

Day 3

A standard model in Lab experiments:

The time * treatment interaction

Why
time * treatment interaction?

In many laboratory experiments we randomly assign the experimental units (cell cultures, animals) to two or more groups with different treatments.

At time zero the groups vary only by random variation. Over time we expect that for both groups the measurements drift apart. Differences between both groups are expected to become larger at least for a certain time window of interest. Thus even if it's justified to expect linear dependency of measurements on time, it is not justified to expect parallel lines.

If linearity is not assumed we also do not expect an identical shape for both group (e.g. a quadratic function which differs only in the linear term).

The scientific hypothesis of a different effect between groups thus reduces to the statistical test for the interaction between treatment and time.

Two standard experiments

Now compare the following experiments:

In the first experiment genetically identical mice are implanted tumour tissue. The mice are assigned randomly to a group with radiation therapy and a group with chemotherapy. Each of four weeks (week = 0,1,2,3) a fixed number of animals in both groups are killed and the weight of the tumor tissue is measured.

In the second experiment genetically identical mice are assigned randomly to a low calories diet or to a high calories diet. Each week the weight of the animals is recorded.

The important difference is that we have to deal with repeated measurements in the second experiment.

A closer look to both experiments

Balancedness:

In both experiments we obtain measurements at four time points $t = 0$, $t = 1$, $t = 2$, and $t = 3$. For each time point we have eight measurements per group. As we will choose the time points identical in both groups we don't have problems to balance treatment and time.

Number of animals:

In experiment 1 we need 64 animals (for each time point and each group 8 animals). We obtain one measurement per animal.

In experiment 2 we only need 16 animals (for each group 8 animals). We obtain four measurements per animal.

Two standard analyses of the experiments

For both experiments, we have two options to parameterize time:

We can use time as a continuous covariate (like in analysis of covariance). Then the main effect of time has one degree of freedom (parameter: slope) and the interaction between time and group the same (parameter: deviations of slopes in both groups from the average slope).

But we can also use time as a factor with four levels. Then we have three degrees of freedom for the main effect and three degrees of freedom for the interaction.

Note:

The parameterization of time has nothing to do with the question whether we have independent or dependent measurements as discussed on the preceding sheets.

Summary and notation of the different approaches

	Independent Data from 64 mice in two groups	Dependent Data from 16 mice in two groups
Time is factor (df=3)	I	II
Time is covariate (df=1)	III	IV

Note:

For each of the four models the focus lies identically on the interaction between time and group.

Further notes (apply to I, II, III, IV)

As noted we don't have problems to balance the factor (treatment) and the covariate (time). It would be a flaw in our experiment to violate this balance.

In other situations we only observe covariates (e.g. baseline measurements of an outcome) and thus we have to accept a certain degree of unbalancedness as randomization will not perfectly produce identical averages of the covariates in both groups.

If we only had to compare two time points within one group we would use the t-test for independent samples in experiment 1 and the t-test for dependent samples in experiment 2.

But now the situation is much more complicated...

The first analysis model for both experiments (I, II)

$$y_{ikt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ikt}$$

Y_{ikt} = observation in group i of animal k at timepoint j

μ : Total mean

α_i : Deviation of group i from total mean ($\alpha_1 + \alpha_2 = 0$)

β_j : Deviation of timepoint j from total mean ($\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$)

γ_{ij} : Interaction of group i at time point j

$$(Y_{11} + Y_{21} = 0, \dots, Y_{14} + Y_{42} = 0, Y_{11} + \dots + Y_{14} = 0, Y_{21} + \dots + Y_{24} = 0)$$

ε_{ijk} : normally distributed, mean 0, constant variance σ^2 and ???

In Experiment 1: independent (all measurements from different animals),

In Experiment 2: partially dependent (if measurements are from the same animal)

this holds exactly if i and k are identical

The second analysis model for both experiments (II, IV)

$$y_{ikt} = \mu + \alpha_i + \beta \times t + \gamma_i \times t + \varepsilon_{ikt}$$

y_{ikt} = observation in group i of animal k at time t

μ : Total mean

α_i : Deviation of group i from total mean ($\alpha_2 = -\alpha_1$)

β : average slope in the linear dependency of y on time

γ_i : Deviation of slope in group i from average slope ($\gamma_2 = -\gamma_1$)

t : time point (continuous!)

ε_{ikt} : normally distributed, mean 0, constant variance σ^2 and ???

