Block 2: Multiple-Choice Questions (Extended)

Inference, Uncertainty & Decision Logic

Data Science and Strategy for Business

Question 1

Was misst der Standardfehler (SE)?

A. Die Varianz in den Rohdaten
B. Die Unsicherheit einer Schätzung über wiederholte Stichproben
C. Die Differenz zwischen zwei Mittelwerten
D. Die Anzahl der Ausreißer in den Daten

Question 1

Was misst der Standardfehler (SE)?

A. Die Varianz in den Rohdaten

B. Die Unsicherheit einer Schätzung über wiederholte Stichproben

C. Die Differenz zwischen zwei Mittelwerten

D. Die Anzahl der Ausreißer in den Daten

Answer: B

Der Standardfehler misst die Variabilität einer Statistik (z.B. Mittelwert) über wiederholte Stichproben. Er berechnet sich als $SE = s/\sqrt{n}$ und wird kleiner mit größerem n.

Question 2

Ein 95%-Konfidenzintervall für den Mittelwert ist [6.5, 7.9]. Was bedeutet das?

A. 95% der Datenpunkte liegen zwischen 6.5 und 7.9
B. Der wahre Mittelwert liegt mit 95% Wahrscheinlichkeit zwischen 6.5 und 7.9
C. Bei wiederholter Stichprobenziehung würden 95% der berechneten Intervalle den wahren Wert enthalten
D. Es gibt eine 5% Chance, dass der Mittelwert außerhalb liegt

Question 2

Ein 95%-Konfidenzintervall für den Mittelwert ist [6.5, 7.9]. Was bedeutet das?

A. 95% der Datenpunkte liegen zwischen 6.5 und 7.9

B. Der wahre Mittelwert liegt mit 95% Wahrscheinlichkeit zwischen 6.5 und 7.9

C. Bei wiederholter Stichprobenziehung würden 95% der berechneten Intervalle den wahren Wert enthalten

D. Es gibt eine 5% Chance, dass der Mittelwert außerhalb liegt

Answer: C

Die korrekte Interpretation: Wenn wir das Experiment viele Male wiederholen und jedes Mal ein 95%-KI berechnen, würden 95% dieser Intervalle den wahren Parameter enthalten.

Question 3

Ein p-Wert von 0.03 bedeutet:

A. Die Nullhypothese ist mit 3% Wahrscheinlichkeit wahr
B. Es gibt eine 3% Chance, die beobachteten Daten (oder extremere) zu sehen, wenn H0 wahr ist
C. Der Effekt ist mit 97% Wahrscheinlichkeit echt
D. Das Ergebnis wird mit 97% Wahrscheinlichkeit repliziert

Question 3

Ein p-Wert von 0.03 bedeutet:

A. Die Nullhypothese ist mit 3% Wahrscheinlichkeit wahr
B. Es gibt eine 3% Chance, die beobachteten Daten (oder extremere) zu sehen, wenn H0 wahr ist
C. Der Effekt ist mit 97% Wahrscheinlichkeit echt
D. Das Ergebnis wird mit 97% Wahrscheinlichkeit repliziert

Answer: B
Der p-Wert ist die Wahrscheinlichkeit der Daten unter H0, NICHT die Wahrscheinlichkeit von H0! Diese häufige Fehlinterpretation führt zu falschen Schlüssen.

Question 4

Was ist das Grundprinzip des Bootstrapping?

A. Aus der Stichprobe neue Stichproben mit Zurücklegen ziehen
B. Immer größere Stichproben aus der Population ziehen
C. Die Daten normal verteilen
D. Ausreißer systematisch entfernen

Question 4

Was ist das Grundprinzip des Bootstrapping?

A. Aus der Stichprobe neue Stichproben mit Zurücklegen ziehen
B. Immer größere Stichproben aus der Population ziehen
C. Die Daten normal verteilen
D. Ausreißer systematisch entfernen

Answer: A
Bootstrap simuliert die Stichprobenvariabilität, indem aus der vorhandenen Stichprobe wiederholt mit Zurücklegen gezogen wird. So entsteht eine empirische Stichprobenverteilung ohne theoretische Annahmen.

Question 5

Wann ist Bootstrap besonders nützlich?

A. Nur bei normalverteilten Daten
B. Wenn theoretische Verteilungen unbekannt oder komplex sind
C. Ausschließlich für kleine Stichproben
D. Nur für kategoriale Daten

Question 5

Wann ist Bootstrap besonders nützlich?

A. Nur bei normalverteilten Daten
B. Wenn theoretische Verteilungen unbekannt oder komplex sind
C. Ausschließlich für kleine Stichproben
D. Nur für kategoriale Daten

Answer: B
Bootstrap ist besonders wertvoll, wenn die theoretische Verteilung einer Statistik unbekannt oder schwierig zu bestimmen ist. Es funktioniert auch bei schiefen Verteilungen und komplexen Statistiken.

