1 Grundproblem der α-Fehler Kumulierung

2 Bonferroni-Korrektur

Exkurs: α-Fehler Kumulierung in korrelativen Studien

3 A posteriori Tests

4 Durchführung der Tests in SPSS

5 A priori Tests

6 Trendtests

ONEWAY ANOVA

av

		SAQ Quadratsumme	*df* df	*MQ* Mittel der Quadrate	*F* F	*p* Signifikanz
zw	Zwischen den Gruppen	1402,625	3	467,542	6,754	,001
in	Innerhalb der Gruppen	1938,250	28	69,223		
	Gesamt	3340,875	31			

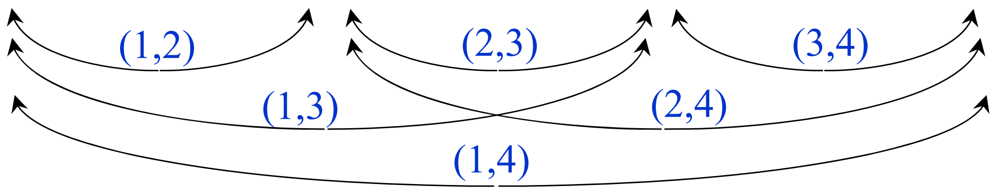
Die vier Lehrmethoden führen zu statistisch signifikanten Unterschieden in dem Wissen der Schüler, $F(3, 28) = 6.75$, $p < .05$.

Einfaktorielle Varianzanalyse: Interpretation

- Haben wir ein statistisch signifikantes Ergebnis in der einfaktoriellen Varianzanalyse erhalten, so wissen wir nur (bei einer Irrtumswahrscheinlichkeit von α), dass nicht alle Mittelwerte (in der Population) gleich sind. Aber: Welche unterscheiden sich jetzt? Bzw.: Unterscheiden sich die Gruppen wie postuliert?
- $\mathrm{H}_0\colon\ \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad (m = 4)$ vs. $\mathrm{H}_1\colon\ \mu_k \neq \mu_l$ für mindestens zwei Gruppen $k \neq l$

Lehrbuch konventionell	Lehrbuch verzweigt	Computer verzweigt	Computer Hyperlink
16	25	34	37
…	…	…	…
21	35	34	47

$\bar{x}_1 = 18.25$	$\bar{x}_2 = 30.00$	$\bar{x}_3 = 35.63$	$\bar{x}_4 = 32.88$



Mögliche Lösung: $m \cdot (m-1)/2$ t-Tests zwischen allen Paaren von Gruppen bestimmen.

Multiple Mittelwertsvergleiche: α-Fehler Kumulierung

- Aber: Geht man so vor, so ergibt sich ein Problem mit der Kontrolle des α-Fehlers, dass man auch als α-Fehler Kumulierung bezeichnet.

- Wenn die (vollständige, Overall-, Omnibus-) Nullhypothese gilt, also $\mu_1 = \mu_2 = \ldots = \mu_m$ ist, dann ist ...
  - bei <u>einem</u> paarweisen Test (z.B. bezüglich der Mittelwerts-unterschiede der beiden ersten Gruppen) die Wahrscheinlichkeit dessen H_0 fälschlicherweise zurückzuweisen gleich unserem gesetzten α. Wir bezeichnen den Fehler hier auch als Irrtumswahrscheinlichkeit pro Vergleich $\alpha_V = \alpha$.
  - bei <u>mehreren</u> paarweisen Tests die Wahrscheinlichkeit, mindestens einmal eine H_0 fälschlicherweise zurückzuweisen höher als das gesetzte α. Wir bezeichnen diesen Fehler auch als Irrtumswahrscheinlichkeit für eine Familie von Vergleichen $\alpha_F > \alpha$ (engl.: experiment- oder family-wise error rate)

Vergleich von Gruppen 1 und 2:

H_0: $\mu_1 = \mu_2$

Vergleich von allen Paaren von Gruppen:

[1] H_0: $\mu_1 = \mu_2$

[2] H_0: $\mu_1 = \mu_3$

[3] H_0: $\mu_1 = \mu_4$

[4] H_0: $\mu_2 = \mu_3$

[5] H_0: $\mu_2 = \mu_4$

[6] H_0: $\mu_3 = \mu_4$

Multiple Mittelwertsvergleiche: α-Fehler Kumulierung

- Wenn die vollständige Nullhypothese gilt, also $\mu_1 = \mu_2 = \ldots = \mu_m$ ist, dann ist beim ersten und beim zweiten (und allen weiteren) Tests die Wahrscheinlichkeit, sich korrekt für die Nullhypothese zu entscheiden, immer gleich $1-\alpha$, also z.B. $1-0.05=0.95$.

- Die Wahrscheinlichkeit, sich in zwei Tests richtig zu entscheiden, wenn sie unabhängig voneinander sind, beträgt $(1-\alpha)\cdot(1-\alpha)=(1-\alpha)^2=0.95\cdot 0.95=0.9025$. Entsprechend ist die Wahrscheinlichkeit, nicht beide Male richtig zu entscheiden, also mindestens einen Fehler zu machen, gleich $1-0.9025=0.0975$.

- Verallgemeinert ist bei einer Familie von q Tests die Wahrscheinlichkeit für mindestens einen α-Fehler

$$\alpha \leq \alpha_F = 1-(1-\alpha)^q$$

- Sind die $q \geq 2$ Tests nicht unabhängig voneinander, so gilt

$$\alpha < \alpha_F < 1-(1-\alpha)^q$$

- Es gilt die folgende (Bonferroni)-Ungleichung:

$$1-(1-\alpha)^q < q\cdot\alpha$$

m	q	α_F	$q\cdot\alpha$
3	3	0.14	0.15
4	6	0.26	0.30
5	10	0.40	0.50
10	45	0.90	2.25

für $\alpha = 0.05$

$\Rightarrow$ Die familienweise Fehlerrate α_F ist immer kleiner als $q\cdot\alpha$ (die Summe aller Einzel-α´s)

Multiple Mittelwertsvergleiche: Bonferroni-Korrektur

- Man kann sich nun die obige Beziehung zu Nutze machen und das α_V so festlegen, dass man ein bestimmtes α_F erreicht (Bonferroni-Korrektur) . Dazu lösen wir nach α_V auf:

$$\alpha_F = 1-(1-\alpha_V)^q$$
$$\Leftrightarrow \quad \alpha_V = 1-(1-\alpha_F)^{1/q}$$

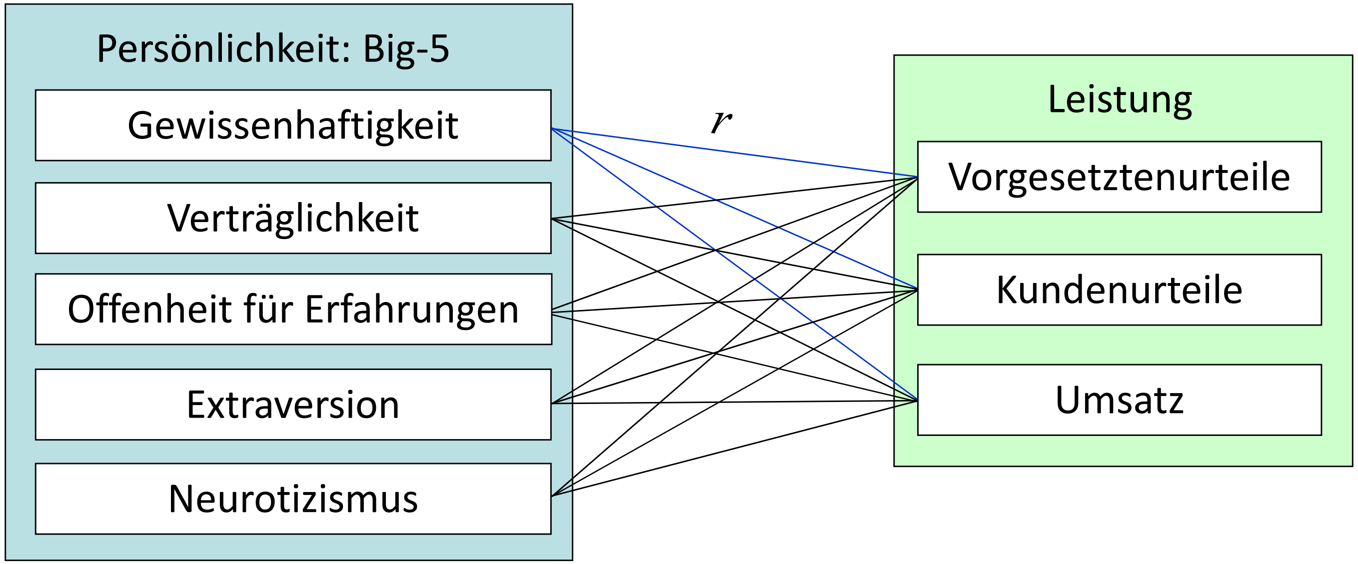
oder angenähert:

$$\alpha_F \approx q \cdot \alpha_V$$
$$\Leftrightarrow \quad \alpha_V \approx \alpha_F / q$$

		$\alpha_F = 0.05$		$\alpha_F = 0.01$	
m	q	α_V	α_F / q	α_V	α_F / q
3	3	0.017	0.017	0.003	0.003
4	6	0.009	0.008	0.002	0.002
5	10	0.005	0.005	0.001	0.001

Wenn wir also insgesamt das Fehlerrisiko auf $\alpha_F = 0.05$ festsetzen wollen, müssen wir bei $m = 4$ Gruppen (mit $q = 6$ paarweisen Einzelvergleichen) jeweils mit einer Irrtumswahrscheinlichkeit von $\alpha_V = 0.009$ testen.

- Exkurs: Problem der α-Fehler Kumulierung im Kontext korrelativer Studien.
- Beispiel-Fragestellung: Gibt es einen Zusammenhang zwischen der Persönlichkeit und der Leistung im Beruf? Untersuchung an einer Stichprobe von Versicherungsmaklern.



- Werden alle $q = 5 \cdot 3 = 15$ Korrelationen r berechnet, so lässt sich auch hier abschätzen, dass das Risiko, mindestens einen α-Fehler zu machen mit $q \cdot \alpha = 15 \cdot 0.05 = 0.75$ sehr hoch ist.
- Um α_F auf 0.05 zu halten, müssten entsprechend alle Korrelationen auf einem Signifikanzniveau von $\alpha / q = 0.05 / 15 = 0.003$ getestet werden.

Multiple Mittelwertsvergleiche: Gliederung

1 Grundproblem der α-Fehler Kumulierung

2 Bonferroni-Korrektur

Exkurs: α-Fehler Kumulierung in korrelativen Studien

3 A posteriori Tests

4 Durchführung der Tests in SPSS

5 A priori Tests

6 Trendtests

Multiple Mittelwertsvergleiche: Tukey-Kramer Test

- Wendet man die Bonferroni-Korrektur an, so ist das Risiko, einen β-Fehler zu machen, sehr hoch. Dieses Vorgehen der α-Adjustierung hat also eine geringe Power, d.h. vorhandene Unterschiede werden häufiger übersehen.

