

Midterm-Examen 20.12.2022

Zeit: 45 Minuten

Name: _____

Matr. Nummer: _____

Hinweise:

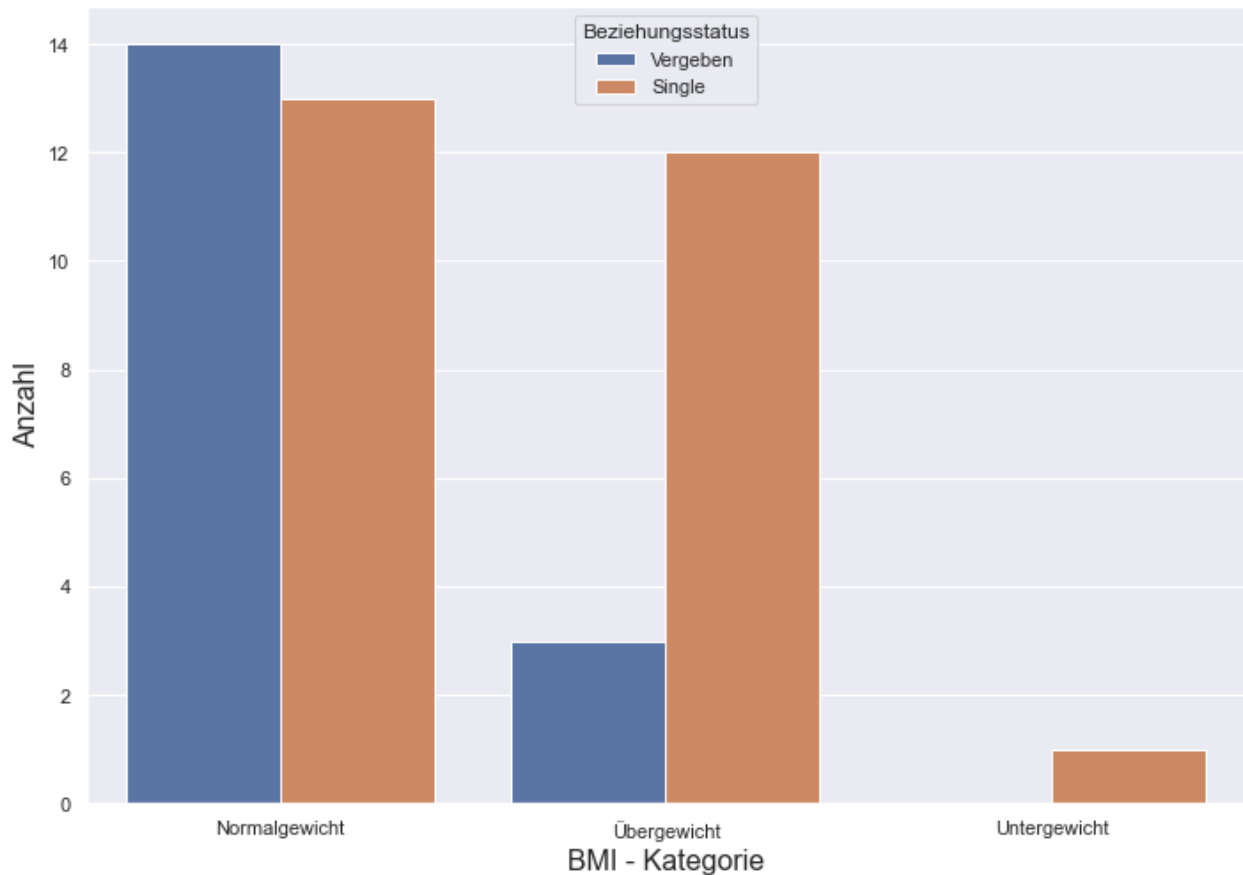
1. Zugelassene Hilfsmittel: Open-Book: Aufschriebe, Formelsammlung, Skript, Taschenrechner (keine gespeicherten Formeln etc.!), Notizen.
2. Jede Antwort muss hinreichend begründet werden. Antworten ohne Begründung ergeben 0 Punkte.
3. Unleserliche Ergebnisse werden nicht gewertet. Nutzen Sie bei weiterem 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	5	
2	25	
3	15	
Gesamt	45	

Aufgabe 1. Interpretation Countplot (5 Punkte)

Bei den Daten Ihres Kurses wurde der Body-Mass-Index berechnet und in die zugehörigen Kategorien eingeteilt (≤ 18 : Untergewicht, 19 – 24: Normalgewicht, > 24 : Übergewicht).

Das folgende Bild zeigt einen Countplot der Daten (BMI – Kategorie vs. Beziehungsstatus):



- Gibt es augenscheinlich Unterschiede zwischen den Gruppen in Bezug auf deren Beziehungsstatus? Begründen Sie Ihre Antwort (Bezug zum Bild).
- Warum könnte es eine Abhängigkeit zwischen der BMI-Kategorie und dem Beziehungsstatus geben?

