Vergleich von Verlaufskurven
Standard–Auswertung mit SPSS–GLM

H.Hecker

1. Allgemeine Beschreibung der Fragestellung

Zur Auswertung von Verlaufskurven bietet SPSS (analog auch SAS) die Prozedur GLM (General Linear Models) an mit der Option "Repeated Measurement". Im folgenden werden die Modellannahmen beschrieben, die dieser Prozedur zugrundeliegen, sowie die Nullhypothesen, die mit ihr überprüft werden können. Dabei wird von der einfachen Situation ausgegangen, daß nur *ein* "Between"–Faktor (die Gruppenvariable) und *ein* "Within"–Faktor (die Zeit, realisiert durch Meßwiederholungen) vorliegt.

2. Voraussetzungen und Bezeichnungen

- Es wird angenommen, daß der Meßparameter Y für jede Beobachtungseinheit zu J verschiedenen Zeitpunkten (oder Bedingungen) gemessen wird. Die Beobachtungen seien als die Variablen $Y_1, Y_2, \ldots, Y_J$ gekennzeichnet. (Es handelt sich also um "verbundene Stichproben").
- Die Stichprobe setzt sich aus den Beobachtungen von K verschiedenen Gruppen zusammen.
- Die Variablen $Y_1, Y_2, \ldots, Y_J$ seien in jeder Gruppe (angenähert) normalverteilt mit gleichen Varianzen und Kovarianzen *zwischen den Gruppen.*
- Zusätzliche Voraussetzungen über die Kovarianzmatrix werden zunächst nicht gemacht, müssen aber bei einigen Tests bedacht werden.
- Es interessieren die Erwartungswerte von Y :

$$\mu_{kj} = E(Y) \text{ in Gruppe } k \text{ zum Zeitpunkt j } (k = 1, \ldots, K; \quad j = 1, \ldots, J)$$

- Zusammensetzung dieser Erwartungswerte als die Summe verschiedener "Effekte" (Allgemeines lineares Modell Ω):

$$\Omega: \ \mu_{kj} = \mu + \alpha_k + \beta_j + \gamma_{kj}$$

  mit der Interpretation

$$\begin{aligned} \mu &= \text{Allgemeines Mittel ("Constant", "Intercept")} \\ \alpha_k &= \text{Gruppen-Effekt (Effekt der Gruppe k)} \\ \beta_j &= \text{Zeit-Effekt (Effekt des Zeitpunktes j)} \\ \gamma_{kj} &= \text{Wechselwirkung Gruppe} \times \text{Zeit} \end{aligned}$$

  Die Gruppen-Effekte werden auch "*between-subject effect*" genannt und die Zeiteffekte entsprechend "*within-subject effect*".
  Damit die Darstellung eindeutig ist, werden die folgenden Nebenbedingungen eingeführt:

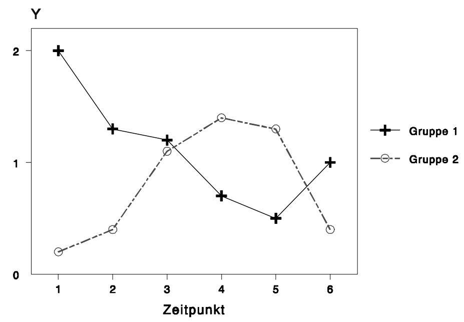
$$\begin{aligned} \sum_{k=1}^{K} \alpha_k &= 0 \\ \sum_{j=1}^{J} \beta_j &= 0 \\ \sum_{k=1}^{K} \gamma_{kj} &= 0 \quad \text{für alle j, und } \sum_{j=1}^{J} \gamma_{kj} = 0 \quad \text{für alle k} \end{aligned}$$

- Tabellarische Darstellung (K=2; p=3):

Gr.	Zeitpunkt 1	2	3	$\sum$
1	$\mu_{11} = \mu + \alpha_1 + \beta_1 + \gamma_{11}$	$\mu_{12} = \mu + \alpha_1 + \beta_2 + \gamma_{12}$	$\mu_{13} = \mu + \alpha_1 + \beta_3 + \gamma_{13}$	$\mu_{1.} = 3\mu + 3\alpha_1$
2	$\mu_{21} = \mu + \alpha_2 + \beta_1 + \gamma_{21}$	$\mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22}$	$\mu_{12} = \mu + \alpha_2 + \beta_3 + \gamma_{23}$	$\mu_{2.} = 3\mu + 3\alpha_2$
$\sum$	$\mu_{.1} = 2\mu \quad + 2\beta_1$	$\mu_{.2} = 2\mu \quad + 2\beta_2$	$\mu_{.2} = 2\mu \quad + 2\beta_3$	$\mu_{..} = 6\mu$