- Es gibt Testverfahren, die ebenfalls erlauben, α_F zu kontrollieren, aber eine höhere Power aufweisen. Dazu zählt der Tukey Test (auch als WSD = „wholly significant difference“ oder HSD = „honestly significant difference“ test bezeichnet).

- Im Tukey-Test werden alle Mittelwerte paarweise mit dem gewünschten familienweisen α auf Signifikanz geprüft. Zur Überprüfung der $\mathrm{H}_0 : \mu_k = \mu_l$ lautet die Prüfgröße für zwei Gruppen k und l bei gleichem Stichprobenumfang in beiden Gruppen $n_\bullet$

$$q_{T_{kl}} = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{\frac{MQ_{in}}{n_\bullet}}}$$

Prüfgröße des t-Tests für unabhängige Gruppen (bei $n_\bullet = n_k = n_l$):

$$t = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{\frac{s_k^2 + s_l^2}{n_\bullet}}} = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{\frac{2 \cdot MQ_{in}}{n_\bullet}}}$$

- Sind beide Stichprobenumfänge ungleich, so ist $n_\bullet$ zu ersetzen durch das harmonische Mittel beider Stichprobenumfänge $2/(1/n_k + 1/n_l)$. Dieser Test wird dann häufig als Tukey-Kramer Test bezeichnet.

Multiple Mittelwertsvergleiche: Alternativen zum Tukey-Kramer Test

- Die Prüfgrößen folgen unter den Voraussetzungen der Varianzanalyse (Unabhängigkeit, Varianzhomogenität, Normalverteilung) einer studentisierten Range-Verteilung mit m Gruppen und $n-m$ Freiheitsgraden. Der kritische Wert $q_{crit} = q_{m;n-m;1-\alpha}$ kann entsprechenden Tabellen entnommen werden (z.B. Tabelle H in Diehl & Arbinger, 2001).

  Ist $\left|q_{T_{kl}}\right| > q_{crit}$, so wird die zweiseitig formulierte Nullhypothese H_0: $\mu_k = \mu_l$ bei einer Irrtumswahrscheinlichkeit von α zurückgewiesen.

  Über diese Verteilung erfolgt die Kontrolle des familienweisen Signifikanzniveaus.

- Es gibt noch eine Reihe von Alternativen zum Tukey- (Kramer-) Test (vgl. Diehl & Arbinger, 2001). Studien haben aber gezeigt, dass bei Gültigkeit der Voraussetzungen dieser eine sehr gute Kontrolle des α_F-Risikos bei gleichzeitig hoher Power erzielt.

- Daneben gibt es auch einige Test, die bei speziellen, im Anschluss an die ANOVA durchgeführten multiplen Mittelwertsvergleichen geeignet sind (vgl. Kirk, 1995), z.B. wenn ...
  - nicht alle Mittelwerte paarweise verglichen werden sollen, sondern nur eine Kontrollbedingung mit allen anderen Gruppen vergleichen werden soll (Dunnett-Test).
  - komplexere Vergleiche betrachtet werden, z.B. ob sich die beiden Lehrbuch-Methoden von den beiden computergestützten Methoden unterscheiden (Scheffé-Test).

Multiple Mittelwertsvergleiche: Tukey-Kramer Test

Beispiel

$$q_{T_{kl}} = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{\frac{MQ_{in}}{n_\bullet}}}$$

$n_\bullet = 8$
$m = 4$

Lehrbuch konventionell	Lehrbuch verzweigt	Computer verzweigt	Computer Hyperlink
$\bar{x}_1 = 18.25$	$\bar{x}_2 = 30.00$	$\bar{x}_3 = 35.63$	$\bar{x}_4 = 32.88$

ONEWAY ANOVA

av

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	1402,625	3	467,542	6,754	,001
Innerhalb der Gruppen	1938,250	28	69,223		
Gesamt	3340,875	31			

$$q_{T_{12}} = \frac{18.25 - 30.00}{\sqrt{\frac{69.223}{8}}} = -4.00$$

$q_{T_{13}} = -5.91$

$q_{T_{14}} = -4.97$

$p < .05$

$q_{T_{23}} = -1.91$

$q_{T_{24}} = -0.98$

$q_{T_{34}} = 0.94$

ns

Kritischer q-Wert für $\alpha = \alpha_F = 0.05$ aus studentisierter Range-Verteilung mit $m = 4$ und $df = n - m = 32 - 4) = 28$ ist $q_{crit} = 3.86$.

Die Methode mit dem konventionellen Lehrbuch führt zu einer statistisch signifikant schlechteren Leistung als alle anderen Lehrmethoden, die sich untereinander nicht statistisch signifikant unterscheiden.

Multiple Mittelwertsvergleiche: Games-Howell Test

- Wird (nur) gegen die Annahme der Varianzhomogenität verstoßen, so kann statt des Tukey-(Kramer-) Tests der Games-Howell Test eingesetzt werden (vgl. Kirk, 1995, S. 147f). Die wiederum einer studentisierten Range-Verteilung folgende Prüfgröße lautet allgemein:

$$q_{GH_{kl}} = \frac{\bar{x}_k - \bar{x}_l}{\sqrt{\left(\frac{s_k^2}{n_k} + \frac{s_l^2}{n_l}\right) \Big/ 2}} \quad \text{mit } m \text{ Gruppen und} \quad df_{GH} = \frac{\left(a_k + a_l\right)^2}{\frac{a_k^2}{n_k - 1} + \frac{a_l^2}{n_l - 1}} \quad \text{wobei} \quad a_t = \frac{s_t^2}{n_t}$$

Die Prüfung erfolgt wie beim Tukey-Kramer Test.

- Der Games-Howell Test wird also dann eingesetzt, wenn z.B. der Levene-Test Varianzheterogenität anzeigt und stellt also das Pendant zum Welch- bzw. Brown-Forsyth Test dar.

- Bzgl. der Robustheit der Verfahren lassen sich die bei der ANOVA gemachten Aussagen weitgehend übertragen.

Multiple Mittelwertsvergleiche: Games-Howell Test

- Uneinigkeit herrscht in der scientific community darüber, ob man immer zunächst prüfen sollte, ob sich ein signifikanter Effekt in der ANOVA ergeben hat und nur, wenn dies der Fall ist, die dargestellten auch als Anschlusstests bezeichneten Verfahren einsetzen sollte. Von der Logik der dargestellten Verfahren könnte man auch auf die ANOVA verzichten.

- Praktisch spielt dies kaum eine Rolle, da die Fälle, das die ANOVA einen statistisch signifikanten Effekt anzeigt, aber keiner der Einzelvergleiche bzw. die ANOVA keinen Effekt anzeigt, wohl aber mindestens einer der Einzelvergleiche, äußerst selten vorkommen.

Multiple Mittelwertsvergleiche: Gliederung

1 Grundproblem der α-Fehler Kumulierung

2 Bonferroni-Korrektur

Exkurs: α-Fehler Kumulierung in korrelativen Studien

3 A posteriori Tests

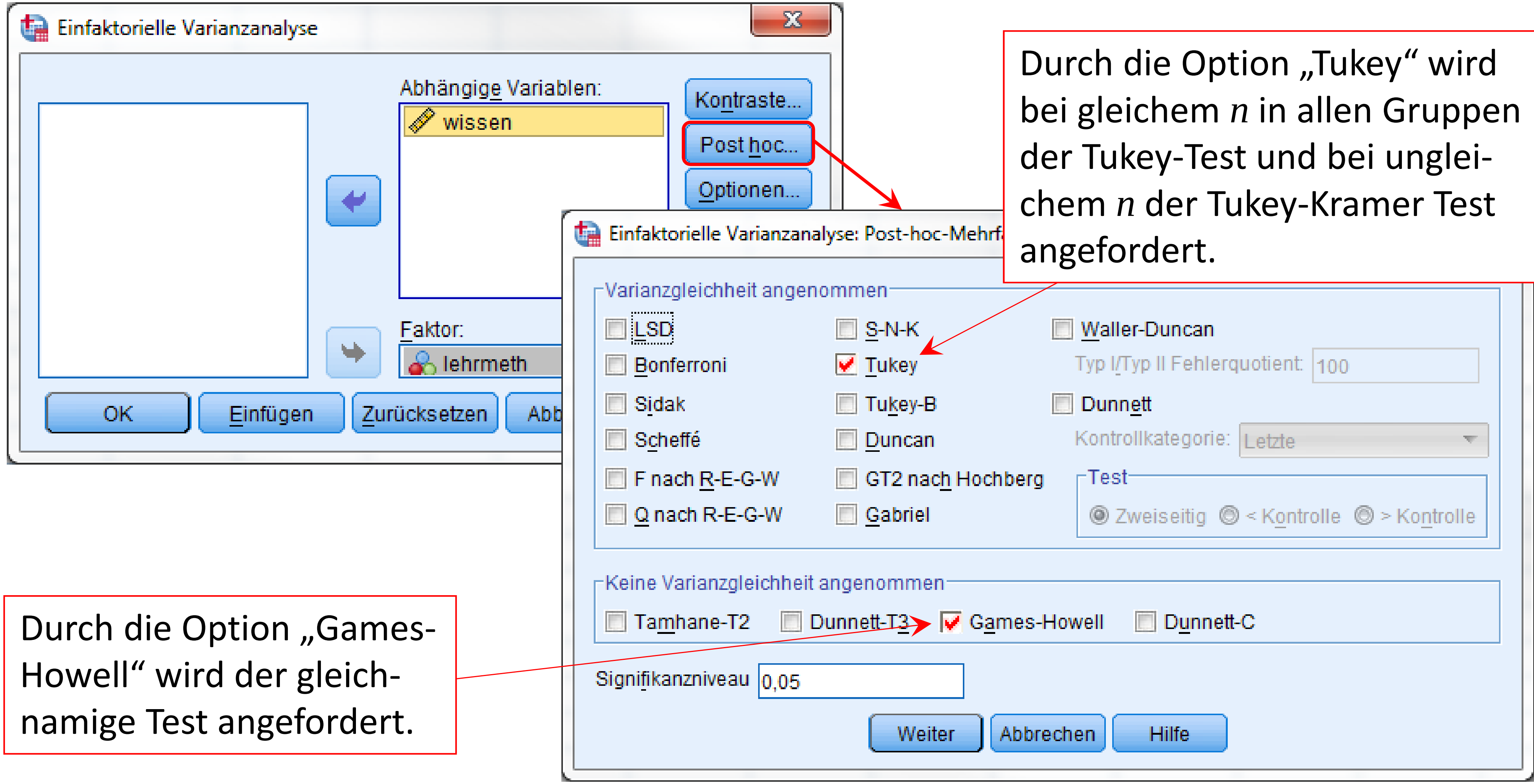
4 Durchführung der Tests in SPSS

5 A priori Tests

6 Trendtests

Multiple Mittelswertsvergleiche: Tukey- (Kramer-) Test in SPSS

Den Tukey (-Kramer) Test erhält man unter: `Analysieren/Mittelwerte_vergleichen/Einfaktorielle_ANOVA…` nach dem Drücken des Buttons (`Post_Hoc…`):



Durch die Option „Tukey“ wird bei gleichem *n* in allen Gruppen der Tukey-Test und bei ungleichem *n* der Tukey-Kramer Test angefordert.

