Assumptions underlying analysis of variance

Sanne Berends

Assumptions can pertain to:

- Measurement scale
- Method of sampling and/or assigning subjects to treatments
- Selection of factor levels
- etc

Assumptions in analysis of variance:

1. The normality assumption
2. The homogeneity of variance assumption
3. The independence assumption

What can be done?

- In general:
  - Analysis of variance is robust.
    - Nominal α level ≈ actual α level
      - Nonnormality
      - Heterogeneous variances (cell frequencies large and equal)
  - But: small violations of the independence assumption can have dramatic effects.
    - This is a requirement of the design

Normality of variances and homogeneity

- Normality of variances
  - The distribution is skewed
  - The dependent variable is not an interval variable
  - Kolmogorov-Smirnov and Shapiro-Wilk normality tests
- Homogeneity of variances
  - Levene's test => Welch's test

Example

- Dyslexia
- Communicative Developmental Inventory
- Closed-class words
- 17m, 23m & 29m

Example

Descriptives

FWRDSCHT

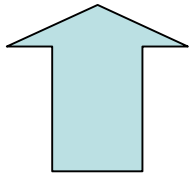
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1,00	206	1,8155	2,36333	,16466	1,4909	2,1402	,00	12,00
2,00	206	13,7039	12,17253	,84810	12,0318	15,3760	,00	53,00
3,00	206	39,4320	22,85549	1,59242	36,2924	42,5717	,00	89,00
Total	618	18,3172	21,71451	,87349	16,6018	20,0325	,00	89,00

Example

Test of Homogeneity of Variances

FWRDSCHT

Levene Statistic	df1	df2	Sig.
235,876	2	615	,000

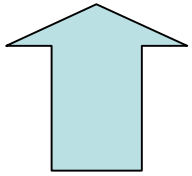


Example

ANOVA

FWRDSCHT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	152321,4	2	76160,681	337,927	,000
Within Groups	138606,5	615	225,376		
Total	290927,8	617			



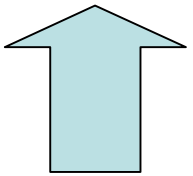
Example

Robust Tests of Equality of Means

FWRDSCHT

	Statistic[a]	df1	df2	Sig.
Welch	363,686	2	286,030	,000

a. Asymptotically F distributed.



Assumptions not met:

- Transformation
  - Log
  - Arcsine
- Non-paramatric alternative
  - Kruskal-Wallis
  - Friedman
- Randomization test

Repeated measures

- Disadvantage Repeated measures:
  - Independency assumption is violated
- Additional assumption:
  - Assumption of *sphericity*

Repeated measures

- What is sphericity?
  - More general condition of *compound symmetry*
- How is it measured?
- SPSS:
  - Mauchly's test

variance A-B ≈ variance A-C ≈ variance B-C

Group A	Group B	Group C	A – B	A – C	B – C
10	12	8	-2	2	5
15	15	12	0	3	3
25	30	20	-5	5	10
35	30	28	5	7	2
30	27	20	3	10	7
		Variance:	15.7	10.3	10.3

Assessing and correcting for sphericity

- The Greenhouse-Geisser estimate (ε)
- $1/(k - 1) \leq \varepsilon \leq 1$ with k = number of repeated measures conditions

Example

Within-Subjects Factors

Measure: MEASURE_1

AGE	Dependent Variable
1	M17
2	M23
3	M29

Descriptive Statistics

	Mean	Std. Deviation	N
M17	1,8155	2,36333	206
M23	13,7039	12,17253	206
M29	39,4320	22,85549	206

Example

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AGE	,514	135,591	2	,000	,673	,676	,500

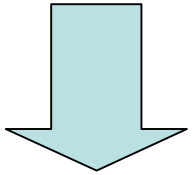
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