(Tabelle 1)

- Im allgemeinen ("vollständigen") Modell Ω werden keine weiteren Einschränkungen über die Parameter α_k, β_j und γ_{kj} gemacht. Dann kann jede Konstellation von Erwartungswerten μ_{kj} möglich sein:



- Im Modell ohne Wechselwirkung zwischen Gruppe und Zeit, Ω_γ, sind die Zeitverläufe "parallel":

$$\Omega_\gamma : \ \mu_{kj} = \mu + \alpha_k + \beta_j$$

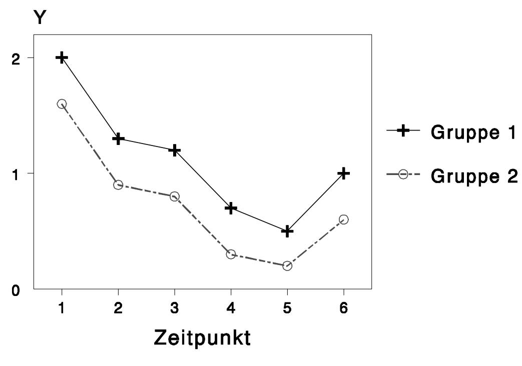
mit der Interpretation

$$\begin{aligned} \mu &= \text{Allgemeines Mittel ("Constant", "Intercept")} \\ \alpha_k &= \text{Gruppen-Effekt (Effekt der Gruppe k)} \\ \beta_j &= \text{Zeit-Effekt (Effekt des Zeitpunktes j)} \end{aligned}$$

und den Nebenbedingungen

$$\sum_{k=1}^{K} \alpha_k = 0, \qquad \sum_{j=1}^{J} \beta_j = 0$$

Graphisch Darstellung:



Ist diese Modellannahme gerechtfertigt, sind die verbleibenden Parameter (insbesondere Gruppeneffekt und Zeiteffekt) sehr viel einfacher zu interpretieren als im allgemeinen Modell:

> *Der Gruppeneffekt* (die Veränderung der Erwartungswertes z.B. beim Wechsel von Gruppe 1 zu Gruppe 2) *ist für jeden Zeitpunkt derselbe,*

und analog:

> *Der Zeiteffekt* (die Veränderung der Erwartungswertes z.B. beim Wechsel von Zeitpunkt 1 zu Zeitpunkt 2) *ist innerhalb jeder Gruppe identisch.*

Im allgemein Modell mit Wechselwirkungen sind solche Regelmäßigkeiten nicht gegeben. Der *Gruppeneffekt* beispielswiese zeigt sich dort nur *im Mittel über alle Zeitpunkte* , (siehe dazu die letzte Spalte der Tabelle 1), und analog der *Zeiteffekt* nur *im Mittel über alle Gruppen* (siehe dazu die letzte Zeile der Tabelle 1).

3. Hypothesen und Signifikanztests

Es werden im folgenden diejenigen Nullhypothesen beschrieben, die mit dem Programm GLM getestet werden. Dabei ist darauf zu achten, welche der beiden Modellannahmen (Ω oder Ω_γ) zugrundegelegt werden. Bei der Anwendung von Signifikanztests ist außerdem darauf zu achten, ob zusätzliche Annahmen über die Kovarianzstruktur gemacht werden.

3.1. Signifikanztests ohne Zusatzannahmen

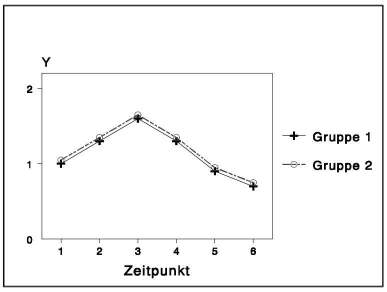
1. Unter dem allgemeinen linearen Modell Ω wird die folgende zusammengesetzte Nullhypothese getestet:
   Kein Gruppeneffekt: weder Haupteffekt (α) noch Wechselwirkungseffekte zum Zeitverlauf (γ):

$$H_{\alpha\gamma} \;:\; \alpha_k = 0,\ \gamma_{kj} = 0 \quad (k = 1, \ldots, K, \quad j = 1, \ldots, J) \qquad \text{Gleichwertig:}$$

