Repeated Measures (RM-ANOVA)

Andrew Zieffler

Educational Psychology

University of Minnesota

Driven to Discover[SM]

Read in the minneapolis.csv data

```
## Read in the data
> mpls = read.csv("http://www.tc.umn.edu/~zief0002/Data/minneapolis.csv")
```

Packages Needed

- **ez**
- **ggplot2**
- **reshape2**

	studentID	read.5	read.6	read.7	read.8	atRisk	female	minority	ell	sped	att
1	1	172	185	179	194	1	1	1	0	0	0.94
2	2	200	210	209	NA	1	1	1	0	0	0.91
3	3	191	199	203	215	1	0	1	0	0	0.97
4	4	200	195	194	NA	1	1	1	0	0	0.88
5	5	207	213	212	213	1	1	1	0	0	0.85
6	6	191	189	206	195	1	0	1	0	0	0.90
7	7	199	208	213	218	1	0	1	1	0	0.97
8	8	191	194	194	NA	1	1	1	1	1	0.97
9	9	149	154	174	177	1	1	1	0	1	0.97
10	10	200	212	213	NA	1	1	1	0	0	0.96
11	11	218	231	233	239	1	1	0	0	0	0.98
12	12	228	232	248	246	1	1	0	0	1	0.96
13	13	228	236	228	239	0	1	0	0	0	0.99
14	14	199	210	225	235	0	0	1	1	0	0.99
15	15	218	223	236	NA	0	1	0	0	0	1.00
16	16	228	226	234	227	0	0	0	0	0	0.97
17	17	201	210	208	219	0	0	0	0	0	0.98
18	18	218	220	217	221	0	0	0	0	0	1.00
19	19	215	216	221	NA	0	1	0	0	0	0.96
20	20	204	215	219	214	0	1	1	0	0	0.95
21	21	237	241	243	NA	0	0	0	0	0	0.98
22	22	219	233	236	NA	0	1	1	0	0	0.96

Most repeated measures data are entered in the wide-format

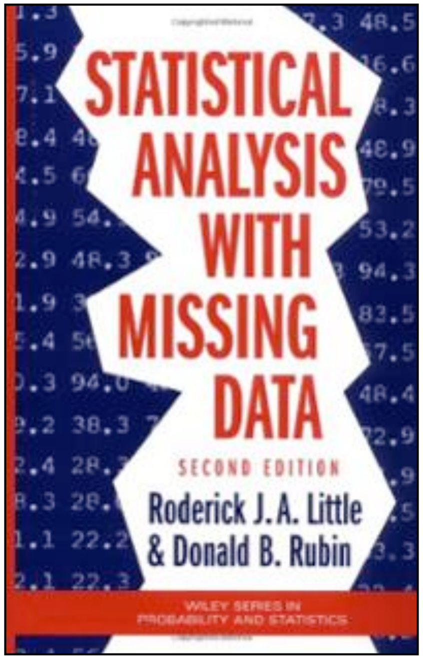
Each of the repeated measures is in its own column.

Missing Data

Missing data is a problem for most data analysis

There are several ways to deal with missing data.

- Remove cases with missing data (*case-wise deletion*)
- Replace missing values with an actual value (*imputation*)



No matter how you treat missing data, there may be a problem for your inferences....Are the cases you deleted an unbiased sample of the full set of cases?

Missing data is problematic for RM-ANOVA.

We will delete *any* cases with missing data.

	studentID	read.5	read.6	read.7	read.8	atRisk	female	minority	ell	sped	att
1	1	172	185	179	194	1	1	1	0	0	0.94
~~2~~	~~2~~	~~200~~	~~210~~	~~209~~	~~NA~~	~~1~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0.91~~
3	3	191	199	203	215	1	0	1	0	0	0.97
~~4~~	~~4~~	~~200~~	~~195~~	~~194~~	~~NA~~	~~1~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0.88~~
5	5	207	213	212	213	1	1	1	0	0	0.85
6	6	191	189	206	195	1	0	1	0	0	0.90
7	7	199	208	213	218	1	0	1	1	0	0.97
~~8~~	~~8~~	~~191~~	~~194~~	~~194~~	~~NA~~	~~1~~	~~1~~	~~1~~	~~1~~	~~1~~	~~0.97~~
9	9	149	154	174	177	1	1	1	0	1	0.97
~~10~~	~~10~~	~~200~~	~~212~~	~~213~~	~~NA~~	~~1~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0.96~~
11	11	218	231	233	239	1	1	0	0	0	0.98
12	12	228	232	248	246	1	1	0	0	1	0.96
13	13	228	236	228	239	0	1	0	0	0	0.99
14	14	199	210	225	235	0	0	1	1	0	0.99
~~15~~	~~15~~	~~218~~	~~223~~	~~236~~	~~NA~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~1.00~~
16	16	228	226	234	227	0	0	0	0	0	0.97
17	17	201	210	208	219	0	0	0	0	0	0.98
18	18	218	220	217	221	0	0	0	0	0	1.00
~~19~~	~~19~~	~~215~~	~~216~~	~~221~~	~~NA~~	~~0~~	~~1~~	~~0~~	~~0~~	~~0~~	~~0.96~~
20	20	204	215	219	214	0	1	1	0	0	0.95
~~21~~	~~21~~	~~237~~	~~241~~	~~243~~	~~NA~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0~~	~~0.98~~
~~22~~	~~22~~	~~219~~	~~233~~	~~236~~	~~NA~~	~~0~~	~~1~~	~~1~~	~~0~~	~~0~~	~~0.96~~

Use indexing and `complete.cases()` to remove any cases with NAs

```
> mpls2 = mpls[complete.cases(mpls), ]

> mpls2

   studentID read.5 read.6 read.7 read.8 atRisk female minority ell sped  att
1          1    172    185    179    194      1      1        1   0    0 0.94
3          3    191    199    203    215      1      0        1   0    0 0.97
5          5    207    213    212    213      1      1        1   0    0 0.85
6          6    191    189    206    195      1      0        1   0    0 0.90
7          7    199    208    213    218      1      0        1   1    0 0.97
9          9    149    154    174    177      1      1        1   0    1 0.97
11        11    218    231    233    239      1      1        0   0    0 0.98
12        12    228    232    248    246      1      1        0   0    1 0.96
13        13    228    236    228    239      0      1        0   0    0 0.99
14        14    199    210    225    235      0      0        1   1    0 0.99
16        16    228    226    234    227      0      0        0   0    0 0.97
17        17    201    210    208    219      0      0        0   0    0 0.98
18        18    218    220    217    221      0      0        0   0    0 1.00
20        20    204    215    219    214      0      1        1   0    0 0.95
```

We removed 8 of the original 22 cases (36% were removed!).

