

## Block 3: Multiple-Choice Questions (Extended)

Complexity, Causality & Generalization

Data Science and Strategy for Business

## Question 1

**Welche Eigenschaft gehoert NICHT zu den BLUE-Eigenschaften eines OLS-Schaetzers?**

- A. Best (kleinste Varianz unter allen linearen unverzerrten Schaetlern)
- B. Linear (lineare Funktion der abhaengigen Variable)
- C. Unbiased (erwartungstreu)
- D. Normalverteilt (Normal distribution)

## Question 1

Welche Eigenschaft gehoert NICHT zu den BLUE-Eigenschaften eines OLS-Schaetzers?

- A. Best (kleinste Varianz unter allen linearen unverzerrten Schaetlern)
- B. Linear (lineare Funktion der abhaengigen Variable)
- C. Unbiased (erwartungstreu)
- D. Normalverteilt (Normal distribution)

**Answer: D**

BLUE steht fuer Best Linear Unbiased Estimator. Normalverteilung ist keine BLUE-Eigenschaft, sondern eine zusaetzliche Annahme fuer Inferenz.

## Question 2

**Welche Annahme ist fuer die Unverzerrtheit der OLS-Schaetzer am wichtigsten?**

- A. Homoskedastizitaet der Residuen
- B. Normalverteilung der Residuen
- C. Exogenitaet:  $E[\epsilon|X] = 0$
- D. Keine perfekte Multikollinearitaet

Welche Annahme ist fuer die Unverzerrtheit der OLS-Schaetzer am wichtigsten?

- A. Homoskedastizitaet der Residuen
- B. Normalverteilung der Residuen
- C. Exogenitaet:  $E[\epsilon|X] = 0$
- D. Keine perfekte Multikollinearitaet

**Answer: C**

Die Exogenitaetsannahme  $E[\epsilon|X] = 0$  garantiert Unverzerrtheit. Homoskedastizitaet und Normalverteilung beeinflussen Effizienz und Inferenz, nicht Unverzerrtheit.

## Question 3

**Was passiert, wenn die Annahme der Homoskedastizitaet verletzt ist?**

- A. Die OLS-Schaetzer werden verzerrt
- B. Die Standardfehler sind nicht mehr korrekt
- C. Die Koeffizienten koennen nicht mehr berechnet werden
- D. R-Quadrat ist nicht mehr interpretierbar

**Was passiert, wenn die Annahme der Homoskedastizitaet verletzt ist?**

- A. Die OLS-Schaetzer werden verzerrt
- B. Die Standardfehler sind nicht mehr korrekt
- C. Die Koeffizienten koennen nicht mehr berechnet werden
- D. R-Quadrat ist nicht mehr interpretierbar

**Answer: B**

Bei Heteroskedastizitaet bleiben OLS-Schaetzer unverzerrt, aber die Standardfehler sind inkorrekt. Man verwendet dann robuste Standardfehler.

### Was ist Multikollinearität?

- A. Starke Korrelation zwischen abhängiger Variable und Prädiktoren
- B. Starke Korrelation zwischen zwei oder mehr Prädiktoren
- C. Nicht-lineare Beziehung zwischen Variablen
- D. Heteroskedastizität der Residuen

### Was ist Multikollinearitaet?

- A. Starke Korrelation zwischen abhaengiger Variable und Praediktoren
- B. Starke Korrelation zwischen zwei oder mehr Praediktoren
- C. Nicht-lineare Beziehung zwischen Variablen
- D. Heteroskedastizitaet der Residuen

### Answer: B

Multikollinearitaet liegt vor, wenn Praediktoren stark miteinander korreliert sind, was zu instabilen Koeffizientenschaetzungen fuehrt.

## Question 5

**Ein VIF (Variance Inflation Factor) von 15 fuer eine Variable bedeutet:**

- A. Die Variable erklart 15% der Varianz
- B. Die Varianz des Koeffizienten ist 15-mal groesser als ohne Multikollinearitaet
- C. Die Variable hat 15 fehlende Werte
- D. Der Koeffizient ist 15-mal zu gross

**Ein VIF (Variance Inflation Factor) von 15 fuer eine Variable bedeutet:**

- A. Die Variable erklart 15% der Varianz
- B. Die Varianz des Koeffizienten ist 15-mal groesser als ohne Multikollinearitaet
- C. Die Variable hat 15 fehlende Werte
- D. Der Koeffizient ist 15-mal zu gross

**Answer: B**

VIF = 15 bedeutet, dass die Varianz des Schaetzers 15-mal groesser ist als bei unabhaengigen Praediktoren. Werte  $> 10$  deuten auf problematische Multikollinearitaet hin.