In Experiment 1: independent (all measurements from different animals),

In Experiment 2: partially dependent (if measurements are from the same animal)

this holds exactly if for two measurements i and k are identical and only t differs.

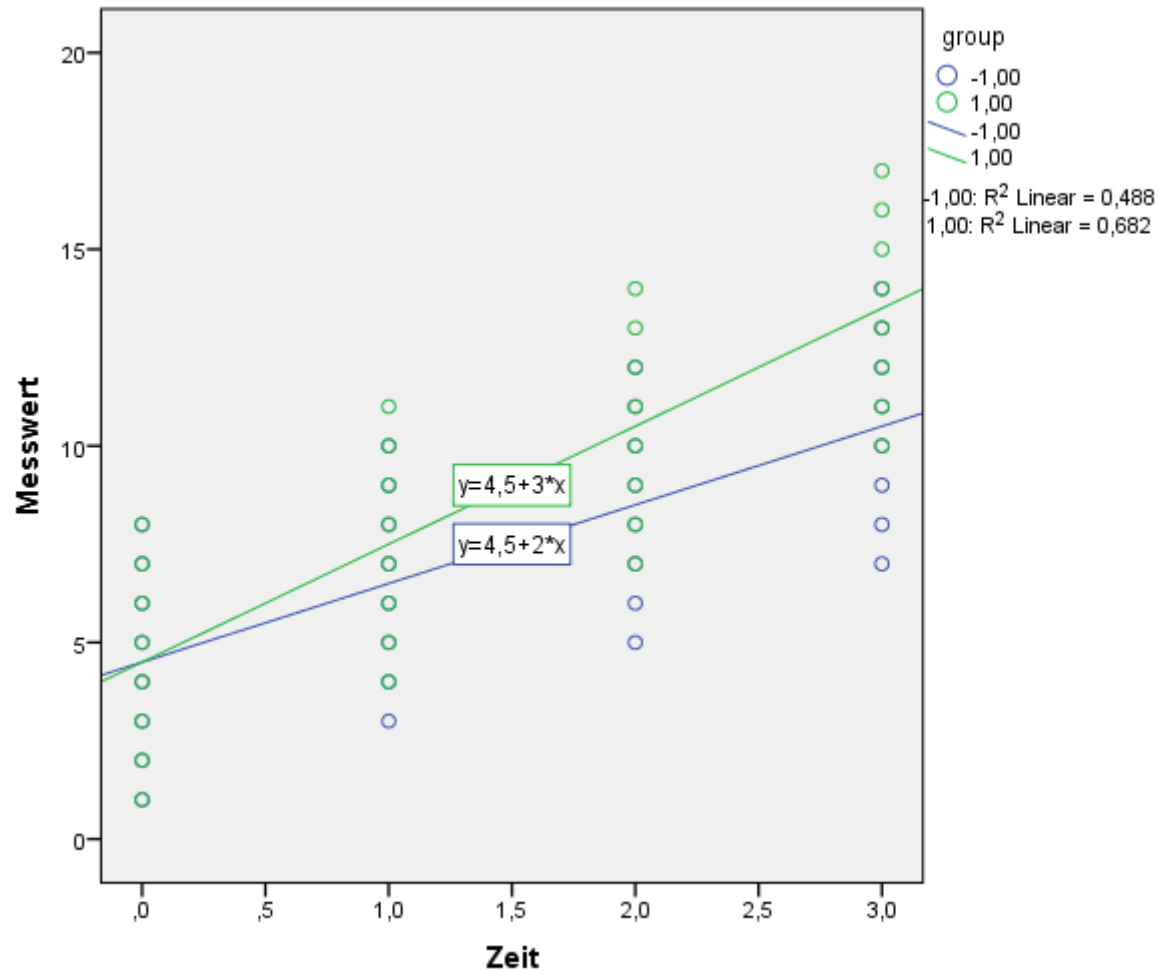
The data **group 1** – each value from a different animal (I, III)

Group	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Time	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Value	1	2	3	4	5	6	7	8	4	5	6	7	8	9	10	11
Group	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Time	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
Value	7	8	9	10	11	12	13	14	10	11	12	13	14	15	16	17