Durch die Option „Games-Howell“ wird der gleichnamige Test angefordert.

Post-Hoc-Tests

Mehrfachvergleiche

wissen
Tukey-HSD

(I) lehrmeth	(J) lehrmeth	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall Untergrenze	95%-Konfidenzintervall Obergrenze
1 Lb konventionell	2 Lb verzweigt	-11,750*	4,160	,041	-23,11	-,39
	3 Web verzweigt	-17,375*	4,160	,001	-28,73	-6,02
	4 Web Hyperlink	-14,625*	4,160	,008	-25,98	-3,27
2 Lb verzweigt	1 Lb konventionell	11,750*	4,160	,041	,39	23,11
	3 Web verzweigt	-5,625	4,160	,539	-16,98	5,73
	4 Web Hyperlink	-2,875	4,160	,900	-14,23	8,48
3 Web verzweigt	1 Lb konventionell	17,375*	4,160	,001	6,02	28,73
	2 Lb verzweigt	5,625	4,160	,539	-5,73	16,98
	4 Web Hyperlink	2,750	4,160	,911	-8,61	14,11
4 Web Hyperlink	1 Lb konventionell	14,625*	4,160	,008	3,27	25,98
	2 Lb verzweigt	2,875	4,160	,900	-8,48	14,23
	3 Web verzweigt	-2,750	4,160	,911	-14,11	8,61

*. Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.

Gruppen-Paare

Mittelwerts-differenzen

p-Werte

Prüfgrößen werden nicht ausgegeben

1 Grundproblem der α-Fehler Kumulierung

2 Bonferroni-Korrektur

Exkurs: α-Fehler Kumulierung in korrelativen Studien

3 A posteriori Tests

4 Durchführung der Tests in SPSS

5 A priori Tests

6 Trendtests

- Bisher haben wir im Anschluss an die ANOVA mit einem statistisch signifikanten Ergebnis a posteriori (post-hoc) explorativ alle paarweisen Mittelwerte gegeneinander getestet.
- Interessiert man sich hypothesengeleitet nur für bestimmte Unterschiede (und nicht den Omnibus-Test), dann kann man sog. geplante Vergleiche (Kontraste) durchführen. Solche Vergleiche benötigen die Varianzanalyse nicht, sondern können a priori direkt getestet werden. Der Vorteil des a priori-Testens besteht in der größeren Power. Allerdings können damit nicht alle eventuell vorhandenen Unterschiede gefunden werden.
- Dabei können nicht notwendigerweise immer nur zwei Gruppen verglichen werden, sondern auch Paare von zusammengefassten Gruppen (Obergruppen, komplexe Vergleiche).
- Im Beispiel könnte eine komplexere Hypothese z.B. lauten: „Computergestützte Verfahren sind besser als Lehrbuch-Methoden“. Formal ausgedrückt lauten die zweiseitig formulierten Hypothesen dann:

$$H_0: \frac{\mu_1+\mu_2}{2} = \frac{\mu_3+\mu_4}{2} \quad \Leftrightarrow \quad H_0: +\tfrac{1}{2}\cdot\mu_1 + \tfrac{1}{2}\cdot\mu_2 - \tfrac{1}{2}\cdot\mu_3 - \tfrac{1}{2}\cdot\mu_4 = 0$$

$$H_1: \frac{\mu_1+\mu_2}{2} \neq \frac{\mu_3+\mu_4}{2} \quad \Leftrightarrow \quad H_1: +\tfrac{1}{2}\cdot\mu_1 + \tfrac{1}{2}\cdot\mu_2 - \tfrac{1}{2}\cdot\mu_3 - \tfrac{1}{2}\cdot\mu_4 \neq 0$$

Lehrbuch konventionell	μ_1
Lehrbuch verzweigt	μ_2
Computer verzweigt	μ_3
Computer Hyperlink	μ_4

Multiple Mittelwertsvergleiche

- Jeder Vergleich zwischen Populationsmitteln lässt sich als linearer Kontrast Λ („lambda") in Form einer gewichtete Summe (Linearkombination) der Mittelwerte darstellen:

$$\Lambda = c_1 \cdot \mu_1 + c_2 \cdot \mu_2 + \ldots + c_m \cdot \mu_m = \sum_{j=1}^{m} c_j \cdot \mu_j \qquad \text{mit} \quad \sum_{j=1}^{m} c_j = 0$$

- Die Gewichte (Koeffizienten) wählt man so, dass sie den gewünschten Vergleich ausdrücken:
  - Beispiel 1: Vergleich der beiden Lehrbuch-Methoden einerseits mit den beiden computergestützten Verfahren auf der anderen Seite wie folgt:

$$\Lambda_1 = +\tfrac{1}{2} \cdot \mu_1 + \tfrac{1}{2} \cdot \mu_2 - \tfrac{1}{2} \cdot \mu_3 - \tfrac{1}{2} \cdot \mu_4$$

  - Beispiel 2: Vergleich der unverzweigten Lehrmethode einerseits mit den drei verzweig-ten Lehrmethoden andererseits:

$$\Lambda_2 = -1 \cdot \mu_1 + \tfrac{1}{3} \cdot \mu_2 + \tfrac{1}{3} \cdot \mu_3 + \tfrac{1}{3} \cdot \mu_4$$

  - Beispiel 3: Vergleich der verzweigten Methoden ohne und mit Hyperlink:

$$\Lambda_3 = 0 \cdot \mu_1 + \tfrac{1}{2} \cdot \mu_2 + \tfrac{1}{2} \cdot \mu_3 - 1 \cdot \mu_4$$

Lehrbuch konventionell	Lehrbuch verzweigt	Computer verzweigt	Computer Hyperlink
μ_1	μ_2	μ_3	μ_4

- Die Gewichte wurden bei obigen Kontrasten nach der folgenden Regel festgelegt: Weise beiden zu vergleichenden Obergruppen Gewichte mit unterschiedlichem Vorzeichen zu und setze den Betrag des Gewichts jeweils gleich 1 dividiert durch die Zahl der Gruppen in der Obergruppe. Setze die Gewichte bei allen nicht beteiligten Gruppen 0 (Standardform).

- Im Prinzip ist es aber nicht erforderlich, dass sich die positiven und negativen Einzelsummen zu 1 addieren (nur, dass die Summe aller c gleich 0 ist). So gilt z.B. für das obige Beispiel:

$$\begin{aligned}\Lambda_2 &= -1\cdot\mu_1 + \tfrac{1}{3}\cdot\mu_2 + \tfrac{1}{3}\cdot\mu_3 + \tfrac{1}{3}\cdot\mu_4 \\ &= -3\cdot\mu_1 + 1\cdot\mu_2 + 1\cdot\mu_3 + 1\cdot\mu_4\end{aligned}$$

- Meist wollen wir prüfen, ob ein Kontrast signifikant von 0 verschieden ist. Unsere Hypothesen lauten dann
  - zweiseitig: $H_0\colon \Lambda = 0$ und $H_1\colon \Lambda \neq 0$
  - einseitig z.B.: $H_0\colon \Lambda \leq 0$ und $H_1\colon \Lambda > 0$
  - Im Beispiel: Die verzweigten Lehrmethoden unterscheiden sich von der unverzweigten Lehrmethode (zweiseitig) bzw. sind besser als die unverzweigte Lehrmethode (einseitig) in Bezug auf die Lernleistung der Studierenden.

Die Richtung in der Hypothese muss mit den Vorzeichen der Gewichte korrespondieren.

Multiple Mittelwertsvergleiche

- Zur Prüfung der Hypothese erhalten wir zunächst eine Schätzung des Kontrastes, indem wir die Populationsmittel durch die Stichprobenmittel ersetzen:

$$\hat{\Lambda} = L = \sum_{j=1}^{m} c_j \cdot \bar{x}_j$$

- Ferner benötigen wir die Quadratsumme für den Kontrast (Die Normierung im Nenner sorgt dafür, dass die Quadratsumme unabhängig von der Größe der Koeffizienten wird):

$$QS_L = \frac{L^2}{\sum_{j=1}^{m} \frac{c_j^2}{n_j}} = \frac{\left(\sum_{j=1}^{m} c_j \cdot \bar{x}_j\right)^2}{\sum_{j=1}^{m} \frac{c_j^2}{n_j}}$$ mit $df_L = 1$

Bei gleichem $n_{\bullet}$ in allen Gruppen gilt vereinfacht: $$QS_L = \frac{n_{\bullet} \cdot L^2}{\sum_{j=1}^{m} c_j^2}$$

- Unter den Annahmen der Varianzanalyse ist dann die Prüfgröße

$$F = \frac{MQ_L}{MQ_{in}} = \frac{L^2}{MQ_{in} \cdot \sum_{j=1}^{m} \frac{c_j^2}{n_j}}$$

F-verteilt mit $df_1 = 1$ und $df_2 = n - m$.

Bei gleichem $n_{\bullet}$ in allen Gruppen gilt vereinfacht: $$F = \frac{n_{\bullet} \cdot L^2}{MQ_{in} \cdot \sum_{j=1}^{m} c_j^2}$$

Multiple Mittelwertsvergleiche

- Man vergleicht also bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ den empirischen F-Wert mit dem kritischen F-Wert mit $df_1 = 1$ und $df_2 = n - m$, entweder bei α (zweiseitig) oder $2 \cdot \alpha$ (einseitig, wenn die Richtung – d.h. das Vorzeichen des Kontrasts – stimmt).

- Alternativ kann statt des F-Tests auch der t-Test rechts herangezogen werden, wobei bei $df_1 = 1$ wieder gilt, dass $F = t^2$ und beide Tests zu identischen Entscheidungen führen. (Dabei prüft man den t-Wert gegen den kritischen t-Wert $t_{n-m;1-\alpha/2}$.)

$$t = \frac{L}{\sqrt{MQ_{in} \cdot \sum_{j=1}^{m} \frac{c_j^2}{n_j}}}$$

- Gilt Varianzhomogenität nicht, so muss MQ_{in} durch die Aggregierung über die Varianzen in den Gruppen s_j^2 ersetzt werden und es ergibt sich die Welch-Korrektur:

$$t_W = \frac{L}{\sqrt{\sum_{j=1}^{m} \frac{s_j^2 \cdot c_j^2}{n_j}}} \quad \text{mit} \quad df_W = \frac{\left(\sum_{j=1}^{m} \frac{s_j^2 \cdot c_j^2}{n_j} \right)^2}{\sum_{j=1}^{m} \frac{c_j^4 \cdot s_j^4}{n_j^2 \cdot (n_j - 1)}}$$

Multiple Mittelwertsvergleiche: A priori Kontraste

- **Beispiel**: Wir wollen zweiseitig prüfen, ob die beiden computergestützten Lehrformen den beiden konventionellen überlegen sind.