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: AGE

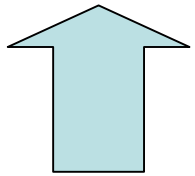
Example



Tests of Within-Subjects Effects

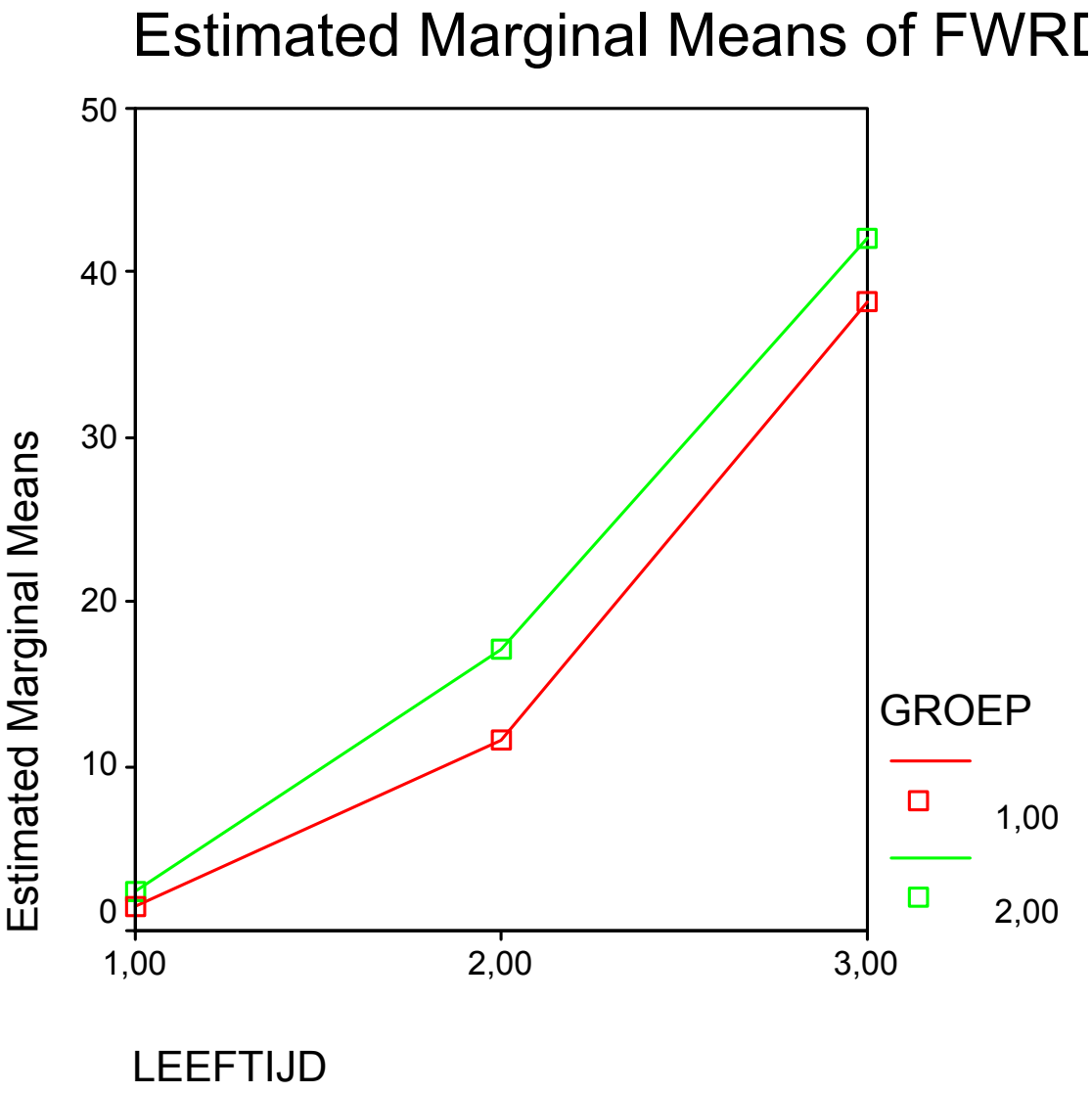
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AGE	Sphericity Assumed	152321,362	2	76160,681	509,877	,000
	Greenhouse-Geisser	152321,362	1,346	113140,769	509,877	,000
	Huynh-Feldt	152321,362	1,352	112664,590	509,877	,000
	Lower-bound	152321,362	1,000	152321,362	509,877	,000
Error(AGE)	Sphericity Assumed	61241,971	410	149,371		
	Greenhouse-Geisser	61241,971	275,991	221,898		
	Huynh-Feldt	61241,971	277,158	220,964		
	Lower-bound	61241,971	205,000	298,741		



References

- Field, A. (2000) *Discovering Statistics using SPSS for Windows.* Sage: London.
- Moore, D. S. and McCabe, G. (1993) *Introduction to the Practice of Statistics* 5th edition. Freeman: New York.
- Rietveld, T. and Van Hout, R. (1993) *Statistical Techniques for the Study of Language and Language Behavior.* Mouton De Gruyter: Berlin.
- Rietveld, T. and Van Hout, R. (2005) *Statistics in Language Research: Analysis of Variance*. Mouton De Gruyter: Berlin.


Estimated Marginal Means of FWRI
Estimated Marginal Means
50
40
30
20
10
0
1,00
2,00
3,00
LEEFTIJD
GROEP
1,00
2,00