

Statistik – Übungen SS 2026

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

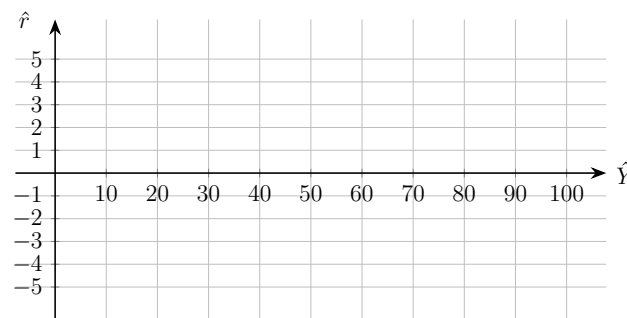
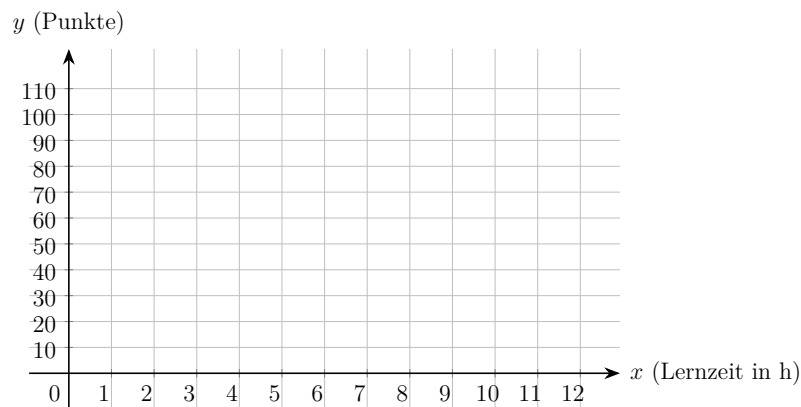
Analysieren Sie die Abhängigkeit des Umsatzes von der Verkaufsfläche.

- 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 beiden Werte 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 eine Filiale mit 1250 m^2 Verkaufsfläche erzielen?
- Führen Sie die Regressionsanalyse in EXCEL durch.

2. **P 41** Ein Online-Lernplattform-Betreiber möchte den Zusammenhang zwischen der investierten Lernzeit und dem Testerfolg untersuchen. In einem Pilotkurs wurden für $n = 6$ zufällig ausgewählte Teilnehmerinnen die wöchentliche Lernzeit x (in Stunden) und die erreichte Punktezahl y (von 0 bis 100 Punkten) gemessen:

Teilnehmerin	1	2	3	4	5	6
Lernzeit X (h/Woche)	2	4	5	7	8	10
Punkte Y	38	55	63	74	78	90

- Tragen Sie die Datenpaare in nachstehendes Koordinatensystem ein.
- Schätzen Sie die Regressionsparameter $\hat{\beta}_0$ und $\hat{\beta}_1$ mithilfe der Methode der kleinsten Quadrate.
- Zeichnen Sie die geschätzte Regressionsgerade in das Koordinatensystem ein.
- Interpretieren Sie den Wert $\hat{\beta}_1$ inhaltlich.
- Wie viele Punkte erreicht eine Teilnehmerin voraussichtlich, die 6 Stunden pro Woche lernt?
- Berechnen Sie das Bestimmtheitsmaß und interpretieren Sie den erhaltenen Wert.
- Zeichnen Sie im zweiten Diagramm den zugehörigen Residuenplot (Residuen gegen die geschätzten Werte).
- Ist es sinnvoll, das Modell zur Prognose der Punktezahl bei 20 Stunden Lernzeit pro Woche zu verwenden. Begründen Sie Ihre Antwort.



3. P 42 In einer Studie soll untersucht werden, ob die Anzahl der Schlafstunden pro Nacht einen Einfluss auf die Punktzahl in einem Konzentrationstest hat. Es wird ein einfaches lineares Regressionsmodell der Form

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

angenommen, wobei

- Y : Punktzahl in einem standardisierten Konzentrationstest,
- X : durchschnittliche Schlafdauer pro Nacht (in Stunden),
- ε : Störterm.

Es wurden $n = 7$ Personen untersucht und deren Schlafdauer gemessen. Auf Basis dieser Daten ergeben sich die geschätzten Regressionskoeffizienten $\hat{\beta}_0 = -1,3$, $\hat{\beta}_1 = 2,4$. Die geschätzte Standardabweichung des Störterms beträgt $\hat{\sigma} = 2,1$.

Weiters wurde $\sum_{i=1}^7 (x_i - \bar{x})^2 = 28$ berechnet.

- a) Berechnen Sie die Standardabweichung des Koeffizienten $\hat{\beta}_1$.
- b) Testen Sie zum Niveau $\alpha = 0,05$, ob die Schlafdauer einen signifikanten Einfluss auf die Punktzahl im Konzentrationstest hat.
- c) Betrachten Sie zwei Personen:

Person A schläft im Durchschnitt 5 Stunden pro Nacht.