**Wie kann man Multikollinearität bekämpfen?**

- A. Mehr Daten sammeln
- B. Stark korrelierte Variablen entfernen oder kombinieren
- C. Ridge-Regression oder Lasso verwenden
- D. Alle Antworten sind korrekt

**Wie kann man Multikollinearität bekämpfen?**

- A. Mehr Daten sammeln
- B. Stark korrelierte Variablen entfernen oder kombinieren
- C. Ridge-Regression oder Lasso verwenden
- D. Alle Antworten sind korrekt

**Answer: D**

Alle genannten Methoden können Multikollinearität reduzieren: mehr Daten, Variablenselektion, oder Regularisierungsmethoden wie Ridge/Lasso.

### Was beschreibt den Bias-Variance Tradeoff?

- A. Den Kompromiss zwischen Modellkomplexitaet und Generalisierungsfahigkeit
- B. Den Kompromiss zwischen Trainings- und Testzeit
- C. Den Kompromiss zwischen Datenqualitaet und -quantitaet
- D. Den Kompromiss zwischen linearen und nicht-linearen Modellen

## Question 7

### Was beschreibt den Bias-Variance Tradeoff?

- A. Den Kompromiss zwischen Modellkomplexitaet und Generalisierungsfahigkeit
- B. Den Kompromiss zwischen Trainings- und Testzeit
- C. Den Kompromiss zwischen Datenqualitaet und -quantitaet
- D. Den Kompromiss zwischen linearen und nicht-linearen Modellen

### Answer: A

Der Bias-Variance Tradeoff beschreibt den Zielkonflikt zwischen Underfitting (hoher Bias) und Overfitting (hohe Varianz).  $MSE = \text{Bias}^2 + \text{Varianz} + \text{irreduzibler Fehler}$ .

## Question 8

**Ein Modell mit hohem Bias und niedriger Varianz zeigt typischerweise:**

- A. Overfitting - perfekt auf Trainingsdaten, schlecht auf Testdaten
- B. Underfitting - schlecht auf Trainings- und Testdaten
- C. Perfekte Generalisierung
- D. Hohe R-Quadrat-Werte

## Question 8

**Ein Modell mit hohem Bias und niedriger Varianz zeigt typischerweise:**

- A. Overfitting - perfekt auf Trainingsdaten, schlecht auf Testdaten
- B. Underfitting - schlecht auf Trainings- und Testdaten
- C. Perfekte Generalisierung
- D. Hohe R-Quadrat-Werte

**Answer: B**

Hoher Bias bedeutet, dass das Modell zu einfach ist (Underfitting) und sowohl Trainings- als auch Testdaten schlecht fittet. Niedrige Varianz bedeutet stabile, aber ungenaue Schätzungen.

### Welches Problem entsteht bei Overfitting?

- A. Das Modell ist zu einfach
- B. Das Modell lernt Rauschen statt Signal
- C. Die Trainingsgenauigkeit ist zu niedrig
- D. Die Koeffizienten sind alle null

### Welches Problem entsteht bei Overfitting?

- A. Das Modell ist zu einfach
- B. Das Modell lernt Rauschen statt Signal
- C. Die Trainingsgenauigkeit ist zu niedrig
- D. Die Koeffizienten sind alle null

### Answer: B

Overfitting bedeutet, dass das Modell zu komplex ist und zufällige Schwankungen (Rauschen) in den Trainingsdaten als Muster lernt, was zu schlechter Generalisierung führt.

**Was ist der Hauptunterschied zwischen Lasso und Ridge Regression?**

- A. Lasso verwendet L1-Penalty, Ridge verwendet L2-Penalty
- B. Lasso ist schneller als Ridge
- C. Ridge kann nur mit linearen Modellen verwendet werden
- D. Lasso benoetigt mehr Trainingsdaten

**Was ist der Hauptunterschied zwischen Lasso und Ridge Regression?**

- A. Lasso verwendet L1-Penalty, Ridge verwendet L2-Penalty
- B. Lasso ist schneller als Ridge
- C. Ridge kann nur mit linearen Modellen verwendet werden
- D. Lasso benoetigt mehr Trainingsdaten

**Answer: A**

Lasso (L1:  $\lambda \sum |\beta_j|$ ) kann Koeffizienten auf exakt null setzen (Variablenselektion). Ridge (L2:  $\lambda \sum \beta_j^2$ ) schrumpft alle Koeffizienten, aber nie auf null.

