

Statistische Verfahren

Hier erfährst du, wie du das richtige statistische Verfahren auswählst

STATISTISCHE SIGNIFIKANZTESTS

Welches statistische Verfahren brauche ich?

„Welches **statistische Verfahren** brauche ich?“ – Das ist die mit Abstand häufigste Frage, die uns in der statistischen Beratung gestellt wird. Im Wesentlichen hängt die Auswahl des statistischen Verfahrens vom **Skalenniveau** deiner Variablen ab. Das Skalenniveau kann **nominal, ordinal oder metrisch** sein. Nominale Variablen können zudem **dichotom** sein, d.h. dass sie nur zwei verschiedene Ausprägungen haben. Nachfolgend findest du die am häufigsten angewendeten Verfahren und welche Arten von Hypothesen sich mit ihnen jeweils testen lassen. Für ordinal skalierte abhängige Variablen lassen sich jeweils die alternativen Verfahren einsetzen. Welches statistische Verfahren du anwenden solltest, kannst du vermutlich mithilfe der nachfolgenden Auflistung der häufigsten statistischen Verfahren bestimmen.

Zur Erinnerung: Die unabhängige Variable (UV) ist die beeinflussende Variable und die abhängige Variable (AV) ist die beeinflusste Variable.

Bitte beachte:

Die nachfolgende Auswahl an vorgestellten Testverfahren bildet die Basis statistischer Tests ab, mit denen grundlegende Zusammenhänge und Unterschiede auf Signifikanz geprüft werden können. Bitte beachte, dass zur Untersuchung bestimmter Fragestellungen aufwändigere Verfahren genutzt werden müssen. In diesen Fällen könnt ihr uns gerne eine Kontaktanfrage über tnb-statistics.de senden.

MITTELWERTDIFFERENZ VON ZWEI GRUPPEN TESTEN

T-Test für unabhängige Stichproben

Der t-Test für unabhängige Stichproben ist zum Testen von Hypothesen mit einer dichotomen unabhängigen Variable und einer metrischen abhängigen Variable geeignet.

Annahme ist außerdem, dass mindestens eine der folgenden Bedingungen erfüllt ist:

- Normalverteilung der abhängigen Variable
- Beide Stichprobengruppen sind jeweils größer als 30

Beispiel-Hypothese:

Frauen und Männer geben im Durchschnitt unterschiedlich viel Geld für Katzenfutter aus.

Nichtparametrisches Äquivalent: [Mann-Whitney-Test](#)

MITTELWERTDIFFERENZ VON ZWEI MESSZEITPUNKTEN TESTEN

T-Test für gebundene Stichproben

Der t-Test für gebundene Stichproben ist für Hypothesen geeignet, bei denen ein Unterschied zwischen zwei verschiedenen (metrischen) Variablenwerten getestet wird. In den meisten Fällen bezieht sich der Unterschied auf zwei verschiedene Messzeitpunkte derselben Variable.

Annahme ist außerdem, dass mindestens eine der folgenden Bedingungen erfüllt ist:

- Normalverteilung bei der abhängigen Variable
- Stichprobenumfang von über 30

Beispiel-Hypothese:

Die durchschnittlichen Ausgaben für Katzenfutter waren im Jahr 2020 höher als im Jahr 2019.

Nichtparametrisches Äquivalent: [Wilcoxon-Test](#)

MITTELWERTDIFFERENZEN VON MEHR ALS ZWEI GRUPPEN

Einfaktorielle Varianzanalyse (ANOVA)

Die einfaktorielle Varianzanalyse ist für Hypothesen geeignet, die eine nominale unabhängige Variable mit mehr als zwei verschiedenen Ausprägungen (Gruppen) und eine metrische abhängige Variable enthalten. Ein signifikantes Ergebnis weist darauf hin, dass sich die Mittelwerte mehrerer Gruppen unterscheiden. Allerdings kann die Varianzanalyse NICHT prüfen, welche der untersuchten Gruppen sich signifikant unterscheiden. Zu diesem Zweck werden im Zusammenhang mit Varianzanalysen häufig Post-Hoc-Tests eingesetzt.