Lehrbuch konventionell	Lehrbuch verzweigt	Computer verzweigt	Computer Hyperlink
$\bar{x}_1 = 18.25$	$\bar{x}_2 = 30.00$	$\bar{x}_3 = 35.63$	$\bar{x}_4 = 32.88$
$n_1 = 8$	$n_2 = 8$	$n_3 = 8$	$n_4 = 8$

ONEWAY ANOVA

wissen

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	1402,625	3	467,542	6,754	,001
Innerhalb der Gruppen	1938,250	28	69,223		
Gesamt	3340,875	31			

- Die Nullhypothese lautet entsprechend:

$$H_0: \quad -1\cdot\mu_1 - 1\cdot\mu_2 + 1\cdot\mu_3 + 1\cdot\mu_4 = 0$$

- Die Quadratsummen des Kontrastes sind:

$$QS_L = \frac{n_\bullet \cdot \left(\sum_{j=1}^{m} c_j \cdot \bar{x}_j\right)^2}{\sum_{j=1}^{m} c_j^2} = \frac{8\cdot(-18.25 - 30.00 + 35.63 + 32.88)^2}{(-1)^2 + (-1)^2 + 1^2 + 1^2} = \frac{8\cdot 20.25^2}{4} = 820.125 \; = MQ_L$$

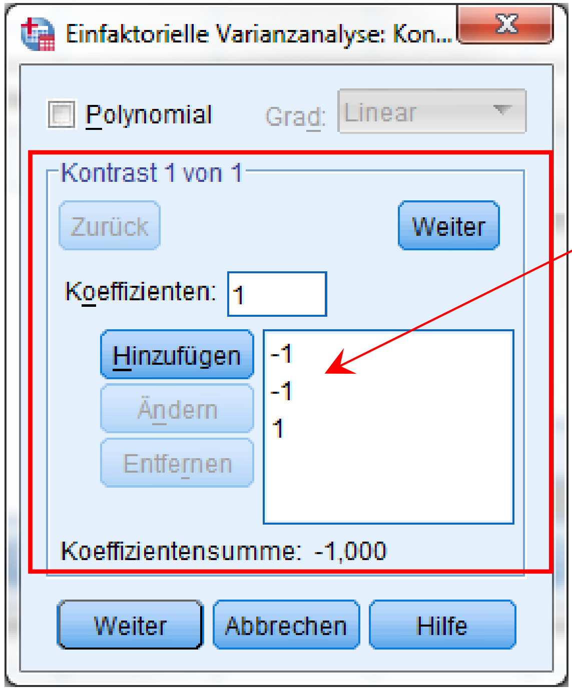
- Da der Levene-Test nicht zur Zurückweisung der Varianzhomogenitätsannahme geführt hat (s.o.), berechnen wir den unkorrigierten F-Wert:

$$F = \frac{MQ_L}{MQ_{in}} = \frac{820.125}{69.223} = 11.85$$

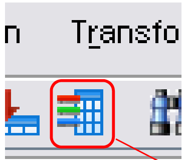
- Da $F_{crit} = F_{1;n-m;1-\alpha} = F_{1;28;0.95} = 4.196 < 11.85 = F$ (und der Tatsache, dass $L = 20.25 > 0$ ist) schließen wir, dass die Computermethoden statistisch signifikant besser sind.

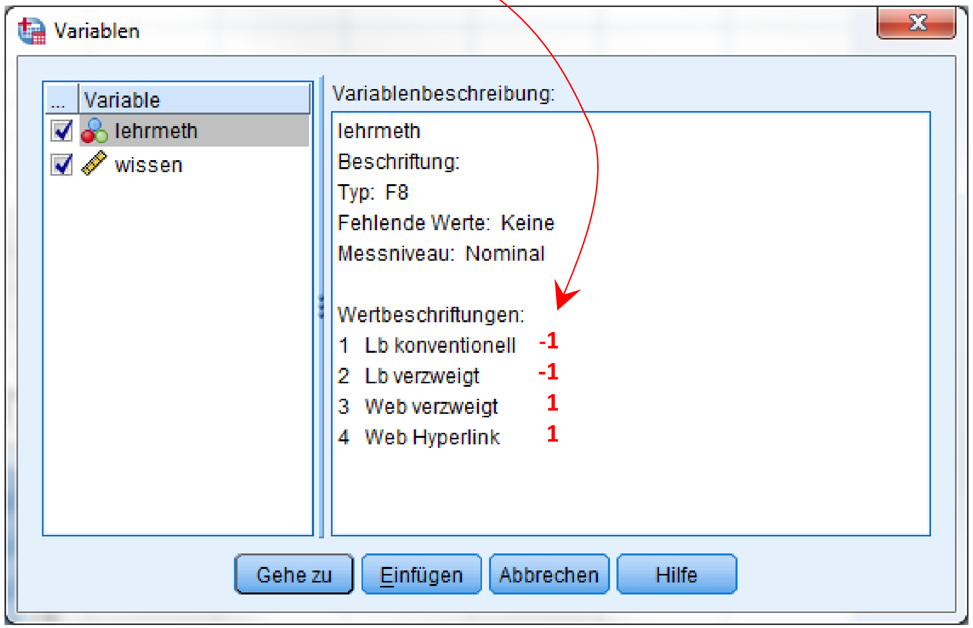
Multiple Mittelwertsvergleiche: A priori Kontraste mit SPSS

In der Prozedur Oneway können unter (`Kontraste…`) a priori Vergleiche angefordert werden:



Hier werden nacheinander die Gewichte des Kontrastes eingegeben und jeweils mit dem Button (`Hinzufügen`) in das Listenfeld eingetragen. Die Gewichte werden den Gruppen entsprechend deren aufsteigender Codierung zugeordnet.





Im Beispiel werden die beiden Faktorstufen mit den beiden kleinsten Codierwerten (hier 1 und 2 für die beiden Lehrbuchmethoden) gegen die beiden anderen Faktorstufen (hier 3 und 4 für die computergestützten Methoden) getestet, formal:

$$\Lambda = -1 \cdot \mu_1 - 1 \cdot \mu_2 + 1 \cdot \mu_3 + 1 \cdot \mu_4$$

Multiple Mittelwertsvergleiche: A priori Kontraste mit SPSS

Kontrast-Koeffizienten

Kontrast	lehrmeth: 1 Lb konventionell	2 Lb verzweigt	3 Web verzweigt	4 Web Hyperlink
1	-1	-1	1	1

Zuordnung der Gewichte (Koeffizienten) zu den Gruppen (Faktorstufen)

Kontrast-Tests

		Kontrast	Kontrastwert	Standardfehler	T	df	Signifikanz (2-seitig)
wissen	Varianzen sind gleich	1	20,25	5,883	3,442	28	,002
	Varianzen sind nicht gleich	1	20,25	5,883	3,442	26,468	,002

Die Tabelle enthält in der oberen Zeile die Prüfung für den Fall der bestehenden Varianz-homogenität (mittels des t-Tests): Der Kontrast beträgt $L = 20.25$. Der Test zeigt mit $t(28) = 3.44$, $p < .05$ einen statistisch signifikanten Unterschied zwischen den computergestützten einerseits und den Lehrbuchmethoden andererseits an. (Es gilt: $t = 3.44 = \sqrt{F} = \sqrt{11.85}$.)

In der Zeile darunter wird für den Fall der Varianzheterogenität die nach Welch korrigierte Prüfgröße t_W mit den entsprechenden Freiheitsgraden $df_W = 26.468$ ausgegeben.

Multiple Mittelwertsvergleiche: A priori Kontraste

- Wir könnten aber a priori nicht nur eine Hypothese aufstellen, sondern uns gleichzeitig für mehrere geplante Kontraste interessieren, etwa im Beispiel für die folgenden vier:

Lehrbuch konventionell	Lehrbuch verzweigt	Computer verzweigt	Computer Hyperlink
μ_1	μ_2	μ_3	μ_4

Nr.	Kontrast	Hypothese
1	$-3\cdot\mu_1+1\cdot\mu_2+1\cdot\mu_3+1\cdot\mu_4$	Verzweigte Verfahren sind besser als unverzweigte.
2	$0\cdot\mu_1-2\cdot\mu_2+1\cdot\mu_3+1\cdot\mu_4$	Verzweigte computergestützte Verfahren sind besser als die verzweigte Lehrbuchmethode.
3	$0\cdot\mu_1+0\cdot\mu_2-1\cdot\mu_3+1\cdot\mu_4$	Bei verzweigten computergestützten Verfahren ist die Hyperlink-Variante besser als die fest verzweigte.
4	$-1\cdot\mu_1-1\cdot\mu_2+1\cdot\mu_3+1\cdot\mu_4$	Computergestützte Verfahren sind besser als lehrbuch-gestützte.

- Will man mehrere Kontraste untersuchen, so hat es bestimmte Konsequenzen, wie die Beziehung der Kontraste untereinander ist. Betrachtet man jeweils ein Paar von Kontrasten, so können beide Vergleiche abhängig oder unabhängig (=orthogonal) sein. Sind alle betrachteten Kontraste unabhängig, so spricht man von einem Satz unabhängiger (orthogonaler) Vergleiche (oder wechselseitiger Unabhängigkeit).

Multiple Mittelwertsvergleiche: Orthogonalität

- Vergleiche sind dann orthogonal, wenn sie keine redundante, überlappende Informationen beinhalten. Anders ausgedrückt: Weiß ich bei orthogonalen Kontrasten etwas darüber, dass der eine ungleich 0 ist, so kann ich daraus nichts für den anderen Kontrast folgern.

- Will ich z.B. $\Lambda_1 = \mu_1 - (\mu_2 + \mu_3)/2 = 0$ und $\Lambda_2 = \mu_1 - \mu_2 = 0$ testen, so sind diese beiden Kontraste nicht unabhängig. Der Kontrast $\Lambda_3 = \mu_2 - \mu_3 = 0$ ist hingegen zu Λ_1 und der Kontrast $\Lambda_4 = \mu_3 - \mu_4 = 0$ zu Λ_2 orthogonal.

- Formal sind zwei Kontraste mit den Gewichten c_j und d_j (bei gleichem $n_\bullet$ in allen Gruppen) orthogonal, wenn die Summe der Produkte korrespondierender Gewichte 0 ist:

$$c_1 \cdot d_1 + c_2 \cdot d_2 + \ldots + c_m \cdot d_m = \sum_{j=1}^{m} c_j \cdot d_j = 0$$

Bei ungleichen n_j muss gelten: $\sum_{j=1}^{m} \frac{c_j \cdot d_j}{n_j} = 0$

- Beispiel: Prüfung von Λ_1 und Λ_2 aus obigem Beispiel (Folie 28) auf Unabhängigkeit:

$$\left.\begin{aligned} \Lambda_1 &= -3 \cdot \mu_1 \; +1 \cdot \mu_2 \; +1 \cdot \mu_3 \; +1 \cdot \mu_4 \\ \Lambda_2 &= \;\; 0 \cdot \mu_1 \; -2 \cdot \mu_2 \; +1 \cdot \mu_3 \; +1 \cdot \mu_4 \end{aligned}\right\}$$

$$\begin{aligned} & c_1 \cdot d_1 + c_2 \cdot d_2 + c_3 \cdot d_3 + c_4 \cdot d_4 \\ &= (-3) \cdot 0 + 1 \cdot (-2) + 1 \cdot 1 + 1 \cdot 1 \\ &= 0 + (-2) + 1 + 1 = 0 \end{aligned}$$

$\Rightarrow$ Die beiden Kontraste sind orthogonal.

