.102

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	TIME1 - TIME2	-2.0000	2.40370	.76012	-3.7195	-.2805	-2.631	9	.027

$t(9) = -2.63, p = .027$

- We could also obtain the same result by conducting a one-sample t-test on the difference between time 1 and time 2:

```
COMPUTE diff = time1 – time2.
T-TEST  /TESTVAL=0
  /VARIABLES=diff.
```

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
DIFF	-2.631	9	.027	-2.0000	-3.7195	-.2805

- Both analyses give the identical results:

  $t(9) = -2.63, p = .027$

- A comparison of the differences of the two analyses:

	Independent Groups	Repeated Measures
Mean Difference	-2.00	-2.00
Standard Error	1.10	0.76
t-value	-1.82	-2.63
p-value	.086	.027

- The mean difference is the same
- Importantly, <u>the standard error is smaller</u> in the repeated measures case. The smaller standard error results in a larger t-value and a smaller p-value

- The greater the correlation between the time 1 and the time 2 observations, the greater the advantage gained by using a repeated measures design.

  Recall that the variance of the difference between two variables is given by the following formula (see p 2-38)

$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2Cov(X_1, X_2) \qquad \text{(Formula 8-1)}$$

  The covariance of X_1 and X_2 is a measure of the association between the two variables. If we standardize the covariance (so that it ranges from –1 to +1), we call it the correlation between X_1 and X_2

$$\rho_{12} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1) * Var(X_2)}}$$

  If we rearrange the terms and substitute into (Formula 8-1), we obtain a formula for the variance of the difference scores:

$$Cov(X_1, X_2) = \rho_{12} * \sqrt{Var(X_1)} * \sqrt{Var(X_2)}$$
$$Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2\rho_{12} * \sqrt{Var(X_1)} * \sqrt{Var(X_2)}$$

$$\sigma_D^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 - 2\rho_{12}\sigma_{X_1}\sigma_{X_2}$$

  And finally, we can obtain the standard error of the difference scores:

  In the population:
$$\sigma_{\overline{D}}^2 = \frac{\sigma_{X_1}^2}{n} + \frac{\sigma_{X_2}^2}{n} - \frac{2\rho_{12}\sigma_{X_1}\sigma_{X_2}}{n}$$
$$\sigma_{\overline{D}} = \sqrt{\frac{1}{n}\left(\sigma_{X_1}^2 + \sigma_{X_2}^2 - 2\rho_{12}\sigma_{X_1}\sigma_{X_2}\right)}$$

  In the sample :
$$s_{\overline{D}} = \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n} - \frac{2r_{12}s_1 s_2}{n}}$$
$$s_{\overline{D}} = \sqrt{\frac{1}{n}\left(s_1^2 + s_2^2 - 2r_{12}s_1 s_2\right)}$$

⇒ If $r_{X_1X_2} = 0$, then the variance of the difference scores is equal to $\sigma_{X_1}^2 + \sigma_{X_2}^2$

⇒ As $r_{X_1X_2}$ becomes greater, then the variance of the difference scores becomes smaller than $\sigma_{X_1}^2 + \sigma_{X_2}^2$

So the greater the correlation between time 1 and time 2 scores, the greater the advantage of using a paired t-test.

- Some important observations concerning the paired t-test
  - The paired t-test is calculated on the difference scores, not on the actual time 1 and time 2 data points. In fact, even s_D can be computed directly from the difference scores

$$s_D = \sqrt{\frac{SS_D}{n}}$$

$$SS_D = \sum_{i=1}^{n} \left(D_i - \overline{D}\right)^2$$

  - Individual differences in the DV are removed by taking the difference of the two observations.
    - (When we turn to repeated-measures ANOVA, the individual differences in the DV will be captured in a Sum of Squares due to subjects.)
  - In other words, for a repeated measures analysis, we do not care about variability in the DV itself (That variability is due to individual differences and is removed by taking a difference score). What we do care about is the variability in the difference scores.

- An example might help clarify this issue. Suppose we add 10 to the pre- and post-test scores of the first five participants

	Original Data			Modified Data		
Participant	Pre-test	Post-Test	Difference	Pre-test	Post-Test	Difference
1	6	9	-3	**16**	**19**	-3
2	4	6	-2	**14**	**16**	-2
3	6	5	-1	**16**	**15**	-1
4	7	10	-3	**17**	**20**	-3
5	4	10	-6	**14**	**20**	-6
6	5	8	-3	5	8	-3
7	5	7	-2	5	7	-2
8	12	10	2	12	10	2
9	6	6	0	6	6	0
10	1	5	-4	1	5	-4
Average	5.4	7.4	-2	10.4	12.4	-2

- We have increased the amount of individual differences in the pre- and post-test data, but we have not changed the difference scores.

- As a result, a between-subject analysis of this data is greatly affected by the addition of extra noise to the data, but a within-subject analysis of this data is unchanged.

	Original Data	Modified Data
Between-Subjects	$t(18) = -1.82$ $p = .086$ $SE = 1.10$	$t(18) = -0.76$ $p = .459$ $SE = 2.64$
Within-Subjects	$t(9) = -2.63$ $p = .027$ $SE = 0.760$	$t(9) = -2.63$ $p = .027$ $SE = 0.760$

- To compute an effect-size for a paired t-test, we can use Cohen's d

$$d = \frac{\overline{D}}{\sigma_D}$$

Where σ_D is the standard deviation of the difference scores.

- Assumptions of the paired t-test
    - Normality (actually symmetry) of the difference scores
    - Participants are randomly selected from the population
    - Equality of variances between time 1 and time 2 scores is NOT required

- Options when the normality assumption of the paired t-test is violated
    - Find a suitable transformation of the difference scores
    - Use the non-parametric Wilcoxon Signed-Rank test
        - Similar to the Mann-Whitney U test for independent groups
        - Tests the null hypothesis that the rank of the time 1 data is the same as the rank of the time 2 data

i. Calculate the difference scores for each pair
ii. (Ignore difference scores of zero)
iii. Take the absolute value of all difference scores
iv. Rank the absolute value of the difference scores
v. Attach the sign (+ or –) of the difference score to each rank
vi. Compute the test statistic, W

Compute the sum of the positive signed ranks: $W+$
Compute the sum of the negative signed ranks: $W-$

$$W = Minimum(W+, W-)$$

vii. Look up the tabled critical value for W
If $W \leq W_{crit}$ we reject the null hypothesis

- For our original example data:

	Original Data			Modified Data		
Participant	Pre-test	Post-Test	Difference	|D|	Rank of |D|	Signed Rank
1	6	9	-3	3	6	-6
2	4	6	-2	2	3	-3
3	6	5	1	1	1	1
4	7	10	-3	3	6	-6
5	4	10	-6	6	9	-9
6	5	8	-3	3	6	-6
7	5	7	-2	2	3	-3
8	12	10	2	2	3	3
9	6	6	0			
10	1	5	-4	4	8	-8

Sum of positive ranks = 4
Sum of negative rank = 41 $W = 4$

For a two-tailed test for $n = 10$ and with $\alpha = .05$, $W_{crit} = 8$
We reject the null hypothesis

- In practice, you will probably use SPSS:

```
NPAR TEST
 /WILCOXON=time1  WITH time2 (PAIRED)
 /STATISTICS DESCRIPTIVES.
```

Ranks

		N	Mean Rank	Sum of Ranks
TIME2 - TIME1	Negative Ranks	2[a]	2.00	4.00
	Positive Ranks	7[b]	5.86	41.00
	Ties	1[c]		
	Total	10		

a. TIME2 < TIME1
b. TIME2 > TIME1
c. TIME1 = TIME2

Test Statistics[b]

	TIME2 - TIME1
Z	-2.207[a]
Asymp. Sig. (2-tailed)	.027

a. Based on negative ranks.
b. Wilcoxon Signed Ranks Test

$z = -2.21, p = .03$

4. Analyzing paired data: An example

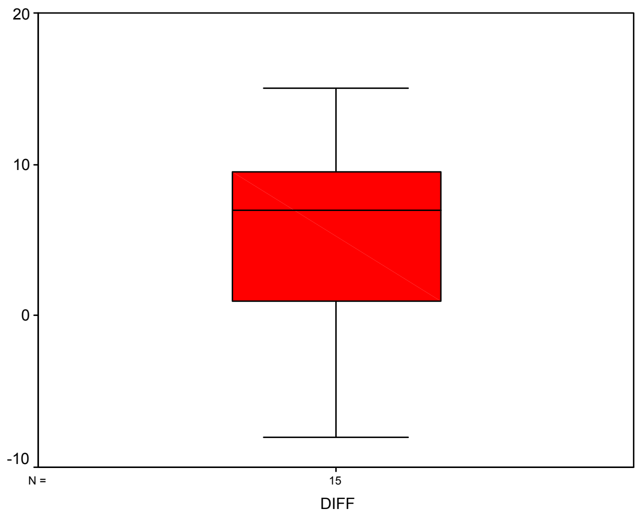
- To investigate the effectiveness of a new method of mathematical instruction, 30 middle-school students were matched (in pairs) on their mathematical ability. Within the pair, students were randomly assigned to receive the traditional mathematics instruction or to receive a new method of instruction. At the end of the study, all participants took the math component of the California Achievement Test (CAT). The following data were obtained:

	Method of Instruction	
Pair	Traditional	New
1	78	74
2	55	45
3	95	88
4	57	65
5	60	64
6	80	75
7	50	41
8	83	68
9	90	80
10	70	64
11	50	43
12	80	82
13	48	55
14	65	57
15	85	75

- First, let's check assumptions necessary for statistical testing
  - The participants who received the old and new methods of instruction are not independent of each other (they are matched on ability), so an independent samples t-test is not appropriate. We can use a paired t-test.