Is there an effect of time (i.e., a longitudinal effect) on reading scores?

$$H_0 : \mu_{\text{Grade 5}} = \mu_{\text{Grade 6}} = \mu_{\text{Grade 7}} = \mu_{\text{Grade 8}}$$

Examine this descriptively before any testing...

```
## Examine the means at each measurement wave
> summary(mpls2[2:5])

    read.5           read.6           read.7           read.8
 Min.   :149.0   Min.   :154.0   Min.   :174.0   Min.   :177.0
 1st Qu.:193.0   1st Qu.:201.2   1st Qu.:206.5   1st Qu.:213.2
 Median :202.5   Median :211.5   Median :215.0   Median :218.5
 Mean   :202.4   Mean   :209.1   Mean   :214.2   Mean   :218.0
 3rd Qu.:218.0   3rd Qu.:224.5   3rd Qu.:227.2   3rd Qu.:233.0
 Max.   :228.0   Max.   :236.0   Max.   :248.0   Max.   :246.0
```

The sample means suggest an increase in reading scores over time, on average

It would be great to plot this as well.

To plot the reading scores over time using **ggplot**, we need to reshape the data from the wide format to the long format

long-format data

In the long-formatted data, each row is not a different student, but a different student/grade combination.

Each student is associated with multiple rows

This is similar to the long format of the NBA data where teams (groups) were associated with multiple rows

In the NBA data, players (each row) were nested in teams (which had multiple rows)

In repeated measures data, time points (each row) are nested in the subjects (having multiple rows)...subjects are the groups in these models!

	studentID	variable	value
1	1	read.5	172
2	1	read.6	185
3	1	read.7	179
4	1	read.8	194
5	3	read.5	191
6	3	read.6	199
7	3	read.7	203
8	3	read.8	215
9	5	read.5	207
10	5	read.6	213
	⋮	⋮	⋮
47	17	read.7	208
48	17	read.8	219
49	18	read.5	218
50	18	read.6	220
51	18	read.7	217
52	18	read.8	221
53	20	read.5	204
54	20	read.6	215
55	20	read.7	219
56	20	read.8	214

one column for time predictor

one column for outcome

Reshape Wide to Long Data

```
## Use the reshape2 package
> library(reshape2)

## Melt the data to the long format
> mplsLong = melt(
    mpls2,
    id = c("studentID"),
    measure = c("grade.5", "grade.6", "grade.7", "grade.8")
    )
```

The id= argument

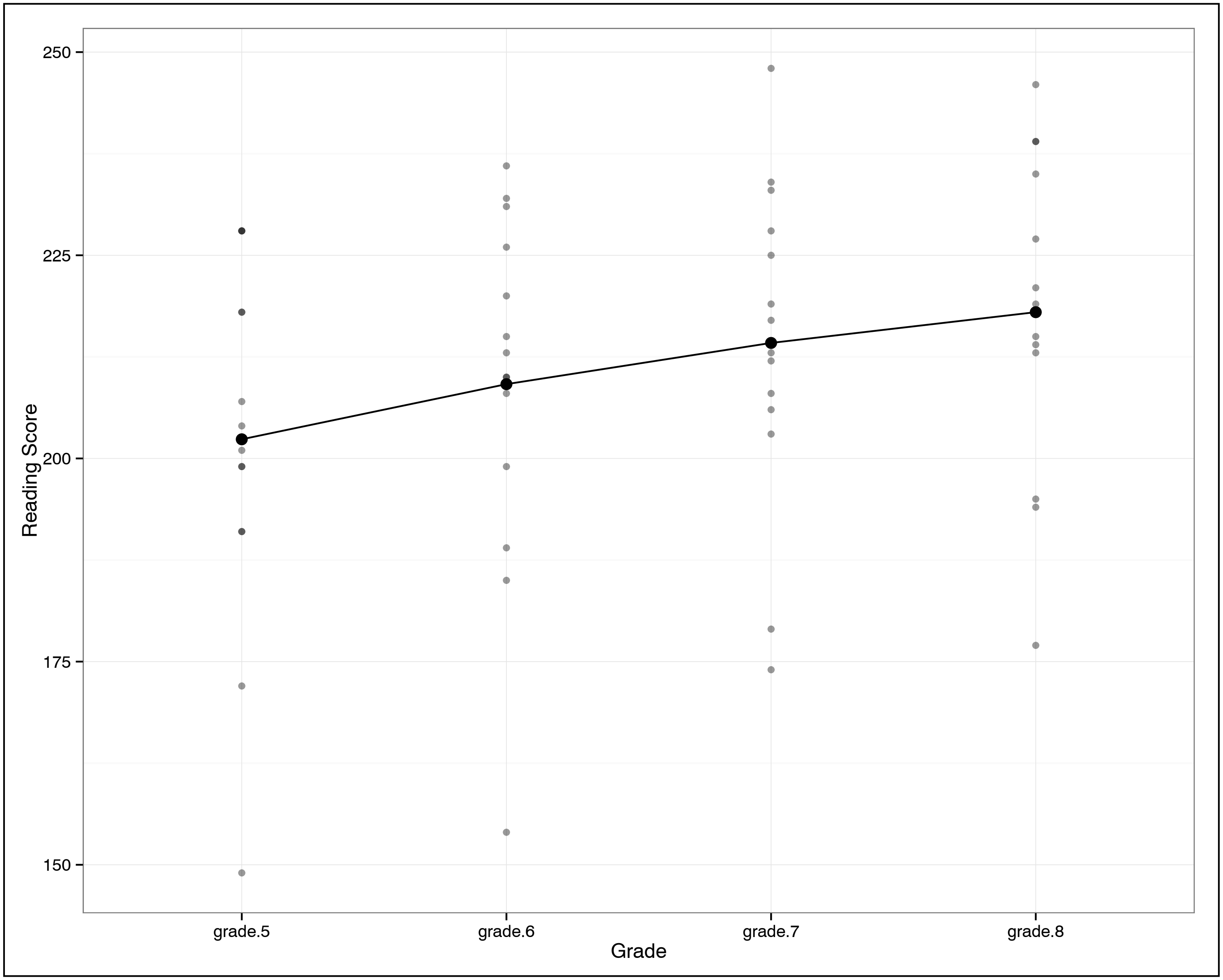
keep these variables 'as is'