**Welche Regularisierungsmethode eignet sich besser fuer automatische Variablenselektion?**

- A. Ridge Regression
- B. Lasso Regression
- C. Beide gleich gut
- D. Keine der beiden

**Welche Regularisierungsmethode eignet sich besser fuer automatische Variablenselektion?**

- A. Ridge Regression
- B. Lasso Regression
- C. Beide gleich gut
- D. Keine der beiden

**Answer: B**

Lasso setzt irrelevante Koeffizienten auf exakt null und fuehrt so automatisch Variablenselektion durch. Ridge schrumpft nur, eliminiert aber keine Variablen.

## Question 12

**Der Regularisierungsparameter  $\lambda$  in Ridge/Lasso kontrolliert:**

- A. Die Lernrate des Algorithmus
- B. Die Staerke der Bestrafung grosser Koeffizienten
- C. Die Anzahl der Iterationen
- D. Die Trainings-Test-Split Ratio

**Der Regularisierungsparameter  $\lambda$  in Ridge/Lasso kontrolliert:**

- A. Die Lernrate des Algorithmus
- B. Die Staerke der Bestrafung grosser Koeffizienten
- C. Die Anzahl der Iterationen
- D. Die Trainings-Test-Split Ratio

**Answer: B**

$\lambda$  kontrolliert die Balance zwischen Model Fit und Koeffizientengroesse. Grosses  $\lambda$  = starke Regularisierung (kleinere Koeffizienten), kleines  $\lambda$  = schwache Regularisierung.

### Was ist der Hauptzweck von Cross-Validation?

- A. Trainingsdaten zu vergrößern
- B. Modellperformance auf ungesehenen Daten zu schätzen
- C. Fehlende Werte zu imputieren
- D. Ausreißer zu identifizieren

### Was ist der Hauptzweck von Cross-Validation?

- A. Trainingsdaten zu vergrössern
- B. Modellperformance auf ungesehenen Daten zu schätzen
- C. Fehlende Werte zu imputieren
- D. Ausreisser zu identifizieren

### Answer: B

Cross-Validation testet das Modell auf mehreren Trainings-Test-Splits, um eine robuste Schätzung der Generalisierungsperformance zu erhalten und Overfitting zu vermeiden.

**Bei 5-Fold Cross-Validation wird:**

- A. Das Modell 5-mal auf allen Daten trainiert
- B. Der Datensatz in 5 Teile geteilt; 4 Teile Training, 1 Teil Test (5 Durchläufe)
- C. Der Datensatz in 5% Training und 95% Test aufgeteilt
- D. Das Modell mit 5 verschiedenen Algorithmen verglichen

**Bei 5-Fold Cross-Validation wird:**

- A. Das Modell 5-mal auf allen Daten trainiert
- B. Der Datensatz in 5 Teile geteilt; 4 Teile Training, 1 Teil Test (5 Durchläufe)
- C. Der Datensatz in 5% Training und 95% Test aufgeteilt
- D. Das Modell mit 5 verschiedenen Algorithmen verglichen

**Answer: B**

Bei k-Fold CV wird der Datensatz in k Teile (Folds) geteilt. In jedem Durchlauf dienen k-1 Folds als Training und 1 Fold als Test. Die k Ergebnisse werden gemittelt.

**Welche Aussage ueber Leave-One-Out Cross-Validation (LOOCV) ist korrekt?**

- A. LOOCV ist schneller als k-Fold CV
- B. LOOCV trainiert  $n$  Modelle ( $n = \text{Anzahl Beobachtungen}$ )
- C. LOOCV kann nur bei kleinen Datensuetzen verwendet werden
- D. LOOCV garantiert kein Overfitting

Welche Aussage ueber Leave-One-Out Cross-Validation (LOOCV) ist korrekt?

- A. LOOCV ist schneller als k-Fold CV
- B. LOOCV trainiert  $n$  Modelle ( $n = \text{Anzahl Beobachtungen}$ )
- C. LOOCV kann nur bei kleinen Datensatzen verwendet werden
- D. LOOCV garantiert kein Overfitting

**Answer: B**

Bei LOOCV wird jede Beobachtung einmal als Testset verwendet, daher  $n$  Trainings-Durchlaeufer. Sehr rechenintensiv, aber maximale Nutzung der Trainingsdaten. Hohe Varianz moeglich.