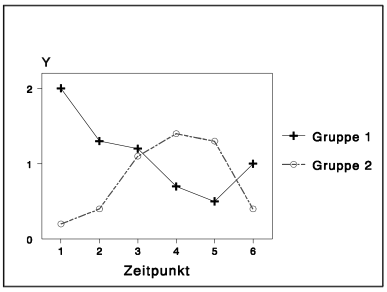
$$H_{\alpha\gamma} \;:\; \mu_{kj} = \mu_{k'j} \qquad \text{für alle } k \neq k' \text{ , } j = 1, \ldots, p$$

Graphische Darstellung: Test von

$H_{\alpha\gamma}$:



versus Ω :



Der Test dieser Nullhypothese wird durch die folgende SPSS–Prozedur erzeugt:

```
GLM
      Y1 to Y6 BY gruppe
      /METHOD = SSTYPE(3)
      /DESIGN = gruppe.
```

Das Ergebnis:

Multivariate Tests[d]

Effekt		Wert	F	Hypothese df	Fehler df	Signifikanz
Intercept	Pillai-Spur	.987	715.923[b]	6.000	55.000	.000
	Wilks-Lambda	.013	715.923[b]	6.000	55.000	.000
	Hotelling-Spur	78.101	715.923[b]	6.000	55.000	.000
	Größte charakteristische Wurzel nach Roy	78.101	715.923[b]	6.000	55.000	.000
GRUPPE	Pillai-Spur	.584	3.845	12.000	112.000	.000
	Wilks-Lambda	.439	4.661[b]	12.000	110.000	.000
	Hotelling-Spur	1.223	5.503	12.000	108.000	.000
	Größte charakteristische Wurzel nach Roy	1.178	10.997[c]	6.000	56.000	.000

b. Exakte Statistik

c. Die Statistik ist eine Obergrenze auf F, die eine Untergrenze auf dem Signifikanzniveau ergibt.

d. Design: Intercept+GRUPPE

Tabelle 2

Es ist unter dem Effekt "Gruppe" bei Wilks-Lambda abzulesen (hier: Wert= .439, Signifikanz = .000).
Bei diesem Test wird über die Struktur der Kovarianzmatrix keine Voraussetzung gemacht (grundsätzlich müssen aber die Kovarianzmatrizen in den Gruppen identisch sein).

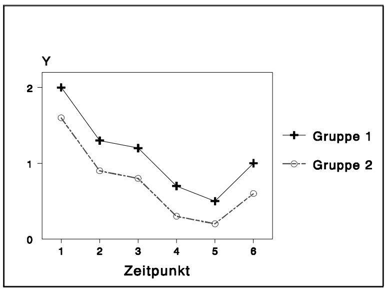
2. Keine Wechselwirkung zwischen Gruppe und Zeitverlauf (γ) :

$$H_\gamma \quad : \quad \gamma_{kj} = 0 \quad (k = 1, \ldots, K, \quad j = 1, \ldots, J) \quad \text{Gleichwertig:}$$

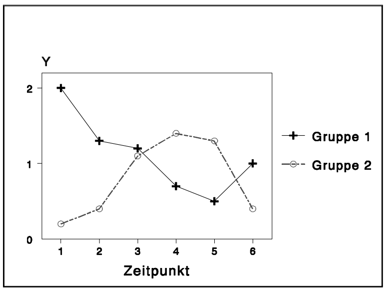
$$H_\gamma \quad : \quad \mu_{kj} - \mu_{k'j} = \text{ konstant für alle Zeitpunkte j } \quad (k, k' = 1, \ldots, K)$$

Graphische Darstellung: Test von

H_γ :



versus Ω :



Dies ist also gleichzeitig ein Test darüber, ob das Modell ohne Wechselwirkung, Ω_γ, als gültig angesehen werden kann.
Diese Hypothese H_γ wird unter dem allgemeinen Modell Ω getestet. Die *Alternative* zu H_γ ist also das allgemeine Modell Ω mit beliebigen Parameterwerten. Ist das Testergebnis "signifikant", so ist dies so zu interpretieren, daß die Kurven in den K Gruppen *nicht alle parallel zueinander verlaufen*.
H_γ kann nur mit der Prozedur für Meßwiederholungen behandelt werden. Der Aufruf dazu lautet:

```
GLM
      Y1 to Y6 BY gruppe
      /WSFACTOR = zeit 6 Polynomial
      /MEASURE = Y
```

```
/METHOD = SSTYPE(3)
/PLOT = PROFILE( zeit*gruppe )
/WSDESIGN = zeit
/DESIGN = gruppe.
```