The measure= argument

Change these variables into two new ones...variable and value

```
> head(mplsLong)

  studentID variable value
1         1   read.5   172
2         3   read.5   191
3         5   read.5   207
4         6   read.5   191
5         7   read.5   199
6         9   read.5   149
```

```
## Rename the variable and value columns
> names(mplsLong)[2] = "grade"
> names(mplsLong)[3] = "read"

> head(mplsLong)

  studentID  grade read
1         1 read.5  172
2         3 read.5  191
3         5 read.5  207
4         6 read.5  191
5         7 read.5  199
6         9 read.5  149


## Rename the levels of the grade variable
> levels(mplsLong$grade)

[1] "read.5" "read.6" "read.7" "read.8"

> levels(mplsLong$grade)[1] = "grade.5"
> levels(mplsLong$grade)[2] = "grade.6"
> levels(mplsLong$grade)[3] = "grade.7"
> levels(mplsLong$grade)[4] = "grade.8"
```



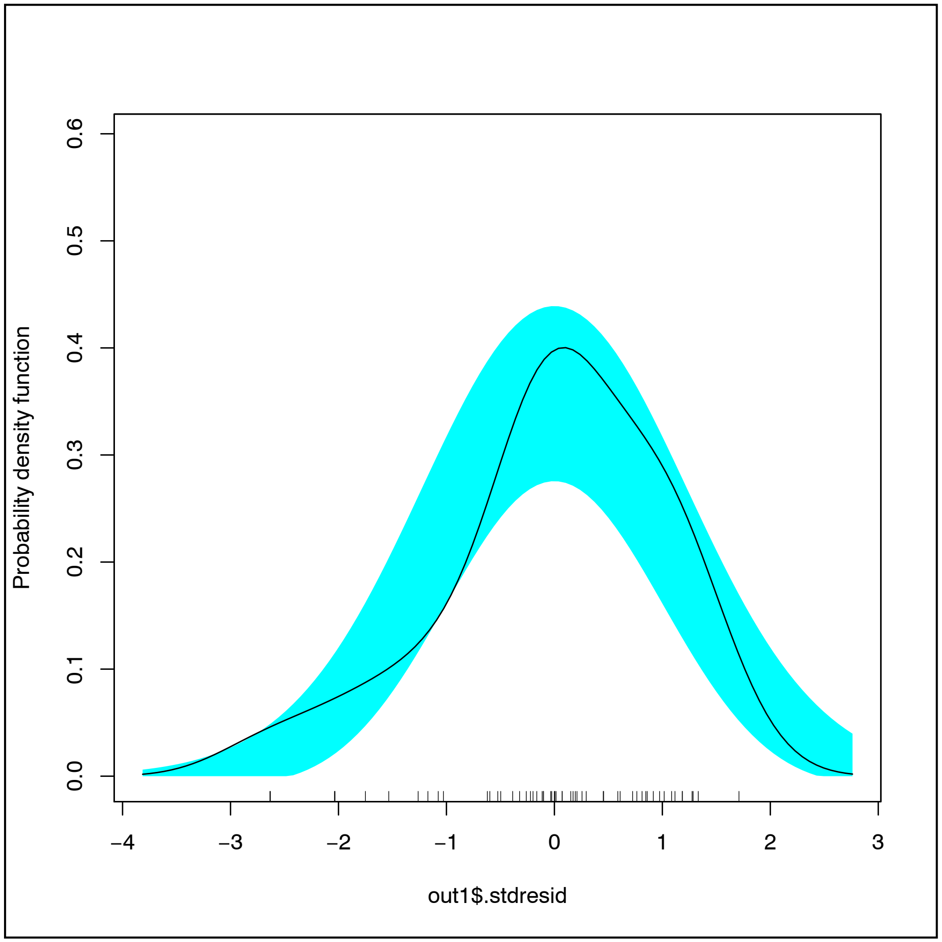
The plot shows the same increasing trend that we observed in the summaries

