

Aufgaben zu Lineare Einfachregression

Aufgabe 1

Der Ausstoß an Kohlenwasserstoffen hängt bei Automotoren von der Drehzahl ab. Versuche ergaben folgende Werte, mit deren Hilfe Sie eine lineare Einfachregression durchführen sollen.

Drehzahl (in 1 000 Umdrehungen/min)	2	4	6	8
Kohlenwasserstoffe (in mg/min)	1	2	5	10

1. Beschreiben Sie den Vorgang der Datenerhebung, der diesem Datenmaterial zugrunde liegt!
2. Zeichnen Sie die Beobachtungswerte in ein Koordinatensystem ein!
3. Schätzen Sie mithilfe der Methode der kleinsten Quadrate die Parameter a und b der Regressionsgeraden!
4. Interpretieren Sie den erhaltenen Wert von \hat{b} inhaltlich!
5. Zeichnen Sie die Regressionsgerade in das obige Koordinatensystem ein!
6. Mit welchem Ausstoß an Kohlenwasserstoffen (in mg/min) rechnen Sie bei einer Drehzahl von
 - a) 3500 Umdrehungen pro Minute
 - b) 12000 Umdrehungen pro Minute?
7. Berechnen Sie das Stichprobenbestimmtheitsmaß R^2 !
8. Interpretieren Sie den unter g) erhaltenen Wert!

Aufgabe 2

In einem Windkanal wird der Prototyp einer Windkraftanlage getestet. Die Abhängigkeit der erzeugten Leistung (in kW/h) von der Windgeschwindigkeit (in m/sec) soll mit Hilfe einer linearen Einfachregressionsanalyse untersucht werden. Folgende Messwerte wurden ermittelt:

Windgeschwindigkeit (in m/sec)	2	3	4	6	10
Leistung (in kW/h)	50	80	80	90	100

1. Beschreiben Sie den Vorgang der Datenerfassung, d. h. wie zum Zwecke der Regressionsanalyse diese Daten gesammelt wurden und erläutern Sie den Unterschied zur Datenerfassung bei der Korrelationsanalyse.
2. Zeichnen Sie die Werte in ein Streuungsdiagramm ein! Ist der lineare Ansatz gerechtfertigt?
3. Bestimmen Sie mit der Methode der kleinsten Quadrate die Parameter der Regressionsgeraden! Zeichnen Sie sodann die Regressionsgerade in das Streuungsdiagramm ein! Durch welche Punkte verläuft eine so ermittelte Regressionsgerade stets?
4. Interpretieren Sie den erhaltenen Wert von \hat{b} inhaltlich!
5. Welche Leistung erwarten Sie, wenn die Windgeschwindigkeit 8 m/sec bzw. 14 m/sec beträgt?
6. Berechnen Sie das Stichprobenbestimmtheitsmaß R^2 und interpretieren Sie den erhaltenen Wert!

Aufgabe 3

Ein Bäcker möchte ein neues Backpulver ausprobieren. Speziell möchte er wissen, ob die Höhe seiner Kuchen linear von der Menge des verwendeten Backpulvers abhängt. Zu diesem Zweck backt er je zwei Kuchen mit Backpulvermengen von 2g, 4g, 6g und 8g. Die Backergebnisse sind in folgender Tabelle dargestellt:

Backpulvermenge in g	2	2	4	4	6	6	8	8
Höhe des Kuchens in mm	20	10	40	40	60	50	60	40

1. Beschreiben Sie die Art der Datenerhebung im Modell der Regressionsrechnung!
2. Zeichnen Sie die Wertepaare in ein Koordinatensystem ein! Ist der lineare Ansatz gerechtfertigt?
3. Schätzen Sie die Koeffizienten \hat{a} und \hat{b} der Regressionsgerade!
4. Zeichnen Sie die Regressionsgerade in das Koordinatensystem ein! Durch welche beiden Punkte verläuft diese!
5. Interpretieren Sie den erhaltenen Wert von \hat{b} inhaltlich!
6. Schätzen Sie die Höhe des Kuchens, wenn für den Teig 3g, 6g bzw. 10g Backpulver verwendet werden!
7. Berechnen und interpretieren Sie das Stichprobenbestimmtheitsmaß R^2 für den vorliegenden Fall!