Question 6

Was macht ein Permutationstest?

A. Er verändert die Reihenfolge der Datenpunkte zufällig
B. Er mischt Gruppenlabels zufällig, um die Null-Verteilung zu simulieren
C. Er sortiert die Daten nach Größe
D. Er entfernt zufällig Datenpunkte

Question 6

Was macht ein Permutationstest?

A. Er verändert die Reihenfolge der Datenpunkte zufällig

B. Er mischt Gruppenlabels zufällig, um die Null-Verteilung zu simulieren

C. Er sortiert die Daten nach Größe

D. Er entfernt zufällig Datenpunkte

Answer: B

Unter der Nullhypothese (kein Gruppenunterschied) spielt die Gruppenzugehörigkeit keine Rolle. Der Permutationstest mischt die Labels und prüft, ob der beobachtete Unterschied extrem ist.

Question 7

Welche R-Funktion aus dem infer-Paket erzeugt Bootstrap-Stichproben?

A. `calculate()`
B. `generate(type = "bootstrap")`
C. `specify()`
D. `hypothesize()`

Question 7

Welche R-Funktion aus dem infer-Paket erzeugt Bootstrap-Stichproben?

A. `calculate()`

B. `generate(type = "bootstrap")`

C. `specify()`

D. `hypothesize()`

Answer: B

Die Funktion `generate(reps = 1000, type = "bootstrap")` erzeugt 1000 Bootstrap-Stichproben. Das infer-Paket bietet eine konsistente Syntax für Inferenzaufgaben.

Question 8

Was testet ein t-Test?

A. Ob zwei Varianzen gleich sind
B. Ob ein oder zwei Mittelwerte signifikant von einem hypothetischen Wert abweichen
C. Ob Daten normalverteilt sind
D. Ob Korrelationen signifikant sind

Question 8

Was testet ein t-Test?

A. Ob zwei Varianzen gleich sind
B. Ob ein oder zwei Mittelwerte signifikant von einem hypothetischen Wert abweichen
C. Ob Daten normalverteilt sind
D. Ob Korrelationen signifikant sind

Answer: B
Der t-Test vergleicht Mittelwerte: Ein-Stichproben-t-Test prüft $\bar{x}$ gegen einen Wert, Zwei-Stichproben-t-Test vergleicht zwei Gruppen, gepaarter t-Test für abhängige Messungen.

Question 9

Was ist der Unterschied zwischen ANOVA und t-Test?

A. ANOVA ist für kategoriale, t-Test für numerische Daten
B. ANOVA vergleicht mehr als zwei Gruppen, t-Test nur zwei
C. ANOVA benötigt Normalverteilung, t-Test nicht
D. Es gibt keinen Unterschied

Question 9

Was ist der Unterschied zwischen ANOVA und t-Test?

A. ANOVA ist für kategoriale, t-Test für numerische Daten
B. ANOVA vergleicht mehr als zwei Gruppen, t-Test nur zwei
C. ANOVA benötigt Normalverteilung, t-Test nicht
D. Es gibt keinen Unterschied

Answer: B
ANOVA (Analysis of Variance) testet, ob sich die Mittelwerte von drei oder mehr Gruppen unterscheiden. Bei nur zwei Gruppen liefern ANOVA und t-Test äquivalente Ergebnisse.

Question 10

In einer Regressionstabelle zeigt der Koeffizient für “Werbung” einen Wert von 2.5 mit p ¡ 0.001. Was bedeutet das?

A. Pro 1 CHF mehr Werbung steigt die abhängige Variable um 2.5 Einheiten
B. Die Korrelation zwischen Werbung und Y ist 2.5
C. 2.5% der Varianz wird durch Werbung erklärt
D. Der Effekt ist zu klein, um relevant zu sein

Question 10

In einer Regressionstabelle zeigt der Koeffizient für “Werbung” einen Wert von 2.5 mit p ¡ 0.001. Was bedeutet das?

A. Pro 1 CHF mehr Werbung steigt die abhängige Variable um 2.5 Einheiten
B. Die Korrelation zwischen Werbung und Y ist 2.5
C. 2.5% der Varianz wird durch Werbung erklärt
D. Der Effekt ist zu klein, um relevant zu sein

Answer: A
Der Regressionskoeffizient gibt die Änderung in Y bei einer Einheit Erhöhung in X an, ceteris paribus (alle anderen Variablen konstant). Der p-Wert zeigt, dass dieser Effekt statistisch signifikant ist.

Question 11

Was ist ein Typ-I-Fehler?