Multiple Mittelwertsvergleiche: Orthogonalität

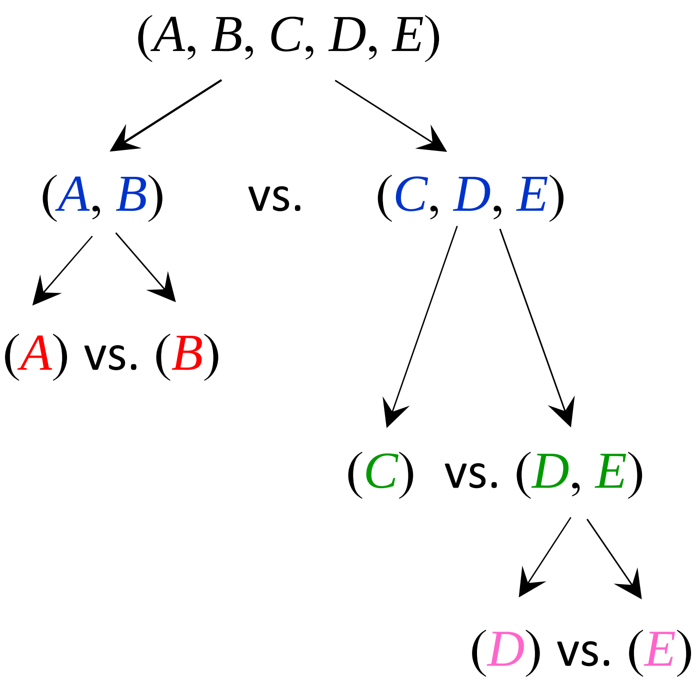
- Prüfen wir die Orthogonalität für alle Paare der vier Kontraste oben, so resultiert:

	Kontrast 1				Kontrast 2				
	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4	$\Sigma\, c_j \cdot d_j$
Λ_1 vs. Λ_2	-3	$+1$	$+1$	$+1$	0	-2	$+1$	$+1$	0
Λ_1 vs. Λ_3	-3	$+1$	$+1$	$+1$	0	0	-1	$+1$	0
Λ_1 vs. Λ_4	-3	$+1$	$+1$	$+1$	-1	-1	$+1$	$+1$	4
Λ_2 vs. Λ_3	0	-2	$+1$	$+1$	0	0	-1	$+1$	0
Λ_2 vs. Λ_4	0	-2	$+1$	$+1$	-1	-1	$+1$	$+1$	4
Λ_3 vs. Λ_4	0	0	-1	$+1$	-1	-1	$+1$	$+1$	0

- Man erkennt, dass die Kontraste Λ_1, Λ_2 und Λ_3 einen Satz orthogonaler Kontraste bilden. Kontrast Λ_4 ist hingegen nicht unabhängig zu Λ_1 und zu Λ_2.

- Man kann zeigen, dass bei m Gruppen nicht mehr als $m - 1$ ($=df_{zw}$) wechselseitig orthogonale Kontraste möglich sind. Sind $m - 1$ Kontraste unabhängig, so spricht man von einem vollständigen Satz orthogonaler Kontraste. Es gibt immer verschiedene solcher Sätze.

- Wie kann man einen vollständigen Satz $m-1$ orthogonaler Kontraste erzeugen?
  - Eine Möglichkeit besteht darin, im ersten Kontrast die erste Gruppe gegen alle anderen zu testen, im zweiten Kontrast die erste Gruppe nullzusetzen und die zweite gegen die verbleibenden usw. (sog. Helmert Kontraste)
  - Eine andere Möglichkeit besteht darin, sich die möglichen Vergleiche über die Aufspaltung in einem Baumdiagramm zu visualisieren (Howell, 2010, S. 355f), z.B. wie folgt bei $m=5$ Gruppen = Faktorstufen, hier bezeichnet mit A bis E:



$$+1/2\cdot A + 1/2\cdot B - 1/3\cdot C - 1/3\cdot D - 1/3\cdot E$$

$$+1\cdot A - 1\cdot B + 0\cdot C + 0\cdot D + 0\cdot E$$

$$0\cdot A + 0\cdot B + 1\cdot C - 1/2\cdot D - 1/2\cdot E$$

$$0\cdot A + 0\cdot B + 0\cdot C + 1\cdot D - 1\cdot E$$

Multiple Mittelwertsvergleiche: Orthogonalität

- Für einen vollständigen Satz orthogonaler Vergleiche gilt, dass die Summe aller $m-1$ Kontrast-Quadratsummen gleich der Treatment-Quadratsumme ist:

$$QS_{zw} = \sum_{q=1}^{m-1} QS_{L_q}$$

- Die gesamte Varianz, die durch das Modell der einfaktoriellen Varianzanalyse erklärt werden kann, kann also in die nicht überlappenden Anteile zu Lasten der verschiedenen Kontraste zerlegt werden. Dies gilt bei abhängigen Vergleichen nicht, die auch den Nachteil aufweisen, in ihren Aussagen partiell redundant zu sein.

- Trotz dieser attraktiven Eigenschaft ist die Konstruktion orthogonaler Kontraste natürlich kein Selbstzweck. Letztlich sollte man die Kontraste testen, die auch wirklich von Interesse sind.

„When I first started teaching and writing about statistics, orthogonal contrasts were a big deal, just as was the distinction between a priori and post hoc tests. Authors went out of their way to impress on you the importance of orthogonality and the need to feel somewhat guilty if you ran comparisons that were not orthogonal. That attitude changed over times. While it is nice to have a set of orthogonal contrasts, in part because they sum up to SS_{treat}, people are far more willing to run nonorthogonal contrasts. I would certainly not suggest that you pass up an important contrast just because it is not orthogonal to others that you ran.“ (Howell, 2010, S. 376f).

Multiple Mittelwertsvergleiche: A priori Kontraste

- Die Empfehlungen, wie bei mehr als einem geplanten Kontrast mit dem erhöhten Risiko des familienweisen α umgegangen werden sollten, differieren (vgl. Keppel & Wickens, 2004, S. 112ff):
  - Viele Autoren vertreten die Position, dass bei (wenigen) aus der Theorie abgeleiteten, geplanten Vergleichen keine α-Adjustierung vorgenommen werden sollte, auch um die Power für diese Testungen hoch zu halten. Jeder einzelne Kontrast wird also auf dem gewählten Signifikanzniveau (z.B. $\alpha = 0.05$) getestet. Einige Autoren empfehlen dieses Vorgehen bei a priori Tests nur dann, wenn diese orthogonal sind (z.B. Kirk, 1995; Bortz & Schuster, 2010), andere generell (Keppel & Wickens, 2004).
  - Aber, was ist, wenn jemand sagt: „Ich plane a priori, alle Gruppen paarweise gegeneinander zu testen“?
  - Es gibt auch Autoren, die generell eine α-Adjustierung empfehlen (z.B. Stevens, 2007, S. 79-93).
  - Einige Autoren setzen dann, wenn sie eine α-Adjustierung vornehmen, das familienweise α höher an (z.B. $\alpha = 0.10$).

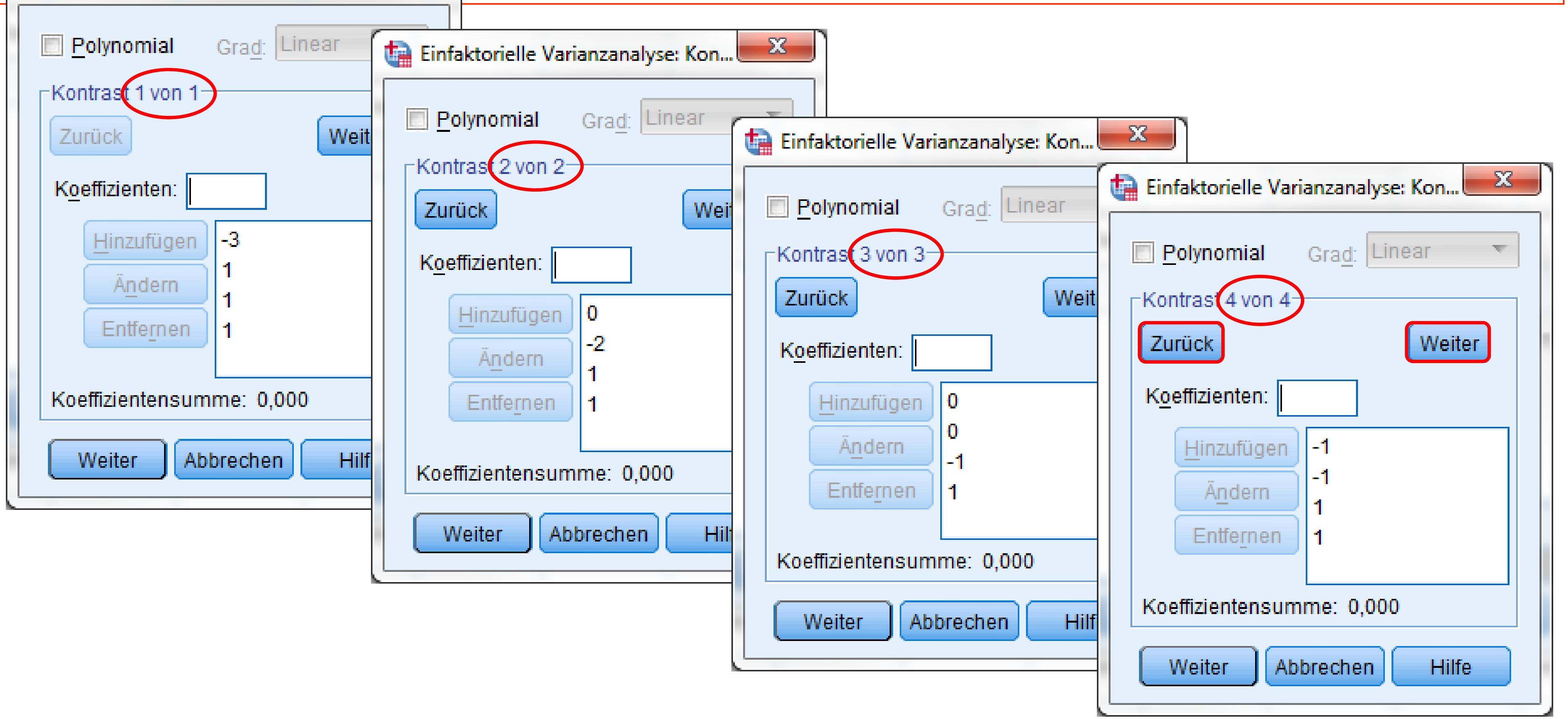
- Will man eine Adjustierung vornehmen, die sicherstellt, dass das gesetzte α auch für die gesamte Familie von Vergleichen eingehalten wird, so kann man bei a priori Vergleichen den oben dargestellten F- bzw. t-Test (gegebenenfalls mit Welch Korrektur) und dann eine der folgenden drei Strategien heranziehen, die in ihrer Power (und in ihrem Aufwand) der Reihe nach zunehmen:

  - Im auch als Bonferroni- oder Dunn-Test bezeichneten Vorgehen prüft man jeden der v Kontraste gegen α / v.