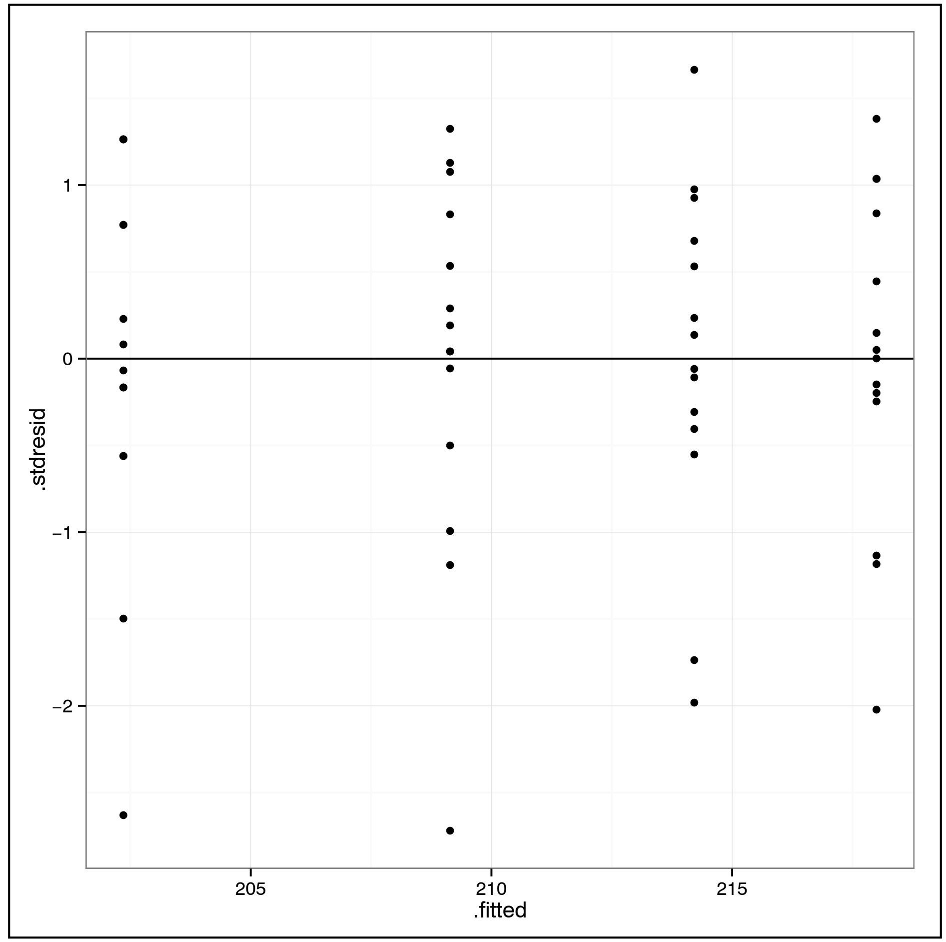
ANALYZING THE DATA UNDER THE ASSUMPTION OF INDEPENDENCE

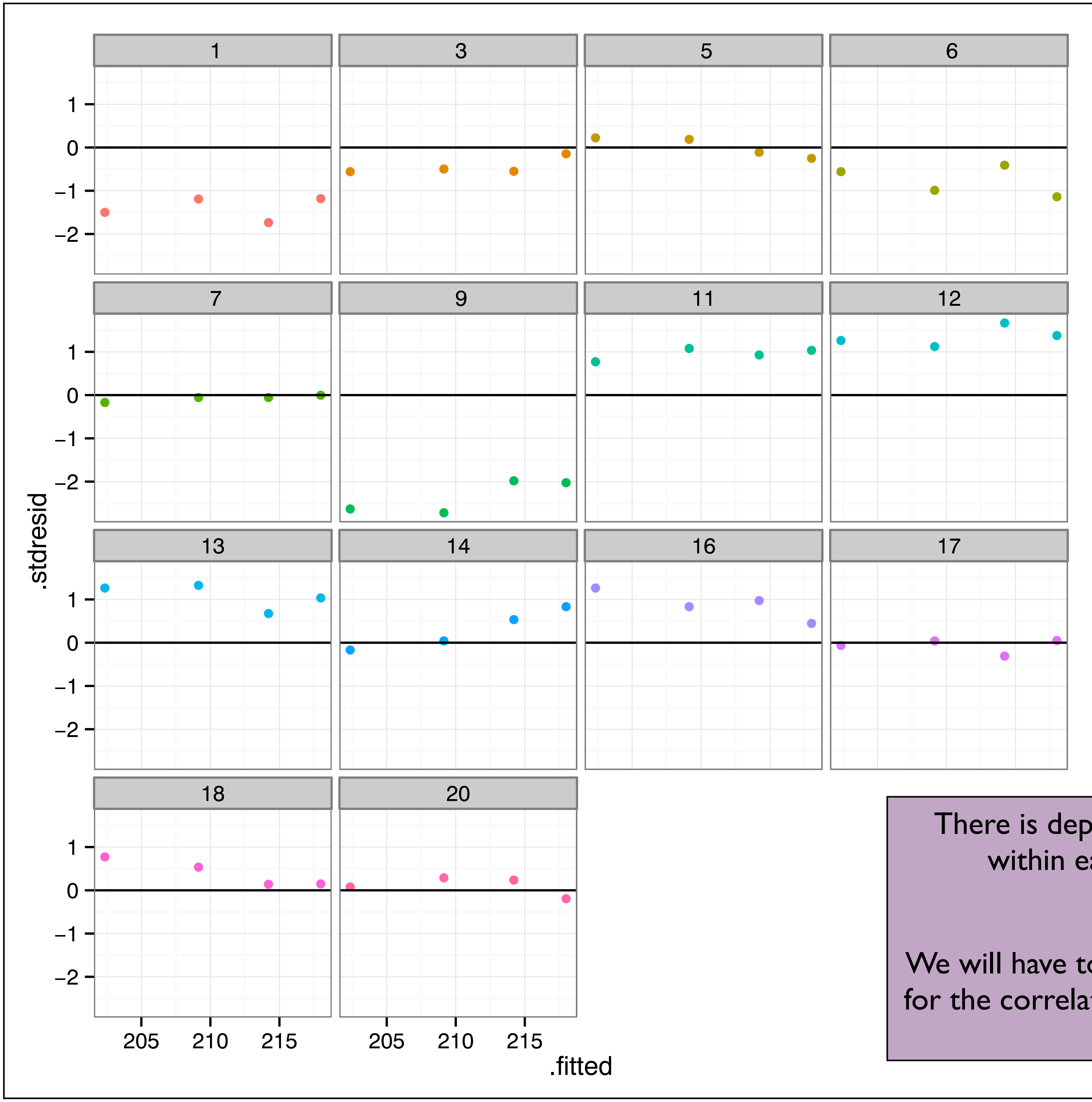
```
## Fit the regression model
> lm.1 = lm(read ~ grade, data = mplsLong)

## Examine anova results
> anova(lm.1)

Analysis of Variance Table

Response: read
          Df  Sum Sq Mean Sq F value Pr(>F)
grade      3  1924.4  641.48  1.4474 0.2397
Residuals 52 23045.3  443.18
```

There does **not** appear to be a main effect of time, $F(3, 52) = 1.45$, $p = 0.240$. This suggests that there are **no differences** in the average reading scores across grades.

Check Assumptions







Remember, when we have groups (nesting), we should check the residuals for each group

There is dependence in the residuals within each group (student).

We will have to fit a model that accounts for the correlation of the residuals within subjects.

ALTERNATIVE ANALYSES

Now What?

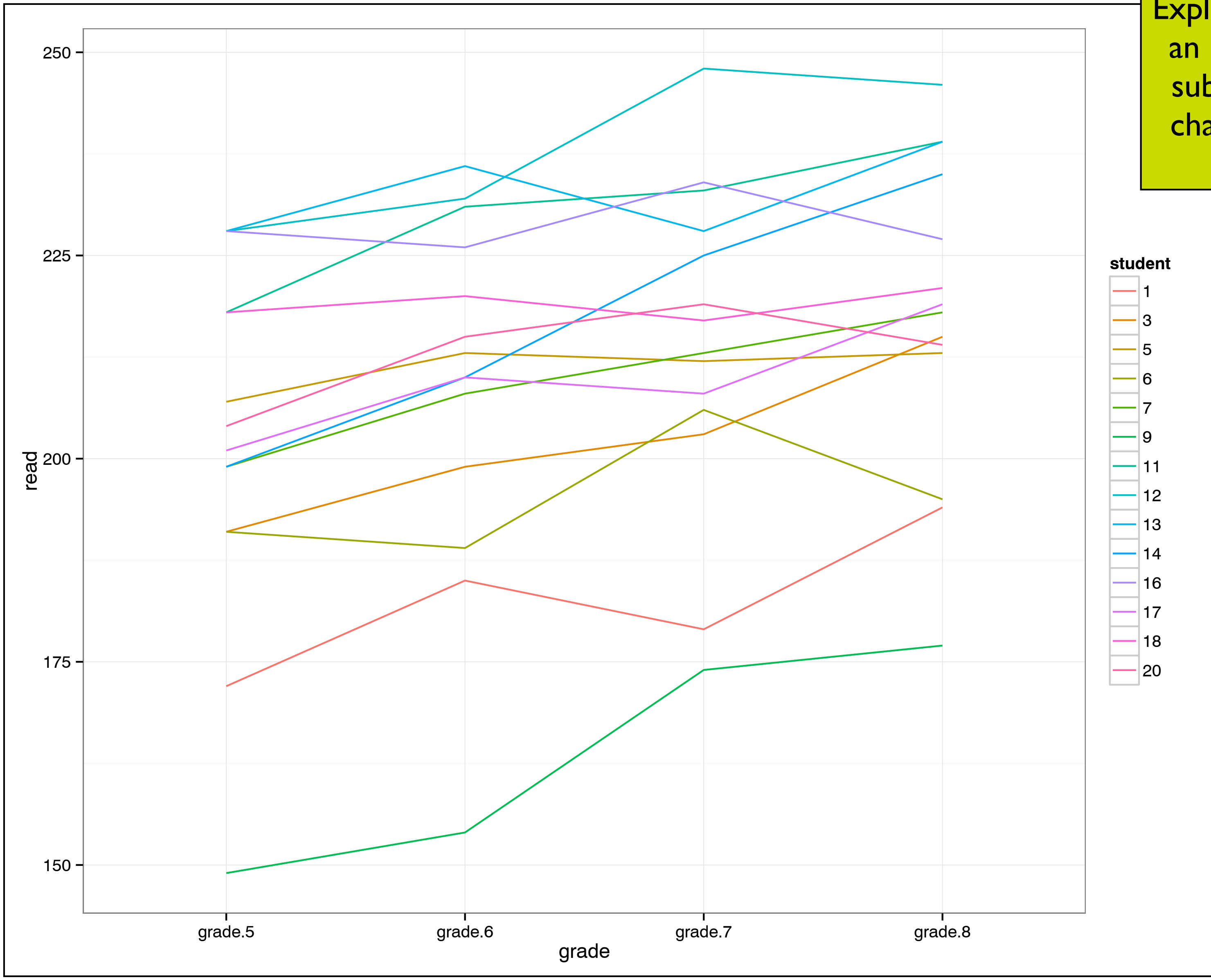
- Remove the correlation by 'transforming' the data
  - ✓ Randomly select one of the two time points for each subject
  - ✓ Compute a composite score
- Or...add subject as a factor into the model explicitly
  - ✓ Two-factor ANOVA
  - ✓ Main effect of time and main effect of subject

