STECKBRIEF

Einfaktorielle Varianzanalyse

WORUM GEHT'S?
Um Mittelwerts-Unterschiede bezüglich einer AV bei mindestens drei Gruppen, d. h. um Unterschiede in der AV abhängig von einem Faktor mit mindestens drei Faktorstufen.

Dabei wird bei jeder Person nur einmal gemessen bzw. die AV erhoben.
Es handelt sich daher um ein Between-Subjects-Design ohne Messwiederholung.

IN WELCHER WELT DER STATISTIK SIND WIR?
Inferenzstatistik – in der Welt der Unterschiede, d. h. bei den Unterschiedshypothesen.

BEISPIELE
Drei verschiedene Feinmotorik-Trainings unterscheiden sich in ihrer Wirksamkeit. (ungerichtet)

Die Art des Tiers bei einer tiergestützten Therapie hat einen Einfluss auf den Rückgang von ADHS-Symptomen. (ungerichtet)

Es bestehen Unterschiede in der Arbeitszufriedenheit bei fünf verschiedenen Berufsgruppen. (ungerichtet)

HINWEISE:
- Es können **ausschließlich ungerichtete Hypothesen** getestet werden!
- Das Ergebnis ist ein empirischer *F*-Wert, der in die *F*-Verteilung fällt, die dazugehörige Testverteilung

VORAUSSETZUNGEN
- Die UV ist eine kategoriale Variable
- Die AV ist metrisch und in jeder Faktorstufe normalverteilt
- Unkorrelierte Fehler
- Die Gruppengrößen sollten idealerweise annähernd gleich sein
- Varianzhomogenität der AV in allen Gruppen

WICHTIG:
Im Allgemeinen ist die einfaktorielle ANOVA robust gegenüber Voraussetzungs-Verletzungen. Dies gilt insbesondere dann, wenn du mindestens 30 Personen in jeder Faktorstufe hast und die Personenanzahl in jeder Stufe relativ ähnlich ist.

Wenn du eine kleine Stichprobe hast (unter $N = 30$) und / oder nicht ganz klar ist, ob deine AV wirklich metrisch und normalverteilt ist, verwendest du den Kruskal-Wallis-Test.

Bei ungleichen Varianzen kannst du auch eine robuste ANOVA nach Welch rechnen.

Bei ungleichen Varianzen plus unterschiedlichen Gruppengrößen empfiehlt sich der Brown-Forsythe-Test.

ALTERNATIVES NON-PARAMETRISCHES VERFAHREN
Kruskal-Wallis-Test = *H*-Test = Rangvarianzanalyse

BERECHNUNG PER HAND

$$F = \frac{QS_{zw} : df_{zw}}{QS_{inn} : df_{inn}} = \frac{MQ_{zw}}{MQ_{inn}}$$

QUADRATSUMME ZWISCHEN

$$QS_{zw} = \sum_{j=1}^{k} n_j \cdot (\bar{x}_j - \bar{x})^2$$

QUADRATSUMME INNERHALB

$$QS_{inn} = \sum_{j=1}^{k} \sum_{m=1}^{n_j} (x_{mj} - \bar{x}_j)^2$$

QUADRATSUMME TOTAL

$$QS_{total} = QS_{zw} + QS_{inn}$$

$$QS_{total} = \sum_{j=1}^{k} \sum_{m=1}^{n_j} (x_{mj} - \bar{x})^2$$

Bestimmung der Freiheitsgrade:

$$df_{zw} = k - 1$$

$$df_{inn} = n - k$$

EFFEKTSTÄRKE ETA-QUADRAT

Formel:

$$\eta^2 = \frac{QS_{zw}}{QS_{total}}$$

Konventionen für die Interpretation (Cohen, 1988):

- Kleiner Effekt = 0,01
- Mittlerer Effekt = 0,06
- Starker Effekt = 0,14

EFFEKTSTÄRKE *f*

Neben Eta-Quadrat kann auch die Effektstärke *f* berechnet werden, die du z. B. für die Poweranalyse mit G*Power brauchst.

