Repeated measures

Statistics II (LIX002X05)

University of Groningen, Faculty of Arts, Information Science

Wilbert Heeringa

Introduction

- **Repeated measurements** or **within-subjects design** are several measurements of the same variable for the same *subject* or *observational unit*.
- The observational units are what you take measurements on.
- **Several points of time**: when patients are repeatedly measured in an follow-up period, i.e. monitoring a person's health over time after treatment.
- **Several locations**: measurements at several locations in the body of the same person (left and right eye, several slices in a MRI image).
- **Several conditions**: when the same patient is measured under two or more different conditions (for example before and after a treatment).

Introduction

- Repeated measures ANOVA looks like a paired-samples t test, but with three or more conditions.
- Repeated measures ANOVA is available only for continuous dependent variables which are normally distributed.
- All of the subjects are simultaneously measured at fixed times or under fixed conditions.

Example

- Example: we study the change of 25 local dialects in the Netherlands and Flanders.
- The locations are the observational units.
- For each location we consider the change in:
  - Lexis, for example in Slochteren *ontdaan* 'upset' is translated as *veraldereerd* by the elderly and as *ontdoan* by young people.
  - Morphology, for example in Slochteren *ineens* 'suddenly' is translated as *inainen* by the elderly and as *inains* by young people.
  - Sound components, for example in Slochteren *vraagt* 'asked' is translated as *vragt* by the elderly and as *vroagt* by young people.

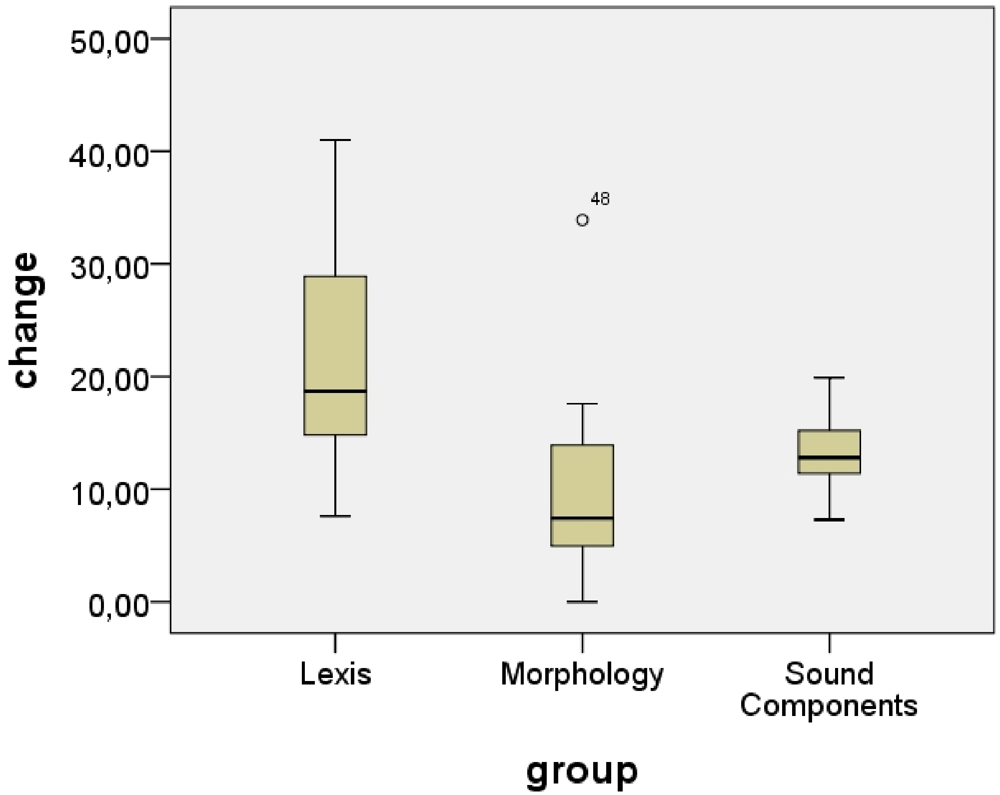
Example

- The measurements are obtained on the basis of a sample of 125 words which are translated by speakers in each of the 25 locations in the period 2008–2011.
- In each location older men (60 years and older) and young female (between 20 and 40 years old) are recorded. For each location we measure:
  - Lexical change: percentage of words in which young people chose another lexeme than the elderly.
  - Morphological change: percentage of words which younger people inflected differently from the elderly.
  - Change in the sound components: percentage of sounds in words which younger people pronounced differently from the elderly.

