

**One-Way Repeated Measures Designs**  
Maxwell & Delaney

**Review of one-way anova for independent groups**

**Example.** A researcher wanted to determine whether gender is a significant predictor of preferred number of children. The responses are given below. Use an alpha of .05 to determine whether gender is a significant predictor of preferred number of children.

<b>Null Hypothesis</b>	<b>Reduced (Null) Model</b>
	$Y_{ij} = \mu + e_{ij}$
<b>Alternative Hypothesis</b>	<b>Full (Alternative) Model</b>
	$Y_{ij} = \mu + \alpha_j + e_{ij}$

*Note. 5 observations were used per group-- should have had at least 15 per group!*

Gender	Kids	(score – reduced model prediction) <sup>2</sup>	(score – full model prediction) <sup>2</sup>
Female	0	$(0-3)^2 = 9$	$(0-2)^2 = 4$
Female	2	$(2-3)^2 = 1$	$(2-2)^2 = 0$
Female	1	$(1-3)^2 = 4$	$(1-2)^2 = 1$
Female	5	$(5-3)^2 = 4$	$(5-2)^2 = 9$
Female	2	$(2-3)^2 = 1$	$(2-2)^2 = 0$
Male	4	$(4-3)^2 = 1$	$(4-4)^2 = 0$
Male	3	$(3-3)^2 = 0$	$(3-4)^2 = 1$
Male	4	$(4-3)^2 = 1$	$(4-4)^2 = 0$
Male	5	$(5-3)^2 = 4$	$(5-4)^2 = 1$
Male	4	$(4-3)^2 = 1$	$(4-4)^2 = 0$
		<b><math>E_R = 26</math></b>	<b><math>E_F = 16</math></b>

### Comparing the Prediction Errors of the Full and Reduced Models

$$df_R = N - 1 = 10 - 1 = 9$$

$$df_F = N - a = 10 - 2 = 8$$

$$E_R = 26$$

$$E_F = 16$$

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/df_F} = \frac{(26 - 16)/(9 - 8)}{16/8} = 5$$

Obtain the critical value from the  $F$  distribution to determine whether the  $F$  statistics is significantly greater than 1.

$$df_{\text{Numerator}} = a - 1 = 2 - 1 = 1$$

$$df_{\text{Denominator}} = N - a = 10 - 2 = 8$$

$$\text{Critical Value} = 5.318$$

Interpret the results of the test.

Using an alpha of .05, the one-way anova indicated that gender (female, male) was not a significant predictor of the preferred number of children,  $F(1, 8) = 5.00$ ,  $MSE = 2.00$ ,  $p > .05$ , omega-squared = .29.

### Oneway

#### Descriptives

KIDS

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Female	5	2.0000	1.87083	.83666	-.3229	4.3229	.00	5.00
Male	5	4.0000	.70711	.31623	3.1220	4.8780	3.00	5.00
Total	10	3.0000	1.69967	.53748	1.7841	4.2159	.00	5.00

#### ANOVA

KIDS

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10.000	1	10.000	5.000	.056
Within Groups	16.000	8	2.000		
Total	26.000	9			

### Assumptions and Robustness of the One-Way ANOVA

---

Three assumptions must be met for  $F$  statistic to follow an  $F$  distribution.

**1. The residuals of the dependent variable must be normal within each group.**

- ANOVA generally controls the  $p(\text{type I})$  when the residuals are non-normal.
- ANOVA is generally robust in terms of  $\text{power}$  when the residuals are non-normal. An exception is when the data come from a mixed-normal distribution. The power of the ANOVA test would be greatly reduced if the data are mixed-normal.

**2. The population variances of the residuals for each group should be the same:**

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$$

Equal n

- With *moderate cell sizes*, the ANOVA is robust to fairly large differences in population variances. With *small sample sizes*, the *Welch's Anova F test should be used instead of the one-way anova*.

Unequal n

- If the **small sample sizes** are associated with the **smaller variances** then the actual  $p(\text{type I})$  is drastically **less** than the alpha set by the researcher.
- If the **small sample sizes** are associated with the **larger variances** then the actual  $p(\text{type I})$  may be drastically **greater** than the alpha set by the researcher.
- With unequal sample sizes, the *Welch's Anova F test should be used instead of the one-way anova*.

