

Statistik – Übungen WS 2025

Blatt 4: Regressionsrechnung

Einfache lineare Regression

1. Die folgende Tabelle enthält Kennzahlen für 6 zufällig ausgewählte Filialen einer österreichischen Supermarktkette:

Filiale	Mitarbeiter	Verkaufsfläche (in m^2)	Ausgaben für lokales Marketing (in Tsd. €)	Umsatz (in Tsd. €)
1	9	1150	23	540
2	7	1000	21	510
3	12	1500	35	680
4	10	1200	28	620
5	10	1100	29	620
6	8	950	26	550

Analysieren Sie die Abhängigkeit des Umsatzes von der Verkaufsfläche.

- Erstellen Sie ein Streudiagramm für die beiden Merkmale „Verkaufsfläche“ und „Umsatz“.
- Bestimmen Sie den Pearson-Korrelationskoeffizienten.
- Geben Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit des Umsatzes von der Verkaufsfläche an.
- Berechnen Sie die Residuen.
- Berechnen Sie einen Schätzwert für die Standardabweichung des Merkmals „Umsatz“ und die Standardabweichung des Störterms. Vergleichen Sie diese miteinander!
- Wie viel Prozent der Streuung des Merkmals „Umsatz“ werden durch das Merkmal „Verkaufsfläche“ erklärt?
- Ist das Merkmal „Verkaufsfläche“ auf dem 5 %-Niveau signifikant?
- Welchen Umsatz würde ein Unternehmen mit 1250 m^2 Verkaufsfläche erzielen?
- Führen Sie die Regressionsanalyse in EXCEL durch.

2. **P 41** Ein Schulpsychologe hat an 75 Vorschulkindern die folgenden Kennwerte eines Schuleignungstests ermittelt: Mittelwert $\bar{x} = 41$, Standardabweichung $s_x = 4,5$. Nach dem ersten Schuljahr wurden die tatsächlichen schulischen Leistungen derselben Kinder erfasst. Diese weisen folgende Kennwerte auf: Mittelwert $\bar{y} = 31,6$, Standardabweichung $s_y = 3,7$. Die Kovarianz zwischen den Ergebnissen des Eignungstest und der schulischen Leistung beträgt $Cov(x, Y) = 12,25$.
- Stellen Sie die Regressionsgleichung zur Vorhersage der schulischen Leistungen (Y) auf Basis der Ergebnisse im Schuleignungstest (X) auf.
 - Erläutern Sie, wie der Regressionskoeffizient der erklärenden Variable zu interpretieren ist.
 - Mit welchem Wert für die schulische Leistung ist bei einer Schülerin zu rechnen, die im Schuleignungstest einen Wert von $x = 53$ erzielt hat?
 - Ermitteln Sie das Bestimmtheitsmaß zur Beurteilung der Güte des Regressionsmodells und interpretieren Sie dessen Aussage.
3. **P 42** Im Folgenden wird eine Regressionsanalyse dargestellt, die den Zusammenhang zwischen der Arbeitslosenzahl und der Anzahl offener Stellen in Österreich für die Jahre 2016 bis 2024 untersucht. Das Modell soll die Abhängigkeit der Arbeitslosenzahl (in Tausend) von der Anzahl offener Stellen (in Tausend) schätzen.

Die Analyse ergab folgenden Output:

Regressions-Statistik	
Bestimmtheitsmaß	0,016
Standardfehler	56,202
Beobachtungen	9

	Koeffizient	Standardfehler	t-Statistik	p-Wert	Untere 95%	Obere 95%
Schnittpunkt	318,687	86,436	3,687	0,008	114,299	523,075
offene Stellen	-0,413	1,221			-3,301	2,474

- Stellen Sie die geschätzte Regressionsgleichung zur Vorhersage der Arbeitslosenzahl (Y) auf Basis der Anzahl offener Stellen (X) auf.
- Welche Arbeitslosenzahl prognostiziert das Modell für das Jahr 2025, in dem 81.000 offene Stellen erwartet werden?
- Interpretieren Sie das Bestimmtheitsmaß in Bezug auf die Güte des Modells.
- Hat die Anzahl der offenen Stellen einen signifikanten Einfluss auf die Arbeitslosenzahlen? Führen Sie den entsprechenden Test zum Niveau $\alpha = 5\%$ durch und beantworten Sie die Frage.