Eta-Quadrat lässt sich mit dieser Formel ganz leicht in *f* umwandeln:

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

ANOVA TABELLE BZW. TAFEL DER ANOVA

Varianz-quelle	QS	*df*	MQ	*F*-Wert	*p*-Wert	η^2
Zwischen	QS_{zw}	$df_{zw} = k - 1$	$MQ_{zw} = \frac{QS_{zw}}{df_{zw}}$	$F = \frac{MQ_{zw}}{MQ_{inn}}$	p	$\frac{QS_{zw}}{QS_{total}}$
Innerhalb	QS_{inn}	$df_{inn} = n - k$	$MQ_{inn} = \frac{QS_{inn}}{df_{inn}}$			
Total	QS_{total}	$N - 1$				

BERECHNUNG MIT SPSS

H_1: Die Paarungswilligkeit auf einer Ü30-Party unterscheidet sich abhängig vom konsumierten Getränk. (ungerichtet)

Deskriptive Statistik

Paarungswilligkeit

	N	Mittelwert	Std.-Abweichung	Std.-Fehler	95% Konfidenzintervall des Mittelwerts Untergrenze	Obergrenze	Minimum	Maximum
kein Alkohol	10	3,6000	1,71270	,54160	2,3748	4,8252	,00	6,00
Prosecco	11	8,6364	1,12006	,33771	7,8839	9,3888	7,00	10,00
Wodka	9	2,7778	1,64148	,54716	1,5160	4,0395	1,00	6,00
Gesamt	30	5,2000	3,04450	,55585	4,0632	6,3368	,00	10,00

Tests der Varianzhomogenität

		Levene-Statistik	df1	df2	Sig.
Paarungswilligkeit	Basiert auf dem Mittelwert	,616	2	27	,548
	Basiert auf dem Median	,267	2	27	,767
	Basierend auf dem Median und mit angepaßten df	,267	2	22,385	,768
	Basiert auf dem getrimmten Mittel	,563	2	27	,576

ANOVA

Paarungswilligkeit

	Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zwischen den Gruppen	208,299	2	104,149	46,479	<,001
Innerhalb der Gruppen	60,501	27	2,241		
Gesamt	268,800	29			

ANOVA-Effektgrößen[a]

		Punktschätzung	95% Konfidenzintervall Unterer	Oberer
Paarungswilligkeit	Eta-Quadrat	,775	,570	,845
	Epsilon-Quadrat	,758	,539	,834
	Omega-Quadrat, fester Effekt	,752	,530	,829
	Omega-Quadrat, Zufallseffekt	,603	,361	,708

a. Eta-Quadrat und Epsilon-Quadrat werden basierend auf dem Modell mit festen Effekten geschätzt.

SPSS WIRFT DIE FOLGENDEN BEREICHE AUS:

- Zunächst die deskriptiven Statistiken, wovon uns besonders die Gruppenmittelwerte interessieren.
- Dann folgt der Levene-Test zur Überprüfung, ob die Voraussetzung der Varianzgleichheit gegeben ist.
- An dritter Stelle siehst du die Ausgabe für die ANOVA, d. h. den globalen *F*-Test.
- Zuletzt folgen die verfügbaren Effektgrößen.

SO LIEST DU DIE ERGEBNISSE AB

1. Schritt: Mittelwerte ansehen

Bei „Deskriptive Statistik" kannst du dir in der Spalte **Mittelwert** ein Bild davon machen, ob und wie sich die Mittelwerte voneinander unterscheiden.

Im Beispiel siehst du, dass die durchschnittliche Paarungswilligkeit bei den Personen, die Prosecco konsumierten, am höchsten ist (8,64), und bei den Wodka-Konsumenten am geringsten (2,78).