A. H0 wird abgelehnt, obwohl H0 wahr ist (falsch positiv)
B. H0 wird beibehalten, obwohl H1 wahr ist (falsch negativ)
C. Die Daten sind fehlerhaft
D. Der Test hat zu wenig Power

Question 11

Was ist ein Typ-I-Fehler?

A. H0 wird abgelehnt, obwohl H0 wahr ist (falsch positiv)

B. H0 wird beibehalten, obwohl H1 wahr ist (falsch negativ)

C. Die Daten sind fehlerhaft

D. Der Test hat zu wenig Power

Answer: A

Typ-I-Fehler (Alpha-Fehler): Wir sehen einen Effekt, der nicht existiert. Beispiel: Marketingkampagne wird gestartet, obwohl sie nicht wirkt. Die Wahrscheinlichkeit ist das Signifikanzniveau α (typisch 5%).

Question 12

Was ist ein Typ-II-Fehler?

A. H0 wird abgelehnt, obwohl H0 wahr ist
B. H0 wird beibehalten, obwohl H1 wahr ist (falsch negativ)
C. Der Stichprobenumfang ist zu groß
D. Die Effektgröße ist zu groß

Question 12

Was ist ein Typ-II-Fehler?

A. H0 wird abgelehnt, obwohl H0 wahr ist

B. H0 wird beibehalten, obwohl H1 wahr ist (falsch negativ)

C. Der Stichprobenumfang ist zu groß

D. Die Effektgröße ist zu groß

Answer: B

Typ-II-Fehler (Beta-Fehler): Ein echter Effekt wird übersehen. Beispiel: Wirksame Kampagne wird nicht gestartet. Die Wahrscheinlichkeit ist β, typisch 20% (Power = 1-β = 80%).

Question 13

Was ist die statistische Power eines Tests?

A. Die Wahrscheinlichkeit, einen Typ-I-Fehler zu machen
B. Die Wahrscheinlichkeit, einen echten Effekt zu finden (1-β)
C. Die Größe des Effekts
D. Die Anzahl der Stichproben

Question 13

Was ist die statistische Power eines Tests?

A. Die Wahrscheinlichkeit, einen Typ-I-Fehler zu machen
B. Die Wahrscheinlichkeit, einen echten Effekt zu finden (1-β)
C. Die Größe des Effekts
D. Die Anzahl der Stichproben

Answer: B
Power = 1-β ist die Wahrscheinlichkeit, H0 korrekt abzulehnen, wenn H1 wahr ist. Standard ist 80% Power. Power hängt von Stichprobengröße, Effektgröße, Signifikanzniveau und Varianz ab.

Question 14

Wie berechnet man näherungsweise die benötigte Stichprobengröße für einen t-Test mit Power 80%?

A. $n \approx 8/d^2$ pro Gruppe
B. $n \approx 16/d^2$ pro Gruppe
C. $n \approx d^2/16$ pro Gruppe
D. $n \approx 100$ unabhängig von d

Question 14

Wie berechnet man näherungsweise die benötigte Stichprobengröße für einen t-Test mit Power 80%?

A. $n \approx 8/d^2$ pro Gruppe

B. $n \approx 16/d^2$ pro Gruppe

C. $n \approx d^2/16$ pro Gruppe

D. $n \approx 100$ unabhängig von d

Answer: B

Die Faustregel für 80% Power bei $\alpha = 0.05$ ist $n \approx 16/d^2$ pro Gruppe. Für eine mittlere Effektgröße (d=0.5) braucht man ca. 64 Probanden pro Gruppe. Für d=0.2 sind es 400 pro Gruppe.

Question 15

Was ist p-Hacking?

A. Das Hacken von statistischen Datenbanken
B. Die Manipulation der Datenanalyse, bis p ¡ 0.05 erreicht wird
C. Eine Methode zur p-Wert-Berechnung
D. Die Korrektur für multiple Tests

Question 15

Was ist p-Hacking?

A. Das Hacken von statistischen Datenbanken
B. Die Manipulation der Datenanalyse, bis p ¡ 0.05 erreicht wird
C. Eine Methode zur p-Wert-Berechnung
D. Die Korrektur für multiple Tests

Answer: B
p-Hacking umfasst Praktiken wie: viele Tests durchführen und nur signifikante berichten, selektiv Ausreißer entfernen, Subgruppen bilden bis signifikant, oder Daten sammeln bis p ¡ 0.05. Dies erhöht die Falsch-Positiv-Rate massiv.

Question 16

Was macht die Bonferroni-Korrektur bei 20 Tests mit $\alpha = 0.05$?

A. Setzt das Signifikanzniveau auf $\alpha = 0.05/20 = 0.0025$
B. Multipliziert alle p-Werte mit 20
C. Verwendet nur die ersten 5 Tests
D. Berechnet den Durchschnitt aller p-Werte

Question 16

Was macht die Bonferroni-Korrektur bei 20 Tests mit $\alpha = 0.05$?