Das Ergebnis:

Multivariate Tests[d]

Effekt		Wert	F	Hypothese df	Fehler df	Signifikanz
ZEIT	Pillai-Spur	.301	4.812[b]	5.000	56.000	.001
	Wilks-Lambda	.699	4.812[b]	5.000	56.000	.001
	Hotelling-Spur	.430	4.812[b]	5.000	56.000	.001
	Größte charakteristische Wurzel nach Roy	.430	4.812[b]	5.000	56.000	.001
ZEIT * GRUPPE	Pillai-Spur	.507	3.869	10.000	114.000	.000
	Wilks-Lambda	.507	4.528[b]	10.000	112.000	.000
	Hotelling-Spur	.945	5.195	10.000	110.000	.000
	Größte charakteristische Wurzel nach Roy	.915	10.428[c]	5.000	57.000	.000

b. Exakte Statistik

c. Die Statistik ist eine Obergrenze auf F, die eine Untergrenze auf dem Signifikanzniveau ergibt.

d. Design: Intercept+GRUPPE
Innersubjekt-Design: ZEIT

Tabelle 3

Es ist als Effekt ZEIT*GRUPPE unter Wilks Lambda abzulesen (hier: Wert= .507, Signifikanz = .000)

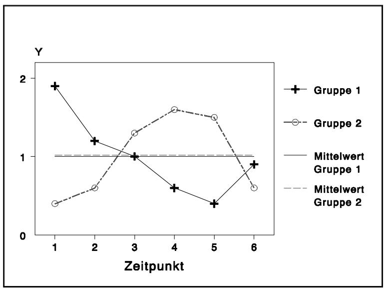
3. Kein Gruppen-Haupteffket (α) :

$$H_\alpha \quad : \quad \alpha_k = 0, \ (k = 1, \dots, K) \qquad \text{Gleichwertig:}$$

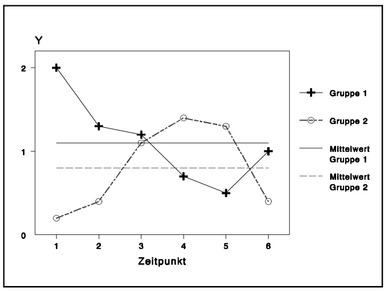
$$H_\alpha \quad : \quad \sum_{j=1}^{J} \mu_{kj} = \sum_{j=1}^{J} \mu_{k'j} \ \text{für alle } k \neq k'$$

Je nachdem, ob man von dem allgemeinen Modell Ω oder dem Modell ohne Wechselwirkung, Ω_γ, ausgehen kann, wird getestet:

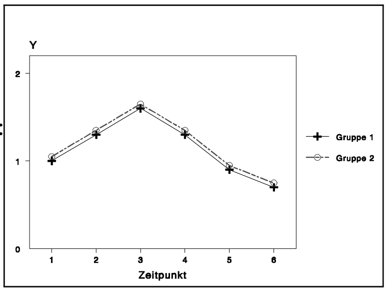
H_α :

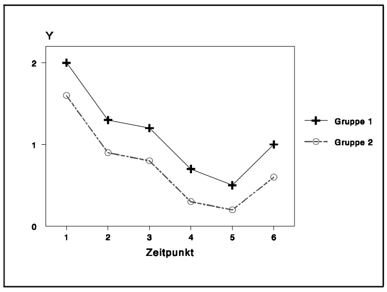


versus Ω :



oder

H_α : 

versus Ω_γ : 

Der Test hierzu ist wiederum nur in der GLM-Prozedur mit Meßwiederholung zu erhalten. (Aufruf siehe oben). Das Ergebnis:

Tests der Zwischensubjekteffekte

Maß: Y

Transformierte Variable: Mittel

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
Intercept	6697086.233	1	6697086	3588.395	.000
GRUPPE	37931.339	2	18965.669	10.162	.000
Fehler	111979.095	60	1866.318		

