

Midterm-Examen Kurs A

Zeit: 45 Minuten

Name: _____

Matr. Nummer: _____

Hinweise:

1. Zugelassene Hilfsmittel: Open-Book: Aufschriebe, Formelsammlung, Skript, Taschenrechner (keine gespeicherten Formeln etc.!), Notizen.
2. Jede Antwort muss hinreichend begründet werden. Antworten ohne Begründung ergeben 0 Punkte.
3. Unleserliche Ergebnisse werden nicht gewertet. Nutzen Sie bei weiterem Platzbedarf bitte auch die Rückseiten der Klausurblätter!
4. Die geschätzte Bearbeitungszeit (in Minuten) für eine Aufgabe entspricht der Punktzahl. Somit sind die Aufgaben insgesamt 45 Punkte wert.
5. **Viel Glück!!!**

Frage	Punkte	Erreichte Punkte
1	25	
2	20	
Gesamt	45	

Aufgabe 1. Interpretation Logistische Regression (25 Punkte)

Wir möchten untersuchen inwieweit der Intelligenzquotient (IQ) und die Vorbereitungszeit die Wahrscheinlichkeit beeinflussen ein Statistik-Examen zu bestehen.

Hierfür führen wir eine logistische Regression mit den folgenden Variablen durch:

- Intelligenzquotient (IQ): 1,2,3,... (ganze positive Zahl)
- Vorbereitungszeit: gemessen in Lernstunden

Die Regression ergibt folgende Koeffizienten Tabelle:

		coef	exp(coef)	std err	z	P> z	[0.025	0.975]
logistic reg.	intercept	-14.062	7.815396e-07	9.177	-5.324	0.000	-30.867	-10.321
coefficients	iq	0.101	1.106277e+00	0.025	2.355	0.019	0.900	1.250
	vorbereitungszeit	0.300	1.349859e+00	0.053	4.759	0.000	0.149	0.358

- Schreiben Sie die geschätzte Regressionsgleichung in Bezug auf den Kontext auf.
- Berechnen Sie die Wahrscheinlichkeit mit welcher ein Studierender mit einem IQ von 100 und 12 Stunden Vorbereitungszeit das Examen besteht.
- Schreiben Sie die Null- und die Alternativhypothese des Signifikanztests für einen der beiden Koeffizienten auf (in Bezug auf den Kontext).
- Welche der unabhängigen Variablen beeinflussen signifikant die Wahrscheinlichkeit ein Statistik-Examen zu bestehen? Begründen Sie Ihre Antwort.
- Berechnen Sie wie viele Stunden Vorbereitungszeit ein Studierender mit einem IQ von 100 braucht, um mit genau 50% das Examen zu bestehen.
- Geben Sie einen Grund an, warum ein lineares Regressionsmodell in der vorliegenden Fragestellung zu fehlerhaften Ergebnissen führen kann.
- Berechnen Sie wie viel Stunden ein Studierender mit einem IQ von 100 mehr lernen muss, um die gleiche Wahrscheinlichkeit zu erzielen wie ein Studierender mit einem IQ von 120.

Lösung:

a) Geschätzte Regressionsgleichung:

$$P(Y = \text{Bestehen}) = \frac{e^z}{1 + e^z}$$

wobei $z = -14.062 + 0.121 \cdot iq + 0.3 \cdot \text{zeit}$

b) $iq = 100$, $\text{Zeit} = 12$:

$$z = -14.062 + 0.101 \cdot 100 + 0.3 \cdot 12 \approx -0.362$$

$$P = 0.41 = 41\%$$

c) $H_0: \beta_{iq} = 0$ (Variable hat keinen Einfluss), $H_1: \beta_{iq} \neq 0$ (Variable hat einen Einfluss)

d) Beide: die p-Werte 0.019 und 0 sind beide kleiner als 5%. Somit kann jeweils die Nullhypothese abgelehnt werden.

e) Es muss gelten:

$$\frac{e^z}{1 + e^z} = 0.5$$

Auflösen nach z ergibt:

$$z = \ln(1) = 0$$

Also:

$$-14.062 + 0.121 \cdot 100 + 0.3 \cdot \text{zeit} = 0$$

$$\text{zeit} \approx 13.21$$

f) Beispiel: Die lineare Regression kann Wahrscheinlichkeiten $> 100\%$ und $< 0\%$ schätzen. Solche Wahrscheinlichkeiten gibt es nicht.

g) Es soll gelten:

$$-14.062 + 0.101 \cdot 100 + 0.3 \cdot \text{zeit1} = -14.062 + 0.101 \cdot 120 + 0.3 \cdot \text{zeit2}$$

Also:

$$\text{zeit1} - \text{zeit2} = \frac{0.101 \cdot 20}{0.3} = 6.73$$

Aufgabe 2. Interpretation T-Test (20 Punkte)

Ein Pharmaunternehmen möchte überprüfen, ob ein experimentelles Medikament beim Abnehmen hilft. Hierzu erhalten Probanden der Gruppe *med* das Medikament und Probanden der Gruppe *placebo* ein Placebo.

Zur Überprüfung der Hypothese, dass das Medikament wirkt führt das Unternehmen einen Zweistichproben t-Test mit folgendem Resultat durch. Sie dürfen annehmen, dass die Varianzen in den Gruppen in etwa gleich sind (Varianzhomogenität):

	var	group	mean	variances	dof	alternative	p-val	CI95%	cohen-d	BF10
Two-Sample	weight	med	73.0230	equal	48	greater	0.042	[2.32,12.43]	0.82	3.86
t-Test		placebo	77.7558							

- Schreiben Sie kontextbezogen die Null- und die Alternativhypothese des t-Tests auf.
- Wie groß war die gesamte Stichprobe (n)? Begründen Sie Ihre Antwort.
- Interpretieren Sie das Resultat des t-Tests. Besteht Evidenz für eine Wirksamkeit des Medikaments?
- Erklären Sie kurz den Unterschied zwischen Effektstärke (Cohen's d) und Signifikanz (p-Wert). Interpretieren Sie den Wert von Cohen's d.
- Beschreiben Sie kurz wie sich der Bayes-Factor 10 (BF10) berechnet. Interpretieren Sie den Wert des Bayes-Factors.

Lösung:

- H_0 : Gewicht *med* \leq Gewicht *placebo*
 H_a : Gewicht *med* $>$ Gewicht *placebo*
- $n = \text{dof} + 2 = 50$
- Der p-Wert ist $4.2\% < 5\%$. Damit können wir die Nullhypothese ablehnen. Es besteht Evidenz für eine Wirksamkeit des Medikaments.
- Die Signifikanz sagt uns, ob wir ein Ergebnis verallgemeinern dürfen. Die Effektstärke sagt uns wie stark das Ergebnis ist, z.B. ein hoher Mittelwertunterschied. Cohens d ist größer als 0.8 und legt einen starken Effekt nahe.
- Der Bayes-Factor ist das Verhältnis der bedingten Wahrscheinlichkeit die Daten unter der Alternativhypothese zu beobachten geteilt durch die bedingte Wahrscheinlichkeit die Daten unter der Nullhypothese zu beobachten. Im vorliegenden Fall ist es 3.86 mal wahrscheinlicher die Daten in der Alternativwelt zu beobachten.