

## Antworten zu Kapitel 8: Grundlagen der Inferenzstatistik

(1) Wie ist beim Nullhypothesentest nach Fisher der  $p$ -Wert zu interpretieren? Kreuzen Sie die richtige(n) Antwort(en) an:

- ☐  $p$  ist die Wahrscheinlichkeit, mit der die Nullhypothese wahr ist.
- ☐  $p$  ist die Wahrscheinlichkeit, mit der die Nullhypothese falsch ist.
- ☐  $p$  ist die bedingte Wahrscheinlichkeit für die Gültigkeit der Nullhypothese, gegeben ein empirisches Ergebnis.
- ☒  $p$  ist die Wahrscheinlichkeit für ein empirisches (oder jedes noch extremer gegen die  $H_0$  sprechende) Ergebnis unter der Annahme, dass die  $H_0$  gilt.
- ☐  $p$  ist die Wahrscheinlichkeit, mit der ein Populationseffekt gleich 0 ist.
- ☐  $p$  ist die Wahrscheinlichkeit, mit der Ablehnung der Nullhypothese eine falsche Entscheidung zu treffen.
- ☐  $p$  ist die Wahrscheinlichkeit, mit der Ablehnung der Alternativhypothese eine falsche Entscheidung zu treffen.

(2) Kreuzen Sie die korrekten Antworten an:

Die Teststärke ...

- ☐ gibt die Wahrscheinlichkeit an, sich fälschlicherweise für die  $H_0$  zu entscheiden.
- ☐ gibt die Wahrscheinlichkeit an, sich fälschlicherweise für die  $H_1$  zu entscheiden.
- ☒ gibt die Wahrscheinlichkeit an, einen Effekt einer vorher definierten Größe zu finden, falls dieser tatsächlich existiert.
- ☒ entspricht der Fläche  $1 - \beta$  der Stichprobenkennwerteverteilung unter der Alternativhypothese.
- ☒ hängt u. a. von der Größe des spezifizierten Populationseffektes ab.

Ein großer Stichprobenumfang ...

- ☐ ist notwendig, um große Effekte aufzudecken.
- ☒ führt dazu, dass auch kleinere Effekte mit größerer Wahrscheinlichkeit aufgedeckt werden.
- ☐ bringt dem Test eine kleine Teststärke ein.
- ☐ führt dazu, dass die Standardabweichung der Stichprobenkennwerteverteilung groß wird.
- ☐ führt dazu, dass die Fehlerwahrscheinlichkeiten  $\alpha$  und  $\beta$  kleiner werden.

*Anmerkung: Diese Antwort ist richtig, wenn zuvor ein bestimmtes Verhältnis zwischen  $\alpha$  und  $\beta$  festgelegt wurde. Im Regelfall wird jedoch  $\alpha$  vorher festgelegt und ist dann natürlich auch nicht durch den Stichprobenumfang beeinflusst. Wenn  $\alpha$  vorher festgelegt wurde, wird – unter ansonsten gleichen Bedingungen –  $\beta$  umso kleiner, je größer der Stichprobenumfang ist.*

Ein kleiner postulierter Effekt in der Population ...

- ☐ führt immer dazu, dass die Fehlerwahrscheinlichkeit  $\alpha$  sehr groß ist.
- ☒ kann am besten dann aufgedeckt werden, wenn die Standardabweichung der Stichprobenkennwerteverteilung klein ist.
- ☐ führt dazu, dass der statistische Test eine große Teststärke hat.
- ☐ führt dazu, dass man sich fälschlicherweise für die Alternativhypothese entscheidet.
- ☐ verringert von vornherein die Irrtumswahrscheinlichkeit  $\beta$ .

Beim  $t$ -Test wird bei konstantem Effekt und konstanter Stichprobengröße ...

- ☒  $\alpha$  umso größer, je kleiner  $\beta$  ist.
- ☒ die Teststärke umso größer, je kleiner  $\beta$  ist.
- ☐  $\beta$  umso größer, je größer  $\alpha$  ist.
- ☒  $\alpha$  umso kleiner, je kleiner die Teststärke ist.
- ☒  $\beta$  umso kleiner, je größer die Wahrscheinlichkeit ist, einen Effekt zu finden, falls dieser tatsächlich existiert.

**(3) In welchem Fall ist die Teststärke – bei gegebenem  $\alpha$ , gegebenem Effekt und gegebenem Stichprobenumfang  $n$  – größer: wenn der Test einseitig oder zweiseitig durchgeführt wird?**

Da man beim zweiseitigen Testen das Signifikanzniveau  $\alpha$  auf die beiden Seiten der  $t$ -Verteilung aufteilt und der Flächenanteil, der zu einer Seite hin abgeschnitten wird, dementsprechend nur  $\alpha/2$  beträgt, während die Teststärke sich demgegenüber immer nur auf eine Richtung des Effekts bezieht, ist die Teststärke beim zweiseitigen Test kleiner als beim einseitigen Test.