  - Im auch als Dunn-Šidák Test bezeichneten Vorgehen prüft man jeden der v Vergleiche gegen $1-(1-\alpha)^{1/v}$.

  - In dem sequentiell vorgehenden Holm-Test werden nicht alle v Kontraste gegen das gleiche adjustierte Signifikanzniveau getestet, sondern zunächst alle Kontraste ihrer Größe nach abfallend geordnet. Der größte Kontrast $q_{(v)}$ wird dann gegen α / v getestet. Ist dieser nicht signifikant, wird abgebrochen. Ist er signifikant, so bleiben nur noch $v-1$ Nullhypothesen, die falsch sein können, und entsprechend kann man den zweitgrößten Kontrast gegen $\alpha / (v-1)$ testen. Ist dieser signifikant, wird für den nächsten Kontrast mit einem erneut verringerten Signifikanzniveau von $\alpha / (v-2)$ fortgefahren usw.

In der Prozedur Oneway können unter (`Kontraste…`) auch mehrere apriori-Kontraste auf einmal angefordert werden. Dazu sind alle Gewichte des ersten Kontrastes wie gehabt einzugeben, dann auf den Button (`Weiter`) zu klicken, den zweiten Kontrast einzugeben usw. Sind mehrere Kontraste eingegeben, kann mit den Buttons (`Zurück`) und (`Weiter`) zwischen ihnen hin- und hergewechselt werden.



Multiple Mittelwertsvergleiche: A priori Kontraste mit SPSS

Kontrast-Koeffizienten

Kontrast	lehrmeth: 1 Lb konventionell	2 Lb verzweigt	3 Web verzweigt	4 Web Hyperlink
1	-3	1	1	1
2	0	-2	1	1
3	0	0	-1	1
4	-1	-1	1	1

Der Output entspricht dem im Falle von einem Kontrast. Die vier Kontraste wer-den jeweils untereinander dargestellt.

Es zeigt sich, dass der erste und der vierte Kontrast statistisch signifikant ausfallen (ohne α-Adjustierung).

Kontrast-Tests

		Kontrast	Kontrastwert	Standardfehler	T	df	Signifikanz (2-seitig)
wissen	Varianzen sind gleich	1	43,75	10,190	4,293	28	,000
		2	8,50	7,205	1,180	28	,248
		3	-2,75	4,160	-,661	28	,514
		4	20,25	5,883	3,442	28	,002
	Varianzen sind nicht gleich	1	43,75	8,814	4,964	16,027	,000
		2	8,50	7,868	1,080	12,974	,300
		3	-2,75	4,282	-,642	13,909	,531
		4	20,25	5,883	3,442	26,468	,002

- Wollte man eine α-Adjustierung bei den $q = 4$ Tests vornehmen, so muss man jeweils auf der Basis des gewünschten α_F – hier 0.05 – für die Familie der vier Kontraste das α_V für die einzelnen Vergleiche bestimmen und dann den p-Wert damit vergleichen.

Für Holm-Test absteigend sortiert

Nr.	L	p	Dunn	Dunn-Šidák	Holm
1	$L_1 = -3 \cdot \bar{x}_1 + 1/3 \cdot \bar{x}_2 + 1/3 \cdot \bar{x}_3 + 1/3 \cdot \bar{x}_4 = 43.75$	.000	.0125	.0127	.0125
4	$L_4 = -1 \cdot \bar{x}_1 - 1 \cdot \bar{x}_2 + 1 \cdot \bar{x}_3 + 1 \cdot \bar{x}_4 = 20.25$	.002	.0125	.0127	.0167
2	$L_2 = 0 \cdot \bar{x}_1 - 2 \cdot \bar{x}_2 + 1 \cdot \bar{x}_3 + 1 \cdot \bar{x}_4 = 8.50$	.248	.0125	.0127	.0250
3	$L_3 = 0 \cdot \bar{x}_1 + 0 \cdot \bar{x}_2 - 1 \cdot \bar{x}_3 + 1 \cdot \bar{x}_4 = -2.75$	.514	.0125	.0127	

Kontrast	Kontrastwert	Standardfehler	T	df	Signifikanz (2-seitig)
1	43,75	10,190	4,293	28	,000
2	8,50	7,205	1,180	28	,248
3	-2,75	4,160	-,661	28	,514
4	20,25	5,883	3,442	28	,002

$$\alpha / q = 0.05 / 4 = 0.0125$$

$$1 - (1 - \alpha)^{1/q} = 1 - \sqrt[4]{(1 - 0.05)} = 0.0127$$

$$\alpha / s = 0.05 / 3 = 0.0167 \quad (s = q - 1 = 3)$$
$$\alpha / s = 0.05 / 2 = 0.0250 \quad (s = q - 2 = 2)$$

Multiple Mittelwertsvergleiche: Gliederung

1 Grundproblem der α-Fehler Kumulierung

2 Bonferroni-Korrektur

Exkurs: α-Fehler Kumulierung in korrelativen Studien

3 A posteriori Tests

4 Durchführung der Tests in SPSS

5 A priori Tests

6 Trendtests

- **Beispiel** (aus Keppel & Wickens, 2004, S. 89ff): Es soll untersucht werden, ob Personen besser lernen, wenn das Training über die Zeit verteilt wird, als wenn es massiert (ohne Pausen) erfolgt.

- Die Aufgabe der Vpn besteht darin, auf einem Bildschirm einem zufällig sich bewegenden Kreuz möglichst akkurat zu folgen und mittels eines control sticks einen Kreis über das Kreuz zu positionieren.

  In der Lernphase werden die Pausen (Intertrial Intervalle) zwischen 10 Versuchsdurchgängen (Trials, mit je einer Minute Dauer) in vier Abstufungen variiert: 0 s (massiert), 20 s, 40 s und 60 s. Je 5 Personen werden den vier Bedingungen per Zufall zugewiesen. Nach den 10 Übungs-trials wird in einem Testtrial die AV erhoben: Innerhalb von 30 Sekunden wird gemessen, wie viele Sekunden die Vpn den Kreis mit dem Kreuz zur Deckung bringen können (Tracking-Zeit).

Intertrial Intervall			
0 s	20 s	40 s	60 s
4	18	24	16
6	13	19	17
10	15	21	13
9	11	16	23
11	13	15	21

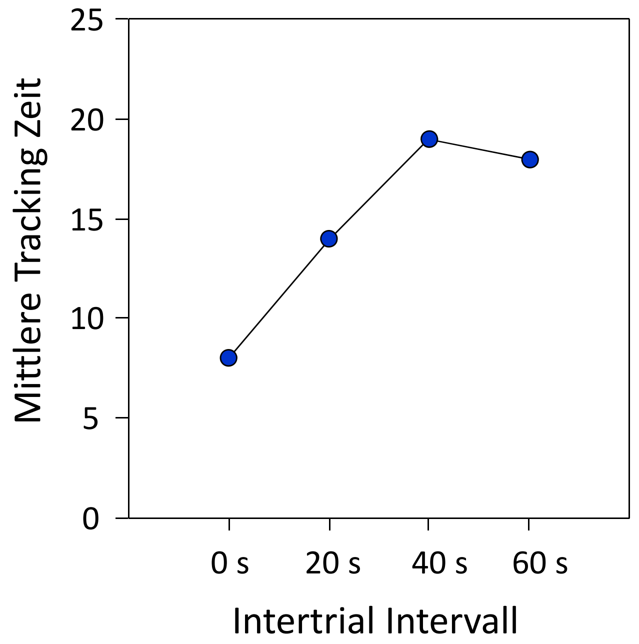
Multiple Mittelwertsvergleiche: Trends

➢ Zunächst können wir prüfen, ob sich das Ausmaß der Pausen (Intertrial Intervall) auf die Lernleistung (Tracking-Zeit) auswirkt. Wir führen eine einfaktorielle Varianzanalyse durch und erhalten:

ONEWAY deskriptive Statistiken

track

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
0 0 s	5	8,00	2,915	1,304	4,38	11,62	4	11
1 20 s	5	14,00	2,646	1,183	10,71	17,29	11	18
2 40 s	5	19,00	3,674	1,643	14,44	23,56	15	24
3 60 s	5	18,00	4,000	1,789	13,03	22,97	13	23
Gesamt	20	14,75	5,399	1,207	12,22	17,28	4	24



Test der Homogenität der Varianzen

track

Levene-Statistik	df1	df2	Signifikanz
,523	3	16	,673

ONEWAY ANOVA

track

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	373,750	3	124,583	11,074	,000
Innerhalb der Gruppen	180,000	16	11,250		
Gesamt	553,750	19			

Multiple Mittelwertsvergleiche: Trendtests

- Das besondere hier ist, das nicht nur die AV (Tracking-Zeit) metrisch ist, sondern auch die UV (Intertrial Intervall in Sekunden). Dies bedeutet, dass die Faktorstufen nicht nur nominal sind, sondern in der Weise geordnet sind, dass ihr Abstand eine Bedeutung hat.

- Weitere Beispiele für solche metrischen UVs könnten sein: Menge eines Medikaments; Lärm in Dezibel; Größe eines Reizes; Dauer, mit der eine Liste von Worten memoriert wird ...

- Nur bei solchen metrischen UVs kann man sich dafür interessieren, wie der funktionale Zusammenhang (Trend) zwischen der metrischen UV und AV ist. Potentielle Formen solcher Zusammenhänge haben wir schon im Kontext bivariater Zusammenhänge kennengelernt.

- Ein einfacher Zusammenhang ist der positive oder negative lineare Trend, der durch die Funktion $Y = a + b \cdot X$ beschrieben wird.

- Andere Formen des Zusammenhangs sind U-förmig oder umgekehrt U-förmig. Eine Funktion, die dies beschreiben kann, ist die Parabel, die durch die Funktion $Y = b \cdot X^2$ beschrieben wird. Man spricht deshalb auch von einem quadratischen Trend.

