

Kapitel III

Regressionsanalyse

Aufgaben

A. Rechenaufgaben

1.

Die nachfolgende Tabelle zeigt Daten über das Einkommen und die Ausgaben für Lebensmittel [in 100 \$] von 7 US-amerikanischen Familien in einem bestimmten Monat:

Einkommen	Ausgaben für Lebensmittel
35	9
49	15
21	7
39	11
15	5
28	8
25	9

Die Ausgaben für Lebensmittel sollen als eine lineare Funktion des Einkommens dargestellt werde.

1. Stellen Sie die Daten als Streudiagramm dar.
2. Bestimmen Sie die Regressionsfunktion. Stellen Sie sie graphisch dar.
3. Berechnen Sie die Residuen und die quadrierten Residuen.
4. Interpretieren Sie die Regressionskoeffizienten a_0 und a_1 .
5. Interpretieren Sie die wichtigsten Voraussetzungen des Regressionsmodells konkret anhand der vorliegenden Aufgabe.
6. Berechnen Sie die Standardabweichung der Residuen.
7. Berechnen und interpretieren Sie das Bestimmtheitsmaß.

8.

Ermitteln Sie die Standardfehler der Regressionsparameter a_0 und a_1 . Wie weit sind sie annehmbar?

9.

Bestimmen Sie 95%-Konfidenzintervalle für die Regressionsparameter A_0 und A_1 .

10.

Testen Sie mit einem Signifikanzniveau von 1% die Steigung der Regressionsfunktion auf Positivität.

2.

Ein Forscher möchte die Auswirkung der Fahrerfahrungen (Jahre) und der Anzahl der Verkehrsdelikte in den letzten drei Jahren auf dem monatlichen Versicherungsbeitrag (\$) in einer Großstadt der USA untersuchen. Dazu wurde eine Stichprobe von 12 Autofahrern gewählt. Die nachfolgende Tabelle zeigt die gewonnenen Informationen:

Monatlicher Beitrag [\$]	Fahrerfahrung [Jahr]	Verkehrsdelikte [Anzahl]
148	5	2
76	14	0
100	6	1
126	10	3
194	4	6
110	8	2
114	11	3
86	16	1
198	3	5
92	9	1
70	19	0
120	13	3

1. Berechnen und interpretieren Sie die Koeffizienten der entsprechenden linearen multiplen Regressionsfunktion:

$$y^* = a_0 + a_1x_1 + a_2x_2$$

mit

x_1 : Fahrerfahrung [Jahre]

x_2 : Verkehrsdelikte [Anzahl].

2. Berechnen und interpretieren Sie das Bestimmtheitsmaß der multiplen Regression.
3. Welchen monatlichen Beitrag muss ein Kunde mit 12 Jahren Fahrerfahrung und vier Verkehrsdelikten in den vergangenen 3 Jahren leisten?

4. Konstruieren Sie Konfidenzintervalle für die Regressionskoeffizienten mit einem Signifikanzniveau von 0.05.
5. Führen Sie einen F-Test mit einem Signifikanzniveau von 5% aus.
6. Testen Sie mit einem Signifikanzniveau von 5% den Koeffizienten A_1 auf Negativität und den Koeffizienten A_2 auf Positivität.

Hinweis:

$$(X^T X)^{-1} = \begin{pmatrix} 1.449 & -0.089 & -0.216 \\ -0.089 & b & 0.011 \\ -0.216 & 0.011 & c \end{pmatrix}.$$

3.

Eine zufällige Stichprobe von 5 Familien ergab folgende Ergebnisse:

Ersparnisse [Tsd. \$]	Einkommen [Tsd. \$]	Anzahl der Kinder
2,1	15	2
3,0	28	4
1,6	20	4
2,1	22	3
1,2	10	2

1. Berechnen und interpretieren Sie die Koeffizienten der entsprechenden linearen multiplen Regressionsfunktion:

$$y^* = a_0 + a_1 x_1 + a_2 x_2$$

mit

y^* : Ersparnisse [Tsd.]

x_1 : Einkommen [Tsd.]

x_2 : Anzahl der Kinder [Tsd.].

2. Berechnen und interpretieren Sie das Bestimmtheitsmaß der multiplen Regression.
3. Konstruieren Sie Konfidenzintervalle für die Regressionskoeffizienten A_1 und A_2 mit einem Signifikanzniveau von 0.05.
4. Testen Sie mit einem Signifikanzniveau von 5% den Koeffizienten A_1 auf Positivität und den Koeffizienten A_2 auf Negativität.

Hinweis:

$$(X^T X)^{-1} = \begin{pmatrix} 2.50493 & -0.03139 & -0.56951 \\ -0.03139 & b & -0.10314 \\ -0.56951 & -0.10314 & c \end{pmatrix}.$$

4.

Eine zufällige Stichprobe von 5 Familien ergab folgende Ergebnisse:

Ersparnisse [Tsd. \$]	Einkommen [Tsd. \$]	Anzahl der Kinder
2,1	15	2
3,0	28	4
1,6	20	4
2,1	22	3
1,2	10	2

1. Berechnen und interpretieren Sie die Koeffizienten der entsprechenden linearen multiplen Regressionsfunktion:

$$y^* = a_0 + a_1x_1 + a_2x_2$$

mit

y^* : Ersparnisse [Tsd.]

x_1 : Einkommen [Tsd.]

x_2 : Anzahl der Kinder [Tsd.].