  - The assumptions of the pair t-test are:
    - Participants are randomly selected from the population
    - Normality (actually symmetry) of the difference scores

```
Compute diff = old - new.
EXAMINE VARIABLES=diff
  /PLOT BOXPLOT STEMLEAF NPPLOT.
```



Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
DIFF	.899	15	.091

- There are no apparent problems with the normality assumption.

○ Test the instruction hypothesis using a paired t-test.

```
T-TEST PAIRS old new.
```

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	OLD	69.7333	15	15.77732	4.07369
	NEW	65.0667	15	14.53796	3.75369

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	OLD & NEW	15	.902	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	OLD - NEW	4.6667	6.82084	1.76113	.8894	8.4439	2.650	14	.019

- We find that the new method of instruction significantly lowers math performance. (Good thing we did not use a one-tailed test!)

$$d = \frac{\overline{D}}{\hat{\sigma}_D} = \frac{4.667}{6.821} = .68$$

$$t(14) = 2.65, p = .02, d = .68$$

- We could have also analyzed these data as a randomized block design
  - (Be careful! The data must be entered differently in SPSS for this analysis.)

```
UNIANOVA cat  BY instruct block
  /DESIGN = instruct block.
```

Tests of Between-Subjects Effects

Dependent Variable: CAT

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6281.533[a]	15	418.769	18.002	.000
Intercept	136282.800	1	136282.800	5858.626	.000
INSTRUCT	163.333	1	163.333	7.021	.019
BLOCK	6118.200	14	437.014	18.787	.000
Error	325.667	14	23.262		
Total	142890.000	30			
Corrected Total	6607.200	29			

a. R Squared = .951 (Adjusted R Squared = .898)

  - The test for the method instruction is identical to the paired t-test.

$$F(1,14) = 7.02, p = .02$$

  - In this case, we also obtain a direct test for the effect of the matching on the math scores.

$$F(14,14) = 18.79, p < .01$$

- If we are concerned about the normality/symmetry of the difference scores, then we can turn to a non-parametric test.

```
NPAR TEST
  /WILCOXON=old WITH new (PAIRED)
  /STAT DESC.
```

Ranks

		N	Mean Rank	Sum of Ranks
NEW - OLD	Negative Ranks	11	9.09	100.00
	Positive Ranks	4	5.00	20.00
	Ties	0		
	Total	15		

Test Statistics[b]

	NEW - OLD
Z	-2.276[a]
Asymp. Sig. (2-tailed)	.023

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

$$z = -2.28, p = .02$$

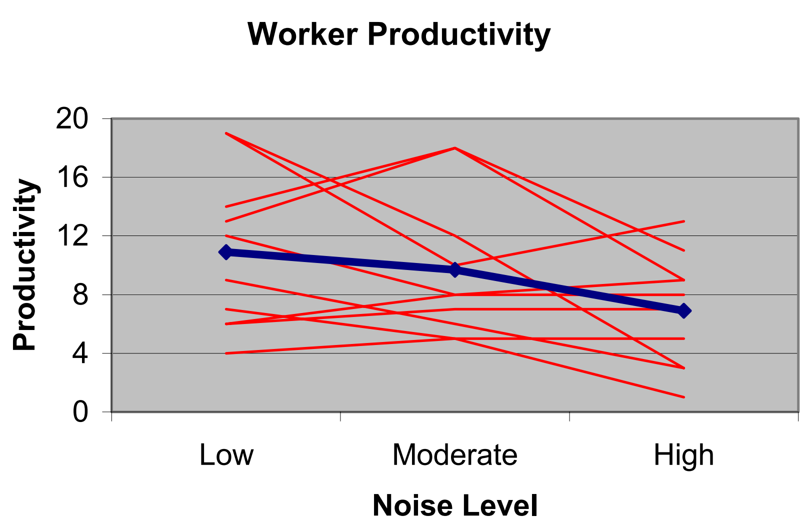
One-Factor Repeated Measures ANOVA

- If we observe participants at more than two time-points, then we need to conduct a repeated measures ANOVA

5. An Initial Example
   - Productivity of factory workers at different noise levels

     We examine the productivity of factory workers under three different noise levels: low, moderate, and high. Each worker experiences each of the noise levels in a random order.

	Noise Level			
Worker	Low	Moderate	High	Average
1	12	8	9	9.7
2	9	6	3	6.0
3	13	18	11	14.0
4	6	8	8	7.3
5	19	12	3	11.3
6	7	5	1	4.3
7	4	5	5	4.7
8	6	7	7	6.7
9	14	18	9	13.7
10	19	10	13	14.0
Average	10.9	9.7	6.9	



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

- What we would like to do is to decompose the variability in the DV into
  - Variability due to individual differences in the participants

    A random effect
  - Variability due to the factor (low, moderate, or high)

    A fixed effect

  - The effect of participants is always a random effect
  - We will only consider situations where the factor is a fixed effect

- The structural model for a one-way within-subjects design

$$Y_{ij} = \mu + \alpha_j + \pi_{\sigma i} + \varepsilon_{ij}$$

or

$$Y_{ij} = \mu + \alpha_j + \pi_{\sigma i} + (\alpha\pi)_{\sigma ij}$$

μ = Grand population mean

$\hat{\mu} = \overline{Y}..$

α_j = The treatment/time effect:

The effect of being in level j of the factor

Or the effect of being at time j

$\sum \alpha_j = 0$

$\hat{\alpha}_j = \overline{Y}._j - \overline{Y}..$

$\pi_{\sigma i}$ = The participant effect:

The random effect due to participant i

$\pi_{\sigma i} \sim N(0, \sigma_\pi)$

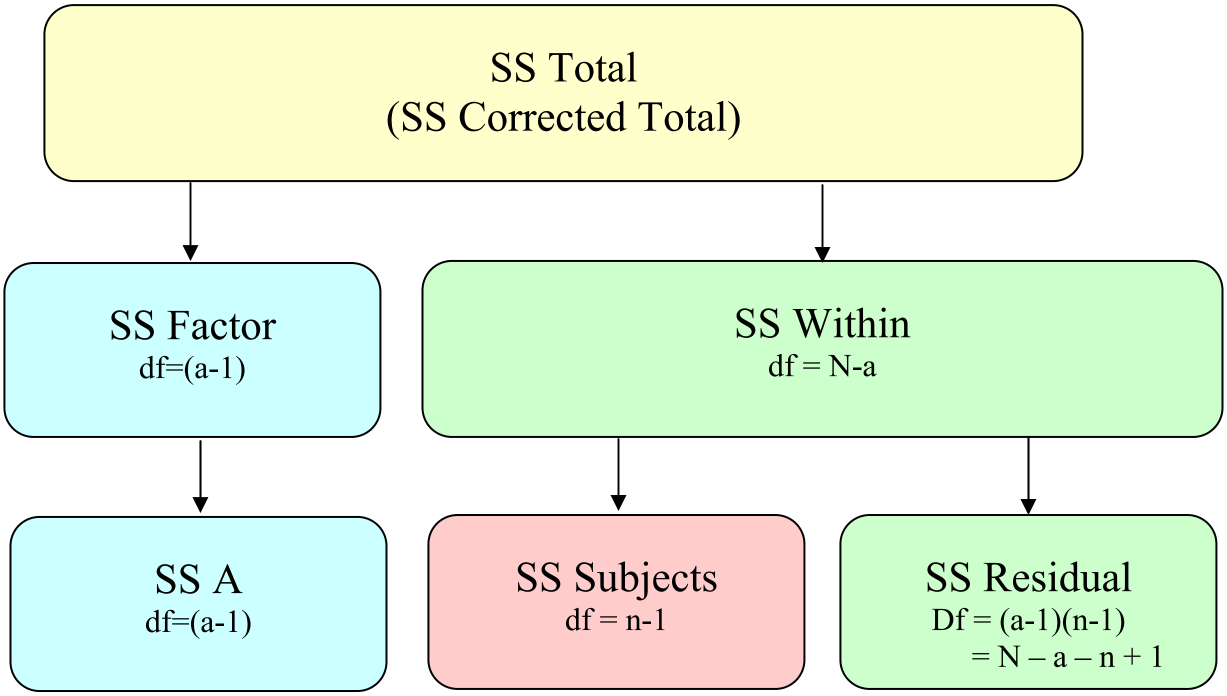
ε_{ij} or $(\alpha\pi)_{\sigma ij}$ = The unexplained error associated with Y_{ij}

$\hat{\varepsilon}_{ij} = \overline{Y}_{ij} - \overline{Y}_i. - \overline{Y}._j + \overline{Y}..$

  - Note that with only one observation on each participant at each level of the factor (or at each time), we can not estimate the participant by factor/time interaction

  - This structural model is just like the model for a randomized block design, with participants as a random blocking effect

* Now, we can identify SS due to the factor of interest, and divide the error term into a SS due to individual differences in the participants, and a SS residual (everything we still can not explain)



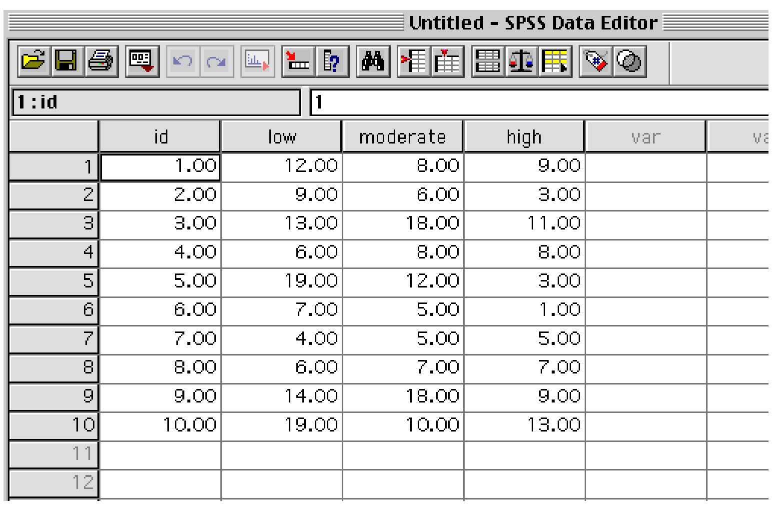
* Sums of squares decomposition and ANOVA table for a repeated measures design:

Source	SS	df	MS	E(MS)	F
Treatment/Time	*SSA*	*a-1*	*MSA*	$\sigma_\varepsilon^2 + \frac{n\sum \alpha_j^2}{a-1}$	$\frac{MSA}{MSE}$
Subjects	*SS(Subject)*	*n-1*	*MS(Sub)*	$\sigma_\varepsilon^2 + n\sigma_\pi^2$	$\frac{MS(Sub)}{MSE}$
Error	*SSError*	*(a-1)(n-1)*	*MSE*	σ_ε^2	
Total	*SST*	*N-1*			

- Conducting the statistical tests in SPSS:

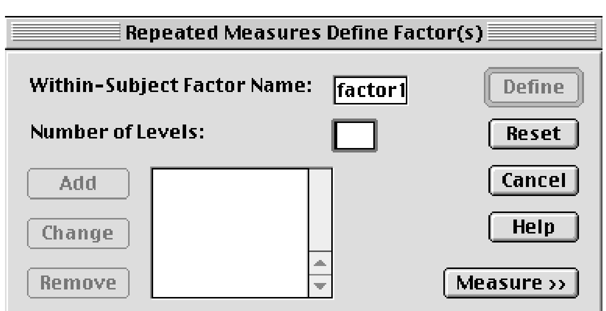
  - Remember, you need to have the data entered with one row per participant.

```
data list free
  /id low moderate high.
begin data.
1 12  8  9
2 9  6  3
3 13  18  11
4 6  8  8
5 19  12  3
6 7  5  1
7 4  5  5
8 6  7  7
9 14  18  9
10 19  10  13
end data.
```

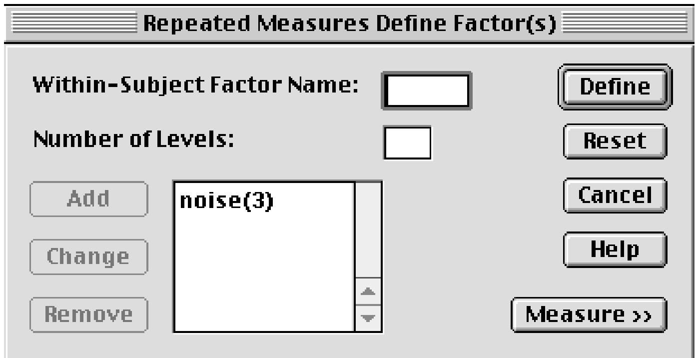


	id	low	moderate	high	var
1	1.00	12.00	8.00	9.00	
2	2.00	9.00	6.00	3.00	
3	3.00	13.00	18.00	11.00	
4	4.00	6.00	8.00	8.00	
5	5.00	19.00	12.00	3.00	
6	6.00	7.00	5.00	1.00	
7	7.00	4.00	5.00	5.00	
8	8.00	6.00	7.00	7.00	
9	9.00	14.00	18.00	9.00	
10	10.00	19.00	10.00	13.00	
11					
12					

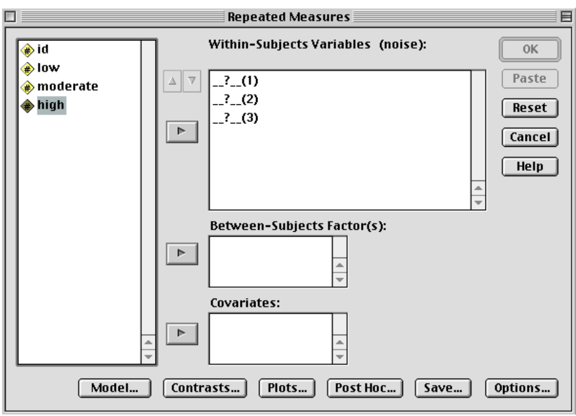
  - When you indicate a repeated measures analysis, the following dialog box will open:

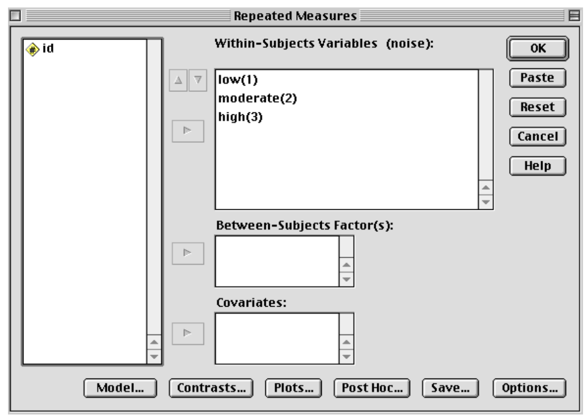


- You need to enter the name of the repeated variable (this can be any label), and the number of levels of the repeated variable.



- When you click define, you need to specify the three variables in your data file that correspond to the three levels of time





- Or you can also use the following syntax

```
GLM  low moderate high
  /WSFACTOR = noise 3
  /PRINT = DESC.
```


- Here is the (unedited) output file:

General Linear Model

Repeats the levels of the repeated measure variable that you entered

Within-Subjects Factors

Measure: MEASURE_1

NOISE	Dependent Variable
1	LOW
2	MODERATE
3	HIGH

Produced by the /PRINT = DESC command

Descriptive Statistics

	Mean	Std. Deviation	N
LOW	10.9000	5.38413	10
MODERATE	9.7000	4.87739	10
HIGH	6.9000	3.84274	10

These are multivariate tests of repeated measures. We will ignore these tests.

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
NOISE	Pillai's Trace	.422	2.918[a]	2.000	8.000	.112
	Wilks' Lambda	.578	2.918[a]	2.000	8.000	.112
	Hotelling's Trace	.730	2.918[a]	2.000	8.000	.112
	Roy's Largest Root	.730	2.918[a]	2.000	8.000	.112

a. Exact statistic

b. Design: Intercept
Within Subjects Design: NOISE

This box will help us test one of the key assumptions for repeated-measures

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

					Epsilon[a]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
NOISE	.945	.456	2	.796	.947	1.000	.500

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

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: NOISE

The test listed under "Sphericity Assumed" is the omnibus repeated measures test of the time factor

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
NOISE	Sphericity Assumed	84.267	2	42.133	3.747	.044
	Greenhouse-Geisser	84.267	1.895	44.469	3.747	.047
	Huynh-Feldt	84.267	2.000	42.133	3.747	.044
	Lower-bound	84.267	1.000	84.267	3.747	.085
Error(NOISE)	Sphericity Assumed	202.400	18	11.244		
	Greenhouse-Geisser	202.400	17.054	11.868		
	Huynh-Feldt	202.400	18.000	11.244		
	Lower-bound	202.400	9.000	22.489		

These are contrasts specified by the command "Polynomial" which are performed on the time variable

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	NOISE	Type III Sum of Squares	df	Mean Square	F	Sig.
NOISE	Linear	80.000	1	80.000	5.806	.039
	Quadratic	4.267	1	4.267	.490	.502
Error(NOISE)	Linear	124.000	9	13.778		
	Quadratic	78.400	9	8.711		

These are tests on the between-subjects factors in the design (in this case there are none)

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	2520.833	1	2520.833	55.949	.000
Error	405.500	9	45.056		

- Let's fill in a standard ANOVA table for a repeated measures design, using the SPSS output labeled "Tests of Within-Subjects Effects"

Source	*SS*	*df*	*MS*	*F*	*p*
Noise	84.267	2	42.133	3.747	.044
Subjects					
Error (Noise)	202.4	18	11.244		
Total					

  - SPSS does not print a test for the random effect of subject. It also does not print SST.
  - We showed previously that for a one-factor repeated measures design, the error term for *Subjects* is the same as the error term for *Noise*. If we had the SS (Subjects), we could construct the test ourselves.
  - In general, however, we are not interested in the test of the effect of individual differences due to subjects.

  - The noise effect compares the marginal noise means, using an appropriate error term [MSE(noise)]

Noise Level		
Low	Moderate	High
10.9	9.7	6.9

$$H_0 : \mu_{1} = \mu_{\cdot 2} = \mu_{\cdot 3}$$
$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$F(2,18) = 3.75, p = .04$

We conclude that productivity is not the same at all three noise levels. We need to conduct follow-up tests to determine exactly how these means differ.

* Suppose we incorrectly treated the data as a between-subjects design. We can compare this incorrect analysis with the correct within-subjects design.

	Between-Subjects	Within-Subjects
MSE	22.51	11.24
F-value	$F(2,27) = 1.87$	$F(2,18) = 3.75$
p-value	p=.173	p=.044

  - By conducting the proper within-subjects analysis, we decreased the error term and increased our power to detect the important noise effect.

  - Note: This example is only to show how a within-subjects design decreases the error term. You do not get to choose between conducting a between-subjects or a within-subjects analysis. Your design will determine the analysis.

7. Repeated measures as a randomized block design

* We can also recognize a repeated measures design as a special case of a randomized block design. In this framework, the participants are a random blocking factor.

* We previously discussed only how to analyze fixed blocking factors. I will not go into the details, but SPSS can easily handle random blocks

  In general, you will NOT analyze a repeated measures design as a randomized block design. However, it is statistically valid and helps place this design within a (relatively) familiar framework

```
UNIANOVA dv BY subj group
  /RANDOM = subj
  /DESIGN = subj group.
```

Tests of Between-Subjects Effects

Dependent Variable: DV

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	2520.833	1	2520.833	55.949	.000
	Error	405.500	9	45.056[a]		
SUBJ	Hypothesis	405.500	9	45.056	4.007	.006
	Error	202.400	18	11.244[b]		
NOISE	Hypothesis	84.267	2	42.133	3.747	.044
	Error	202.400	18	11.244[b]		

a. MS(SUBJ)

b. MS(Error)

- (Main) Effect of the noise level: $F(2,18) = 3.75$, p = .044
  - Worker productivity is affected by noise level
  - This is the exact same test result we obtained from the repeated-measures analysis

- (Main) Effect of participant: $F(9,18) = 4.01$, $p = .006$
  - Worker productivity varies by participant
  - This is the subject effect that SPSS (and most computer programs) does not print for a repeated-measures analysis

8. Assumptions of a repeated measures design

- For a repeated measures design, we start with the same assumptions as a paired t-test
  - Participants are independent and randomly selected from the population
  - Normality (actually symmetry)
    - In general, people check the normality of the scores at each time/treatment level.
    - Technically, this is overly restrictive, but it is sufficient.