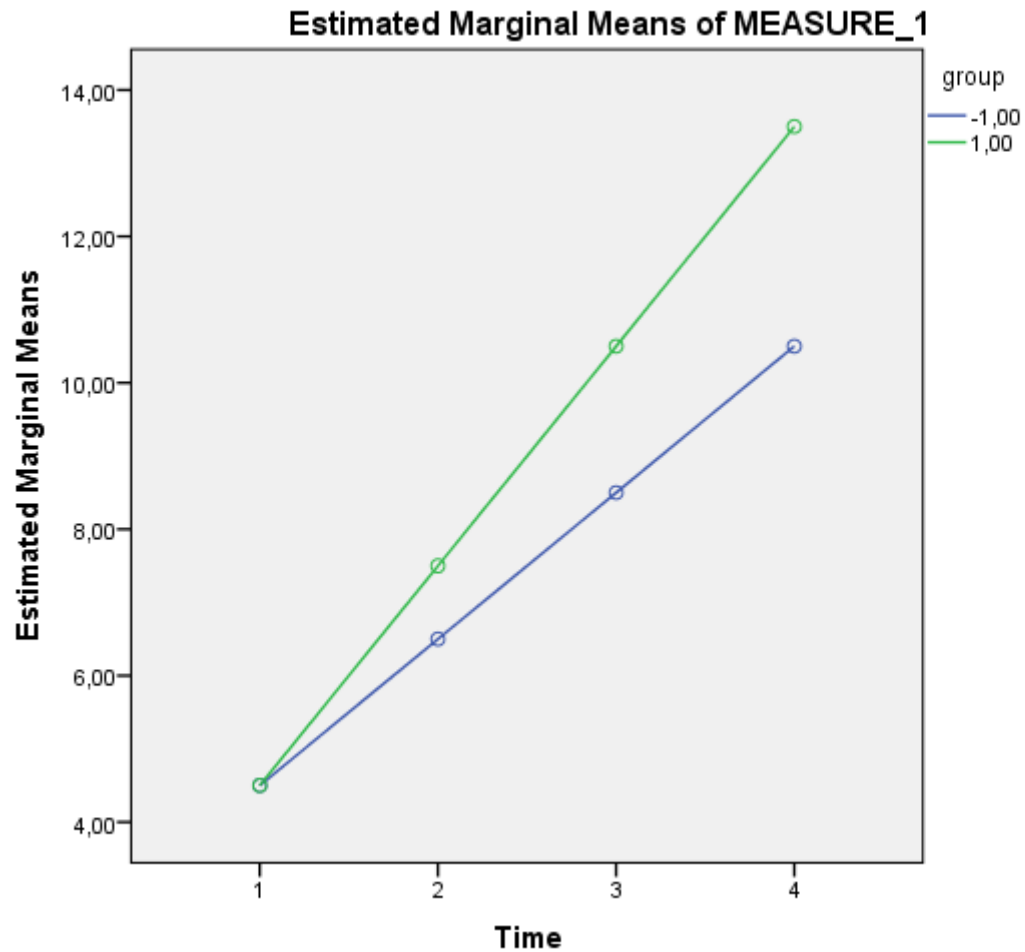
The data **group 2** – each value from a different animal (I, III)

Group	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Time	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Value	1	2	3	4	5	6	7	8	3	4	5	6	7	8	9	10
Group	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Time	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
Value	5	6	7	8	9	10	11	12	7	8	9	10	11	12	13	14

Individual data – each value from a different animal (I, III)



Averaged data – each value from a different animal



Note

As you have seen on the preceding sheet, to make calculations a bit more easy, I have constructed the data in a way, that the effect of time is perfectly linear in both groups if we consider the 8 means (two groups, four time points).

We will see, that we loose power, if we model time as factor instead of a continuous covariate.

The assumption of perfect linearity in the data of course is not realistic in real experiments. But mostly we will accept linearity if the shape is at least monotone and not „obviously“ quadratic.

The analysis with independent data (each value from a different animal), time factor (I)

Tests of Between-Subjects Effects

Dependent Variable: Messwert

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	556.000 ^a	7	79.429	13.238	0.000
Intercept	4356.000	1	4356.000	726.000	0.000
Gruppe	36.000	1	36.000	6.000	0.017
Zeit	500.000	3	166.667	27.778	0.000
Gruppe * Zeit	20.000	3	6.667	1.111	0.352
Error	336.000	56	6.000		
Total	5248.000	64			
Corrected Total	892.000	63			

a. R Squared = 0.623 (Adjusted R Squared = 0.576)

The analysis with independent data (each value from a different animal), time covariate (III)

Tests of Between-Subjects Effects

Dependent Variable: value

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	556,000 ^a	3	185,333	33,095	,000	,623
Intercept	4356,000	1	4356,000	777,857	,000	,928
time	500,000	1	500,000	89,286	,000	,598
group	36,000	1	36,000	6,429	,014	,097
group * time	20,000	1	20,000	3,571	,064	,056
Error	336,000	60	5,600			
Total	5248,000	64				
Corrected Total	892,000	63				

a. R Squared = ,623 (Adjusted R Squared = ,604)

The data in the second experiment (II, IV)

We use identical data for both experiments to demonstrate the differences of the analysis.