**3. The residuals must be statistically independent of each other.**

- Positively correlated residuals would lead to an unacceptably **high  $p(\text{type I})$** .
- Negatively correlated residuals would lead to an unacceptably **low  $p(\text{type I})$** .
- If the residuals are correlated, a different statistical procedure must be used.

**Three Situations that Lead to a Repeated Measures Design**

**1. Each subject is observed on  $a$  different treatment conditions.** The dependent variable is the same for all treatments.

	SHOCK	VERBAL FEEDBACK
S1	Y = # errors in word recall	
S2		
S3		
S4		
S5		

**2. Each subject is measured on  $a$  different tests that have the same scale.** The same scale is necessary for psychologically meaningful interpretations. This kind of design is also known as profile analysis.

	GRE-VERBAL	GRE-QUANTITATIVE
S1		
S2		
S3		
S4		
S5		

**3. Each subject is measured at two or more different times (longitudinally).**

	Time 1 (Pre-Test)	Time 2 (Post-Test)
S1	Y = Anxiety	
S2		
S3		
S4		
S5		

If  $a = 2$  (as it is in the above examples), we can perform a **two correlated sample  $t$  (F) test**.

**Repeated Measures Designs with Two Levels**

	Time 1	Time 2
S1	8	10
S2	3	6
S3	12	13
S4	5	9
S5	7	8
S6	13	14

If we treat the above data as if it were a two-independent-sample ANOVA, the

$$E_F = \sum_{i=1}^n \sum (Y_{ij} - \bar{Y}_j)^2 \text{ where } \bar{Y}_1 = 8 \text{ and } \bar{Y}_2 = 10.$$

	$e_{i1} = Y_{i1} - \bar{Y}_1$ $= Y_{i1} - 8$	$e_{i2} = Y_{i2} - \bar{Y}_2$ $= Y_{i2} - 10$
S1	8 - 8 = 0	10 - 10 = 0
S2	3 - 8 = -5	6 - 10 = -4
S3	12 - 8 = 4	13 - 10 = 3
S4	5 - 8 = -3	9 - 10 = -1
S5	7 - 8 = -1	8 - 10 = -2
S6	13 - 8 = 5	14 - 10 = 4

The correlation between  $e_{i1}$  and  $e_{i2} = .96$ . If a subject is below average in *Time 1*, they are quite likely to be below average on *Time 2*.

**Univariate Analysis of Variance**

**Between-Subjects Factors**

	N
TIME 1.00	6
2.00	6

**Tests of Between-Subjects Effects**

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12.000 <sup>a</sup>	1	12.000	.984	.345
Intercept	972.000	1	972.000	79.672	.000
TIME	12.000	1	12.000	.984	.345
Error	122.000	10	12.200		
Total	1106.000	12			
Corrected Total	134.000	11			

<sup>a</sup>. R Squared = .090 (Adjusted R Squared = -.001)

The  $e_{ij}$  are not independent . The one-way anova for independent groups is not robust to violation of the independence assumption! We must use a repeated measures anova to analyze these data.

2. Correct Approach for RMDs when  $a = 2$ : Two correlated sample  $F$  test.

	Time 1	Time 2	d = Time 2 - Time 1
S1	8	10	2
S2	3	6	3
S3	12	13	1
S4	5	9	4
S5	7	8	1
S6	13	14	1

$$\bar{d} = 12/6 = 2 \quad S_d^2 = 1.6$$

$$\begin{array}{ll}
 H_o: \mu_d = 0 & \text{RESTRICT: } d_i = 0 + e_i \\
 & = 0 \\
 H_1: \mu_d \neq 0 & \text{FULL: } d_i = \mu_d + e_i \\
 & = \bar{d}
 \end{array}$$

$$\begin{aligned}
 E_F &= \sum_{i=1}^n (d_i - \bar{d})^2 \\
 &= (2 - 2)^2 + (3 - 2)^2 + (1 - 2)^2 + (4 - 2)^2 + (1 - 2)^2 + (1 - 2)^2 \\
 &= 8
 \end{aligned}$$

$$\begin{aligned}
 E_R &= \sum_{i=1}^n (d_i - 0)^2 = \sum_{i=1}^n d_i^2 \\
 &= 2^2 + 3^2 + 1^2 + 4^2 + 1^2 + 1^2 \\
 &= 32
 \end{aligned}$$

$$\begin{aligned}
 F &= \frac{(E_R - E_F) / (df_R - df_F)}{E_F / df_F} = \frac{(32 - 8) / (6 - 5)}{8 / 5} = 15 \\
 &= \frac{n(\bar{d})^2}{S_d^2} = t^2 = \frac{\bar{d}^2}{S_d^2 / n}
 \end{aligned}$$