2. Schritt: Varianzhomogenität mit dem Levene-Test prüfen

Als nächstes überprüfst du bei **Tests der Varianzhomogenität**, ob die Voraussetzung der Varianzgleichheit in jeder Gruppe vorliegt.

Diese testen die Nullhypothese, dass die Varianzen in allen Gruppen gleich sind.
Ein signifikantes Ergebnis (p < 0,05) bedeutet, dass KEINE Varianzhomogenität vorliegt!

Der p-Wert des Levene-Tests basierend auf dem Median ist 0,77 und somit nicht signifikant. Daher liegt Varianzgleichheit in allen Gruppen vor.

Dies sagt dir auch, dass du bei den Post-hoc-Tests bei demjenigen Verfahren ablesen musst, das für Varianzhomogenität geeignet ist.

Läge keine Varianzgleichheit vor, müsstest du entsprechend bei dem Post-hoc-Test nachschauen, der für Varianzheterogenität geeignet ist, wie z. B. Games-Howell.

3. Schritt: Ergebnis der ANOVA ablesen

Du liest das **Ergebnis des *F*-Tests** in der ersten Zeile bei **Sig.** ab.

Das p < 0,001 sagt dir, dass sich mindestens zwei der drei Mittelwerte signifikant voneinander unterscheiden. Die Angaben zum empirischen F-Wert (46,48) sowie zu den Freiheitsgraden (df = 2, 27) findest du links vom p-Wert.

4. Schritt: Bei signifikantem Ergebnis die Effektstärke bestimmen

Die Effektstärke Eta-Quadrat lässt sich in der ersten Zeile der Effektgrößen unter **Punktschätzung** ablesen.

Ein Eta-Quadrat von 0,78 ist ein sehr starker Effekt.

POST-HOC-TESTS

Mehrere Vergleiche

Abhängige Variable: Paarungswilligkeit

	(I) Alkoholart	(J) Alkoholart	Mittelwertdifferenz (I-J)	Std.-Fehler	Sig.	95% Konfidenzintervall Untergrenze	Obergrenze
Tukey-HSD	kein Alkohol	Prosecco	-5,03636*	,65405	<,001	-6,6580	-3,4147
		Wodka	,82222	,68779	,466	-,8831	2,5275
	Prosecco	kein Alkohol	5,03636*	,65405	<,001	3,4147	6,6580
		Wodka	5,85859*	,67282	<,001	4,1904	7,5268
	Wodka	kein Alkohol	-,82222	,68779	,466	-2,5275	,8831
		Prosecco	-5,85859*	,67282	<,001	-7,5268	-4,1904
Games-Howell	kein Alkohol	Prosecco	-5,03636*	,63827	<,001	-6,6910	-3,3817
		Wodka	,82222	,76988	,546	-1,1537	2,7981
	Prosecco	kein Alkohol	5,03636*	,63827	<,001	3,3817	6,6910
		Wodka	5,85859*	,64299	<,001	4,1710	7,5461
	Wodka	kein Alkohol	-,82222	,76988	,546	-2,7981	1,1537
		Prosecco	-5,85859*	,64299	<,001	-7,5461	-4,1710

*. Die Mittelwertdifferenz ist in Stufe 0.05 signifikant.

SO LIEST DU DIE ERGEBNISSE AB

1. Schritt: In der Spalte „Sig.“ signifikante *p*-Werte suchen
Hier sehen wir in der ersten und vierten Zeile signifikante Ergebnisse: zwischen kein Alkohol und Prosecco sowie zwischen Prosecco und Wodka, jeweils $p < 0{,}001$.

2. Schritt: In der Spalte „Mittelwertdifferenz“ die Größe des signifikanten Unterschieds ablesen
Eine Mittelwertsdifferenz von -5,04 in der ersten Zeile bedeutet, dass die durchschnittliche Paarungswilligkeit der Personen, die keinen Alkohol konsumierten, um 5,04 Punkte niedriger liegt als bei den Prosecco-Trinkenden.