- Linearer und quadratischer Trend sind die beiden am häufigsten untersuchten Formen von Zusammenhängen; sie sind Spezialfälle der polynomischen Funktionen (Polynome)

$$Y = b_0 + b_1 \cdot X + b_2 \cdot X^2 + b_3 \cdot X^3 + b_4 \cdot X^4 + \ldots$$

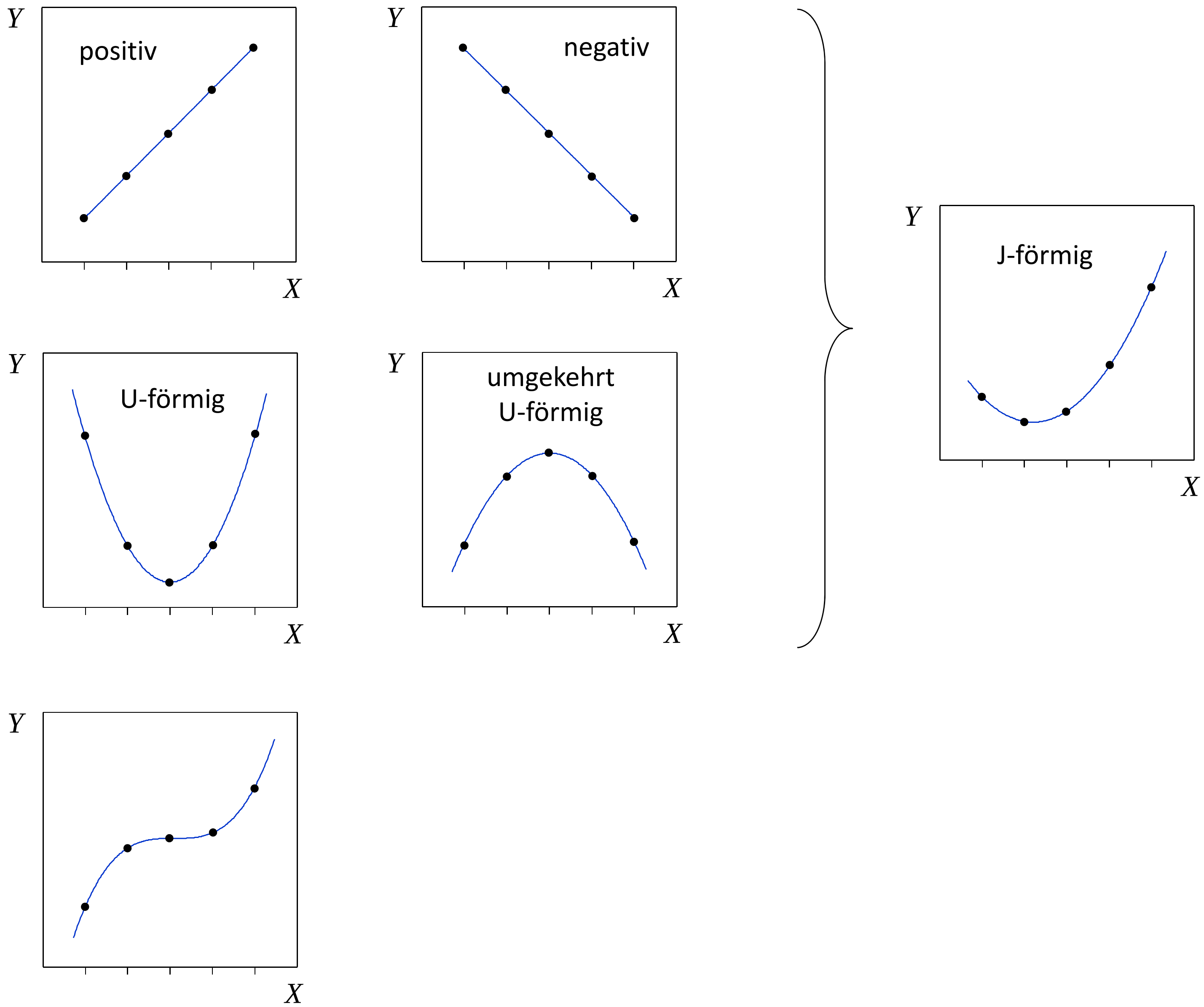
Multiple Mittelwertsvergleiche: Trendtests

Linearer Trend

Quadratischer Trend

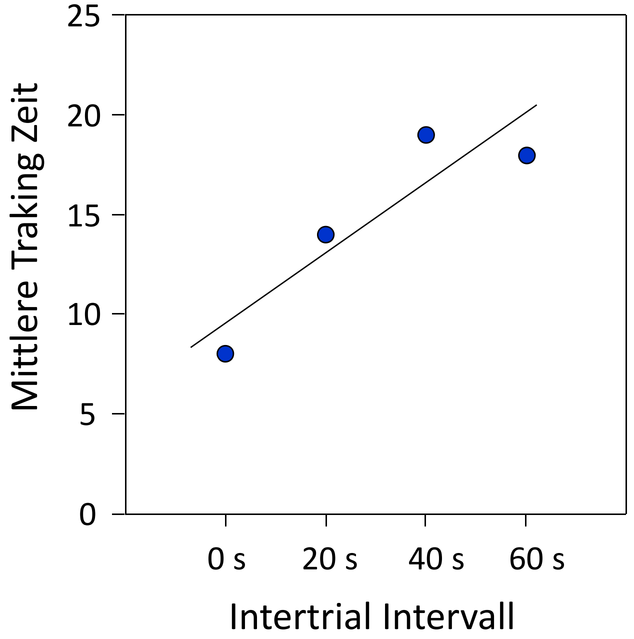
Kubischer Trend

...



Multiple Mittelwertsvergleiche: Trendtests

- Im Beispiel könnte man vermuten, dass ein positiver linearer Trend zwischen dem Intertrial Intervall und der Tracking-Zeit besteht, d.h. je länger die Pausen, desto besser die Lern-leistung. (Es ist einsichtig, dass dies theoretisch nur innerhalb eines bestimmten Werte-bereichs der UV eine sinnvolle Hypothese darstellt.)



- Die Hypothese, dass ein linearer Trend existiert, kann an-hand eines geplanten Kontrasts geprüft werden. Dabei kann die Quadratsumme wie allgemein bei a priori Kon-trasten bestimmt werden, d.h. (für gleiches $n_\bullet$)

$$QS_{linear} = \frac{n_\bullet \cdot \left(\sum_{j=1}^{m} c_j \cdot \bar{x}_j \right)^2}{\sum_{j=1}^{m} c_j^2} \quad \text{mit} \quad df_{linear} = 1$$

wobei hier Koeffizienten für die (vier) Mittel verwendet werden müssen, die unsere Hypothese eines linearen Trends adäquat widerspiegeln. Im Beispiel könnten dies z.B. die Koeffizienten $c_1 = -3$, $c_2 = -1$, $c_3 = 1$ und $c_4 = 3$ sein. Für sie gilt, dass sie linear ansteigen und dass ihre Summe gleich 0 ist.

Multiple Mittelwertsvergleiche: Trendtests

- Dann ist wieder die F-Prüfgröße zu bestimmen:

$$F = \frac{MQ_{linear}}{MQ_{in}} \quad \text{mit} \quad df_1 = 1 \quad \text{und} \quad df_2 = n - m$$

Intertrial Intervall			
0 s	20 s	40 s	60 s
$\bar{x}_1 = 8$	$\bar{x}_2 = 14$	$\bar{x}_3 = 19$	$\bar{x}_4 = 18$

ONEWAY ANOVA

track

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	373,750	3	124,583	11,074	,000
Innerhalb der Gruppen	180,000	16	11,250		
Gesamt	553,750	19			

- Im Beispiel:

$$QS_{linear} = \frac{n_{\bullet} \cdot \left(\sum_{j=1}^{m} c_j \cdot \bar{x}_j \right)^2}{\sum_{j=1}^{m} c_j^2}$$

$$= \frac{5 \cdot (-3 \cdot 8 - 1 \cdot 14 + 1 \cdot 19 + 3 \cdot 18)^2}{(-3)^2 + (-1)^2 + 1^2 + 3^2} = \frac{5 \cdot 35^2}{20} = 306.25 = MQ_{linear}$$

$$F = \frac{MQ_{linear}}{MQ_{in}} = \frac{306.25}{11.25} = 27.22 \qquad F_{crit} = F_{1;n-m;1-\alpha} = F_{1;16;0.95} = 4.49$$

- Es besteht also ein statistisch signifikanter linearer Trend. An den Mittelwerten bzw. der Grafik erkennt man, dass es sich um einen positiven linearen Trend handelt.

Multiple Mittelwertsvergleiche: Trendtests

- Man kann nun prüfen, ob der lineare Trend die erklärbare Varianz erschöpfend erklärt oder – anders formuliert – wie viel der potentiell erklärbaren Treatmentvarianz noch verbleibt, wenn man den linearen Kontrast (man sagt auch: die lineare Komponente) eliminiert hat:

$$QS_{res} = QS_{zw} - QS_{linear} = 373.75 - 306.25 = 67.50 \quad \text{mit} \quad df_{res} = df_{zw} - df_{linear} = (m-1) - 1 = 2$$

- Eine inferenzstatistische Prüfung der verbliebenen, *res*idualen Varianz ergibt kein statistisch signifikantes Ergebnis ($F < F_{crit}$):

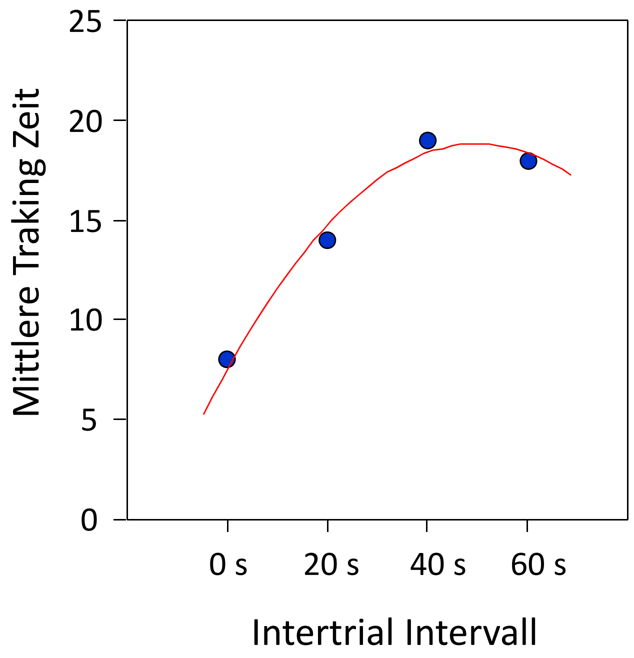
$$F = \frac{MQ_{res}}{MQ_{in}} = \frac{\frac{QS_{res}}{df_{res}}}{MQ_{in}} = \frac{\frac{67.5}{2}}{11.25} = 3.00 \qquad F_{crit} = F_{df_{res};n-m;1-\alpha} = F_{2;16;0.95} = 3.63$$

- Dieses Ergebnis ist so deutbar, dass die noch verbliebenen zum linearen Kontrast orthogonalen Kontraste <u>zusammen</u> keinen statistisch signifikanten Beitrag mehr leisten.

- Bei $df_{zw} = m - 1 = 3$ möglichen orthogonalen Kontrasten sind dies im Beispiel noch zusätzlich der quadratische und der kubische Trend.

- Da es aber theoretisch plausibel ist, dass zu lange Pausen möglicherweise wieder von Nachteil sein könnten (und zudem auch der letzte Mittelwert wieder einen kleinen Abfall zeigt), prüfen wir analog auch noch den quadratischen Trend.

- Dabei stellt sich wieder die Frage, wie die geeigneten Gewichte aussehen. Diese lassen sich einschlägigen Tabellen entnehmen (vgl. nächste Folie bzw. umfassender Fisher & Yates, 1979): In diesem Fall sind $c_1 = 1$, $c_2 = -1$, $c_3 = -1$ und $c_4 = 1$.



- Die Orthogonalitätsprüfung zeigt, dass der quadratische Kontrast unabhängig von dem linearen Kontrast ist:

$$(-3)\cdot 1 + (-1)\cdot(-1) + 1\cdot(-1) + 3\cdot 1 = 0$$

(Auch der entsprechende kubische Kontrast ist wiederum orthogonal zu beiden, so dass die drei Vergleiche einen Satz orthogonaler Vergleiche bilden.)