## General Linear Model

### Within-Subjects Factors

Measure: MEASURE\_1

TIMES	Dependent Variable
1	TIME1
2	TIME2

### Descriptive Statistics

	Mean	Std. Deviation	N
TIME1	8.0000	3.89872	6
TIME2	10.0000	3.03315	6

### Multivariate Tests<sup>b</sup>

Effect	Value	F	Hypothesis df	Error df	Sig.
TIMES Pillai's Trace	.750	15.000 <sup>a</sup>	1.000	5.000	.012
Wilks' Lambda	.250	15.000 <sup>a</sup>	1.000	5.000	.012
Hotelling's Trace	3.000	15.000 <sup>a</sup>	1.000	5.000	.012
Roy's Largest Root	3.000	15.000 <sup>a</sup>	1.000	5.000	.012

a. Exact statistic

b.

Design: Intercept  
 Within Subjects Design: TIMES

### Mauchly's Test of Sphericity<sup>b</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIMES	1.000	.000	0	.	1.000	1.000	1.000

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: TIMES

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
TIMES	Sphericity Assumed	12.000	1	12.000	15.000	.012
	Greenhouse-Geisser	12.000	1.000	12.000	15.000	.012
	Huynh-Feldt	12.000	1.000	12.000	15.000	.012
	Lower-bound	12.000	1.000	12.000	15.000	.012
Error(TIMES)	Sphericity Assumed	4.000	5	.800		
	Greenhouse-Geisser	4.000	5.000	.800		
	Huynh-Feldt	4.000	5.000	.800		
	Lower-bound	4.000	5.000	.800		

**Tests of Within-Subjects Contrasts**

Measure: MEASURE\_1

Source	TIMES	Type III Sum of Squares	df	Mean Square	F	Sig.
TIMES	Linear	12.000	1	12.000	15.000	.012
Error(TIMES)	Linear	4.000	5	.800		

**Tests of Between-Subjects Effects**

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	972.000	1	972.000	41.186	.001
Error	118.000	5	23.600		

### Univariate (Mixed-Model) Approach to Repeated Measures Designs

---

A one-way RMD is treated as a two-way ANOVA where “Subjects” is the second factor. The one-way RMD anova is different than a true two-way anova because

- There is only one observation per cell because each subject represents a different level of the “Subjects” factor. Therefore, the A\*S interaction and the within cell variability are confounded.
- “Subjects” is considered a *random* factor.

		A Factor		
		Time 1	Time 2	
“Subjects” Factor	S1	8	10	$\bar{Y}_{S1} = 9$
	S2	3	6	$\bar{Y}_{S2} = 4.5$
	S3	12	13	$\bar{Y}_{S3} = 12.5$
	S4	5	9	$\bar{Y}_{S4} = 7$
	S5	7	8	$\bar{Y}_{S5} = 7.5$
	S6	13	14	$\bar{Y}_{S6} = 13.5$
		$\bar{Y}_{T1} = 8$	$\bar{Y}_{T2} = 10$	$\bar{Y} = 9$

#### Null Hypothesis:

Reduced Model:  $Y_{ij} = \mu + \pi_i + e_{ij}$     where  $\hat{\mu} = \bar{Y}$  and  $\hat{\pi}_i = \bar{Y}_i - \bar{Y}$

Predicted  $\hat{Y}_{ij} = \bar{Y}_i$

$df_R = n a - n$

#### Alternative Hypothesis:

Full Model:  $Y_{ij} = \mu + \alpha_j + \pi_i + e_{ij}$     where  $\hat{\mu} = \bar{Y}$      $\hat{\pi}_i = \bar{Y}_i - \bar{Y}$      $\hat{\alpha}_j = \bar{Y}_j - \bar{Y}$