Zudem ist die durchschnittliche Paarungswilligkeit der Prosecco-Fraktion um 5,9 Punkte höher als bei den Wodka-Trinkenden.

BERECHNUNG MIT R

SO LIEST DU DIE ERGEBNISSE AB:

1. Schritt: Mittelwerte ablesen

```
 Descriptive statistics by group
group: Kein Alkohol
   vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
X1    1 10  3.6 1.71      4    3.75 1.48   0   6     6 -0.63    -0.49 0.54
---------------------------------------------------------------------------
group: Prosecco
   vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
X1    1 11 8.64 1.12      9    8.67 1.48   7  10     3 -0.12    -1.54 0.34
---------------------------------------------------------------------------
group: Wodka
   vars n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 9 2.78 1.64      2    2.78 1.48   1   6     5 0.61    -0.96 0.55
```

Bei **mean** kannst du dir ein Bild davon machen, ob und wie sich die Mittelwerte voneinander unterscheiden.

Im Beispiel siehst du, dass die durchschnittliche Paarungswilligkeit bei den Personen, die Prosecco konsumierten, am höchsten ist (8,64), und bei den Wodka-Konsumenten am geringsten (2,78).

2. Schritt: Varianzhomogenität mit dem Levene-Test prüfen

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.2675 0.7673
      27
```

Als nächstes überprüfst du mit dem **Levene-Test basierend auf dem Median,** ob die Voraussetzung der Varianzgleichheit in jeder Gruppe vorliegt.

Dieser testet die Nullhypothese, dass die Varianzen in allen Gruppen gleich sind.
Ein signifikantes Ergebnis ($p < 0{,}05$) bedeutet, dass KEINE Varianzhomogenität vorliegt!

Der *p*-Wert des Levene-Tests basierend auf dem Median ist 0,77 und somit nicht signifikant. Daher liegt Varianzgleichheit in allen Gruppen vor.

Dies sagt dir auch, dass du später beim Ablesen der Post-hoc-Tests bei demjenigen Verfahren ablesen musst, das für Varianzhomogenität geeignet ist.
Läge keine Varianzgleichheit vor, müsstest du entsprechend bei dem Post-hoc-Test nachschauen, der für Varianzheterogenität geeignet ist, wie z. B. Games-Howell.

3. Schritt: Ergebnis der ANOVA ablesen

```
Anova Table (Type 3 tests)

Response: paarw
  Effect    df  MSE         F  ges p.value
1    alc 2, 27 2.24 46.48 *** .775  <.001
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘+’ 0.1 ‘ ’ 1
```

Du liest das **Ergebnis des *F*-Tests** in der Zeile ab, in der deine Gruppenvariable steht („alc“).
Den *p*-Wert findest du direkt unter „p.value“.

Das $p < 0{,}001$ sagt dir, dass sich mindestens zwei der drei Mittelwerte signifikant voneinander unterscheiden. Die Angaben zum empirischen *F*-Wert (46,48) sowie zu den Freiheitsgraden ($df = 2, 27$) findest du links von Eta-Quadrat.

4. Schritt: Bei signifikantem Ergebnis die Effektstärke bestimmen
Die Effektstärke **Eta-Quadrat** lässt sich unter „ges“ für *generalized eta-squared* ablesen: $\eta^2 = 0{,}78$ ist ein sehr starker Effekt.

POST-HOC-TESTS

SO LIEST DU DIE ERGEBNISSE AB

1. Schritt: Signifikante *p*-Werte suchen

```
 contrast                estimate    SE df t.ratio p.value
 Kein Alkohol - Prosecco   -5.036 0.654 27  -7.700 <.0001
 Kein Alkohol - Wodka       0.822 0.688 27   1.195  0.4660
 Prosecco - Wodka           5.859 0.673 27   8.708 <.0001