Create a subject factor

```
## Coerce studentID into a factor
> mplsLong$student = as.factor(mplsLong$studentID)
```

Should we include an interaction effect between subject and time?

```
## Examine data for possible interaction b/w student and grade
> ggplot(data = mplsLong, aes(x = grade, y = read, group = student)) +
    geom_line(aes(color = student)) +
    theme_bw()
```



Exploratory plot suggests an interaction between subjects and time... the change trajectories are not parallel.

One More Thorn[†,‡]

- All ANOVA analyses to this point have assumed the effect is fixed
  - ✓ Inferences are only drawn to the levels of the factor included in the sample

- In most analyses, we would like to draw inferences to a broader population of subjects (i.e., not just the 22 students in the sample!)
- It is possible to draw inferences to a broader population of levels if we assume that the levels of the factor included in the sample were indeed randomly sampled (or, at least treated a such)
  - ✓ We treat the effect as random
  - ✓ Need to account for the sampling variation that arises in making estimates from a subset of levels

[†] *"Every rose has its thorn" (Michaels, DeVille, Dall & Rockett, 1988).*
[‡] *"Every ANOVA has its thorn" (Zieffler, 2013).*

CONSEQUENCES OF RANDOM-EFFECT IN AN ANOVA: AN EXAMPLE

18 Therapists

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
Psychodynamic	7	6	5	7	6	5	4	4	4	1	2	3	4	4	4	1	2	3	4
Behavioral	4	4	4	1	2	3	7	6	5	7	6	5	1	2	3	4	4	4	4
Rogerian	1	2	3	4	4	4	1	2	3	4	4	4	7	6	5	7	6	5	4
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	

Three psychotherapeutic modes

Mean ratings of effectiveness

All 18 therapists are, on average, equally effective

All 3 modes are, on average, equally effective

No main effect of therapist nor mode

Is this the case when we randomly sample levels of therapist (rather than using all 18 of them) to use in our study

What if we randomly sample therapists?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
Psychodynamic	7	6	5	7	6	5	4	4	4	1	2	3	4	4	4	1	2	3	5
Behavioral	4	4	4	1	2	3	7	6	5	7	6	5	1	2	3	4	4	4	3.33
Rogerian	1	2	3	4	4	4	1	2	3	4	4	4	7	6	5	7	6	5	3.67
	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	

All 3 therapists are, on average, equally effective

All 3 modes are **NOT**, on average, equally effective

There appears to be a main effect of mode (fixed-effect) because of the random sampling of the other factor.

Looking at the cell means, it is obvious that the appearance of a main-effect of therapy mode is because therapy mode and therapist interact.

If we use the same *F*-test that are used for testing fixed-effects models, we will reject H_0 far more often than we should (increased type I error rate)!

A Reminder of the *F*-test for Fixed-Effects

Source	*df*	EMS
A	$a - 1$	$\sigma^2_\epsilon + n\sigma^2_\alpha$
B	$b - 1$	$\sigma^2_\epsilon + n\sigma^2_\beta$
AB	$(a - 1)(b - 1)$	$\sigma^2_\epsilon + n\sigma^2_{\alpha\beta}$
Error	$N - ab$	σ^2_ϵ

EMS is influenced by variation in *Y* and variation in the particular treatment effects

Null hypothesis assumes there are no treatment effects

$$H_0 : \sigma^2_\alpha = 0$$

$$\mathrm{EMS_A} = \sigma^2_\epsilon$$

Under the null hypothesis, the ratio of $\mathrm{EMS_A}$ and $\mathrm{EMS_{Error}}$ is 1.

Compute estimate from data

$$F_\mathrm{A} = \frac{\mathrm{MS_A}}{\mathrm{MS_{Error}}}$$

This ratio for the sample data will likely not be 1, but only because of sampling error.

Examine estimate within what is expected under random sampling

Under the null hypothesis, this ratio has an *F*-distribution with (*a*–*1*, *N*–*ab*) degrees of freedom.

General Process to Test for an Effect

Source	*df*	EMS
A	$a - 1$	$\sigma^2_\epsilon + n\sigma^2_\alpha$
B	$b - 1$	$\sigma^2_\epsilon + n\sigma^2_\beta$
AB	$(a - 1)(b - 1)$	$\sigma^2_\epsilon + n\sigma^2_{\alpha\beta}$
Error	$N - ab$	σ^2_ϵ

1. Find two EMS that are the same if the null hypothesis is true and would differ by the effect only if the null hypothesis is false
2. Put the mean square with the larger EMS in the numerator.
3. Compute the ratio of the mean squares.
4. Examine computed ratio within *F*-distribution defined by appropriate *df*.