However, we now don't assume perfect linearity of the time dependency for each animal.

The data group 1 – repeated measurements in identical animals

(II, IV)

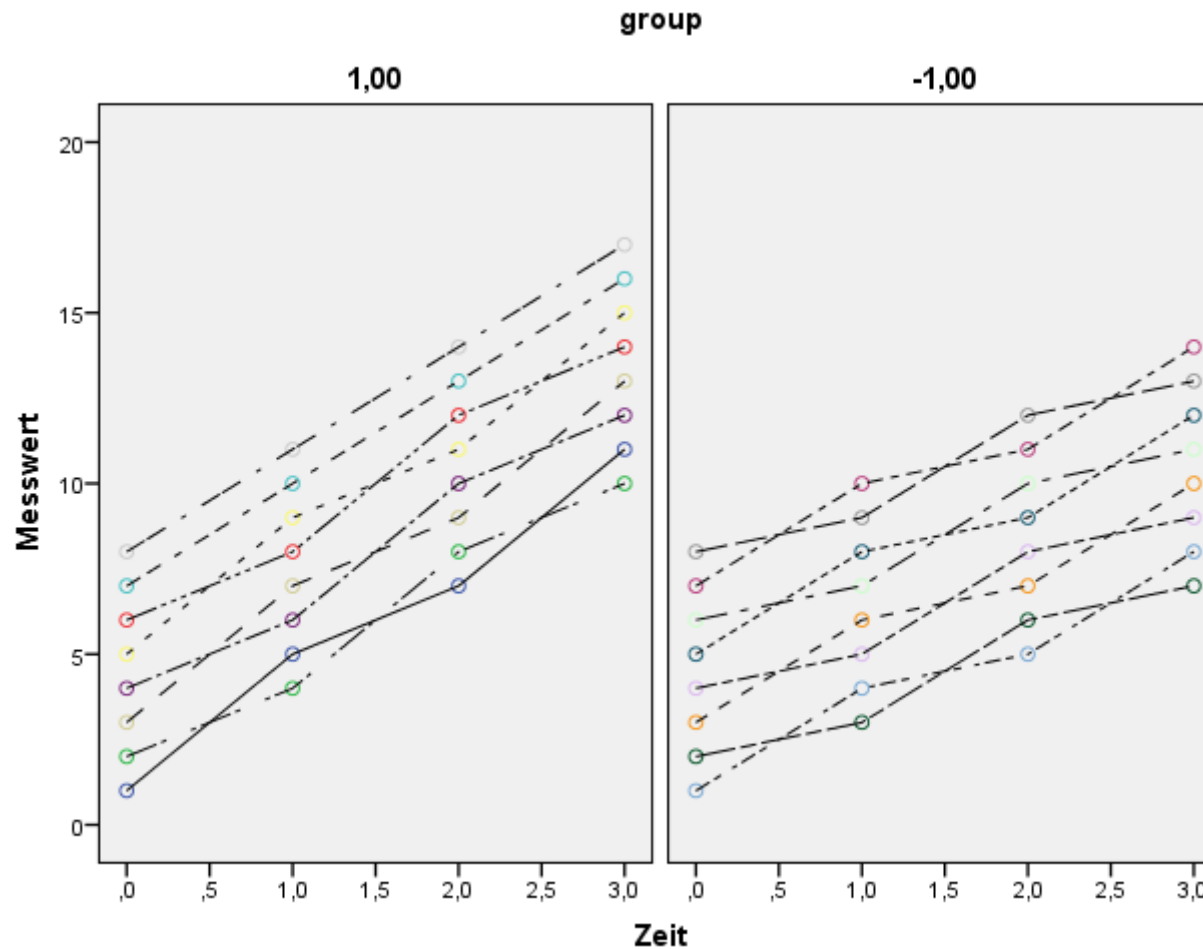
Group	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Time	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Animal	1	2	3	4	5	6	7	8	2	1	4	3	6	5	7	8
Value	1	2	3	4	5	6	7	8	4	5	6	7	8	9	10	11
Group	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Time	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
animal	1	2	3	4	5	6	7	8	2	1	4	3	6	5	7	8
Value	7	8	9	10	11	12	13	14	10	11	12	13	14	15	16	17