- Due to having more than two measurements on each participant, we have an additional assumption on the variances

- Before we delve into the details of the variance assumption, let's look at the variances in a design with three repeated measures:
  - We have the variance at time 1, at time 2, and at time 3:

$$Var(Time_1) = \sigma_1^2$$
$$Var(Time_2) = \sigma_2^2$$
$$Var(Time_3) = \sigma_3^2$$

  - But these variances are not independent of each other. Because the same person responds at time 1, at time 2, and at time 3, the responses of a participant at each of these times will be correlated with each other. We can examine these relationships by looking at the covariances of the measures at each time

$$Cov(Time_1, Time_2) = \sigma_{12} = \rho_{12}\sigma_1\sigma_2$$
$$Cov(Time_1, Time_3) = \sigma_{13} = \rho_{13}\sigma_1\sigma_3$$
$$Cov(Time_2, Time_3) = \sigma_{23} = \rho_{23}\sigma_2\sigma_3$$

ρ_{ij} is the correlation between measurements at time i and time j

σ_i is standard deviation of measurements at time i

  - Putting all this information together, we can construct a variance-covariance matrix of the observations:

	$Time_1$	$Time_2$	$Time_3$
$Time_1$	σ_1^2	σ_{12}	σ_{13}
$Time_2$	σ_{12}	σ_2^2	σ_{23}
$Time_3$	σ_{13}	σ_{23}	σ_3^2

The variances of responses at each time are on the diagonal
The covariances of responses between two times are off the diagonal
The matrix is symmetric (the top triangle = the bottom triangle), so sometimes we only write half of the matrix:

	$Time_1$	$Time_2$	$Time_3$
$Time_1$	σ_1^2		
$Time_2$	σ_{12}	σ_2^2	
$Time_3$	σ_{13}	σ_{23}	σ_3^2

- Now, we can examine the assumption of <u>compound symmetry</u>. Compound symmetry specifies a specific structure for the variance/covariance matrix where:

  - All of the variances are equal: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ $\qquad \sigma^2$
  - All of the covariances are equal: $\sigma_{12} = \sigma_{13} = \sigma_{23}$ $\qquad \sigma_c$

$$\begin{array}{cccc} & Time_1 & Time_2 & Time_3 \\ Time_1 & \sigma^2 & \sigma_c & \sigma_c \\ Time_2 & \sigma_c & \sigma^2 & \sigma_c \\ Time_3 & \sigma_c & \sigma_c & \sigma^2 \end{array}$$

  - If we divide by the common variance, we can state the assumption in terms of the correlation between measurements at different time points:

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

- To have compound symmetry, we must have the correlations between observations at each time period equal to each other!

- Let's consider an example to see how we examine compound symmetry in real data:

	Age (Months)				
Subject	30	36	42	48	**Mean**
1	108	96	110	122	**109**
2	103	117	127	133	**120**
3	96	107	106	107	**104**
4	84	85	92	99	**90**
5	118	125	125	116	**121**
6	110	107	96	91	**101**
7	129	128	123	128	**127**
8	90	84	101	113	**97**
9	84	104	100	88	**94**
10	96	100	103	105	**101**
11	105	114	105	112	**109**
12	113	117	132	130	**123**
Mean	**103**	**107**	**110**	**112**	**108**

- The estimation of the fixed components of the structural model parameters is straightforward:

$$Y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$$

$$\hat{\mu} = \bar{X}_{..} = 108$$

$$\hat{\alpha}_j = \bar{X}_{.j} - \bar{X}_{..}$$

$$\hat{\alpha}_1 = 103 - 108 = -5$$
$$\hat{\alpha}_2 = 107 - 108 = -1$$
$$\hat{\alpha}_3 = 110 - 108 = 2$$
$$\hat{\alpha}_4 = 112 - 108 = 4$$

- To look at the across-time correlation and covariance matrix, we have to use the *CORRELATION* command.

```
CORRELATIONS
  /VARIABLES=time1 time2 time3 time4
  /STATISTICS XPROD.
```

Correlations

		TIME1	TIME2	TIME3	TIME4
TIME1	Pearson Correlation	1	.795	.696	.599
	Sig. (2-tailed)	.	.002	.012	.040
	Sum of Squares and Cross-products	2068.000	1698.000	1401.000	1333.000
	Covariance	188.000	154.364	127.364	121.182
	N	12	12	12	12
TIME2	Pearson Correlation	.795	1	.760	.466
	Sig. (2-tailed)	.002	.	.004	.127
	Sum of Squares and Cross-products	1698.000	2206.000	1580.000	1072.000
	Covariance	154.364	200.545	143.636	97.455
	N	12	12	12	12
TIME3	Pearson Correlation	.696	.760	1	.853
	Sig. (2-tailed)	.012	.004	.	.000
	Sum of Squares and Cross-products	1401.000	1580.000	1958.000	1849.000
	Covariance	127.364	143.636	178.000	168.091
	N	12	12	12	12
TIME4	Pearson Correlation	.599	.466	.853	1
	Sig. (2-tailed)	.040	.127	.000	.
	Sum of Squares and Cross-products	1333.000	1072.000	1849.000	2398.000
	Covariance	121.182	97.455	168.091	218.000
	N	12	12	12	12

- The covariance of a variable with itself is the variance of that variable (the diagonal elements). The variances look reasonably similar

$$\sigma_1^2 = 188.0 \qquad \sigma_3^2 = 178.0$$
$$\sigma_2^2 = 200.5 \qquad \sigma_4^2 = 218.0$$

- However from the correlations/covariances, we see that the correlations between the measurements at time four vary more.

$$\sigma_{34} = 168.09 \qquad \sigma_{24} = 97.46$$
$$r_{34} = .85 \qquad r_{24} = .47$$

- In 1970, several researchers discovered that compound symmetry is actually an overly restrictive assumption.

  - Recall that for the paired t-test, we made no assumption about the homogeneity of the two variances. In fact, we did not use the variances of the time 1 and time 2 data. However, we did compute a variance of the difference of the paired data.

  - The same logic applies to a more general repeated measures design. We are actually concerned about the variance of the difference of observations across the time periods.

  - Earlier, we used the following formula for the variance of the difference of two variables:

$$\sigma^2_{X_1 - X_2} = \sigma^2_{X_1} + \sigma^2_{X_2} - 2\rho_{12}\sigma_{X_1}\sigma_{X_2}$$

  - If we have compound symmetry, then $\sigma^2_{X_i} = \sigma^2_{X_j} = \sigma^2$ and $\rho_{ij} = \rho$ so that we can write:

$$\sigma^2_{X_i - X_j} = \sigma^2 + \sigma^2 - 2\rho\,\sigma\,\sigma$$

    There are no subscripts on the right side of the equation. In other words, under compound symmetry the variance of the difference of any two variables is the same for all variables.

  - The assumption that the variance of the difference of all variables is a constant is known as sphericity. Technically, it is the assumption of sphericity that we need to satisfy for repeated measures ANOVA.

  - If you satisfy the assumptions of compound symmetry, then you automatically satisfy the sphericity assumption. However it is possible (but rather rare) to satisfy sphericity but not the compound symmetry assumption.

    (For you math-heads, compound symmetry is a sufficient condition, but not a necessary condition)

- Here is an example of a variance/covariance matrix that is spherical, but does not display compound symmetry

$$\begin{bmatrix} 2 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 6 \end{bmatrix}$$

$$\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}$$

$$\sigma_{12}^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} = 2 + 4 - 2(1) = 4$$

$$\sigma_{13}^2 = \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} = 2 + 6 - 2(2) = 4$$

$$\sigma_{23}^2 = \sigma_2^2 + \sigma_3^2 - 2\sigma_{23} = 4 + 6 - 2(3) = 4$$

- Why wasn't sphericity listed as an assumption for the paired t-test?

  When you only have two repeated measures, then the assumption of sphericity is automatically satisfied:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

  With only two groups, there is only one difference of variances so this difference must satisfy the sphericity assumption

$$\sigma_{12}^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} = c$$

- Sphericity is difficult to check, so in practice, we tend to check the compound symmetry assumption.

- To recap, the assumptions we have to check for a repeated measures design are:
  - Participants are independent and randomly selected from the population
  - Normality (actually symmetry)
  - Sphericity (In practice, compound symmetry)
    - SPSS prints a few tests of sphericity, but due to the assumptions required of these tests, they are essentially worthless and you should not rely upon them at all.

- What can I do when the compound symmetry assumption is violated?
  - Transformations are of little use due to the complicated nature of the variance/covariance matrix

  - One possibility is the non-parametric, rank-based Friedman test. The Friedman test is appropriate for a one-factor within-subjects design. (However, non-parametric alternatives are not available for multi-factor within-subjects designs or for designs with both within and between factors)

    - Consider *n* participants measured at *a* different times.
    - For each subject, replace the observation by the rank of the observation

Original Data

	Age (Months)			
Subject	30	36	42	48
1	108	96	110	122
2	103	117	127	133

Ranked Data

	Age (Months)			
Subject	30	36	42	48
1	2	1	3	4
2	1	2	3	4

- Conceptually, you then perform a repeated-measures ANOVA on the ranked data
- SPSS performs this test and gives a chi-square statistic and p-value

- In multi-factor designs, the rank-based test less useful. Thus, in many situations you are left with three options:

  - Play dumb and ignore the violation of the variance assumption. Many people use this option. Unfortunately, when the sphericity assumption is violated, the actual Type I error rate will be inflated

  - Use an adjusted test. When the variances across groups in a oneway design were not equal, we used a Brown-Forsythe correction for unequal variances. The BF adjustment corrected the degrees of freedom to correct for the unequal variances. Two such adjustments exist for the repeated measures case: $\tilde{\varepsilon}$ and $\hat{\varepsilon}$.