Tabelle 4

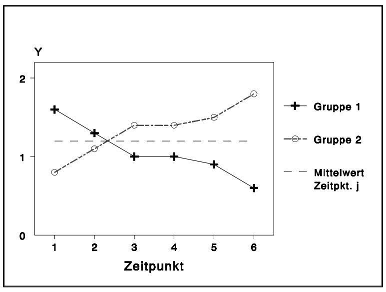
Es ist unter der Zeile "GRUPPE" abzulesen (hier: F=10.162; Signifikanz = .000). Der Test wird einfach dadurch gebildet, daß fallweise aus den Messungen aller Zeitpunkten der Mittelwert berechnet wird und anschließend mit diesen Verlaufsmittelwerten eine einfache Varianzanalyse über die Gruppen durchgeführt wird. Daraus wird auch deutlich, daß über die Kovarianzstruktur hierbei keine Zusatzvoraussetzungen gemacht werden müssen.

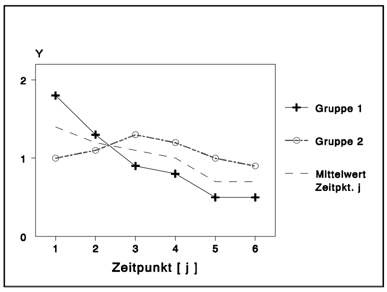
4. Kein Zeiteffekt (β) :

$$H_\beta \quad : \quad \beta_j = 0 \; (j = 1, \ldots, J) \qquad \text{Gleichwertig:}$$

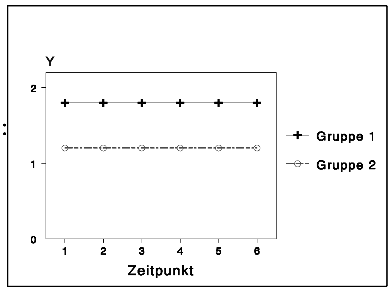
$$H_\beta \quad : \quad \mu_{kj} = \mu_{kj'} \qquad \text{für alle } k = 1, \ldots, K, \quad j \neq j'$$

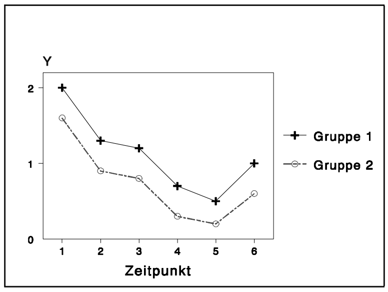
Je nachdem, ob man von dem allgemeinen Modell Ω oder dem Modell ohne Wechselwirkung, Ω_γ, ausgehen kann, wird getestet:

H_β : 

versus Ω: 

oder

H_β : 

versus Ω_γ: 

Hieraus wird deutlich, daß Abweichungen von der Nullhypothese H_β, also eine genereller Zeiteffekt β_j, nur unter dem Modell Ω_γ ohne Wechselwirkungen gut zu interpretieren ist.
Der Test hierzu ist wiederum nur in der Prozedur mit Meßwiederholungen zu erhalten. Im vorliegenden Beispiel ist das Ergebnis als Effekt "ZEIT" in der Zeile mit Wilks Lambda dargestellt (Tabelle 3: Wilks Lambda = .699; Signifikanz = 0.001).

5. Wenn der Zeitverlauf *Differenzen* darstellt, z.B. jeweils die Differenz zum Ausgangswert, interessiert häufig auch, ob die Werte im Mittel gleich Null sind:

$$H_\mu \quad : \quad \mu = 0 \quad \text{Äquivalent:}$$

$$H_\mu \quad : \quad \sum_{k,j} \mu_{kj} = 0 \text{ (Allgemeines Mittel, "Intercept" = 0):}$$

Der Test hierzu wird in der GLM-Prozedur mit Meßwiederholung berechnet und findet sich hier in der Tabelle 4 in der Zeile "Intercept" (Ergebnis dort: F=3588.395; Signifikanz = 0.000).
Der in Tabelle 2 wiedergegebene Test für das "Intercept" überprüft hingegen im Rahmen der allgemeinen Multivariaten Varianzanalyse (MANOVA), ob "*für jeden Zeitpunkt* der Mittelwert über alle Gruppen gleich Null ist". In der Terminologie von Tabelle 1 bedeutet dies den Test der Hypothese:

$$H_{(\mu.)} : \sum_{k} \mu_{kj} = \mu_{.j} = 0 \qquad \text{für alle Zeitpunkte } j = 1, \ldots, J$$

Diese Nullhypothese ist also "strenger" als H_μ .
Im vorliegenden Beispiel findet man das Ergebnis in Tabelle 2, Zeile "Intercept". Unter "Wilks Lambda" ist dort der Wert 0.013 mit der Signifikanz .000 ausgedruckt.