To test for the main effect of A...

$$H_0 : \sigma^2_\alpha = 0$$

$$\frac{\mathrm{EMS_A}}{\mathrm{EMS_{Error}}} = \frac{\sigma^2_\epsilon + n\sigma^2_\alpha}{\sigma^2_\epsilon}$$

We compute *F* as

$$F = \frac{\mathrm{MS_A}}{\mathrm{MS_{Error}}}$$

with a–1 and N–ab degrees of freedom

Random-Effects Models

Source	*df*	EMS
A	$a - 1$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$
B	$b - 1$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$
AB	$(a - 1)(b - 1)$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2$
Error	$N - ab$	σ_ϵ^2

1. Find two EMS that are the same if the null hypothesis is true and would differ by the effect only if the null hypothesis is false
2. Put the mean square with the larger EMS in the numerator.
3. Compute the ratio of the mean squares.
4. Examine computed ratio within *F*-distribution defined by appropriate *df*.

To test for the main effect of A...

$$H_0 : \sigma_\alpha^2 = 0$$

$$\frac{\text{EMS}_\text{A}}{\text{EMS}_\text{AB}} = \frac{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2}{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2}$$

We compute *F* as

$$F = \frac{\text{MS}_\text{A}}{\text{MS}_\text{A x B}}$$

with a–1 and (a–1)(b–1) degrees of freedom

The denominator mean square for *F*-tests in random-effects models will not always be MS_{Error}!

Mixed-Effects Models

Source	*df*	EMS
A (fixed)	$a-1$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$
B (random)	$b-1$	$\sigma_\epsilon^2 + na\sigma_\beta^2$
AB (random)	$(a-1)(b-1)$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2$
Error	$N-ab$	σ_ϵ^2

To test for the main effect of A...

$$H_0 : \sigma_\alpha^2 = 0$$

$$\frac{\mathrm{EMS_A}}{\mathrm{EMS_{AB}}} = \frac{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2}{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2}$$

We compute *F* as

$$F = \frac{\mathrm{MS_A}}{\mathrm{MS_{A\ x\ B}}}$$

with a–1 and (a–1)(b–1) degrees of freedom

To test for the main effect of B...

$$H_0 : \sigma_\beta^2 = 0$$

$$\frac{\mathrm{EMS_B}}{\mathrm{EMS_{Error}}} = \frac{\sigma_\epsilon^2 + na\sigma_\beta^2}{\sigma_\epsilon^2}$$

We compute *F* as

$$F = \frac{\mathrm{MS_B}}{\mathrm{MS_{Error}}}$$

with b–1 and N–ab degrees of freedom

COMPUTE THE EFFECT OF GRADE

For longitudinal data...time is generally a fixed effect.

Thus to test for an effect of time we compute

$$F = \frac{\text{MS}_{\text{Time}}}{\text{MS}_{\text{Time x Subject}}}$$

```
## Fit time by subject by grade interaction model
> lm.1 = lm(read ~ grade + student + grade:student, data = mplsLong)

## Examine the anova results
> anova(lm.2)

Analysis of Variance Table

Response: read
              Df  Sum Sq Mean Sq F value Pr(>F)
grade          3  1924.4  641.48
student       13 21696.2 1668.94
grade:student 39  1349.1   34.59
Residuals      0     0.0
```

$$F = \frac{641.48}{34.59} = 18.55$$

This is tested against the appropriate degrees of freedom, namely 3 and 39.

```
              Df  Sum Sq Mean Sq F value Pr(>F)
grade          3  1924.4  641.48
student       13 21696.2 1668.94
grade:student 39  1349.1   34.59
Residuals      0     0.0

## Compute p-value
> 1 - pf(18.55, df1 = 3, df2 = 39)

[1] 1.22747e-07
```

There is likely a main effect of time, $F(3, 39) = 18.55$, $p < .001$.

SPHERICITY AND COMPOUND SYMMETRY

An Additional Assumption for the Model

- Homogeneity of **treatment-difference** variance
  - ✓ Compute the difference scores between each pair of time points
  - ✓ Each set of difference scores must have the same *population* variance

Henry Rouanet

In 1970, Huynh Huynh and Leonard S. Feldt showed[1] that the homogeneity of treatment-difference variance assumption is equivalent to the **population variance–covariance matrix** has a certain form known as **sphericity**.

Leonard S. Feldt

[1] *This was independently discovered in the same year by Rouanet and Lépine.*

Sphericity and Compound Symmetry

Sphericity can really only be defined through the mathematics of linear algebra.

Sphericity is about the differences.

When there are only two measurement waves, sphericity is always met.

Violation of the sphericity assumption can lead to an increase in the type I error rate by as much as 2 to 3 times!

We will discuss a special case of sphericity called **compound symmetry**.

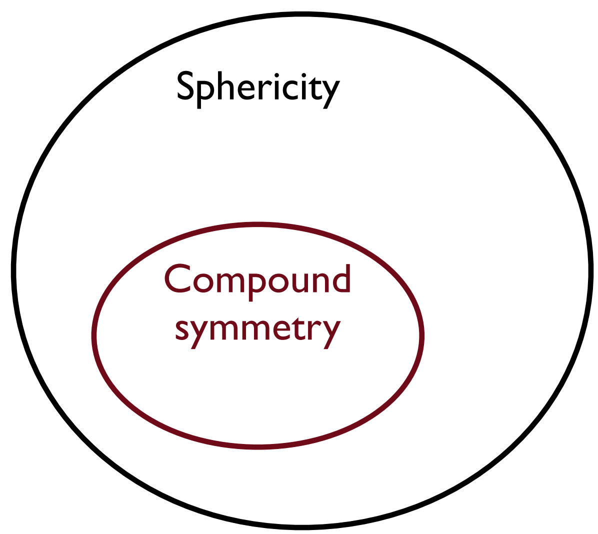
A variance–covariance matrix is said to possess compound symmetry if all of the variances are equal to each other **and** all of the covariances are equal to each other.

All measures must have the same variance and the correlations between every measure must be equal

Compound symmetry is about the measurements.

Compound symmetry is a sufficient condition to satisfy the assumption of sphericity.

However, it is in some ways more difficult to satisfy, so it turns out compound symmetry is not a necessary condition of sphericity.