  - Reject the ANOVA approach to repeated measures altogether and go for a MANOVA approach. This approach makes no assumptions on the variance/covariance matrix and is favored by many statisticians. Unfortunately, the MANOVA approach is beyond the scope of this class.

- Corrected repeated measures ANOVA and when to use them:

  - In 1954 Box derived a measure of how far a variance/covariance matrix departs from sphericity, ε.

    If the data exactly satisfy the sphericity assumption, $\varepsilon = 1$

    If the data are not spherical, then $\varepsilon < 1$.

    The greater ε departs from 1, the greater the departure from sphericity

    Box also showed that it was possible to adjust a standard F-test to correct for the non-sphericity.

    $df_{NUM} = \varepsilon(a-1)$ $\quad$ $a = \textit{Levels of the repeated measure variable}$

    $df_{DEN} = \varepsilon(a-1)(n-1)$ $\quad$ $n = \textit{number of participants}$

    At the time, Box did not know how to estimate ε. Since that time, three estimates have been proposed.

    - The Lower-Bound Adjustment
    - Geisser-Greenhouse's (1959) $\hat{\varepsilon}$
    - Huynh-Feldt's (1976) $\tilde{\varepsilon}$

- In 1958 Geisser-Greenhouse showed that if the repeated-measure variable had a levels, then ε could be no lower than $\frac{1}{(a-1)}$. This is the lower bound estimate of ε. SPSS prints the lower-bound estimate, but more recently, it has become possible to estimate ε. Thus, the lower-bound estimate is out-of-date and should never be used.

- Geisser-Greenhouse (1959) developed a refinement of Box's estimate of ε, called $\hat{\varepsilon}$. $\hat{\varepsilon}$ controls the Type I error rate, but tends to be overly conservative by underestimating ε.

- Huynh-Feldt's (1976) $\tilde{\varepsilon}$ slightly overestimates ε, and in some cases can lead to a slightly inflated Type I error rate, but this overestimate is very small.

- In most cases, $\hat{\varepsilon}$ and $\tilde{\varepsilon}$ give very similar results, but when they differ most statisticians (except Huynh & Feldt) prefer to use $\hat{\varepsilon}$

- Of the three options available in SPSS, the Geisser-Greenhouse $\hat{\varepsilon}$ is the best bet. Remember that $\hat{\varepsilon}$ is an adjustment when sphericity is not a valid assumption in the data. If the data deviate greatly from being spherical, then it may not be possible to correct the F-test. Here is a general rule of thumb:

  $\hat{\varepsilon} > .9$ The sphericity assumption is satisfied, no correction necessary

  $.9 > \hat{\varepsilon} > .7$ The sphericity assumption is not satisfied, use the GG $\hat{\varepsilon}$ correction

$$df_{NUM} = \hat{\varepsilon}(a-1)$$
$$df_{DEN} = \hat{\varepsilon}(a-1)(n-1)$$

  $.7 > \hat{\varepsilon}$ The sphericity assumption is not satisfied, and it is violated so severely that correction is not possible. You should switch to the MANOVA approach to repeated measures or avoid omnibus tests

- Let's look at our age example. By visually examining the variance/covariance matrix, we noticed that the data did not exhibit compound symmetry. Let's see if looking at the estimates of ε confirm our suspicions.

Mauchly's Test of Sphericity

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.243	13.768	5	.018	.610	.725	.333

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

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

DO NOT use Mauchly's test of Sphericity. It is not a reliable test.

In this case both $\hat{\varepsilon}$ and $\tilde{\varepsilon}$ converge to the same conclusion. We do not have sphericity in this data. We can now use $\hat{\varepsilon}$ to construct an adjusted F-test.

- To apply the $\hat{\varepsilon}$ correction, we use the unadjusted F-value, however, we correct the numerator and denominator degrees of freedom

$$df_{NUM} = \hat{\varepsilon}(a-1)$$
$$df_{DEN} = \hat{\varepsilon}(a-1)(n-1)$$

For the age data, SPSS gives us the following analyses:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	552.000	3	184.000	3.027	.043
	Greenhouse-Geisser	552.000	1.829	301.865	3.027	.075
	Huynh-Feldt	552.000	2.175	253.846	3.027	.064
	Lower-bound	552.000	1.000	552.000	3.027	.110
Error(TIME)	Sphericity Assumed	2006.000	33	60.788		
	Greenhouse-Geisser	2006.000	20.115	99.727		
	Huynh-Feldt	2006.000	23.920	83.863		
	Lower-bound	2006.000	11.000	182.364		

The uncorrected test of time: $F(3,33) = 3.03, p = .04$

$$\hat{\varepsilon} = .610 \qquad \begin{aligned} df_{NUM} &= .61(3) &&= 1.83 \\ df_{DEN} &= .61(3)(11) &&= 20.13 \end{aligned}$$

The corrected GG test of time: $F(1.83, 20.13) = 3.03, p = .075$

(The exact p-value can be obtained from SPSS, above, or EXCEL)

- We fail to reject the null hypothesis and conclude that there is no significant difference in the DV across the four time periods.

- As in the between-subjects case, the within-subjects ANOVA is relatively robust to violations of the normality assumption. However, it is not at all robust to violations of sphericity.

- As we will see shortly, contrasts can be conducted without worrying about the sphericity assumption. Contrasts are the way to go!

9. Contrasts

- In a repeated measures design, MSE is an omnibus error term, computed across all the repeated-measures. If there are *a* repeated-measures, then the MSE is an average of *(a-1)* error terms!
  - The *(a-1)* error terms are determined by *(a-1)* orthogonal contrasts run on the *a* repeated-measures. An error term specific to each contrast is computed.
  - The default in SPSS is to conduct polynomial contrasts on the *a* repeated-measures.
  - Let's look at an example with our age data. Here is the within-subject contrast table printed by default in SPSS

```
GLM  time1 time2 time3 time4
 /WSFACTOR = time 4 Polynomial
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Linear	540.000	1	540.000	5.024	.047
	Quadratic	12.000	1	12.000	.219	.649
	Cubic	.000	1	.000	.000	1.000
Error(TIME)	Linear	1182.400	11	107.491		
	Quadratic	604.000	11	54.909		
	Cubic	219.600	11	19.964		

- Note that there are three separate error terms, each with 11 df

  Error (Linear) 107.491
  Error (Quad) 54.909
  Error (Cubic) 19.964

- If we take the average of these three terms, we obtain the MSE with 33 df:

$$\frac{107.491 + 54.909 + 19.964}{3} = 60.788$$

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	552.000	3	184.000	3.027	.043
Error(TIME)	Sphericity Assumed	2006.000	33	60.788		

- The formulae for contrasts are the same as for a one-way ANOVA:

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

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

- The only difference is that MSE has been replaced by MSE'

- We have two choices for MSE'

  - Use the omnibus MSE with $df = (a\text{-}1)(n\text{-}1)$

    When the sphericity assumption is satisfied, each of the separate error terms should be identical. They all estimate the true error variance
    In this case, using the omnibus MSE results in greater power due to the increased denominator degrees of freedom

  - Use the contrast-specific error term with $df = (n\text{-}1)$

    In practice, the loss of power due to a decrease in the degrees of freedom is offset by having a more accurate error term
    Due to the problems of the sphericity assumption and the difficulties in checking this assumption, **most statisticians recommend you always use the contrast-specific error term**

- Unfortunately, calculating the contrast-specific error term is tedious. I will provide a few different methods to avoid any hand-calculations.

- Understanding repeated-measures contrasts

  - These repeated-measures contrasts operate on the marginal repeated-factor means, collapsing across the participants

  - In our infant growth example, we have four repeated measures. With four observations, we can conduct 3 single-df tests. Let's examine the polynomial trends:

	Age (Months)			
	30	36	42	48
Average	103	107	110	112
Linear	-3	-1	1	3
Quad	1	-1	-1	1
Cubic	-1	3	-3	1

$$\hat{\psi}_{lin} = -3(103) + (-1)(107) + (1)(110) + (3)(112) = 30$$

$$SS(\hat{\psi}_{lin}) \frac{30^2}{\frac{(-3)^2}{12} + \frac{(-1)^2}{12} + \frac{(1)^2}{12} + \frac{(3)^2}{12}} = \frac{900}{\left(\frac{20}{12}\right)} = 540$$

$$\hat{\psi}_{quad} = 1(103) + (-1)(107) + (-1)(110) + (1)(112) = -2$$

$$\hat{\psi}_{cub} = -1(103) + (3)(107) + (-3)(110) + (1)(112) = 0$$

- Computing the error term: Method 1 Using SPSS's built-in, brand-name contrasts

  - Remember SPSS's built-in contrasts? If your contrast of interest is one of those contrasts, you can ask SPSS to print the test of the contrast

    - Difference: Each level of a factor is compared to the mean of the previous levels
    - Helmert: Each level of a factor is compared to the mean of subsequent levels
    - Polynomial: Uses the orthogonal polynomial contrasts
    - Repeated: Each level of a factor is compared to the previous level
    - Simple: Each level of a factor is compared to the last level

- The default is to report the polynomial contrasts:

```
GLM  time1 time2 time3 time4
  /WSFACTOR = time 4 Polynomial.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Linear	540.000	1	540.000	5.024	.047
	Quadratic	12.000	1	12.000	.219	.649
	Cubic	.000	1	.000	.000	1.000
Error(TIME)	Linear	1182.400	11	107.491		
	Quadratic	604.000	11	54.909		
	Cubic	219.600	11	19.964		

$$\psi_{lin} : F(1,11) = 5.02, p = .047$$
$$\psi_{quad} : F(1,11) = 0.22, p = .649$$
$$\psi_{cub} : F(1,11) = 0.00, p > .999$$

- But you can ask for any of the brand-name contrasts:

```
GLM  time1 time2 time3 time4
  /WSFACTOR = time 4 repeated.
```

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	192.000	1	192.000	2.405	.149
	Level 2 vs. Level 3	108.000	1	108.000	1.183	.300
	Level 3 vs. Level 4	48.000	1	48.000	.802	.390
Error(TIME)	Level 1 vs. Level 2	878.000	11	79.818		
	Level 2 vs. Level 3	1004.000	11	91.273		
	Level 3 vs. Level 4	658.000	11	59.818		

$$\psi_{1vs2} : F(1,11) = 2.41, p = .15$$
$$\psi_{2vs3} : F(1,11) = 1.18, p = .30$$
$$\psi_{3vs4} : F(1,11) = 0.80, p = .39$$

- <u>Computing the error term: Method 2</u> Using SPSS's "*special*" contrast command

    - Instead of specifying a set of brand-name contrasts, you can enter a set of contrasts.
    - You <u>must</u> first enter a contrast of all 1's
    - You then <u>must</u> enter *(a-1)* contrasts

    - For example, let's ask for the polynomial contrasts using the "*special*" command:

```
GLM  time1 time2 time3 time4
 /WSFACTOR = time 4 special ( 1  1 1 1
                             -3 -1 1 3
                              1 -1 -1 1
                             -1 3 -3 1).
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	L1	10800.000	1	10800.000	5.024	.047
	L2	48.000	1	48.000	.219	.649
	L3	.000	1	.000	.000	1.000
Error(TIME)	L1	23648.000	11	2149.818		
	L2	2416.000	11	219.636		
	L3	4392.000	11	399.273		

L1 is the linear trend: $\psi_{lin} : F(1,11) = 5.02, p = .047$

L2 is the quadratic trend: $\psi_{quad} : F(1,11) = 0.22, p = .649$

L3 is the cubic trend: $\psi_{cub} : F(1,11) = 0.00, p > .99$

- If all of your contrasts are not ortho-normalized (the set of contrasts is orthogonal, and the sum of the squared contrast coefficients equal 1), then the calculation of the SS will be disturbed, but the F-value and the significance test will be ok.

- Now, let's consider a more complicated set of contrasts

```
GLM  time1 time2 time3 time4
  /WSFACTOR = time 4 special (1  1 1 1
                              -1 -1 2 0
                              -1 0 2 -1
                               0 -1 2 -1)
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	L1	1200.000	1	1200.000	3.689	.081
	L2	300.000	1	300.000	1.680	.221
	L3	12.000	1	12.000	.153	.703
Error(TIME)	L1	3578.000	11	325.273		
	L2	1964.000	11	178.545		
	L3	864.000	11	78.545		

L1 is (time 1 and time 2) vs. time 3, $F(1,11) = 3.69, p = .08$
L2 is (time 1 and time 4) vs. time 3, $F(1,11) = 1.68, p = .22$
L3 is (time 2 and time 4) vs. time 3, $F(1,11) = 0.15, p = .70$

- To use the special command, you need to remember to enter the contrast of all 1's and a set of *(a-1)* contrasts.

• <u>Computing the error term: Method 3</u> Compute the value of the contrast as a new variable, and run a one-sample t-test on that variable to test for a difference from zero

- Again, let's start by replicating the polynomial contrasts using this method:

```
compute lin = -3*time1 - time2 + time3 + 3*time4.
compute quad = time1 - time2 - time3 + time4.
compute cubic = -time1 +3* time2 -3*time3 + time4.

T-TEST /TESTVAL=0
  /VARIABLES=lin.
```


One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
LIN	2.241	11	.047	30.0000	.5404	59.4596

ψ_{lin} $t(11) = 2.24, p = .047$ ψ_{lin} $:F(1,11) = 5.02, p = .047$

- The advantage of this procedure is that you can test any contrast one-at-a-time

- This method should make intuitive sense. When we run a contrast, we collapse across participants. With this method, we create a new variable reflecting the value of the contrast for each participant. We then collapse across participants, average all contrast values, and test to see if this average contrast differs from zero!

- So for repeated-measures designs, if we only run contrasts:
  - We can discard the sphericity assumption
  - We can run tests that are easily interpretable

  - The catch is that you need to have strong enough predictions that you can plan contrasts before looking at the data.

10. Planned and post-hoc tests

- Remember that with all the controversy surrounding the assumptions for a repeated-measures design, you should try very hard to avoid omnibus tests and conduct all your analyses using single-df tests.

- The logic for planned and post-hoc tests for a one-factor within-subjects design parallels the logic for a one-factor between-subjects design

    - When there are a repeated-measures, the unadjusted omnibus test has $(a-1)$ dfs

    - If you decide to forgo the omnibus test, then you can use those $(a-1)$ dfs to conduct $(a-1)$ orthogonal, planned contrasts.

      I believe that in the presence of a strong theory, you may conduct $(a-1)$ planned contrasts, even if they are not orthogonal.

- If you first conduct the omnibus test, or if you plan more than $(a-1)$ contrasts, then you need to adjust your p-values to correct for the number of tests you are conducting, using either the Bonferroni or Dunn-Sidák correction:

Dunn/Sidák	Bonferroni
$p_{crit} = 1-(1-\alpha)^{\frac{1}{c}}$	$p_{crit} = \frac{\alpha}{c}$

- When you conduct post-hoc tests, you need to adjust your tests using either Tukey's HSD (for pair-wise comparisons), Scheffé (for complex comparisons), or some other appropriate correction.

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

      Where α = Experimentwise error rate
      a = Number of repeated-measures
      ν = df (error)

        - For single-df tests, df (error) should be $(n-1)$, the df associated with the contrast-specific error estimate.
        - If a person has unwisely decided to use the omnibus MSE term, then the appropriate df error should be $(n-1)(a-1)$.

- To determine significance at the *(1-α)* level,

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

  - To use the Scheffé correction, compute $F_{Crit} = (a-1)F_{\alpha=.05;a-1,\nu}$

    Where α = Experimentwise error rate
    a = Number of repeated-measures
    ν = df(error): $(n\text{ -1})$ for a contrast-specific error estimate.

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

- SPSS only computes post-hoc adjustments for between-subjects factors. We are on our own for within-subjects factors!
  - First, compute the test-statistic for the contrast using one of the previous methods
  - Next, compute the adjusted critical value
  - Finally, compare the observed test-statistic to the critical value to determine significance

- As an example, let's conduct all pairwise comparisons in the age data. I'll use the *simple* command. To conduct all six pairwise comparisons, I need to run three *simple* commands.

```
GLM time1 time2 time3 time4
  /WSFACTOR = age 4 simple (1).
GLM time1 time2 time3 time4
  /WSFACTOR = age 4 simple (2).
GLM time1 time2 time3 time4
  /WSFACTOR = age 4 simple (3).
```


Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	age	Type III Sum of Squares	df	Mean Square	F	Sig.
age	Level 2 vs. Level 1	192.000	1	192.000	2.405	.149
	Level 3 vs. Level 1	588.000	1	588.000	5.284	.042
	Level 4 vs. Level 1	972.000	1	972.000	5.940	.033
Error(age)	Level 2 vs. Level 1	878.000	11	79.818		
	Level 3 vs. Level 1	1224.000	11	111.273		
	Level 4 vs. Level 1	1800.000	11	163.636		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	age	Type III Sum of Squares	df	Mean Square	F	Sig.
age	Level 1 vs. Level 2	192.000	1	192.000	2.405	.149
	Level 3 vs. Level 2	108.000	1	108.000	1.183	.300
	Level 4 vs. Level 2	300.000	1	300.000	1.341	.271
Error(age)	Level 1 vs. Level 2	878.000	11	79.818		
	Level 3 vs. Level 2	1004.000	11	91.273		
	Level 4 vs. Level 2	2460.000	11	223.636		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	age	Type III Sum of Squares	df	Mean Square	F	Sig.
age	Level 1 vs. Level 3	588.000	1	588.000	5.284	.042
	Level 2 vs. Level 3	108.000	1	108.000	1.183	.300
	Level 4 vs. Level 3	48.000	1	48.000	.802	.390
Error(age)	Level 1 vs. Level 3	1224.000	11	111.273		
	Level 2 vs. Level 3	1004.000	11	91.273		
	Level 4 vs. Level 3	658.000	11	59.818		

$$F_{crit} = \frac{(q_{crit}(1-\alpha, a, dfe))^2}{2} = \frac{(q_{crit}(.95,4,11))^2}{2} = \frac{(4.26)^2}{2} = 9.07$$

- None of the pairwise comparisons reach statistical significance.

  30 Months vs. 36 Months: $F(1,11) = 2.41$, *ns*
  30 Months vs. 42 Months: $F(1,11) = 5.28$, *ns*
  30 Months vs. 48 Months: $F(1,11) = 5.94$, *ns*
  36 Months vs. 42 Months: $F(1,11) = 1.18$, *ns*
  36 Months vs. 48 Months: $F(1,11) = 1.34$, *ns*
  42 Months vs. 48 Months: $F(1,11) = 0.80$, *ns*

11. Effect sizes

- The details for computing effect sizes for repeated measures designs have not been entirely worked out. The use of different error terms for each effect creates from special problems for effect size calculations.

- One simple, proposed measure is partial eta-squared:

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

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

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

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

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

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

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

- Research is still being conducted on effect sizes for repeated measures designs. One of the more promising measures is generalized eta squared, η^2_G (see Olejnik & Algina, 2003; Bakeman, 2005).