Predicted  $\hat{Y}_{ij} = \bar{Y}_i + \hat{\alpha}_j$

$df_F = na - n - a + 1$

Subject	Time	Score	(score – reduced model prediction) <sup>2</sup>	(score – full model prediction) <sup>2</sup>
S1	T1	8	(8-9) <sup>2</sup> = 1	(8-8) <sup>2</sup> = 0
S2	T1	3	(3-4.5) <sup>2</sup> = 2.25	(3-3.5) <sup>2</sup> = 0.25
S3	T1	12	(12-12.5) <sup>2</sup> = 0.25	(12-11.5) <sup>2</sup> = 0.25
S4	T1	5	(5-7) <sup>2</sup> = 4	(5-6) <sup>2</sup> = 1
S5	T1	7	(7-7.5) <sup>2</sup> = 0.25	(7-6.5) <sup>2</sup> = 0.25
S6	T1	13	(13-13.5) <sup>2</sup> = 0.25	(13-12.5) <sup>2</sup> = 0.25
S1	T2	10	(10-9) <sup>2</sup> = 1	(10-10) <sup>2</sup> = 0
S2	T2	6	(6-4.5) <sup>2</sup> = 2.25	(6-5.5) <sup>2</sup> = 0.25
S3	T2	13	(13-12.5) <sup>2</sup> = 0.25	(13-13.5) <sup>2</sup> = 0.25
S4	T2	9	(9-7) <sup>2</sup> = 4	(9-8) <sup>2</sup> = 1
S5	T2	8	(8-7.5) <sup>2</sup> = 0.25	(8-8.5) <sup>2</sup> = 0.25
S6	T2	14	(14-13.5) <sup>2</sup> = 0.25	(14-14.5) <sup>2</sup> = 0.25
			<b><math>E_R = 16</math></b>	<b><math>E_F = 4</math></b>

$$F_{Time} = \frac{(E_R - E_F) / (df_R - df_F)}{E_F / df_F} = \frac{(16 - 4) / (1)}{4 / 5} = 15$$

$$F_{Crit} = 6.61$$

### UNIVARIATE ONE-WAY REPEATED MEASURES ANOVA ASSUMPTIONS

1. The residuals are normally distributed.
2. The residuals are independent.
3. We have homogeneity of treatment difference variances, a.k.a. *sphericity*.

**What does *sphericity* mean?** Sphericity means that the population variances of the **difference scores** are equal.

	Sphericity Assumption of the Univariate one-way rmd anova	Variance Assumption of the One-way anova for independent groups
If $a = 2$ levels...		
If $a = 3$ levels...		
If $a = 4$ levels...		

*Compound Symmetry* is a slightly more restrictive assumption than *Sphericity*, but it is easier to demonstrate. *Compound symmetry* requires that all the population variances be the same and that all the population covariances be the same.

An example of compound symmetry being a **valid assumption** is:

	$\sigma_{A1}$	$\sigma_{A2}$	$\sigma_{A3}$	$\sigma_{A4}$
$\sigma_{A1}$				
$\sigma_{A2}$				
$\sigma_{A3}$				
$\sigma_{A4}$				

An example of compound symmetry **not** being a **valid** assumption is:

	$\sigma_{A1}$	$\sigma_{A2}$	$\sigma_{A3}$	$\sigma_{A4}$
$\sigma_{A1}$				
$\sigma_{A2}$				
$\sigma_{A3}$				
$\sigma_{A4}$				

In contrast, the one-way anova for independent groups **equal variances** assumption is:

	$\sigma_{A1}$	$\sigma_{A2}$	$\sigma_{A3}$	$\sigma_{A4}$
$\sigma_{A1}$				
$\sigma_{A2}$				
$\sigma_{A3}$				
$\sigma_{A4}$				

Realistically, the *Sphericity Assumption* is *NOT valid* in the real world. Unfortunately, the Repeated Measures ANOVA is NOT robust to violation of the *sphericity assumption* without modification!

We modify the degrees of freedom of the one-way repeated measures anova to compensate for violation of the sphericity assumption.

$$df'_{numerator} = \varepsilon(a - 1) \quad \text{and} \quad df'_{denominator} = \varepsilon(n - 1)(a - 1)$$

where  $\varepsilon$  is a function of the unequal variances of the difference scores.

$$\frac{1}{a - 1} \leq \varepsilon \leq 1$$

$\varepsilon = 1$  when the population variances of the difference scores are all equal.

