

Midterm-Examen

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	20	
2	22	
3	3	
Gesamt	45	

Aufgabe 1. Interpretation Regression (20 Punkte)

Sie sind Customer Analyst bei Airbnb (private Raumvermietungsplattform) und möchten untersuchen wie stark der Preis eines Raumes vom Raumtyp und dem Gesamt-Rating abhängt.

Hierfür führen Sie eine lineare Regression mit folgenden Variablen durch:

- Preis (in €)
- Raumtyp (Kategorien: Private Room, Entire Appartement, Shared Room)
- Overall Rating (von 0 bis 100)

und erhalten folgendes Resultat:

		coef	stand. coef	std err	t	P> t
linear reg.	intercept	50.5800	1.528859	10.487	4.823	0.000
coefficients	dummy__Room Type_Private room	-39.7258	-0.774881	1.529	-25.981	0.000
	dummy__Room Type_Shared room	-48.5759	-1.000651	8.655	-5.613	0.000
	Overall Rating	0.3187	0.246672	0.110	2.893	0.004

- Schreiben Sie die geschätzte Regressionsgleichung in Bezug auf den Kontext auf. Sie dürfen die Koeffizienten auf zwei Nachkommastellen runden.
- Berechnen Sie den geschätzten Preis eines „Private Room“ mit einem Overall Rating von 82.
- Interpretieren Sie die geschätzten Koeffizienten von intercept, des Dummies für den Private Room und des Overall Ratings.
- Schreiben Sie kontextbezogen die Null- und die Alternativhypothese von einem der t-Tests der Regressionsanalyse auf.
- Welche unabhängigen Variablen haben einen signifikanten Einfluss auf die abhängige Variable? Begründen Sie Ihre Antwort.

Lösung:

a) $Preis = 50.58 - 39.73Private - 48.58Shared + 0.32Rating$

b) $Preis = 50.58 - 39.73 + 0.32 \cdot 82 = 37.09 \text{ €}$

c) Interpretation:

- Intercept: Ein gesamtes Appartement mit einem Rating von 0 kostet 50.58 €
- Dummy Private Room: Ein Private Room kostet im Schnitt 39.73 € weniger als ein ganzes Appartement
- Overall Rating: Pro Rating Punkt steigt der Preis um 0.32 €

d) Beispiel Rating:

$$H_0: \beta_{rating} = 0, H_A: \beta_{rating} \neq 0$$

e) Alle. Der p-Wert aller Variablen ist unter 5%.

Aufgabe 2. Interpretation T-Test (22 Punkte)

Eine Airline möchte überprüfen, ob Flugverzögerungen zu Kundenunzufriedenheit führt. Hierfür wird anhand eines t-Tests überprüft, ob die durchschnittliche Verzögerung in Minuten (Departure Delay in Minutes) bei der Gruppe der unzufriedenen Kunden (dissatisfied) signifikant höher war als in der Gruppe der zufriedenen Kunden (satisfied).

Zur Überprüfung der Annahmen wird ein Levenes Test durchgeführt mit folgendem Resultat:

	var	group		f	dof1	dof2	p-val
Levenes Test	Departure Delay in Minutes	satisfaction	Mean	6.8361	1	187	0.0097
of Equal Variances			Median	2.9726	1	187	0.0863
			Trimmed	19.3263	1	187	0.0000

Der t-Test ergab folgendes Resultat:

	var	group	mean	variances	t	dof	alternative	p-val
Two-Sample	Departure Delay in Minutes	dissatisfied	17.4038	equal	1.7241	187.0000	greater	0.0432
t-Test		satisfied	8.2824	unequal	1.8340	156.1672		0.0343

- Schreiben Sie kontextbezogen die Null- und die Alternativhypothese des Levenes Test auf.
- Es handelt sich nicht um eine schiefe Verteilung. Dürfen wir in diesem Fall Varianzgleichheit annehmen? Interpretieren Sie das Ergebnis des Levenes Tests. Begründen Sie Ihre Antwort.
- Schreiben Sie kontextbezogen die Null- und die Alternativhypothese des t-Tests auf.
- Interpretieren Sie das Resultat des t-Tests. Besteht Evidenz für eine höhere durchschnittliche Abflugverzögerung bei unzufriedenen Kunden?
- Erklären Sie kurz den Unterschied zwischen Effektstärke (Cohen's d) und Signifikanz (p-Wert).
- In der Stichprobe befinden sich
 - n1 = 104 unzufriedenen Kunden (Standardabweichung Verzögerung s1 = 44.6)
 - n2 = 85 zufriedene Kunden (Standardabweichung Verzögerung s2 = 21.9)

Berechnen Sie die Effektstärke (Cohen's d) und interpretieren Sie Ihr Ergebnis. Handelt es sich um einen starken oder schwachen Effekt?

Lösung:

- H0: Varianz Verzögerung zufrieden = Varianz Verzögerung unzufrieden
Ha: Varianz Verzögerung zufrieden ≠ Varianz Verzögerung unzufrieden

- b) Da es sich nicht um eine schiefe Verteilung handelt gelten die Tests basierend auf dem Mittelwert (Mean und Trimmed). In beiden Fällen ist der p-Wert kleiner als 5%. Somit kann die Nullhypothese abgelehnt werden. Wir dürfen also nicht von gleichen Varianzen ausgehen.
- c) $H_0: \text{Verzögerung unzufrieden} \leq \text{Verzögerung zufrieden}$
 $H_a: \text{Verzögerung unzufrieden} > \text{Verzögerung zufrieden}$
- d) Der p-Wert ist $3.4\% < 5\%$. Damit können wir die Nullhypothese ablehnen. Es besteht Evidenz für eine höhere durchschnittliche Flugverzögerung bei unzufriedenen Kunden.
- e) Die Signifikanz sagt uns, ob wir ein Ergebnis verallgemeinern dürfen. Die Effektstärke sagt uns wie stark das Ergebnis ist, z.B. ein hoher Mittelwertunterschied.

f)

$$\text{Cohens } d = \frac{\text{Mittelwert Unzufrieden} - \text{Mittelwert Zufrieden}}{\text{gepoolte Standardabweichung}}$$

$$\begin{aligned} \text{gepoolte Standardabweichung} &= \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \\ &= \sqrt{\frac{103 \cdot 44.6^2 + 84 \cdot 21.9^2}{104 + 85 - 2}} \approx 36.21 \end{aligned}$$

Damit:

$$\text{Cohens } d = \frac{17.4038 - 8.2824}{36.21} \approx 0.25$$

Es handelt sich gemäß Tabelle in den Vorlesungsfolien um einen eher schwachen Effekt.

Aufgabe 3. Unawareness Prof. David Kelsey (3 Punkte)

Erläutern Sie kurz anhand eines selbstgewählten Beispiels was Prof. Kelsey unter Unawareness versteht.