Locations of the 25 dialects. 18 locations are found in the Netherlands and seven locations are found in Flanders.

country	plaats	lexis	morph	sound
1	Budel	9	4	13
2	Deerlijk	21	10	19
1	Den Burg	30	11	12
2	Diepenbeek	29	14	19
1	Dokkum	12	11	12
1	Dordrecht	18	7	8
1	Grolloo	17	10	14
1	Grouw	16	7	9
1	Huizen	19	7	16
1	IJmuiden	26	0	11
1	Kampen	11	7	11
1	Kerkrade	13	15	13
1	Lunteren	23	4	15
2	Maldegem	39	13	20
1	Naaldwijk	31	3	11
1	Nijverdal	12	6	7
2	Oostende	33	3	13
2	Overijse	20	16	12
1	Pannerden	16	3	16
2	Poperinge	30	7	11
2	Rijkevorsel	17	15	16
1	Sint-Oedenrode	41	5	13
1	Slochteren	24	34	14
1	Tegelen	8	18	13
1	Zierikzee	15	14	12

The table shows the percentage of words (lexis, morph) or sounds (sound) that changed for each location.



A boxplot is given for the measurements of each of the linguistic levels.

Example

- Question: is the extent to which local dialects changed the same for all of linguistic levels?
- This question can be answered by means of a one-factor ANOVA test.
- However, the measurements for each of the linguistic levels are **repeated measures**.
- In our example: for each dialect we have a triple: lexical change, morphological change, change in the sound components.
- Generally a repeated measures ANOVA test is more powerful to detect significance than a one-factor ANOVA test.

Repetition: one-factor ANOVA

- Hypotheses:
  H_0: All of groups have the same population mean.
  H_a: Not all of them are the same.

- Test statistic:
$$F = \frac{MSG}{MSE} = \frac{variation\ between\ groups}{variation\ within\ groups}$$

Calculation of MSG

- Variation **between** groups is measured by MSG (Mean Sum of Squares Group): the averaged squared deviation of group means compared to the global mean.
- Group means: 21.1 (lexis), 9.7 (morph), 13.2 (sound). Global mean: 14.7.
- SSG = Sum of Squares Group:

$$SSG = \sum_{i=1}^{I} n_i(\overline{x}_i - \overline{x})^2 =$$

$$(21.1 - 14.7)^2 \times 25 +$$
$$(9.7 - 14.7)^2 \times 25 +$$
$$(13.2 - 14.7)^2 \times 25$$

$$= 1705.3$$

Calculation of MSG

- DFG = Degrees of Freedom Group $= I - 1 = 3 - 1 = 2$.
- MSG = SSG / DFG = 1705.3 / 2 = 852.6.

Calculation of MSE

- Variation **within** the groups is measured by MSE (Mean Sum of Squares Error): average squared deviation of the observations compared to their group means.
- Group means: 21.1 (lexis), 9.7 (morph), 13.2 (sound).
- SSE = Sum of Squares Error:

$$SSE = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij} - \overline{x}_i)^2 =$$

$$(9 - 21.1)^2 + (21 - 21.1)^2 + \ \dots + \\ (4 - 9.7)^2 + (10 - 9.7)^2 + \ \dots + \\ (13 - 13.2)^2 + (19 - 13.2)^2 + \ \dots$$

$$= 3408.0$$

Calculation of MSE

- DFE = Degrees of Freedom Error = $N - I$ = 75-3 = 72
- MSE = SSE / DFE = 3408.0/72 = 47.3

Calculation of test statistic F

- Test statistic:

$$F = \frac{MSG}{MSE} = \frac{852.6}{47.3} = 18.0$$

- ANOVA table in SPSS:

ANOVA

change

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1703,440	2	851,720	17,994	,000
Within Groups	3407,954	72	47,333		
Total	5111,394	74			

Calculation of MSE*

- Variation in SSE may have many causes, for example measurement errors, factors not controlled by the researcher, the weather, etc.
- The larger part of SSE may be the result of individual differences between the observational units.
- However: in a repeated measures design we have **the same observational units** in each group, and therefore the same individual differences!
- We want to remove the effect of **individual differences** on SSE.
- We calculate SSE* (and MSE*): variation within groups minus the differences between the observational units.

Calculation of MSE*

- SSE* = SSE - SSS.
- SSS = Sum of Squares for Subjects. SSS measures the variation between observational units:

$$SSS = I \times \sum_{j=1}^{n_i} \left(\overline{x}_j - \overline{x}\right)^2$$

where $\overline{x}$ is the global mean (14.7) and $\overline{x}_j$ is the mean of the observations per observational unit.