$\varepsilon$  is less than one when the population variances of the difference scores are unequal.

This works to decrease our degrees of freedom and make the test more conservative.

We must estimate  $\varepsilon$  because it is a parameter. There are two well-known sample estimates of  $\varepsilon$ :

Box's  $\hat{\varepsilon}$  (a.k.a. Greenhouse-Geiser, G-G on SAS)

Huynh-Feldt's  $\tilde{\varepsilon}$  (a.k.a. H-F on SAS)

<b>Pairwise Comparisons for the One-way Repeated Measures Anova</b>
---

The research indicates using a pooled error term (like *MSE*) is not a good idea because of the sphericity assumption issue. Instead, the following technique is recommended:

- Conduct pairwise comparisons using the *t* test for dependent samples, and
- Use the Bonferroni (Dunn), Sidak, or Scheffe approaches to maintain the familywise alpha to .05.

## General Linear Model

### Within-Subjects Factors

Measure: MEASURE\_1

TYPE	Dependent Variable
1	FAMILIAR
2	UNFAM
3	NONSENSE

### Descriptive Statistics

	Mean	Std. Deviation	N
FAMILIAR	26.9333	7.05556	15
unfamiliar	17.7333	5.73793	15
NONSENSE	11.0667	5.20256	15

### Multivariate Tests<sup>b</sup>

Effect		Value	F	Hypothesis df	Error df	Sig.
TYPE	Pillai's Trace	.898	57.121 <sup>a</sup>	2.000	13.000	.000
	Wilks' Lambda	.102	57.121 <sup>a</sup>	2.000	13.000	.000
	Hotelling's Trace	8.788	57.121 <sup>a</sup>	2.000	13.000	.000
	Roy's Largest Root	8.788	57.121 <sup>a</sup>	2.000	13.000	.000

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: TYPE

### Mauchly's Test of Sphericity<sup>b</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TYPE	.967	.430	2	.806	.968	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: TYPE

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Sphericity Assumed	1904.178	2	952.089	70.808	.000
	Greenhouse-Geisser	1904.178	1.937	983.085	70.808	.000
	Huynh-Feldt	1904.178	2.000	952.089	70.808	.000
	Lower-bound	1904.178	1.000	1904.178	70.808	.000
Error(TYPE)	Sphericity Assumed	376.489	28	13.446		
	Greenhouse-Geisser	376.489	27.117	13.884		
	Huynh-Feldt	376.489	28.000	13.446		
	Lower-bound	376.489	14.000	26.892		

**Tests of Within-Subjects Contrasts**

Measure: MEASURE\_1

Source	TYPE	Type III Sum of Squares	df	Mean Square	F	Sig.
TYPE	Linear	1888.133	1	1888.133	123.025	.000
	Quadratic	16.044	1	16.044	1.390	.258
Error(TYPE)	Linear	214.867	14	15.348		
	Quadratic	161.622	14	11.544		

**Tests of Between-Subjects Effects**

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	15531.022	1	15531.022	187.393	.000
Error	1160.311	14	82.879		

**T-Test**

**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 FAMILIAR	26.9333	15	7.05556	1.82174
1 unfamiliar	17.7333	15	5.73793	1.48153
Pair 2 FAMILIAR	26.9333	15	7.05556	1.82174
2 NONSENSE	11.0667	15	5.20256	1.34330
Pair 3 unfamiliar	17.7333	15	5.73793	1.48153
3 NONSENSE	11.0667	15	5.20256	1.34330

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 FAMILIAR & unfamiliar	15	.681	.005
Pair 2 FAMILIAR & NONSENSE	15	.629	.012
Pair 3 unfamiliar & NONSENSE	15	.630	.012

**Paired Samples Test**

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	FAMILIAR - unfamiliar	9.2000	5.25357	1.35647	6.2907	12.11	6.782	14	.000
Pair 2	FAMILIAR - NONSENSE	15.8667	5.54033	1.43051	12.7985	18.93	11.092	14	.000
Pair 3	unfamiliar - NONSENSE	6.6667	4.73085	1.22150	4.0468	9.2865	5.458	14	.000