Annahmen:

- Normalverteilung bei der abhängigen Variable ODER alle Gruppen > 30
- Varianzhomogenität (keine signifikanten Unterschiede zwischen den Varianzen der Gruppen)

Beispiel-Hypothese:

Je nach Art des Haustiers (Hund, Katze, Maus) fallen unterschiedlich hohe durchschnittliche Anschaffungskosten an.

Nichtparametrisches Äquivalent: [Kruskal-Wallis-Test](#)

ZUSAMMENHANG VON NOMINALEN MERKMALEN

Chi-Quadrat-Unabhängigkeitstest

Der Chi-Quadrat-Unabhängigkeitstest ist geeignet, um einen Zusammenhang zwischen zwei nominal skalierten Variablen zu untersuchen. Bei diesem Verfahren wird geprüft, ob die in der Stichprobe beobachteten Werte von den Erwartungswerten in einer Kreuztabelle signifikant abweichen. Theoretisch können die untersuchten Variablen beliebig viele Ausprägungen haben. Jedoch ist eine sinnvolle Interpretierbarkeit des Ergebnisses bei größeren Kreuztabellen häufig nicht mehr möglich, da bestimmte Kombinationen an Merkmalsausprägungen in solchen Fällen nur sehr geringe Häufigkeiten aufweisen.

Annahme ist außerdem, dass alle Erwartungswerte mindestens 5 betragen.

Beispiel-Hypothese:

Es besteht ein Zusammenhang zwischen der Hunderasse und der bevorzugten Hundefuttermarke der Hundebesitzer(innen).

ZUSAMMENHANG VON METRISCHEN VARIABLEN

Korrelationstest

Der Korrelationstest nach Pearson ist für Hypothesen geeignet, bei denen sowohl die unabhängige Variable als auch die abhängige Variable metrisch skaliert ist. Eine signifikante Korrelation bedeutet allerdings nicht automatisch, dass ein Kausalzusammenhang besteht.

Annahme ist, dass mindestens eine der folgenden Bedingungen erfüllt ist:

- Stichprobengröße von über 30
- Normalverteilung beider Variablen

Beispiel-Hypothese:

Es besteht ein Zusammenhang zwischen den jährlichen Ausgaben für Katzenfutter und den jährlichen Ausgaben für Katzenstreu.

Nichtparametrisches Äquivalent: Rangkorrelationstest nach Spearman

Für Zusammenhangsanalysen mit mehreren unabhängigen Variablen stehen neben besonderen Varianzanalysen vor allem Regressionsverfahren zur Verfügung. Die beiden wichtigsten sind die **lineare Regression** und die **logistische Regression**, die auf der nächsten Seite vorgestellt werden.

REGRESSIONSVERFAHREN

Lineare Regression

Die lineare Regression kann verwendet werden, um mehrere Hypothesen gleichzeitig zu prüfen, die sich jeweils auf dieselbe (metrische) abhängige Variable beziehen. Die unabhängigen Variablen können dabei metrisch skaliert oder dichotom sein. Zusätzlich können auch Kontrollvariablen im Modell berücksichtigt werden.

Annahmen der linearen Regression sind u.a.:

- Normalverteilung der Variablen ODER Stichprobenumfang von über 30
- Normalverteilung der Residuen mit dem Mittelwert 0
- Homoskedastizität (gleichmäßige Verteilung der Residuen)
- Möglichst geringe Korrelation der unabhängigen Variablen untereinander
- Bestmögliche Approximation wird durch einen linearen Zusammenhang erreicht

REGRESSIONSVERFAHREN

Logistische Regression

Die logistische Regression ist ein spezielles Regressions-Verfahren für dichotome abhängige Variablen. Die unabhängigen Variablen können jedes Skalenniveau annehmen, wobei ordinale und nominale Einflussvariablen zu mehreren Dummy-Variablen umcodiert werden.

Voraussetzungen der logistischen Regression sind:

- Dummy-codierte abhängige Variable (Format: 0 und 1)
- Möglichst geringe Korrelation der unabhängigen Variablen untereinander
- Möglichst großer Stichprobenumfang

Näherungsformel zur Bestimmung des benötigten Stichprobenumfangs einer logistischen Regression:

$$n = \frac{\text{Anzahl unabhängige Variablen} \cdot 10}{\text{Wahrscheinlichkeit, dass AV} = 1}$$