- In our case: for the dialect in location j $\overline{x}_j$ is the mean of the lexical change, the morphological change and the change in the sound components.
- n_i is the number of observational units within the group, which is for all groups the same in a repeated measures design.

country	plaats	lexis	morph	sound	mean
1	Budel	9	4	13	9
2	Deerlijk	21	10	19	16
1	Den Burg	30	11	12	17
2	Diepenbeek	29	14	19	21
1	Dokkum	12	11	12	12
1	Dordrecht	18	7	8	11
1	Grolloo	17	10	14	14
1	Grouw	16	7	9	11
1	Huizen	19	7	16	14
1	IJmuiden	26	0	11	12
1	Kampen	11	7	11	10
1	Kerkrade	13	15	13	14
1	Lunteren	23	4	15	14
2	Maldegem	39	13	20	24
1	Naaldwijk	31	3	11	15
1	Nijverdal	12	6	7	8
2	Oostende	33	3	13	16
2	Overijse	20	16	12	16
1	Pannerden	16	3	16	12
2	Poperinge	30	7	11	16
2	Rijkevorsel	17	15	16	16
1	Sint-Oedenrode	41	5	13	20
1	Slochteren	24	34	14	24
1	Tegelen	8	18	13	13
1	Zierikzee	15	14	12	14

Calculation of the average of lexicale change, morphological change and change in the sound components per location.

Calculation of MSE*

- The total variation between subjects is:

$$SSS = 3 \times [(9 - 14.7)^2 + (16 - 14.7)^2 + \ldots] = 1239.2$$

- DFS = Degrees of Freedom of Subject = number of subjects in each group -1
  DFS = 25 -1 = 24
- SSE* = SSE - SSS
  SSE* = 3408.0 - 1239.2 = 2168.8.
- DFE* = DFE - DFS
  DFE* = 72 - 24 = 48.
- Now we calculate:

$$MSE* = \frac{SSE*}{DFE*} = \frac{2168.8}{48} = 45.2$$

Calculation of test statistic F

- Test statistic:

$$F = \frac{MSG}{MSE*} = \frac{852.6}{45.2} = 18.9$$

- Note that this F is larger than the F we found earlier in the one-factor ANOVA design (where F=18.0)

Assumptions

- Randomness:
  Cases should be derived from a random sample, and scores from different participants should be independent of each other.
- Normality:
  each sample (3 linguistic levels = 3 samples) is drawn from a normally distributed population. Use normal quantile plots and the Shapiro-Wilk test.
- Homogeneity of variance:
  the groups (in our case three groups) should have the same variance. Use Levene's test and Hartley's test.
- Sphericity:
  refers to the equality of variances of the *differences* between the levels.

Sphericity

- In our example, we have three levels:
  lexis, morphology, sound components.
- We form all possible pairs:
  lexis/morphology, lexis/sound components, morphology/sound components.
- For each pair we calculate the differences, these differences should have approximately equal variances.
- **Mauchly's Test of Sphericity**: a statistical test with the null hypothesis that the variances of the differences between the levels are equal.
- If $p < .05$ the condition of sphericity is not met.
- If the condition of sphericity is not met, consider that **Greenhouse-Geisser** correction in the SPSS output.

Mauchly's Test of Sphericity[b]

Measure:LingChange

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
LingLevel	,590	12,151	2	,002	,709	,741	,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: LingLevel

Application of Mauchly's Test of Sphericity to our example in SPSS. The p value (Sig.) is smaller than 0.05, therefore we may not assume that the assumption of sphericity is met.

Tests of Within-Subjects Effects

Measure:LingChange

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
LingLevel	Sphericity Assumed	1703,440	2	851,720	18,849	,000	,440
	Greenhouse-Geisser	1703,440	1,418	1201,255	18,849	,000	,440
	Huynh-Feldt	1703,440	1,481	1149,941	18,849	,000	,440
	Lower-bound	1703,440	1,000	1703,440	18,849	,000	,440
Error(LingLevel)	Sphericity Assumed	2168,907	48	45,186			
	Greenhouse-Geisser	2168,907	34,033	63,729			
	Huynh-Feldt	2168,907	35,552	61,007			
	Lower-bound	2168,907	24,000	90,371			

After *LingLevel/Sphericity Assumed* we find SSG, DFG en MSG. After *Error(LingLevel)/Sphericity Assumed* we find SSE*, DFE* en MSE*.

Since we obtained a signficant result with the *Mauchly's Test of Sphericity*, we used the results of the **Greenhouse-Geisser conversative F test**. Conservative means: H0 is less likely to be rejected.

Effect size

- SPSS also gives the **Eta Squared**. This is the same as the determination coefficient R^2:

$$R^2 = \frac{SSG}{SSG + SSE*} = \frac{1703.440}{1703.440 + 2168.907} = 0.440$$

- About 44% of the dialect change measurements is explained by the distinction in three linguistic levels (lexis, morphology, sound components).
- Conclusion: the measurements of dialect change at the three linguistic levels differ significantly at the 5% level: $F(1.418, 34.033) =$ 18.849, $p < 0.001$. $R^2 = 0.440$.