Multiple lineare Regression

4. Die folgende Tabelle enthält Kennzahlen für 6 zufällig ausgewählte Filialen einer österreichischen Supermarktkette:

Filiale	Mitarbeiter	Verkaufsfläche (in m^2)	Ausgaben für lokales Marketing (in Tsd. €)	Umsatz (in Tsd. €)
1	9	1150	23	540
2	7	1000	21	510
3	12	1500	35	680
4	10	1200	28	620
5	10	1100	29	620
6	8	950	26	550

Analysieren Sie die Abhängigkeit des Umsatzes von der Verkaufsfläche und den Ausgaben für lokales Marketing.

- Geben Sie die Gleichung der Regressionsgeraden zur Beschreibung der Abhängigkeit des Umsatzes von der Verkaufsfläche und den Ausgaben für lokales Marketing an.
- Wie sind die einzelnen Parameter zu interpretieren?
- Sind die Merkmale „Verkaufsfläche“ sowie „Ausgaben für lokales Marketing“ auf dem 5 %-Niveau signifikant?
- Wie groß ist das adjustierte Bestimmtheitsmaß und wie ist es zu interpretieren?
- Wie groß ist die Standardabweichung des Störterms? Wie groß ist die Standardabweichung des Koeffizienten des Merkmals „Verkaufsfläche“ und wie ist sie zu interpretieren?
- Prognostizieren Sie den Umsatz für eine Filiale, die eine Verkaufsfläche von $1250 m^2$ besitzt und Ausgaben für lokales Marketing in Höhe von € 30.000.- veranschlagt.

5. **P 43** Ein Unternehmen möchte den monatlichen Absatz eines Produkts (in Tausend Stück) mithilfe zweier Einflussfaktoren erklären

- x_1 : Preis des Produkts (in Euro)
- x_2 : Werbeausgaben (in Tausend Euro)

und verwendet das Modell

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

mit y_i = Absatz (in Tsd. Stück).

Vier Beobachtungen wurden erfasst:

Die Analyse ergab folgenden Output:

Beobachtung	Preis x_1	Werbung x_2	Absatz y
1	2	1	2
2	3	3	4
3	4	4	6
4	5	6	9

- a) Ein Analyst ermittelt nach der Methode der kleinsten Quadrate folgendes Modell: $\hat{Y} = -1,75 + 1,5x_1 + 0,5x_2$. Berechnen Sie die Residuen und die Summe der quadrierten Residuen (SSE).
- b) Im Rahmen einer weiterführenden Analyse wird für den Schätzer $\beta_1 = 1,5$ auch der Standardfehler berechnet, der den Wert 1,803 ergibt.
- Berechnen Sie den Testwert, wenn Sie prüfen möchten, ob β_1 signifikant von Null abweicht.
 - Wie ist die Teststatistik unter der Nullhypothese $H_0 : \beta_1 = 0$ verteilt? Geben Sie die Verteilung und ihre Parameter an.
 - Beurteilen Sie, ob das Merkmal „Preis des Produkts“ auf einem Signifikanzniveau von 5 % signifikant ist. Begründen Sie Ihre Schlussfolgerung.
- c) Ein anderer Analyst hat ein alternatives Regressionsmodell entwickelt, das neben den bekannten Einflussfaktoren x_1 (Preis) und x_2 (Werbung) eine weitere unabhängige Variable einbezieht. Für die erste Beobachtung liegen die Messwerte $x_1 = 2$ und $x_2 = 1$ sowie $x_3 = 4$ vor. Das erweiterte Regressionsmodell lautet: $\hat{Y} = \beta_0 + 1,5x_1 + 0,5x_2 - 0,25x_3$. Das Residuum der ersten Beobachtung beträgt 0,2. Berechnen Sie mithilfe dieser Information den korrekten Wert für β_0 .