Read more at http://www.statisticshell.com/docs/sphericity.pdf

Mauchly's Test for Sphericity

John Mauchly

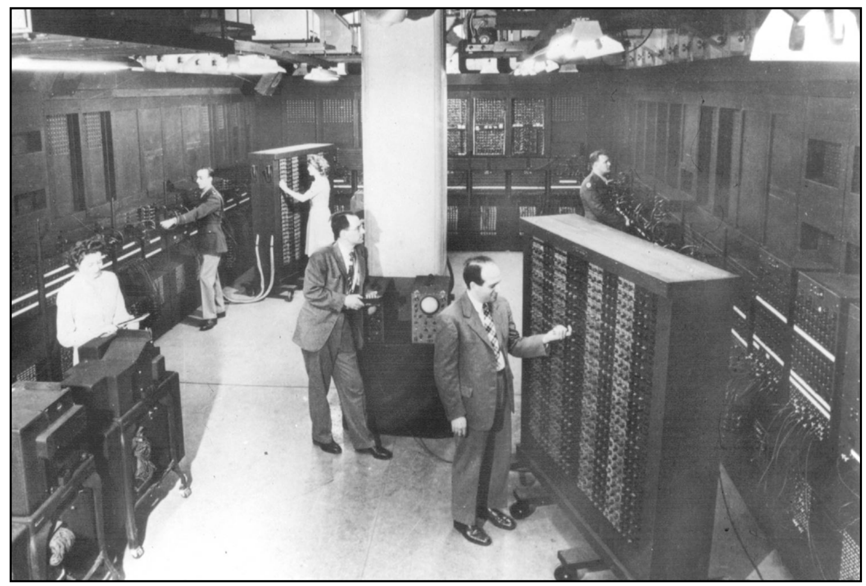
$$H_0 : \text{Sphericity holds in the population.}$$

Test essentially examines whether the variance–covariance matrix is homogenous.

Mauchly's test is generally unreliable and the results should not be trusted.

- Keselman, H. J., Mendoza, J. L., Rogan, J. C., & Breen, L. J. (1980). Testing the validity conditions of repeated measures F tests. *Psychological Bulletin, 87,* 479–481.
- Rogan, J. C., Keselman, H. J., & Mendoza, J. L. (1979). Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology, 32,* 269–286.

Monte carlo studies indicate that the test results are **highly sensitive** to departures from multivariate normality.



Electrical Numerical Integrator and Calculator (ENIAC)

Examine the Variance–Covariance and Correlation Matrix to Assess Compound Symmetry

```
## Examine the variance-covariance matrix (use the wide data)
> var(mpls2[2:5])

         read.5   read.6   read.7   read.8
read.5 502.8626 481.9451 421.0714 383.5385
read.6 481.9451 485.2088 406.8132 396.4615
read.7 421.0714 406.8132 409.1044 361.6923
read.8 383.5385 396.4615 361.6923 375.5385

## Examine the correlation matrix
> cor(mpls2[3:5])

          read.5    read.6    read.7    read.8
read.5 1.0000000 0.9756825 0.9283549 0.8825860
read.6 0.9756825 1.0000000 0.9130901 0.9287729
read.7 0.9283549 0.9130901 1.0000000 0.9227732
read.8 0.8825860 0.9287729 0.9227732 1.0000000
```

The reading score variance seems to be decreasing over measurement waves.

Measurement waves that are further apart are less correlated than measurement waves closer together.

This is very common in longitudinal data!

In applied research it is best to assume that the assumption of compound symmetry will **never** be met.

My Data Have Violated the Assumption of Sphericity...Now What?

There are two methods that are typically used to protect against inflated type I error rate in RM-ANOVA.

- Use an *F*-test with adjusted degrees of freedom (ε-adjustment).
- Use a different method to analyze the data (e.g., Multivariate ANOVA, linear mixed-effects modeling).

Box (1954) showed that under violations of sphericity the *F*-ratios under the null hypothesis would still be distributed in an *F*-distribution, although the *df* would need to be adjusted.

George E. P. Box

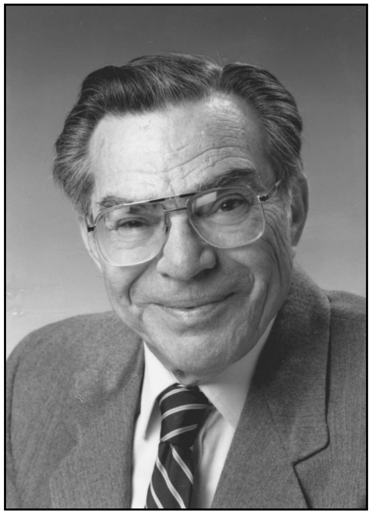
$$df_{\mathrm{A}} = (a-1)\epsilon$$
$$df_{\mathrm{AB}} = (a-1)(b-1)\epsilon$$

Worst case scenario is a distribution of $F(1, b-1)$

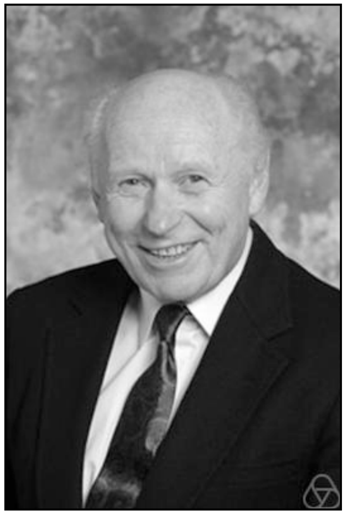
ε is the degree of non-sphericity.

- If sphericity hold, ε = 1.
- If sphericity is violated the lower-bound on ε is $1/(a-1)$.

Greenhouse–Geisser Computation of Epsilon



Samuel W. Greenhouse



Seymour Geisser

Greenhouse and Geisser (1958) extended Box's work and proposed a method to compute the degree to which the data do not meet sphericity (i.e., a computation for epsilon), based on the variance–covariance matrix.

$$\hat{\epsilon}_{GG} = \frac{k^2(\bar{s}_{ii} - \bar{s})^2}{(k-1)(\sum\sum s_{ij}^2 - 2k\sum \bar{s}_{i.}^2 + k^2\bar{s}^2}$$

- $\bar{s}_{ii}$ is the mean of the elements on the main diagonal
- $\bar{s}$ is the mean of all of the elements
- $\bar{s}_{i.}$ is the mean of all of the elements in row *i*
- s_{ij} are the individual elements
- *k* is the number of measurement waves

In 1971, Seymour Geisser became the founding Director of the School of Statistics at the University of Minnesota.

```
## Function to compute Greenhouse-Geisser epsilon estimate
> GG = function(s, k){
     d = k ^ 2 * (mean(diag(s)) - mean(s)) ^ 2
     n1 = sum(s ^ 2)
     n2 = 2 * k * sum(apply(s, 1, mean) ^ 2)
     n3 = k ^ 2 * mean(s) ^ 2
     epsi = d / ((k - 1) * (n1 - n2 + n3))
     return(epsi)
     }

## Assign the variance-covariance matrix (use the wide data)
> s = var(mpls2[2:5])

## Use the function (k is the number of measurement waves)
> GG(s, k = 4)

[1] 0.6774003
```

The Greenhouse–Geisser degree of non-sphericity is 0.677.