Person B schläft im Durchschnitt 8 Stunden pro Nacht.

Bestimmen Sie mithilfe der geschätzten Regressionsgeraden die Differenz der erwarteten Punktzahlen (Punktzahl von Person B minus Punktzahl von Person A) und interpretieren Sie diese Differenz verbal im Kontext der Studie.

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“?
- 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 E-Commerce-Unternehmen wertet die Daten von 40 Produktlistungen aus, um den **monatlichen Umsatz** (in 1.000 Euro) in Abhängigkeit von **Werbebudget**, **Anzahl der Kundenbewertungen** und der **Kategorie** (Mode, Elektronik, Haushalt) zu erklären. Es werden zwei Dummyvariablen verwendet:

- $Mode = 1$, wenn das Produkt in der Kategorie Mode ist, sonst 0,
- $Elektronik = 1$, wenn das Produkt in der Kategorie Elektronik ist, sonst 0,
- Referenzkategorie: Haushalt.

Die Regressionsanalyse liefert folgenden (hier vereinfachten) Output:

Regressions-Statistik	
Bestimmtheitsmaß	0,82
Standardfehler	3,10
Beobachtungen	40

	Koeffizient	Standardfehler	t-Statistik	p-Wert	Untere 95%	Obere 95%
Schnittpunkt	5,20	2,00	2,60	0,013	1,16	9,24
Werbebudget (in 1.000 Euro)	0,80	0,18	4,44	0,000	0,43	1,17
Anzahl Bewertungen	0,01	0,005	2,00	0,052	0,00	0,02
Mode	-2,40	1,20	-2,00	0,052	-4,84	0,04
Elektronik	3,10	1,35	2,30	0,027	0,35	5,85

- Formulieren Sie die geschätzte Regressionsgleichung in Abhängigkeit von Werbebudget, Anzahl der Kundenbewertungen und den Dummyvariablen für die Kategorien.
- Schätzen Sie den Umsatz eines Haushaltsprodukts mit einem Werbebudget von 4.000 Euro und 200 Bewertungen.
- Wie ändert sich der erwartete Umsatz, wenn das Werbebudget um 1.000 Euro erhöht wird, bei ansonsten unveränderten Variablen?
- Interpretieren Sie den Koeffizienten der Dummyvariablen „Elektronik“ im Vergleich zur Referenzkategorie Haushalt.
- Wie groß ist der Anteil der durch das Modell erklärten Varianz?
- Was lässt sich anhand des p-Werts der Variablen „Mode“ über deren Signifikanz auf dem 5 %-Niveau sagen? Begründen Sie Ihre Antwort.

6. **P 44** Ein städtisches Planungsbüro verfügt über Daten zu 10 Wohngebäuden und untersucht, welche Faktoren den jährlichen Wasserverbrauch beeinflussen. Erfasst wurden: jährlicher Wasserverbrauch (in m³), Wohnfläche (in m²), Anzahl der Bewohner, Baujahr (in Jahren seit 1900), Vorhandensein einer Regenwassernutzungsanlage (ja/nein), der Gebäudetyp (Einfamilienhaus (EF)/Mehrfamilienhaus (MF)) sowie die Heizungsart (Gas (G)/Fernwärme (F)/Wärmepumpe (WP)).

Für die folgende Analyse sollen zunächst nur die metrischen Einflussfaktoren Wohnfläche, Anzahl der Bewohner und Baujahr im Regressionsmodell berücksichtigt werden. Die Daten (Auszug) sind:

Haus	Wasserverbrauch (m ³ /Jahr)	Wohnfläche (in m ²)	Bewohner	Baujahr	Regenwasser	Gebäudetyp	Heizungsart
1	120	80	2	1975	nein	EF	G
2	180	120	4	1988	ja	EF	G
3	95	70	2	2005	ja	MF	F
4	210	140	5	1980	nein	EF	G
5	130	95	3	1998	ja	EF	WP
6	160	110	3	1992	nein	MF	F
7	85	60	1	2010	ja	MF	WP
8	190	130	4	1985	nein	EF	G
9	105	75	2	2002	ja	EF	F
10	200	135	5	1978	nein	MF	G

Unterstellen Sie das Modell

$$Y = \beta_0 + \beta_1 \cdot \text{Wohnfläche} + \beta_2 \cdot \text{Bewohner} + \beta_3 \cdot \text{Baujahr} + \varepsilon$$

Analysieren Sie die Abhängigkeit des Wasserverbrauchs von der Wohnfläche, der Anzahl der Bewohner und dem Baujahr. 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 Wasserverbrauchs in Abhängigkeit von Fläche, Bewohnerzahl und Baujahr an.
- Wie sind die Koeffizienten der Variablen „Wohnfläche“ 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 Wasserverbrauch für ein Haus mit einer Fläche von 100 m², 3 Bewohnern und einem Baujahr 1995.

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

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

- 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!