P value adjustment: tukey method for comparing a family of 3 estimates
```

In der Spalte **estimate** siehst du die Mittelwertsdifferenzen: eine Mittelwertsdifferenz von -5,04 in der ersten Zeile bedeutet, dass die durchschnittliche Paarungswilligkeit der Personen, die keinen Alkohol konsumierten, um 5,04 Punkte niedriger liegt als bei den Prosecco-Trinkenden.

Zudem ist die durchschnittliche Paarungswilligkeit der Prosecco-Fraktion um 5,9 Punkte höher als bei den Wodka-Trinkenden.

Den *p*-Wert liest du in der Spalte **p.value** ab: das $p < 0{,}001$ zeigt jeweils signifikante Mittelwertsunterschiede zwischen den Gruppen kein Alkohol und Prosecco sowie zwischen Prosecco und Wodka.

WENN VARIANZHETEROGENITÄT VORLIEGT

Wäre der Levene-Test signifikant geworden, könntest du beispielsweise den Post-hoc-Test nach **Games-Howell** anfordern, der für unser Beispiel so aussieht:

```
  .y.   group1       group2   estimate conf.low conf.high       p.adj p.adj.signif
* <chr> <chr>        <chr>       <dbl>    <dbl>     <dbl>       <dbl> <chr>
1 paarw Kein Alkohol Prosecco    5.04      3.38      6.69 0.00000257  ****
2 paarw Kein Alkohol Wodka      -0.822    -2.80      1.15 0.546       ns
3 paarw Prosecco     Wodka      -5.86     -7.55     -4.17 0.000000999 ****
```

Auch hier liest du wieder die Mittelwertsdifferenz in der Spalte **estimate** ab und den *p*-Wert bei **p.adj**. Zudem wird dir in der letzten Spalte angezeigt, ob die Mittelwertsdifferenz signifikant ist.
„ns“ steht für nicht signifikant.
Das Ergebnis ist das gleiche wie bei Tukey.

ERGEBNIS KORREKT BERICHTEN

Im Gegensatz zur händischen Berechnung kommt bei der Auswertung mit Statistik-Programmen die Angabe des *p*-Wertes hinzu. Zudem werden meist die 95%-igen Konfidenzintervalle der Mittelwertsdifferenz aufgeführt. Die Angaben dazu findest du bei SPSS in den Spalten rechts neben „Sig.“. Bei R ist das nicht immer angegeben.

So sieht der korrekte Bericht aus:

Es bestehen signifikante Unterschiede in der Paarungswilligkeit abhängig vom konsumierten Getränk ($F_{(2,27)} = 46{,}48$, $p < 0{,}001$, $\eta^2 = 0{,}78$).
Der Tukey Post-hoc-Test zeigte einen signifikanten Unterschied ($p < 0{,}001$) in der Paarungswilligkeit zwischen denjenigen, die keinen Alkohol zu sich nahmen, und den Prosecco-Trinkenden: die durchschnittliche Paarungswilligkeit der Alkohol-Abstinenten lag signifikant niedriger (-5,04, 95%-CI[-6,66, -3,41]).
Zudem lag ein signifikanter Unterschied ($p < 0{,}001$) in der Paarungswilligkeit zwischen den Prosecco- und den Wodka-Trinkenden vor: die durchschnittliche Paarungswilligkeit der Prosecco-Trinkenden lag signifikant höher (5,86, 95%-CI[4,19, 7,53]).

INTERPRETATION FÜR TANTE ERNA

Wir haben untersucht, ob sich die Paarungswilligkeit auf einer Ü30-Party abhängig vom konsumierten Getränk unterscheidet.

Unsere Ergebnisse zeigten, dass es sehr starke signifikante Unterschiede gab: Menschen, die auf der Party Prosecco tranken, zeigten eine signifikant höhere Paarungswilligkeit als Menschen, die entweder keinen Alkohol oder Wodka tranken.