

## Midterm-Examen Kurs C

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	
<b>Gesamt</b>	<b>45</b>	

## Aufgabe 1. Interpretation Regression (25 Punkte)

Wir möchten untersuchen inwieweit der Kilometerstand und die Wagenfarbe den Preis eines Gebrauchtwagens beeinflussen.

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

- Preis: in €
- Kilometerstand: in 10'000 km
- Wagenfarbe: Rot, Schwarz/Blau, Weiß, Andere

Zunächst möchten wir überprüfen, wie gut unser Modell zu den Daten passt. Wir erhalten folgende Fit-Tabelle:

	dv	dof resid	dof model	R2	adj. R2	omnibus (F)	omnibus (p-val)	LL
linear reg. fit	Preis	736	4	0.117	?	4.31	0.02	-1578.06

- Schreiben Sie die Null- und die Alternativhypothese des Omnibus-Tests auf. Interpretieren Sie das Resultat des Tests. Passt unser Modell zu den Daten?
- Interpretieren Sie den Wert von R2. Was sagt der Wert 0.117 in Bezug auf den Kontext aus?
- Berechnen Sie die Stichprobengröße, auf welcher unsere Regression basiert.
- Berechnen Sie den adj. R2 (angepassten R2).
- Schreiben Sie die theoretische Regressionsgleichung auf.

### Lösung:

- $H_0: R^2 = 0, H_a: R^2 \neq 0$  (Alternativ:  $H_0$ : Alle betas = 0,  $H_a$ : Mind. 1 beta ungleich 0).

Der Test hat einen p-Wert von 0,02 (2%) < 5%. Wir können also die Nullhypothese ablehnen und haben einen Beleg, dass unser Modell zu den Daten passt.

- Der Kilometerstand in Verbindung mit der Wagenfarbe erklärt 11,7% der Unterschiede im Gebrauchtwagenpreis.

- Dof resid = n – dof model – 1

Einsetzen ergibt:  $736 = n - 5 \rightarrow n = 741$

- $adj. R^2 = \left( R^2 - \frac{dof\ model}{n-1} \right) \cdot \left( \frac{n-1}{dof\ resid} \right)$

Einsetzen ergibt:

$$adj. R^2 = \left( 0.117 - \frac{4}{740} \right) \cdot \left( \frac{740}{736} \right) \approx 0.1122$$

- Theoretische Regressionsgleichung, z.B.:

$$Preis = \beta_0 + \beta_1 Kilometer + \beta_2 rot + \beta_3 schwarz - blau + \beta_4 weiß$$

Von rot, Schwarz/blau, weiß, andere müssen genau 3 in der Regressionsgleichung vorkommen (dummy/one-hot encoding)

### Aufgabe 2. Interpretation Chi2 (20 Punkte)

Wir möchten wissen, ob der selbstangegebene Persönlichkeitstyp (personality) mit den Kategorien

- 'introverted, organized, conscientious (blue type)',
- 'extroverted, self-confident, assertive (red type)',
- 'happy, open, creative (yellow type)',
- 'social, trust seeking, dislike changes (green type)'

und das biologische Geschlecht (sex) [mit den Kategorien männlich und weiblich] abhängig sind.

Hierfür führen wir einen Chi<sup>2</sup>-Test durch und erhalten folgendes Resultat:

	vars	no. categories	test	chi2	dof	p-val	cramer	power
Chi2 Tests	personality	4	pearson	8.489	3	0.037	0.283	0.683
of Independence	sex	2						

- Schreiben Sie die Null- und die Alternativhypothese des Tests in Bezug auf den Kontext auf.
- Interpretieren Sie das Resultat des Tests. Haben wir einen Beleg, dass personality und sex abhängig sind? Begründen Sie Ihre Antwort.
- Berechnen Sie die Freiheitsgrade (dof) des Tests. Geben Sie den Rechenweg an.
- Warum benötigen wir das Maß Cramer neben dem p-Wert? Über was gibt Cramer Auskunft und über was der p-Wert? Interpretieren Sie den Wert des Cramer Maßes.

Betrachten Sie die beobachteten und erwarteten Häufigkeiten:

beobachtet	Blue type	Red type	Yell. type	Green type	gesamt
weiblich	11	20	22	18	71
männlich	11	13	9	2	35
gesamt	22	33	31	20	106

erwartet	Blue type	Red type	Yell. type	Green type
weiblich	14,7	22,1	a	13,4
männlich	b	10,9	10,2	6,6

- Berechnen Sie die erwartete Häufigkeit a und b, welche in der Tabelle ausgeblendet ist.

Lösung:

- a)  $H_0$ : personality und sex sind stochastisch unabhängig  
 $H_a$ : personality und sex sind stochastisch abhängig
- b) P-Wert = 3,7% -> wir können die Nullhypothese ablehnen und haben einen Beleg für eine Abhängigkeit von personality und sex.
- c)  $(4-1) \cdot (2-1) = 3$
- d) Der p-Wert sagt etwas über die Signifikanz (Verallgemeinerbarkeit) aus, Cramer über die Effektstärke (Stärke der Abhängigkeit). Der Wert 0,283 entspricht laut Tabelle aus dem Skript einer moderaten (mittelstarken) Abhängigkeit.
- e)

$$a = \frac{31 \cdot 71}{106} \approx 20,8$$

$$b = \frac{22 \cdot 35}{106} \approx 7,3$$