**Korrelation impliziert Kausalitaet" ist:**

- A. Immer wahr
- B. Wahr, wenn die Korrelation stark genug ist (z.B.  $> 0.8$ )
- C. Falsch - Korrelation ist notwendig, aber nicht hinreichend fuer Kausalitaet
- D. Nur bei experimentellen Daten wahr

**Korrelation impliziert Kausalitaet" ist:**

- A. Immer wahr
- B. Wahr, wenn die Korrelation stark genug ist (z.B.  $> 0.8$ )
- C. Falsch - Korrelation ist notwendig, aber nicht hinreichend fuer Kausalitaet
- D. Nur bei experimentellen Daten wahr

**Answer: C**

Korrelation kann durch Kausalitaet, umgekehrte Kausalitaet, Confounding oder Zufall entstehen. Kausalitaet erfordert zusaetzliche Bedingungen (z.B. zeitliche Reihenfolge, kein Confounding).

**Ein Confounder ist:**

- A. Eine Variable, die sowohl die abhaengige als auch unabhaengige Variable beeinflusst
- B. Ein Messfehler in der abhaengigen Variable
- C. Eine kategoriale Variable mit vielen Auspraegungen
- D. Ein Ausreisser im Datensatz

### Ein Confounder ist:

- A. Eine Variable, die sowohl die abhaengige als auch unabhaengige Variable beeinflusst
- B. Ein Messfehler in der abhaengigen Variable
- C. Eine kategoriale Variable mit vielen Auspraegungen
- D. Ein Ausreisser im Datensatz

### Answer: A

Ein Confounder ist eine Stoervariable, die sowohl X als auch Y beeinflusst und so eine Scheinkorrelation erzeugen kann. Beispiel: Eisverkauf und Ertrinkungsfaelle (Confounder: Temperatur).

**Was ist der Hauptvorteil von randomisierten A/B-Tests gegenüber Beobachtungsstudien?**

- A. Sie sind billiger
- B. Sie benötigen weniger Daten
- C. Randomisierung eliminiert systematische Unterschiede zwischen Gruppen (Confounding)
- D. Sie sind schneller durchzuführen

**Was ist der Hauptvorteil von randomisierten A/B-Tests gegenüber Beobachtungsstudien?**

- A. Sie sind billiger
- B. Sie benötigen weniger Daten
- C. Randomisierung eliminiert systematische Unterschiede zwischen Gruppen (Confounding)
- D. Sie sind schneller durchzuführen

**Answer: C**

Randomisierung stellt sicher, dass Treatment- und Kontrollgruppe im Erwartungswert identisch sind (nur durch Treatment unterschieden), was kausale Interpretation ermöglicht.

## Question 19

Bei einem A/B-Test mit  $p$ -Wert = 0.03 (Signifikanzniveau 5%) sollte man:

- A. Die Nullhypothese verwerfen - es gibt einen signifikanten Effekt
- B. Die Nullhypothese beibehalten
- C. Mehr Daten sammeln
- D. Die Gruppen neu randomisieren

## Question 19

Bei einem A/B-Test mit  $p$ -Wert = 0.03 (Signifikanzniveau 5%) sollte man:

- A. Die Nullhypothese verwerfen - es gibt einen signifikanten Effekt
- B. Die Nullhypothese beibehalten
- C. Mehr Daten sammeln
- D. Die Gruppen neu randomisieren

**Answer: A**

Da  $p = 0.03 < 0.05$ , wird die Nullhypothese verworfen. Der Effekt ist auf dem 5%-Niveau statistisch signifikant. Aber: Signifikanz  $\neq$  praktische Relevanz!

**In einer logistischen Regression bedeutet ein Odds Ratio von 2.5 fuer eine Variable:**

- A. Die Wahrscheinlichkeit steigt um 250%
- B. Die Odds (Chancen) sind 2.5-mal hoeher pro Einheit Zunahme der Variable
- C. Die Variable erklart 2.5% der Varianz
- D. Der Koeffizient betraegt 2.5

In einer logistischen Regression bedeutet ein Odds Ratio von 2.5 fuer eine Variable:

- A. Die Wahrscheinlichkeit steigt um 250%
- B. Die Odds (Chancen) sind 2.5-mal hoeher pro Einheit Zunahme der Variable
- C. Die Variable erklart 2.5% der Varianz
- D. Der Koeffizient betraegt 2.5

**Answer: B**

Odds Ratio = 2.5 bedeutet, dass sich die Odds (Verhaeltnis von  $P(Y=1)$  zu  $P(Y=0)$ ) um Faktor 2.5 aendern bei +1 Einheit der Variable.  $OR = e^{\beta}$  in logistischer Regression.