6. P 44 Für 10 Wohngebäude liegen Informationen zum Energieverbrauch (in Tsd. kWh pro Jahr), Wohnfläche (in m²), Alter (in Jahren), Anzahl der Bewohner, Wärmedämmung (schlecht, gut) und verwendetes Heizsystem (Gas, Öl, Wärmepumpe) vor. Es soll die Abhängigkeit des Energieverbrauchs von verschiedenen Einflussfaktoren untersucht werden. Unterstellen Sie dazu ein lineares Modell der Form

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \varepsilon$$

Die Daten sind in folgender Tabelle zusammengefasst:

Haus	Energie	Fläche	Alter	Bewohner	Dämmung	Heizsystem
1	24	85	30	2	schlecht	Gas
2	32	120	45	4	schlecht	Öl
3	18	90	10	2	gut	Wärmepumpe
4	40	160	50	5	schlecht	Öl
5	22	110	17	3	gut	Gas
6	35	140	35	4	schlecht	Öl
7	16	80	7	1	gut	Gas
8	28	130	25	3	schlecht	Wärmepumpe
9	20	100	13	2	gut	Wärmepumpe
10	38	150	40	5	schlecht	Öl

Analysieren Sie die Abhängigkeit des Energieverbrauchs von der Fläche, dem Alter sowie der Anzahl der Bewohner. Führen Sie dazu die Regression mit Hilfe von EXCEL durch. Beantworten Sie die nachfolgenden Fragen anhand Ihrer Auswertung:

- Geben Sie die Gleichung der Regressionsgeraden zur Beschreibung des Energieverbrauchs in Abhängigkeit von der Fläche, dem Alter sowie der Anzahl der Bewohner an.
- Wie sind die Koeffizienten der Variablen „Fläche“ bzw. „Bewohner“ zu interpretieren?
- Wie groß ist die Standardabweichung des Störterms? Wie groß ist die Standardabweichung des Koeffizienten des Merkmals „Bewohner“?
- Wie groß ist das Bestimmtheitsmaß und wie ist es zu interpretieren?
- Wie groß ist das adjustierte Bestimmtheitsmaß?
- Sind die Merkmale „Fläche“ sowie „Bewohner“ auf dem 5 %-Niveau signifikant?
- Prognostizieren Sie den Energieverbrauch für ein Haus mit einer Fläche von 105 m², das 15 Jahre alt ist und 4 Bewohner besitzt.

7. Die folgende Tabelle enthält Kennzahlen für 10 zufällig ausgewählte Absolventinnen und Absolventen der Wirtschaftswissenschaften:

Einstiegs- gehalt	Alter	Firmengröße (Anzahl Mitarbeiter)	Aus- bildung	Branchenspezifische Fortbildung	Branche
2950	27	1.000	MA	NEIN	Maschinenbau
3540	29	12.500	MA	NEIN	Software
5190	34	18.000	DR	JA	Pharma
2830	30	720	BA	JA	Bank
2900	26	4.000	MA	JA	Software
3110	27	1.400	MA	NEIN	Bank
2740	29	540	BA	JA	Pharma
4043	31	5.200	MA	NEIN	Pharma
6120	36	20.500	DR	JA	Maschinenbau
4100	33	7.700	DR	NEIN	Software

Es soll die Abhängigkeit des durchschnittlichen monatlichen (Brutto-)Einstiegsgehalts von Absolventinnen und Absolventen der Wirtschaftswissenschaften von verschiedenen Einflussfaktoren untersucht werden. Unterstellen Sie dazu ein lineares Modell der Form

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \varepsilon$$

- Analysieren Sie die Abhängigkeit des durchschnittlichen monatlichen (Brutto-)Einstiegsgehalts vom Alter.
- Untersuchen Sie die Abhängigkeit des durchschnittlichen monatlichen (Brutto-)Einstiegsgehalts von Alter und Firmengröße.
- Untersuchen Sie die Abhängigkeit des durchschnittlichen monatlichen (Brutto-)Einstiegsgehalts von Alter, Firmengröße sowie dem Vorliegen einer branchenspezifischen Fortbildung.
- Es wird vermutet, dass es einen Einfluss auf das durchschnittliche Einkommen hat, ob die Absolventinnen und Absolventen einen Bachelor-, Master oder einen Dokortitel vorweisen können. Untersuchen Sie nun die Abhängigkeit des durchschnittlichen monatlichen (Brutto-)Einstiegsgehalts von Alter, Firmengröße, branchenspezifischer Fortbildung und der (höchsten abgeschlossenen) Ausbildung.
- Um wie viel höher ist das durchschnittliche monatliche (Brutto-)Einstiegsgehalt, wenn eine Absolventin über einen Masterabschluss anstelle eines Bachelorabschlusses verfügt und alle übrigen Merkmale unverändert bleiben?

Die mit P gekennzeichneten Beispiele sind von den Studierenden vorzubereiten und nach Aufruf durch die Lehrveranstaltungsleitung zu präsentieren!