- Examples of Effect Size Calculations:

  - Omnibus test of the within-subject factor:

```
GLM  time1 time2 time3 time4
 /WSFACTOR = time 4 Polynomial
```

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	552.000	3	184.000	3.027	.043
	Greenhouse-Geisser	552.000	1.829	301.865	3.027	.075
	Huynh-Feldt	552.000	2.175	253.846	3.027	.064
	Lower-bound	552.000	1.000	552.000	3.027	.110
Error(TIME)	Sphericity Assumed	2006.000	33	60.788		
	Greenhouse-Geisser	2006.000	20.115	99.727		
	Huynh-Feldt	2006.000	23.920	83.863		
	Lower-bound	2006.000	11.000	182.364		

$$\hat{\eta}^2_{Time} = \frac{SS_{Time}}{SS_{Time} + SS_{ErrorTermForTime}} = \frac{552}{552 + 2006} = .22$$

$$F(3,33) = 3.03, p = .04, \eta^2 = .22$$

- Recall that for this example, the sphericity assumption has been severely violated. Thus, we should not report this test. This calculation has been included to show you an example of how to calculate partial eta squared for the omnibus test, but it is not appropriate to report this test.
- If you applied an epsilon adjustment, the effect size calculation would be unaffected.

- Polynomial Trends: Partial Eta-Squared or *r*

```
GLM  time1 time2 time3 time4
  /WSFACTOR = time 4 Polynomial
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	TIME	Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Linear	540.000	1	540.000	5.024	.047
	Quadratic	12.000	1	12.000	.219	.649
	Cubic	.000	1	.000	.000	1.000
Error(TIME)	Linear	1182.400	11	107.491		
	Quadratic	604.000	11	54.909		
	Cubic	219.600	11	19.964		

$$r_{Linear} = \sqrt{\frac{F_{Contrast}}{F_{Contrast} + df_{contrast}}} = \sqrt{\frac{5.02}{5.02 + 11}} = .56$$

$$r_{Quadratic} = \sqrt{\frac{0.22}{0.22 + 11}} = .14$$

$$\eta^2_{Linear} = \frac{SS_{Linear}}{SS_{Linear} + SS_{ErrorTermForLinear}} = \frac{540}{540 + 1182} = .31$$

$$\eta^2_{Quadratic} = \frac{SS_{Quadratic}}{SS_{Quadratic} + SS_{ErrorTermForQuadratic}} = \frac{12}{12 + 604} = .02$$

$\psi_{lin}: F(1,11) = 5.02, p = .05, \eta^2 = .31$

$\psi_{quad}: F(1,11) = 0.22, p = .65, \eta^2 = .02$

$\psi_{cub}: F(1,11) = 0.00, p > .99, \eta^2 < .01$

- Pairwise Contrasts: Partial Eta-Squared or d

```
GLM time1 time2 time3 time4
  /WSFACTOR = age 4 simple (1).
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	age	Type III Sum of Squares	df	Mean Square	F	Sig.
age	Level 2 vs. Level 1	192.000	1	192.000	2.405	.149
	Level 3 vs. Level 1	588.000	1	588.000	5.284	.042
	Level 4 vs. Level 1	972.000	1	972.000	5.940	.033
Error(age)	Level 2 vs. Level 1	878.000	11	79.818		
	Level 3 vs. Level 1	1224.000	11	111.273		
	Level 4 vs. Level 1	1800.000	11	163.636		

$$\eta^2_{1vs2} = \frac{SS_{1vs2}}{SS_{1vs2} + SS_{ErrorTermFor1vs2}} = \frac{192}{192+878} = .18$$

$$\eta^2_{1vs3} = \frac{588}{588+1224} = .32 \qquad \eta^2_{1vs4} = \frac{972}{972+1800} = .35$$

- Thus, with a post-hoc Tukey HSD correction (see 10-48), we have:
  30 Months vs. 36 Months: $F(1,11) = 2.41, ns, \eta^2 = .18$
  30 Months vs. 42 Months: $F(1,11) = 5.28, ns, \eta^2 = .32$
  30 Months vs. 48 Months: $F(1,11) = 5.94, ns, \eta^2 = .35$

```
Compute Cont12 = time2 - time1.
Compute Cont13 = time3 - time1.
Compute Cont14 = time4 - time1.
Descriptives Variables = cont12 cont13 cont14.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Cont12	12	-12.00	20.00	4.0000	8.93410
Cont13	12	-14.00	24.00	7.0000	10.54859
Cont14	12	-19.00	30.00	9.0000	12.79204
Valid N (listwise)	12				

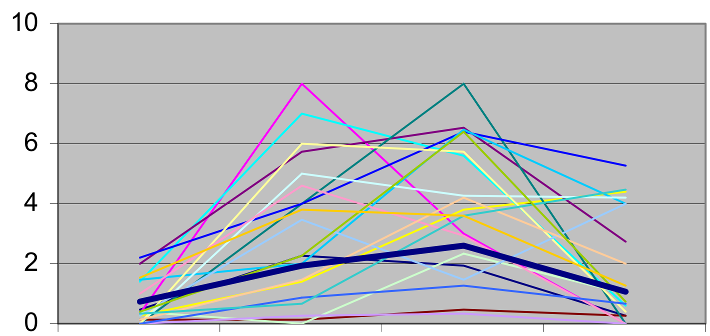
$$d_{1vs2} = \frac{\hat{\psi}}{\hat{\sigma}_\psi} = \frac{4}{8.934} = .45 \qquad d_{1vs3} = \frac{7}{10.549} = .66 \qquad d_{1vs4} = \frac{9}{12.792} = .70$$

- Again, with a post-hoc Tukey HSD correction (see 10-48), we have:
  30 Months vs. 36 Months: $F(1,11) = 2.41, ns, d = .45$
  30 Months vs. 42 Months: $F(1,11) = 5.28, ns, d = .66$
  30 Months vs. 48 Months: $F(1,11) = 5.94, ns, d = .70$

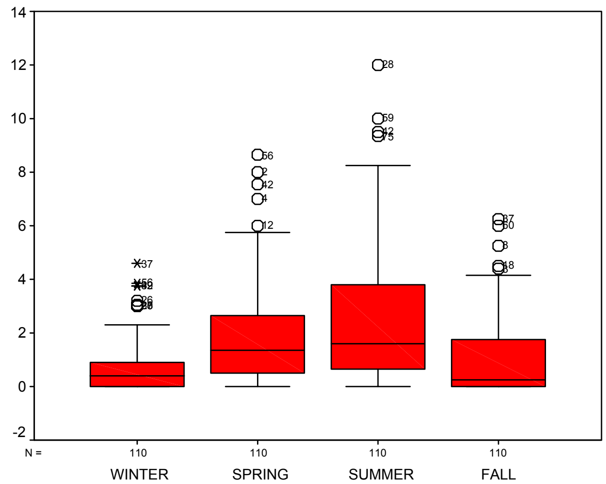
12. A final example

- A researcher is interested in the amount of daylight people are exposed to (She believes that exposure to daylight affects mood). A sample of 110 individuals, all age 60-70, recorded their daily exposure to sunlight over the course of a year. The researcher computed an average daily exposure to sunlight (hours/day) for each of the four seasons, and wanted to test the following hypotheses
  - Are there any trends in exposure across the seasons?
  - Is winter the season with the lowest exposure?
  - Is summer the season with the highest exposure?

- Here is a graph of the data for the first 20 participants
  - The overall average exposure is indicated with the dark blue line



- Each person is observed on more than one occasion. A one factor repeated measures (within-subjects) design would seem to be appropriate.
  - Participants are independent and randomly selected from the population
  - Normality (actually symmetry)
  - Sphericity



Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
WINTER	.723	110	.000
SPRING	.855	110	.000
SUMMER	.854	110	.000
FALL	.755	110	.000

Mauchly's Test of Sphericity

Measure: MEASURE_1

	Epsilon[a]		
Within Subjects Effect	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SEASON	.783	.801	.333

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

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Correlations

	WINTER	SPRING	SUMMER	FALL
WINTER	1	.422	.324	.345
SPRING	.422	1	.349	.190
SUMMER	.324	.349	1	.319
FALL	.345	.190	.319	1

- The data are positively skewed
- The data are also non-spherical

- How should we analyze these data?
  - We could try a transformation to see if we could normalize the data
  - We could conduct Friedman's test

- Let's start with the (omnibus) Friedman non-parametric test.

```
NPAR TESTS
  /FRIEDMAN = winter spring summer fall.
```

Ranks

	Mean Rank
WINTER	1.91
SPRING	2.88
SUMMER	3.13
FALL	2.08

Test Statistics[a]

N	110
Chi-Square	72.264
df	3
Asymp. Sig.	.000

a. Friedman Test

$$\chi^2(3) = 72.26, p < .01$$

- We conclude that there are differences in sunlight exposure across the seasons.
- The trend hypotheses cannot be tested on ranked data, but pairwise hypotheses can be tested with follow-up Wilcoxon Signed-Rank tests

```
NPAR TEST
  /WILCOXON=winter winter winter WITH spring summer fall (PAIRED).
NPAR TEST
  /WILCOXON= summer summer summer WITH winter spring fall (PAIRED).
```

Test Statistics[b]

	SPRING - WINTER	SUMMER - WINTER	FALL - WINTER
Z	-6.645[a]	-7.557[a]	-1.560[a]
Asymp. Sig. (2-tailed)	.000	.000	.119

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Test Statistics[b]

	WINTER - SUMMER	SPRING - SUMMER	FALL - SUMMER
Z	-7.557[a]	-1.568[a]	-6.200[a]
Asymp. Sig. (2-tailed)	.000	.117	.000

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

- These are post-hoc tests and require correction. Our standard corrections all require parametric data (they use the MSE). The only correction that is appropriate for non-parametric post-hoc tests is the Bonferroni adjustment

Let c = the total number of possible pairwise tests = $\frac{a(a-1)}{2} = 6$

$$p_{crit} = \frac{\alpha}{c} = \frac{.05}{6} = .0083$$

- Bonferroni-adjusted post-hoc tests reveal:
  - Daily sunlight exposure is lower in the winter than in the spring or summer.
  - Daily sunlight exposure is higher in the summer than in the fall or winter.

- For pedagogical purposes only, let's image that the data are normally /symmetrically distributed. What would we do then?
  - o We could conduct an epsilon-adjusted omnibus test, and follow it up with post-hoc contrasts.
  - o Alternatively, we could test our hypotheses with contrasts and conduct additional post-hoc contrasts.

- First, let's consider an omnibus test