F should be evaluated in a distribution with 3(0.677) = 2.03 and 39(0.677) = 26.42 degrees of freedom.

There is likely a main effect of time, $F(2.03, 26.42) = 18.55$, $p < .001$.

Huynh-Feldt Computation of Epsilon

Huynh Huynh

Leonard S. Feldt

The Greenhouse-Geisser epsilon tends to underestimate epsilon when epsilon is greater than 0.70.

Huynh and Feldt (1976) proposed a less conservative adjustment to epsilon.

$$\hat{\epsilon}_{HF} = \frac{n(k-1)\hat{\epsilon}_{GG} - 2}{(k-1)\left[n - 1 - (k-1)\hat{\epsilon}_{GG}\right]}$$

- *n* is the number of subjects
- *k* is the number of measurement waves

```
## Function to compute Huynh-Feldt epsilon estimate
> HF = function(epsi, k = 3, n = 30){
    epsiHF = (n * (k - 1) * epsi - 2) / ((k - 1) * ((n - 1) -
      (k - 1) * epsi))
     return(epsiHF)
     }

## Use the function (k is the number of measurement waves; n is the
## number of subjects)
> HF(epsi = 0.6774003, k = 4, n = 14)

[1] 0.8038931
```

The Huynh–Feldt degree of non-sphericity is 0.803.

(Note: This is closer to 1 than the Greenhouse–Geisser epsilon)

F should be evaluated in a distribution with 3(0.803) = 2.41 and 39(0.803) = 31.35 degrees of freedom.

There is likely a main effect of time, $F(2.41, 31.35) = 18.55$, $p < .001$.

Barcikowski, R. S., & Robey, R.R. (1983). Decisions in single group repeated measures analysis: Statistical tests and three computer packages. *American Statistician, 38,* 148–150

Examine Greenhouse–Geisser estimate for ε.

- If $\varepsilon > 0.75$ use the Huynh–Feldt adjustment.
- If $\varepsilon \leq 0.75$ use the Greenhouse–Geisser adjustment.

USING THE EZANOVA FUNCTION

Fit the Mixed-Effects ANOVA using ezANOVA()

```
## Load the ez library
> library(ez)

## Fit the model
> rm.aov = ezANOVA(data = mplsLong,
    dv = read,
    wid = student,
    within = .(grade),
    detailed = TRUE
    )

> rm.aov

$ANOVA
       Effect DFn DFd         SSn       SSd          F            p p<.05        ges
1 (Intercept)   1  13 2491488.286 21696.214 1492.85711 8.380946e-15     * 0.99083516
2       grade   3  39    1924.429  1349.071   18.54429 1.231739e-07     * 0.07707051
```

Unadjusted
Sphericity assumed

Generalized effect size

```
$`Mauchly's Test for Sphericity`
  Effect         W           p p<.05
2  grade 0.2608632 0.007810945     *

$`Sphericity Corrections`
  Effect       GGe        p[GG] p[GG]<.05      HFe        p[HF] p[HF]<.05
2  grade 0.6774003 8.581994e-06         * 0.803893 1.616161e-06         *
```

Mauchly's Test
Sphericity is unlikely to hold in the population.

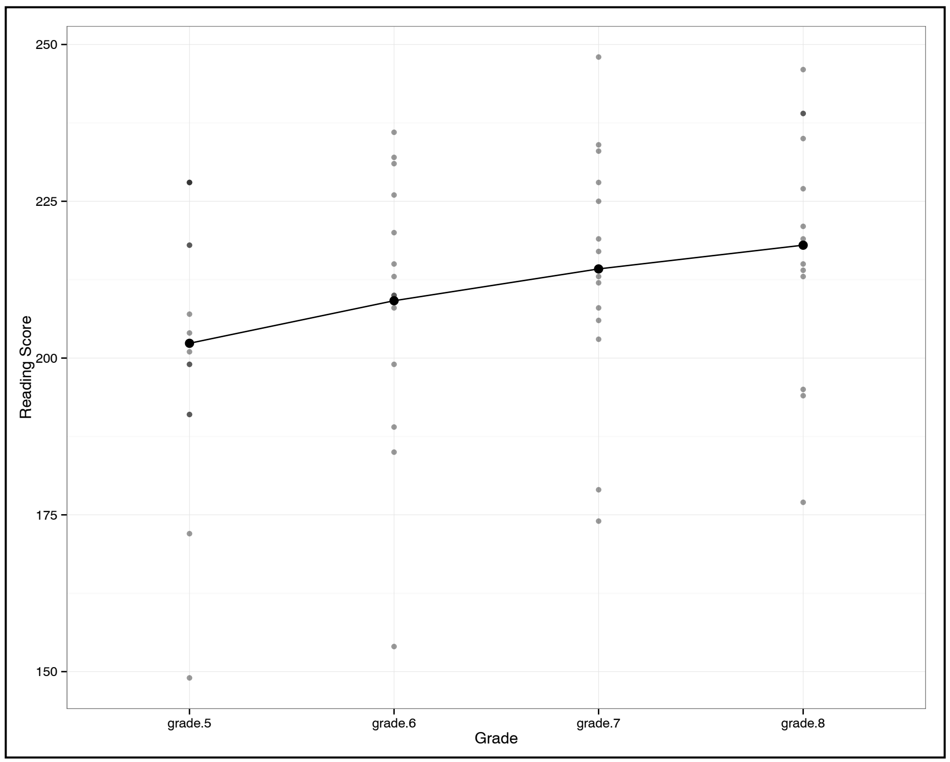
Since the data are repeated measures,
sphericity is assumed to be untenable, regardless of Mauchly's test.

Greenhouse–Geisser Adjustment

Huynh–Feldt Adjustment

WRITING UP THE RESULTS

Students' reading scores were analyzed using a repeated measures analysis of variance (RM-ANOVA) with grade (5th, 6th, 7th, and 8th) as a within-subjects factor. The main effect of grade was significant, $F(3, 39) = 18.54$, Greenhouse-Geisser adjusted $\varepsilon = .677$, $p = < .001$. This suggests that there are statistically reliable differences in the average reading score between 5th and 8th grade. The figure, below, shows the average reading scores by grade.



Other Potential Questions

Which grades are different?

- 5th vs 6th?
- 5th vs 7th?

Is the trend linear? Quadratic?

Which grades are different (5th vs 6th? 5th vs 7th? etc.)

Is there still an effect of grade...

- Controlling for risk?
- Controlling for attendance?
- Between males and females?
- Between minorities and non-minorities?