Lösung:

- Ja, der Anteil der Singles ist bei Über- und Untergewicht deutlich höher.
- Beziehung könnte zu mehr Aktivität und gesünderer Lebensweise führen.

Aufgabe 2. Interpretation Lineare Regression (25 Punkte)

Wir möchten untersuchen inwieweit das BIP (Bruttoinlandsprodukt) pro Kopf (GDP per capita) eines Landes von

- dem Kontinent (Amerika, Europa, Afrika, Asien) und
- der weiblichen Beteiligung am Arbeitsmarkt (Female labour force participation) (0 bis 100) abhängt.

Hierzu wird eine multivariate lineare Regression durchgeführt mit folgenden Ergebnissen:

Fit-Tabelle:

	dependent var	number of obs	fit	omnibus test
Linear Regression (OLS)	GDP p. Capita	132	R-squared 0.32	F-Statistic 14.69
Data Fit			Adj. R-squared 0.30	Prob(F) 0.00

Resultate-Tabelle:

	dependent var		coef	stand. coef.	std err	t	P> t
Linear Regression (OLS)	GDP p. Capita	const	-6311.87	0.00	6055.89	-1.04	0.30
Results		dummy__Region_America	8129.01	2853.51	4558.89	1.78	0.08
		dummy__Region_Asia	13240.00	6042.34	3802.83	3.48	0.00
		dummy__Region_Europe	28880.00	12750.00	3833.92	7.53	0.00
		Female labour force participation	140.58	2215.65	92.91	1.51	0.13

- Schreiben Sie Null- und die Alternativhypothese des Omnibus-Tests auf und interpretieren Sie das Ergebnis (p-Wert). Passt das Modell zu den Daten?
- Schreiben Sie die geschätzte Regressionsgleichung in Bezug auf den Kontext auf.
- Interpretieren Sie die geschätzten Koeffizienten von dummy__Region_Asia und Female labour force participation.
- Berechnen Sie das geschätzte BIP pro Kopf eines afrikanischen Landes mit 100% weiblicher Arbeitsmarktbeteiligung.
- Schreiben Sie die Null- und Alternativhypothese in Bezug auf den Kontext **für einen** der t-Tests bezüglich der Koeffizienten auf.
- Welche Variablen haben einen signifikanten Einfluss auf das BIP pro Kopf? Begründen Sie Ihre Antworten.

Lösung:

- $H_0: R^2 = 0$, $H_a: R^2 \neq 0$. Der p-Wert ist 0. Wir können somit die Nullhypothese ablehnen. Das Modell passt zu den Daten.

b) $BIP = -6311,87 + 8129,01America + 13240Asia + 28880Europa + 140,58FemalePart$

c) Interpretation:

- Asia: Das pro Kopf BIP in Asien ist um 13240 höher als das von Afrika
- FemalePart: Wenn die weibliche Arbeitsmarktbeteiligung um 1% steigt, dann steigt das pro Kopf Einkommen um 140,58.

d) $BIP = -6311,87 + 100 \cdot 140,58 = 7746,13$

e) America: $H_0: \beta_{America} = 0, H_a: \beta_{America} \neq 0$

f) Die Dummies für Asien und Europa sind signifikant (p-Werte = 0 < 5%). Der Rest ist über 5% und damit nicht signifikant.

Aufgabe 3. P-Wert Anpassungen (15 Punkte)

Wir möchten wissen, ob Unterschiede in der Social-Media-Nutzungsdauer zwischen Schülern, Auszubildenden, Studenten und Berufstätigen vorhanden sind.

- Wie viele Tests muss man durchführen, wenn man die Gruppen paarweise vergleicht?
- Berechnen Sie den kumulierten alpha-Fehler (familywise error) für $\alpha = 5\%$ bei unangepassten Tests.

Der Test auf Mittelwertunterschied zwischen Studenten und Berufstätigen ergibt einen unangepassten p-Wert von 1,6%. Dieser p-Wert nimmt Rangplatz 4 unter den p-Werten ein.

- Berechnen Sie die Bonferroni-, Sidak-, Holm-Bonferroni- sowie Holm-Sidak-Korrekturen dieses p-Werts. Bei welchen der Korrekturen ist der Test weiterhin signifikant?

Lösung:

- 6 Tests
- Kumulierter Fehler $= 1 - (1 - 5\%)^6 \approx 26.5\%$
- Bonferroni: $6 \cdot 1,6\% = 9,6\%$
 - Sidak: $1 - (1 - 1,6\%)^6 \approx 9,2\%$
 - Holm-Bonferroni: $3 \cdot 1,6\% = 4,8\%$
 - Holm-Sidak: $1 - (1 - 1,6\%)^3 \approx 4,7\%$

Die beiden Holms-Varianten ergeben weiterhin signifikante p-Werte.