```
GLM  winter spring summer fall
  /WSFACTOR = season 4
  /PRINT = DESC.
```

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SEASON	Sphericity Assumed	244.287	3	81.429	33.294	.000
	Greenhouse-Geisser	244.287	2.349	104.018	33.294	.000
	Huynh-Feldt	244.287	2.403	101.638	33.294	.000
	Lower-bound	244.287	1.000	244.287	33.294	.000
Error(SEASON)	Sphericity Assumed	799.759	327	2.446		
	Greenhouse-Geisser	799.759	255.987	3.124		
	Huynh-Feldt	799.759	261.980	3.053		
	Lower-bound	799.759	109.000	7.337		

$$\eta^2_{Season} = \frac{SS_{Season}}{SS_{Season} + SS_{ErrorTermForSeason}} = \frac{244.287}{244.287 + 799.759} = .23$$

  - o The data are not spherical, so we need to apply the Geisser-Greenhouse $\hat{\varepsilon}$ adjustment.

$$F(2.35, 255.99) = 33.29, p < .01, \eta^2 = .23$$

    - We conclude that there are differences in sunlight exposure across the seasons.

- Next, let's follow-up this omnibus test with post-hoc polynomial contrasts (using a Scheffé correction). There are many ways to test the polynomial trends. The easiest way is to use SPSS's built-in polynomial contrasts.

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SEASON	Type III Sum of Squares	df	Mean Square	F	Sig.
SEASON	Linear	15.106	1	15.106	10.412	.002
	Quadratic	215.936	1	215.936	76.184	.000
	Cubic	13.245	1	13.245	4.340	.040
Error(SEASON)	Linear	158.144	109	1.451		
	Quadratic	308.951	109	2.834		
	Cubic	332.664	109	3.052		

$$F_{Crit} = (a-1)F_{\alpha=.05;a-1,\nu} = 3*F(.95,3,109) = 3*2.688 = 8.06$$

$$\eta^2_{Linear} = \frac{SS_{Linear}}{SS_{Linear} + SS_{ErrorTermForLinear}} = \frac{15.106}{15.106+158.144} = .09$$

$$\eta^2_{Quadratic} = \frac{215.936}{215.936+308.951} = .41 \qquad \eta^2_{Cubic} = \frac{13.245}{13.245+332.664} = .04$$

  - Only the linear and quadratic trends are significant

$$\psi_{lin} : F(1,109) = 10.41, p < .05, \eta^2 = .18$$

$$\psi_{quad} : F(1,109) = 76.18, p < .05, \eta^2 = .41$$

$$\psi_{cub} : F(1,109) = 4.34, ns, \eta^2 = .04$$

  - Be cautious in interpreting these trends. Seasons are cyclical. If we had started our analyses in the summer, we would obtain different results.

- Looking at the data, we also want to know if
  - Is winter the season with the lowest exposure?
  - Is summer the season with the highest exposure?

Descriptive Statistics

	Mean	Std. Deviation	N
WINTER	.7008	.96421	110
SPRING	1.9573	1.95291	110
SUMMER	2.5885	2.58768	110
FALL	1.0428	1.43157	110

- Pairwise post hoc tests can be obtained though the "*simple*" contrast subcommand.
  - Is winter the season with the lowest exposure?

```
GLM  winter spring summer fall
/WSFACTOR = season 4 simple (1)
/PRINT = DESC.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	season	Type III Sum of Squares	df	Mean Square	F	Sig.
season	Level 2 vs. Level 1	173.655	1	173.655	55.031	.000
	Level 3 vs. Level 1	391.987	1	391.987	65.257	.000
	Level 4 vs. Level 1	12.866	1	12.866	6.350	.013
Error(season)	Level 2 vs. Level 1	343.960	109	3.156		
	Level 3 vs. Level 1	654.740	109	6.007		
	Level 4 vs. Level 1	220.864	109	2.026		

$$F_{Tukey-Critical} = \frac{(q(.95,4,109))^2}{2} = \frac{(3.6899)^2}{2} = 6.81$$

$$F_{Tukey-Critical} = \frac{(q(.90,4,109))^2}{2} = \frac{(3.2795)^2}{2} = 5.38$$

  - To compute a *d* effect-size for each contrast, we need to calculate the average value and the standard deviation of each contrast.

```
COMPUTE WinSpr = spring-winter.
COMPUTE WinSum = summer-winter.
COMPUTE WinFall = fall-winter.
DESC VAR = WinSpr WinSum WinFall.
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
WinSpr	110	-2.19	7.67	1.2565	1.77640
WinSum	110	-2.08	10.00	1.8877	2.45088
WinFall	110	-3.75	5.56	.3420	1.42347
Valid N (listwise)	110				

$$d_{W\,int\,ervsSpring} = \frac{\overline{\psi}}{\hat{\sigma}_{\psi}} = \frac{1.2565}{1.7764} = .71 \quad d_{W\,int\,ervsSummer} = \frac{1.8877}{2.45088} = .77 \quad d_{W\,int\,ervsFall} = \frac{0.3420}{1.42347} = .24$$

Tukey HSD post-hoc tests reveal:

- Daily sunlight exposure is significantly lower in the winter than in the spring or summer, *d*s > .70
- Daily sunlight exposure is marginally lower in the winter than in the fall, *d* = .24.

- Is summer the season with the highest exposure?

```
GLM winter spring summer fall
/WSFACTOR = season 4 simple (3)
/PRINT = DESC.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	season	Type III Sum of Squares	df	Mean Square	F	Sig.
season	Level 1 vs. Level 3	391.987	1	391.987	65.257	.000
	Level 2 vs. Level 3	43.836	1	43.836	6.281	.014
	Level 4 vs. Level 3	262.820	1	262.820	41.178	.000
Error(season)	Level 1 vs. Level 3	654.740	109	6.007		
	Level 2 vs. Level 3	760.751	109	6.979		
	Level 4 vs. Level 3	695.693	109	6.383		

$$F_{Tukey-Critical} = \frac{(q(.95,4,109))^2}{2} = \frac{(3.6899)^2}{2} = 6.81$$

$$F_{Tukey-Critical} = \frac{(q(.90,4,109))^2}{2} = \frac{(3.2795)^2}{2} = 5.38$$

```
COMPUTE SumWin = summer-winter.
COMPUTE SumSpr = summer-spring.
COMPUTE SumFall = summer-fall.
DESC VAR = SumWin SumSpr SumFall
```

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SumWin	110	-2.08	10.00	1.8877	2.45088
SumSpr	110	-5.00	10.00	.6313	2.64185
SumFall	110	-3.75	10.00	1.5457	2.52636
Valid N (listwise)	110				

.

$$d_{SummervsWinter} = \frac{\overline{\psi}}{\hat{\sigma}_{\psi}} = \frac{1.8877}{2.45088} = .77 \quad d_{SummervsSpring} = \frac{0.6313}{2.64185} = .23 \quad d_{SummervsFall} = \frac{1.5457}{2.52636} = .61$$

Tukey HSD post-hoc tests reveal:

- Daily sunlight exposure is significantly lower in the winter than in the spring or summer, *d*s > .60.
- Daily sunlight exposure is marginally lower in the winter than in the fall, *d* = .23.

- A final approach to analyzing these data would be to test our hypotheses directly using contrasts (Remember – we had some!)
  - (Planned) Are there any trends in exposure across the seasons?
  - (Post-hoc) Is winter the season with the lowest exposure?
  - (Post-hoc) Is summer the season with the highest exposure?

Descriptive Statistics

	Mean	Std. Deviation	N
WINTER	.7008	.96421	110
SPRING	1.9573	1.95291	110
SUMMER	2.5885	2.58768	110
FALL	1.0428	1.43157	110

  - Previous, we tested the polynomial trends using SPSS's built-in polynomial contrasts.

```
GLM  winter spring summer fall
  /WSFACTOR = season 4 polynomial
  /PRINT = DESC.
```

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	SEASON	Type III Sum of Squares	df	Mean Square	F	Sig.
SEASON	Linear	15.106	1	15.106	10.412	.002
	Quadratic	215.936	1	215.936	76.184	.000
	Cubic	13.245	1	13.245	4.340	.040
Error(SEASON)	Linear	158.144	109	1.451		
	Quadratic	308.951	109	2.834		
	Cubic	332.664	109	3.052		

$$\hat{\eta}^2_{Linear} = \frac{SS_{Linear}}{SS_{Linear} + SS_{ErrorTermForLinear}} = \frac{15.106}{15.106 + 158.144} = .09$$

$$\hat{\eta}^2_{Quadratic} = \frac{215.936}{215.936 + 308.951} = .41 \qquad \hat{\eta}^2_{Cubic} = \frac{13.245}{13.245 + 332.664} = .04$$

    - Because these are planned, we do not need to apply a p-value correction.
    - We find evidence for significant linear, quadratic, and cubic trends

$$\psi_{lin} : F(1,109) = 10.41, p < .01, \eta^2 = .09$$
$$\psi_{quad} : F(1,109) = 76.18, p < .01, \eta^2 = .41$$
$$\psi_{cub} : F(1,109) = 4.34, p = .04, \eta^2 = .04$$

  - The pairwise post hoc tests can be calculated exactly as they were in the previous method of data analysis (using the Tukey HSD correction).