The data group 2 – repeated measurements in identical animals

(II, IV)

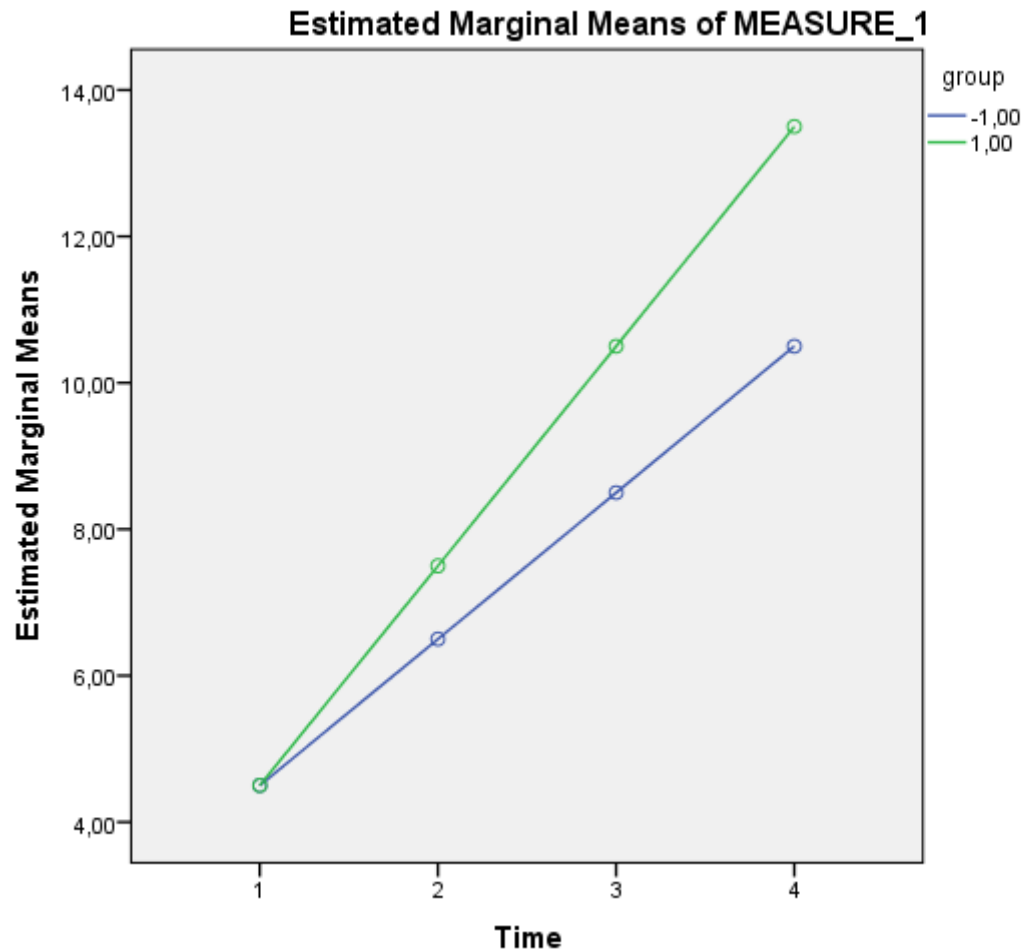
Group	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Time	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Animal	9	10	11	12	13	14	15	16	10	9	12	11	14	13	16	15
Value	1	2	3	4	5	6	7	8	3	4	5	6	7	8	9	10
Group	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Time	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3
animal	9	10	11	12	13	14	15	16	10	9	12	11	14	13	16	15
Value	5	6	7	8	9	10	11	12	7	8	9	10	11	12	13	14

Individual data – repeated measurements (II, IV)



Note: This figure is different to situation I, III.

Averaged data – repeated measures (II, IV)



Note: This figure is identical to situation I, III.