➢ Gewichte für orthogonale Polynome (nur bis kubisch und $m = 6$) bei gleichabständiger UV:

m	Komponente	c_1	c_2	c_3	c_4	c_5	c_6	$\sum c_j^2$
3	linear	–1	0	1				2
	quadratisch	1	–2	1				6
4	linear	–3	–1	1	3			20
	quadratisch	1	–1	–1	1			4
	kubisch	–1	3	–3	1			20
5	linear	–2	–1	0	1	2		10
	quadratisch	2	–1	–2	–1	2		14
	kubisch	–1	2	0	–2	1		10
6	linear	–5	–3	–1	1	3	5	70
	quadratisch	5	–1	–4	–4	–1	5	84
	kubisch	–5	7	4	–4	–7	5	180

Multiple Mittelwertsvergleiche: Trendtests

- Es resultiert

$$QS_{quadratisch} = \frac{n_\bullet \cdot \left(\sum_{j=1}^{m} c_j \cdot \bar{x}_j \right)^2}{\sum_{j=1}^{m} c_j^2} = \frac{5 \cdot (1 \cdot 8 - 1 \cdot 14 - 1 \cdot 19 + 1 \cdot 18)^2}{1^2 + (-1)^2 + (-1)^2 + 1^2} = \frac{5 \cdot (-7)^2}{4} = 61.25 = MQ_{quadratisch}$$

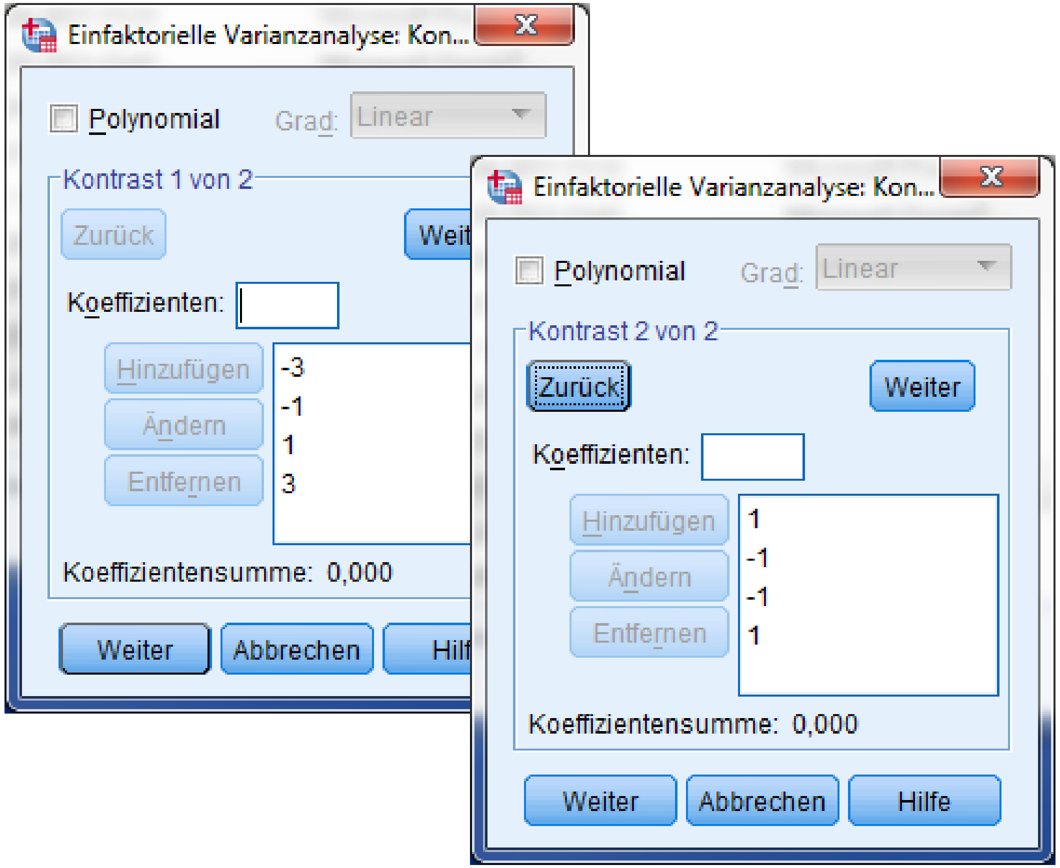
$$F = \frac{MQ_{quadratisch}}{MQ_{in}} = \frac{61.25}{11.25} = 5.44 \qquad F_{crit} = F_{1;n-m;1-\alpha} = F_{1;16;0.95} = 4.49$$

- Erstaunlicherweise – d.h. obwohl der F-Wert der Residualvarianz eben statistisch insignifikant ausfiel – ergibt sich eine statistisch signifikante quadratische Komponente: $F(1, 16) = 5.44 > F_{crit} = 4.49$. Dies ist darauf zurückzuführen, dass der F-Test der Residualvarianz den mittleren Effekt aller verbleibenden orthogonalen Vergleiche prüft. Hier wurde also der Effekt der quadratischen Komponente durch die geringe kubische Komponente maskiert, für die hier gilt:

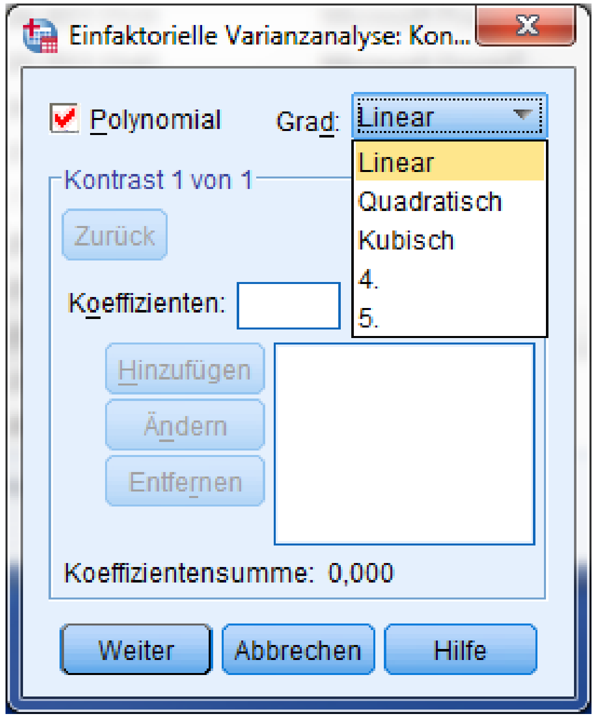
$$QS_{kubisch} = QS_{zw} - QS_{linear} - QS_{quadratisch} = 373.75 - 306.25 - 61.25 = 6.25$$

Multiple Mittelwertsvergleiche: Trendtests per SPSS

In der Prozedur Oneway können Trendtests im Dialog (`Kontraste…`) auf vertraute Weise durch die Eingabe der Kontraste (in diesem Fall die c-Koeffizienten des linearen und quadratischen Kontrasts) angefordert werden.



Alternativ kann man einfacher auch ohne Kenntnis der Koeffizienten die Option „Polynomial“ aktivieren und dann den maximalen Kontrast, den man betrachten möchte unter „Grad“ auswählen.



Multiple Mittelwertsvergleiche: Trendtests per SPSS

Kontrast-Koeffizienten

Kontrast	intertr			
	0 0 s	1 20 s	2 40 s	3 60 s
1	-3	-1	1	3
2	1	-1	-1	1

linear
quadratisch

$$F = t^2 = 5.217^2 = 27.22$$

$$F = t^2 = (-2.333)^2 = 5.44$$

Kontrast-Tests

		Kontrast	Kontrastwert	Standardfehler	T	df	Signifikanz (2-seitig)
track	Varianzen sind gleich	1	35,00	6,708	5,217	16	,000
		2	-7,00	3,000	-2,333	16	,033
	Varianzen sind nicht gleich	1	35,00	6,943	5,041	8,663	,001
		2	-7,00	3,000	-2,333	14,477	,035

linear
quadratisch

Durch die explizite Eingabe der Gewichte erhalten wir den vertrauten Output, der in der oberen Tabelle nochmals die Koeffizienten für beide Vergleiche ausgibt und in der unteren Tabelle die in diesem Beispiel eingegebenen linearen und quadratischen Kontraste anhand des *t*-Tests prüft.

ONEWAY ANOVA

track

			Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	(Kombiniert)		373,750	3	124,583	11,074	,000
	Linearer Term	Kontrast	306,250	1	306,250	27,222	,000
		Abweichung	67,500	2	33,750	3,000	,078
	Quadratischer Term	Kontrast	61,250	1	61,250	5,444	,033
		Abweichung	6,250	1	6,250	,556	,467
Innerhalb der Gruppen			180,000	16	11,250		
Gesamt			553,750	19			

Durch Anwahl der Option „Polynomial" mit der Festlegung „Quadratisch" erhalten wir den obigen Output.

In den Zeilen finden sich von oben nach unten QS, df, MQ, F- und p-Wert des Treatments (zwischen), des linearen Kontrasts, der residualen Varianz nach der Extraktion des linearen Kontrasts („Abweichung"), des quadratischen Kontrasts, der residualen Varianz nach der Extraktion des linearen und quadratischen Kontrasts („Abweichung") sowie innerhalb und Gesamt.

- Plant man eine Studie, bei der der Trend von Interesse ist, so ist zu beachten:
  - Es sollte der Wertebereich, über den eine Aussage gemacht werden soll, durch die Faktorstufen abgedeckt werden. Eine Extrapolation darüber hinaus (im Beispiel über Pausen, die länger als eine Minute sind) sind problematisch.
  - Meist ist eine gleichabständige Wahl der Faktorstufen sinnvoll (wie im Beispiel 20 Sekunden). Man kann davon abweichen, wenn man erwartet, dass sich in einem bestimmten Wertebereich des Faktors besonders starke Änderungen ergeben. Dann legt man dort die Faktorstufen enger. Sind die Faktorstufen nicht mehr gleichabständig, müssen die Koeffizienten auf andere (kompliziertere) Weise bestimmt werden (vgl. Keppel & Wickens, 2004, S. 105f). Die SPSS-Oneway Prozedur berücksichtigt die Abständigkeit der Faktorstufen über die Codierungen der UV.
  - Je komplexer der erwartete Trend, desto mehr Faktorstufen sollte man wählen (meist werden mindestens 5 empfohlen, 3 sind Minimum).

- Neben den hier betrachteten polynomischen Trends sind auch andere Zusammenhänge von Interesse (z.B. monotone oder solche, die sich einer Asymptote annähern). Prinzipiell sind solche Trends auch mittels Kontrastanalyse zu analysieren; dabei ergeben sich aber häufiger Probleme bei der Festlegung der Gewichte (Keppel & Wickens, 2004, S. 107ff).

Zitierte Quellen:

- 🕮 Fisher, R. A. & Yates, F. (1979). *Statistical tables for biological, agricultural and medical research* (6th ed.). London: Longman.
- 🕮 Howell, D. C. (2010). *Statistical methods of psychology* (7th ed.). Belmont, CA: Wadsworth Cengage.
- 🕮 Stevens, J. P. (2007). *Intermediate statistics. A modern approach* (3rd ed.). New York: Erlbaum.