3.2. Signifikanztests mit Zusatzvoraussetzungen über die Kovarianzstruktur

Im "gemischten Modell" macht man die Annahme, daß die Streuung der Daten in der Stichprobe im wesentlichen dadurch zu erklären ist, daß jeder Patient (allgemein: jede Beobachtungseinheit) dem gruppenspezifischen Mittelwertsverlauf einen (*für alle Zeitpunkte identischen*) "Patienteneffekt" überlagert. Dieser wird nur noch von Zufallsfehlern ("Meßfehlern") überlagert, die für jeden Einzelwert unabhängig von allen anderen Messungen sind.

Trifft diese Annahme zu, so hat dies zur Folge, daß die Kovarianzen zwischen je zwei Zeitpunkten alle identisch sind und auch die Varianzen sich zwischen den Zeitpunkten nicht unterscheiden. In diesem Fall kann man einige der genannten Hypothesen noch effektiver (mit größerer Testschärfe) testen als bisher beschrieben. Die Annahme wird überprüft durch den "Mauchly Test auf Sphärizität:"

Mauchly-Test auf Sphärizität[b]

Maß: Y

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Signifikanz	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
ZEIT	.231	85.156	14	.000	.615	.674	.200

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Kann verwendet werden, um die Freiheitsgrade für die gemittelten Signifikanztests anzupassen. Die korrigierten Tests werden in der Standardeinstellung in den Schichten der Tabelle für die Tests der Innersubjekteffekte angezeigt.

b. Design: Intercept+GRUPPE
Innersubjekt-Design: ZEIT

Tabelle 5

Im vorliegenden Fall ist die Annahme der Sphärizität nicht zu halten ("Signifikanz" P = 0.000). Es werden für eine solche Situation von SPSS noch alternative Auswertungen angeboten, die –wie das gemischte Modell– ebenfalls univariat ansetzen und die fehlende Sphärizität durch eine andere Berechnung der Freiheitsgrade approximativ ausgleichen. Dieses geschieht über einen Parameter "Epsilon", der nach drei verschiedenen Verfahren berechnet wird ("Greenhouse–Geisser", "Huyn–Feldt" und "Berechnung einer unteren Grenze"; siehe Tabelle 5). Die zugehörigen Signifikanztests (Tabellen 6a und 6b) beziehen sich inhaltlich auf dieselben Hypothesen wie die Tests in Tabelle 3 der Prozedur mit Meßwiederholungen. (Bis zur Version 7.5 wird von SPSS nur der Test unter Annahme der Sphärizität berechnet; die Tests nach Greenhouse–Geuiser etc. erscheinen erst ab Version 8.0.)

Tests der Innersubjekteffekte

Maß: Y

Sphärizität angenommen

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz
ZEIT	9844.656	5	1968.931	11.518	.000
ZEIT * GRUPPE	12802.915	10	1280.292	7.489	.000
Fehler(ZEIT)	51283.762	300	170.946		

Tabelle 6a

Tests of Within-Subjects Effects

Measure: Y

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
ZEIT	Sphericity Assumed	9844.656	5	1968.931	11.518	.000
	Greenhouse-Geisser	9844.656	3.076	3200.536	11.518	.000
	Huynh-Feldt	9844.656	3.369	2922.023	11.518	.000
	Lower-bound	9844.656	1.000	9844.656	11.518	.001
ZEIT * GRUPPE	Sphericity Assumed	12802.915	10	1280.292	7.489	.000
	Greenhouse-Geisser	12802.915	6.152	2081.139	7.489	.000
	Huynh-Feldt	12802.915	6.738	1900.036	7.489	.000
	Lower-bound	12802.915	2.000	6401.458	7.489	.001
Error(ZEIT)	Sphericity Assumed	51283.762	300	170.946		
	Greenhouse-Geisser	51283.762	184.556	277.876		
	Huynh-Feldt	51283.762	202.147	253.695		
	Lower-bound	51283.762	60.000	854.729		

Tabelle 6b