A. Setzt das Signifikanzniveau auf $\alpha = 0.05/20 = 0.0025$

B. Multipliziert alle p-Werte mit 20

C. Verwendet nur die ersten 5 Tests

D. Berechnet den Durchschnitt aller p-Werte

Answer: A

Bonferroni kontrolliert die familywise error rate durch Division: $\alpha_{adj} = \alpha/m$. Bei 20 Tests: nur p ¡ 0.0025 gilt als signifikant. Nachteil: sehr konservativ, niedrige Power.

Question 17

Was ist der Vorteil der False Discovery Rate (FDR) gegenüber Bonferroni?

A. FDR ist strenger und findet weniger falsch positive Ergebnisse
B. FDR ist weniger konservativ und hat mehr Power bei gleichem Schutz
C. FDR benötigt keine Korrektur
D. FDR funktioniert nur bei 5 Tests

Question 17

Was ist der Vorteil der False Discovery Rate (FDR) gegenüber Bonferroni?

A. FDR ist strenger und findet weniger falsch positive Ergebnisse

B. FDR ist weniger konservativ und hat mehr Power bei gleichem Schutz

C. FDR benötigt keine Korrektur

D. FDR funktioniert nur bei 5 Tests

Answer: B

FDR (Benjamini-Hochberg-Methode) kontrolliert den erwarteten Anteil falscher Entdeckungen unter den signifikanten Ergebnissen. Sie ist weniger konservativ als Bonferroni und hat mehr Power, besonders bei vielen Tests.

Question 18

Was misst Cohen's d?

A. Die statistische Signifikanz
B. Die Effektgröße in Standardabweichungen
C. Die Korrelation zwischen zwei Variablen
D. Die Wahrscheinlichkeit eines Typ-I-Fehlers

Question 18

Was misst Cohen's d?

A. Die statistische Signifikanz
B. Die Effektgröße in Standardabweichungen
C. Die Korrelation zwischen zwei Variablen
D. Die Wahrscheinlichkeit eines Typ-I-Fehlers

Answer: B
Cohen's d $= (\bar{x}_1 - \bar{x}_2)/s_{pooled}$ misst die Differenz zwischen zwei Gruppen in Standardabweichungen. Es ist unabhängig von der Stichprobengröße: d=0.2 (klein), d=0.5 (mittel), d=0.8 (groß).

Question 19

Warum ist bei sehr großen Stichproben (n ¿ 100'000) fast alles statistisch signifikant?

A. Weil große Datenmengen immer Fehler enthalten
B. Weil der Standardfehler mit $1/\sqrt{n}$ sinkt und selbst winzige Effekte signifikant werden
C. Weil große Datensätze immer normalverteilt sind
D. Das ist falsch, große Stichproben ändern nichts an der Signifikanz

Question 19

Warum ist bei sehr großen Stichproben (n ¿ 100'000) fast alles statistisch signifikant?

A. Weil große Datenmengen immer Fehler enthalten
B. Weil der Standardfehler mit $1/\sqrt{n}$ sinkt und selbst winzige Effekte signifikant werden
C. Weil große Datensätze immer normalverteilt sind
D. Das ist falsch, große Stichproben ändern nichts an der Signifikanz

Answer: B
Bei großem n wird SE sehr klein, sodass selbst triviale Unterschiede (z.B. 0.01%) signifikant werden. Deshalb ist die Effektgröße wichtiger als der p-Wert – Signifikanz $\neq$ Relevanz!

Question 20

Ein Online-Shop testet zwei Checkout-Varianten. Variante B hat 2.3% Conversion, Variante A 2.1% (p=0.04, n=5000 pro Gruppe). Cohen's h=0.03 (sehr klein). Was sollte die Entscheidung sein?

A. Ablehnen wegen zu kleiner Effektgröße

B. Implementieren, wenn der absolute Gewinn die Kosten rechtfertigt

C. Mehr Daten sammeln

D. Nur implementieren, wenn d ¿ 0.5

Question 20

Ein Online-Shop testet zwei Checkout-Varianten. Variante B hat 2.3% Conversion, Variante A 2.1% (p=0.04, n=5000 pro Gruppe). Cohen's h=0.03 (sehr klein). Was sollte die Entscheidung sein?

A. Ablehnen wegen zu kleiner Effektgröße

B. Implementieren, wenn der absolute Gewinn die Kosten rechtfertigt

C. Mehr Daten sammeln

D. Nur implementieren, wenn d ¿ 0.5

Answer: B

Trotz kleiner Effektgröße kann bei hohem Volumen (z.B. 100k Besucher/Monat) ein 0.2% Unterschied wirtschaftlich sehr relevant sein. Expected Value und absolute Gewinne zählen, nicht nur Effektgröße!