The analysis with dependent data

Comparison between subjects (II, IV)

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4356,000	1	4356,000	189,391	,000
group	36,000	1	36,000	1,565	,231
Error	322,000	14	23,000		

The analysis with dependent data

Comparison between subjects (II, IV)

The sum of squares and the degrees of freedom for **group** are identical to the analysis with independent data.

The sum of squares for the **error term** changes slightly from 336 to 322, the degrees of freedom change dramatically from 60 to 14.

The sum of squares of error is equal to

$$SSQ_{\text{error independent}} - SSQ_{\text{error(time),dependent}} = 336 - 14 = 322$$

The degrees of freedom of error is equal to

$$\# \text{patients} - 1(\text{df groups}) - 1(\text{grand mean}) = 16 - 2 = 14$$

The **total** sum of squares changes from 892 to 358, the degrees of freedom from 63 to 15.

The remaining sum of squares (534) and degrees of freedom (48) will be used in the within subjects analysis (see following sheets).

The analysis with dependent data

Comparisons **within** subjects (II)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Sphericity Assumed	500,000	3	166,667	500,000	,000
	Greenhouse-Geisser	500,000	1,000	500,000	500,000	,000
	Huynh-Feldt	500,000	1,077	464,286	500,000	,000
	Lower-bound	500,000	1,000	500,000	500,000	,000
Time * group	Sphericity Assumed	20,000	3	6,667	20,000	,000
	Greenhouse-Geisser	20,000	1,000	20,000	20,000	,001
	Huynh-Feldt	20,000	1,077	18,571	20,000	,000
	Lower-bound	20,000	1,000	20,000	20,000	,001
Error(Time)	Sphericity Assumed	14,000	42	,333		
	Greenhouse-Geisser	14,000	14,000	1,000		
	Huynh-Feldt	14,000	15,077	,929		
	Lower-bound	14,000	14,000	1,000		

The analysis with dependent data (II)

Remarks

$$SSQ_{\text{total}} = 500 + 20 + 14 = 534 \quad (= 892 - 358)$$

$$Df_{\text{total}} = 3 + 3 + 42 = 48 \quad (= 63 - 15)$$

The SSQ_{total} in the within analysis is calculated by the formula:

$$\sum_{g=1}^2 \sum_{k=1}^8 \sum_{t=1}^4 (y_{ikt} - \bar{y}_{ik})^2$$

Overall we have 64 measurements and we estimate 16 parameters (for each subject the average over four time points).

We have – as in the independent data case - 3 degrees of freedom for the four time points and 3 degrees of freedom for the interaction of group * time. The remaining $64 - 16 - 3 - 3 = 64 - 22 = 42$ is the error degrees of freedom in the output.

SSQ and DF for independent and dependent data modelling time as factor with three DF (I, II)

Data	Independent		Dependent					
	Between		Within		Between		Total	
	SSQ	DF	SSQ	DF	SSQ	DF	SSQ	DF
Time	500	3	500	3	---	---		
Group	36	1	---	---	36	1		
Interaction	20	3	20	3	---	---		
Model	556	7	520	6	36	1	556	7
Error	336	56	14	42	322	14	336	56
Total	892	63	534	48	358	15	892	63

The analysis with dependent data

Remarks (II)

The four different versions of the output

- sphericity assumed

- Greenhouse-Geisser

- Huyn-Feldt

- Lower Bound

provide different correction for a specific property of the multivariate distribution of data. This property is called the sphericity assumption. It is fulfilled e.g. if not only the variance of measurements at single time points but also the covariances between pairs of time points are equal.

In general this assumption is not fulfilled. It is then recommended to choose the Greenhouse-Geisser correction.

The analysis with dependent data

Contrasts within subjects (IV)

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Linear	500,000	1	500,000	2500,000	,000
	Quadratic	,000	1	,000	.	.
	Cubic	,000	1	,000	,000	1,000
Time * group	Linear	20,000	1	20,000	100,000	,000
	Quadratic	,000	1	,000	.	.
	Cubic	,000	1	,000	,000	1,000
Error(Time)	Linear	2,800	14	,200		
	Quadratic	,000	14	,000		
	Cubic	11,200	14	,800		

The analysis with dependent data

Contrasts within subjects

What is the difference between the test for time ($df = 3$) and the test of the linear contrast for time ($df = 1$)?