Contrasts and multiple comparisons

- Now we know that the measurements of dialect change at the three linguistic levels differ significantly.
- **Which pairs** of population means differ from each other?
- In SPSS it is not possible to specify contrasts in the way we did for one-way ANOVA's. However, **multiple comparisons** (posthoc tests) can be carried out.
- When doing multiple comparisons, the Tukey test is not recommended (Gray & Kinnear 2012, p. 326), but the Bonferroni and Sidak modifications are recommended.
- Both the Bonferroni and the Sidak test are more conservative than the Tukey test (i.e. H0 is less likely rejected, the p values are a bit larger), Bonferroni is even more conservative than Sidak.

Pairwise Comparisons

Measure:LingChange

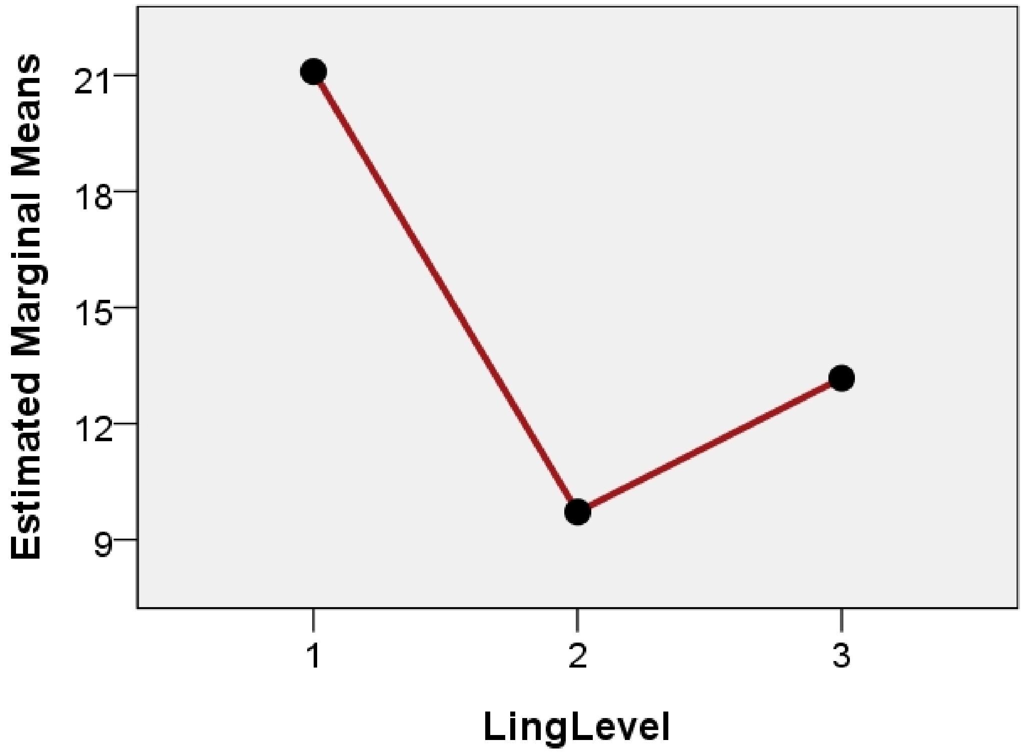
(I) LingLevel	(J) LingLevel	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	11,385*	2,416	,000	5,167	17,603
	3	7,928*	1,751	,000	3,420	12,435
2	1	-11,385*	2,416	,000	-17,603	-5,167
	3	-3,457	1,393	,061	-7,042	,127
3	1	-7,928*	1,751	,000	-12,435	-3,420
	2	3,457	1,393	,061	-,127	7,042

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Multiple comparisons on the basis of the Bonferronie modification. 1 = lexis, 2 = morphology, 3 = sound components. Note that there is not a significant difference between morphology and sound components.



Profile plot. 1 = lexis, 2 = morphology, 3 = sound components. The graph confirms our results: lexis is different compared to the other linguistic levels, but the difference between morphology and sound components is much smaller, we found that they do not differ significantly.

Factorial repeated measures ANOVA

- We looked at **one-way** repeated measures ANOVA, i.e. we have one independent variable.
- In our case the independent variable is *linguistic level* having levels 'lexicale change', 'morphological change' and 'change in the sound components'.
- **Multi-way** or *factorial* repeated measures ANOVA compares several means when there are two or more independent variables, and the same subjects have been used in all experimental conditions.

Non-parametric equivalents

- Non-parametric alternatives for one-way repeated measures ANOVA:
  - Friedman test:
    used for ordinal data. In case this test gives a significant result, multiple comparisons are made by pairwise comparisons of the variables with the Wilcoxon signed-rank test, with the Bonferroni correction.
  - Cochran's Q test:
    can be used for nominal data.
- Both tests are available in SPSS.