**(4) Welchen Vorteil und welchen Nachteil hat Cohens  $\delta$  gegenüber dem Effektstärkemaß  $\varepsilon$ ?**

Cohens  $\delta$  ermöglicht es, Effekte über unterschiedliche Untersuchungen hinweg miteinander vergleichbar zu machen, indem sie an der Standardabweichung des fraglichen Merkmals in der Population relativiert werden. Problematisch ist diese Standardisierung in all denjenigen Fällen, in denen unterschiedliche Populationsstandardabweichungen nichts mit der Metrik des Messinstruments zu tun haben, sondern vielmehr mit »echten«, inhaltlich bedeutsamen Homogenitätsunterschieden zwischen zwei Populationen.

**(5) Was besagt der zentrale Grenzwertsatz?**

Dem zentralen Grenzwertsatz zufolge nähert sich die Stichprobenkennwerteverteilung von Stichprobenmittelwerten aus Stichproben der Größe  $n$  mit zunehmendem  $n$  einer Normalverteilung an – unabhängig davon, wie das Merkmal in der Population verteilt ist. Voraussetzung dabei ist, dass die einzelnen Stichproben unabhängig voneinander gezogen wurden.

**(6) Nennen und erläutern Sie die Kriterien für die Qualität einer Parameterschätzung.**

Die Eignung von Statistiken als Schätzern von Parametern wird an vier Gütekriterien bemessen: Erwartungstreue, Konsistenz, Effizienz und Erschöpftheit.

Eine Statistik (Stichprobenkennwert) schätzt einen Parameter (Populationskennwert) *erwartungstreu*, wenn der Erwartungswert der Stichprobenkennwerteverteilung der Statistik mit dem Parameter identisch ist.

Eine Statistik heißt *konsistent*, wenn sie mit wachsender Stichprobengröße stochastisch gegen den Parameter konvergiert. Mit anderen Worten, die Wahrscheinlichkeit, dass die Statistik beliebig nahe an dem Parameter liegt, strebt mit wachsender Stichprobengröße gegen 1.

Eine Statistik ist als Schätzer eines Populationsparameters *effizient*, wenn sie den geringsten Standardfehler aller erwartungstreuen Schätzer aufweist.

Eine Statistik ist *erschöpfend* (suffizient oder exhaustiv), wenn sie alle in den Daten enthaltenen Informationen nutzt, so dass die Berechnung einer weiteren Statistik keine zusätzliche Information über den Parameter enthält.

**(7) Vervollständigen Sie den folgenden Satz: »Ein zweiseitiges 95 %-Konfidenzintervall für den Mittelwert mit der Untergrenze 2,5 und der Obergrenze 5 besagt, dass ...«**

... ein Bereich zwischen 2,5 und 5 mit einer Wahrscheinlichkeit von 95 % den »wahren« Mittelwert in der Population überdeckt.

**(8) In welchem Zusammenhang stehen der Signifikanztest und die Bestimmung des Konfidenzintervalls für den Mittelwert?**

Das Prinzip des Signifikanztests ist eng mit dem Konzept des Konfidenzintervalls verwandt: Die Nullhypothese  $H_0: \mu = \mu_0$  mit einem zweiseitigen Test auf einem Signifikanzniveau von  $\alpha = 5\%$  zu verwerfen, ist gleichbedeutend mit der Aussage, dass das zweiseitige 95 %-Konfidenzintervall für den Mittelwert den Wert  $\mu_0$  *nicht* beinhaltet.

**(9) Was versteht man unter einem Nonzentralitätsparameter?**

In den Nonzentralitätsparameter  $\lambda$  fließt ein, wie weit der Populationsmittelwert  $\mu$  vom erwarteten Populationsmittelwert unter der Nullhypothese ( $\mu_0$ ) entfernt liegt. Je größer der Nonzentralitätsparameter  $\lambda$ , desto flacher und tendenziell linkssteiler wird die nonzentrale  $t$ -Verteilung.

**(10) Was ist der Unterschied zwischen einer Post-hoc- und einer A-priori-Poweranalyse?**

Bei einer A-priori-Poweranalyse wird der Effekt im Vorhinein spezifiziert, um die Stichprobengröße so zu planen, dass die Power angemessen groß ist, um den spezifizierten Effekt aufzudecken. Bei einer Post-hoc-Poweranalyse wird dagegen die Power eines Tests im Nachhinein abgeschätzt, indem als Schätzer des Effekts, der Effekt verwendet wird, der in der Untersuchung beobachtet wurde.