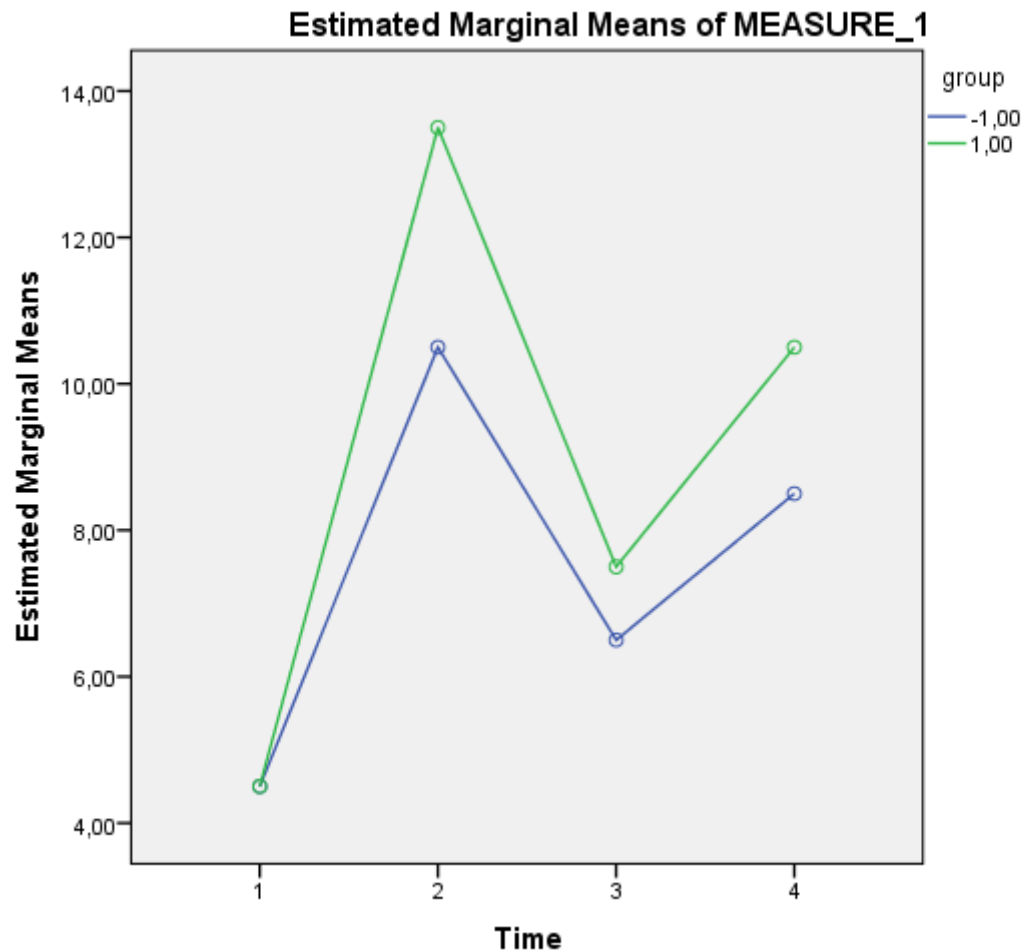
The linear contrast is the analogue approach as the analysis of covariance. We assume a linear dependency of measurement values on time.

The test with $df = 3$ treats time as a factor. Thus, if we permute time points (see next sheet) the result for time ($df = 3$) will not change.

However, the result for time as a linear covariate will not be significant anymore.

The analysis with dependent data

Permutation of time points



The analysis with independent data – permuted time points (time is factor, $df = 3$)

Tests of Between-Subjects Effects

Dependent Variable: Messwert

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	556.000 ^a	7	79.429	13.238	0.000
Intercept	4356.000	1	4356.000	726.000	0.000
Gruppe	36.000	1	36.000	6.000	0.017
zeit_perm	500.000	3	166.667	27.778	0.000
Gruppe * zeit_perm	20.000	3	6.667	1.111	0.352
Error	336.000	56	6.000		
Total	5248.000	64			
Corrected Total	892.000	63			

a. R Squared = 0.623 (Adjusted R Squared = 0.576)

Nothing has changed compared to the original time points

The analysis with independent data – permuted time points (time is covariate, $df = 1$)

Tests of Between-Subjects Effects

Dependent Variable: Messwert

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	119.200 ^a	3	39.733	3.085	0.034
Intercept	1041.429	1	1041.429	80.856	0.000
Gruppe	4.629	1	4.629	0.359	0.551
zeit_perm	80.000	1	80.000	6.211	0.015
Gruppe * zeit_perm	3.200	1	3.200	0.248	0.620
Error	772.800	60	12.880		
Total	5248.000	64			
Corrected Total	892.000	63			

a. R Squared = 0.134 (Adjusted R Squared = 0.090)

Time is due to a slight average increase still significant, but the sum of squares for both, main effect and interaction with groups have been reduced dramatically compared to the original data with perfect linear trend.