2. Berechnen und interpretieren Sie das Bestimmtheitsmaß der multiplen Regression.
3. Konstruieren Sie Konfidenzintervalle für die Regressionskoeffizienten A_1 und A_2 mit einem Signifikanzniveau von 0.05.
4. Testen Sie mit einem Signifikanzniveau von 5% den Koeffizienten A_1 auf Positivität und den Koeffizienten A_2 auf Negativität.

Hinweis:

$$(X^T X)^{-1} = \begin{pmatrix} 2.50493 & -0.03139 & -0.56951 \\ -0.03139 & b & -0.10314 \\ -0.56951 & -0.10314 & c \end{pmatrix}.$$

5.

Folgende Tabelle zeigt den Verkaufspreis, das Alter und die Fläche von 5 Ferienwohnungen in Dänemark:

Preis [1000 €]	Alter [Jahr]	Fläche [m ²]
745	36	66
895	37	68
442	47	64
440	32	53
1598	1	101

1. Ermitteln Sie die entsprechende lineare multiple Regressionsfunktion:

$$y^* = a_0 + a_1x_1 + a_2x_2$$

mit

x_1 : Alter [Jahr]

x_2 : Fläche [m²]

y : Preis [Tsd €].

2. Berechnen und interpretieren Sie das Bestimmtheitsmaß der multiplen Regression.
3. Konstruieren Sie Konfidenzintervalle für die Regressionskoeffizienten A_1 und A_2 mit einem Signifikanzniveau von 0.05.
4. Testen Sie mit einem Signifikanzniveau von 5% den Koeffizienten A_1 auf Negativität und den Koeffizienten A_2 auf Positivität.

Hinweis:

$$(X^T X)^{-1} = \begin{pmatrix} 27.5307 & -0.26579 & -0.27269 \\ -0.26579 & b & 0.00247 \\ -0.27269 & 0.00247 & c \end{pmatrix}.$$

B. SPSS-Aufgaben

1.

Datendatei: *macro.sav*

1.
Bestimmen Sie gleichzeitig die Regressionsgleichungen für folgende Modelle:

Modell 1:

Privater Konsum in Abhängigkeit vom verfügbaren Einkommen und Zinssatz.

Modell 2:

Privater Konsum in Abhängigkeit vom verfügbaren Einkommen, Zinssatz und Lohnquote.

2.
Interpretieren Sie im Modell 2 die Koeffizienten der Regressionsgleichung.

3.
Bestimmen Sie die Signifikanz der Regressionskoeffizienten in beiden Modellen mit einer Irrtumswahrscheinlichkeit von 5%.

4.
(Erklären Sie was die Beta-Koeffizienten sind und welche Bedeutung sie haben.)

5.
Interpretieren Sie das Bestimmtheitsmaß und das korrigierte Bestimmtheitsmaß der Regressionsgleichungen.

6.
Erläutern und interpretieren Sie die Ergebnisse des F-Tests.

7.
Geben Sie Konfidenzintervalle für die Regressionskoeffizienten an.

8.
Welche Kennziffer gibt Auskunft über die Richtung einer linearen Beziehung zwischen zwei Variablen? Illustrieren Sie Ihre Antwort an konkreten Fällen.

9.

Diagnostizieren Sie die Kollinearität der Regressionskoeffizienten anhand des Konditionsindex.

(Hinweis: Als Faustregel gilt, dass bei einem Konditionsindex zwischen 10 und 30 moderate bis starke und über 30 sehr starke Multikollinearität liegt.)

2.

Datendatei: *world95.sav*

Es soll die Lebenserwartung eines Mannes in dem Land (*lifeexpm*) als eine lineare Funktion folgender Faktoren dargestellt werden:

1. Anteil der Bevölkerung, die in Städten lebt (*urban*)
2. Anteil der Bevölkerung, die lesen kann (*literacy*)
3. Brutto sozialprodukt je Einwohner (*gdp_cap*)
4. Durchschnittliche Kalorienaufnahme je Person und Tag (*calories*)

1.

Bestimmen Sie die entsprechende Regressionsgleichung.

2.

Bestimmen Sie die Signifikanz der Regressionskoeffizienten mit einer Irrtumswahrscheinlichkeit von 5%.

3.

Interpretieren Sie das Bestimmtheitsmaß und das korrigierte Bestimmtheitsmaß der Regressionsgleichungen.

4.

Erläutern und interpretieren Sie die Ergebnisse des F-Tests.

5.

Geben Sie Konfidenzintervalle für die Regressionskoeffizienten an.

6.

Welche Kennziffer gibt Auskunft über die Richtung einer linearen Beziehung zwischen zwei Variablen? Illustrieren Sie Ihre Antwort an konkreten Fällen.

9.

Diagnostizieren Sie die Kollinearität der Regressionskoeffizienten anhand des Konditionsindex.

(Hinweis: Als Faustregel gilt, dass bei einem Konditionsindex zwischen 10 und 30 moderate bis starke und über 30 sehr starke Multikollinearität liegt.)