The analysis with dependent data – permuted time points

Comparison between subjects

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4356,000	1	4356,000	189,391	,000
group	36,000	1	36,000	1,565	,231
Error	322,000	14	23,000		

The analysis with dependent data

Comparisons within subjects

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Sphericity Assumed	500,000	3	166,667	500,000	,000
	Greenhouse-Geisser	500,000	1,000	500,000	500,000	,000
	Huynh-Feldt	500,000	1,077	464,286	500,000	,000
	Lower-bound	500,000	1,000	500,000	500,000	,000
Time * group	Sphericity Assumed	20,000	3	6,667	20,000	,000
	Greenhouse-Geisser	20,000	1,000	20,000	20,000	,001
	Huynh-Feldt	20,000	1,077	18,571	20,000	,000
	Lower-bound	20,000	1,000	20,000	20,000	,001
Error(Time)	Sphericity Assumed	14,000	42	,333		
	Greenhouse-Geisser	14,000	14,000	1,000		
	Huynh-Feldt	14,000	15,077	,929		
	Lower-bound	14,000	14,000	1,000		

The analysis with dependent data

Contrasts within subjects

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Linear	80,000	1	80,000	.	.
	Quadratic	100,000	1	100,000	100,000	,000
	Cubic	320,000	1	320,000	.	.
Time * group	Linear	3,200	1	3,200	.	.
	Quadratic	4,000	1	4,000	4,000	,065
	Cubic	12,800	1	12,800	.	.
Error(Time)	Linear	,000	14	,000		
	Quadratic	14,000	14	1,000		
	Cubic	,000	14	,000		

Note:

The sum of squares for time added up over linear quadratic and cubic contrast reproduce exactly the sum of squares of time in the original data ($80+100+320=500$) and the same holds for the interaction with group ($3.2+4+12.8=20$).

Further Approaches for the dependent data situation: GEE

In the GEE (generalized estimating equations) approach we estimate the residuals (observed minus predicted values) in a regression model and determine the correlations of these residuals within subjects.

Then, a standard error is computed which uses all observations (in the example: $n=64$) but corrects this estimate according to the observed correlation.

This approach is very powerful, but it requires knowledge of multiple regression with the generalized linear models (e.g. logistic regression, Poisson regression) which is beyond the scope of this lecture.

Further Approaches: Linear mixed models

In the linear mixed model approach we assume that each subject has its own parameters, e.g. intercept and/or slope if dependency of measurements is linear.

However, we assume a normal distribution of these parameters with mean zero and unknown variance. We speak e.g. of the „random intercept“ model or the „random slope“ model.

We make inferences about the variance of these distributions in our data.

Both approaches, **linear mixed models** and **GEE**, are now mostly used for **clinical and epidemiological** studies.

The (traditional) **repeated measures** approach which was presented in this lecture is used mostly in **planned experiments**.

Last but not least: a specific analysis strategy

$$y_{ikt} = \mu + \alpha_i + \beta \times t + \gamma_i \times t + \varepsilon_{ikt}$$

In many experiments we

- want to compare groups at single time points
- expect the group difference to increase over time (this is e.g. a consequence of difference slopes if $\alpha_1 = \alpha_2 = 0$).
- would like to know from which time point on there is a significant difference.

We obtain measurements at the $T+1$ time points $t=0, t=1, \dots, t=T$.

A specific analysis strategy

$$y_{ikt} = \mu + \alpha_i + \beta \times t + \gamma_i \times t + \varepsilon_{ikt}$$

In this situation it is „allowed“ to use the following strategy:

- (1) Make a pairwise comparison between both groups for the last Timepoint $t = T$.
- (2) If this comparison is not significant $\rightarrow$ stop
- (3) If this comparison is significant $\rightarrow$ proceed to time point $t = T-1$ and repeat the test of significance for this time point **without** any adjustment for multiple testing.
- (4) Proceed till $t = 0$.

This is a general statistical principle called hierarchical testing. It is „risky“ if you are not sure whether the group difference really increases.