Aufgabe 4

Um die Abhängigkeit der „Dauer der abendlichen Heimfahrt (in Minuten)“ von der Entfernung des Wohnsitzes vom Arbeitsplatz (in km)“ der Angestellten eines Betriebes zu schätzen, wollen Sie mit den erhobenen Daten eine lineare Einfachregression durchführen. Dazu ist die Befragung von 5 Angestellten in nachstehender Tabelle zusammengefasst.

An-gestellter	Entfernung des Wohnsitzes (in km)	Dauer der Heimfahrt (in min)
1	2	8
2	6	4
3	8	12
4	10	20
5	14	16

1. Stellen Sie das Befragungsergebnis grafisch dar, d. h. zeichnen Sie das Streudiagramm! Ist der lineare Ansatz gerechtfertigt?
2. Beschreiben Sie den Vorgang der Datenerfassung, d. h. wie Sie zum Zwecke der Regressionsanalyse diese Daten gesammelt haben!
3. Bestimmen Sie im Rahmen einer linearen Regressionsanalyse die Koeffizienten \hat{a} und \hat{b} der Regressionsgerade nach der Methode der kleinsten Quadrate!
4. Zeichnen Sie die Regressionsgerade in das Koordinatensystem ein! Durch welche beiden Punkte verläuft jede so ermittelte Regressionsgerade?
5. Interpretieren Sie den Wert von \hat{b} inhaltlich!
6. Schätzen Sie aufgrund der durchgeführten Analyse die Dauer der Heimfahrt, wenn die Entfernung des Wohnsitzes vom Arbeitsplatz 5 km, 8 km bzw. 20 km beträgt? Halten Sie die Schätzungen für realistisch?
7. Berechnen Sie das Stichprobenbestimmtheitsmaß R^2 und interpretieren Sie den erhaltenen Wert!

4. Regressions- und Korrelationsrechnung

4.1 Regression

4.1.1 Kleinste - Quadrate - Schätzer

$$\hat{b} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2}$$

$$\frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

4.1.2 Bestimmtheitsmaß

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i \hat{y}_i^2 - n\bar{y}^2}{\sum_i y_i^2 - n\bar{y}^2}$$

Aufgabe 3

Ein Bäcker möchte ein neues Backpulver ausprobieren. Speziell möchte er wissen, ob die Höhe seiner Kuchen linear von der Menge des verwendeten Backpulvers abhängt. Zu diesem Zweck backt er je zwei Kuchen mit Backpulvermengen von 2g, 4g, 6g und 8g. Die Backergebnisse sind in folgender Tabelle dargestellt:

Backpulvermenge in g	2	2	4	4	6	6	8	8
Höhe des Kuchens in mm	20	10	40	40	60	50	60	40

1. Beschreiben Sie die Art der Datenerhebung im Modell der Regressionsrechnung!
2. Zeichnen Sie die Wertepaare in ein Koordinatensystem ein! Ist der lineare Ansatz gerechtfertigt?
3. Schätzen Sie die Koeffizienten \hat{a} und \hat{b} der Regressionsgerade!
4. Zeichnen Sie die Regressionsgerade in das Koordinatensystem ein! Durch welche beiden Punkte verläuft diese!
5. Interpretieren Sie den erhaltenen Wert von \hat{b} inhaltlich!
6. Schätzen Sie die Höhe des Kuchens, wenn für den Teig 3g, 6g bzw. 10g Backpulver verwendet werden!
7. Berechnen und interpretieren Sie das Stichprobenbestimmtheitsmaß R^2 für den vorliegenden Fall!

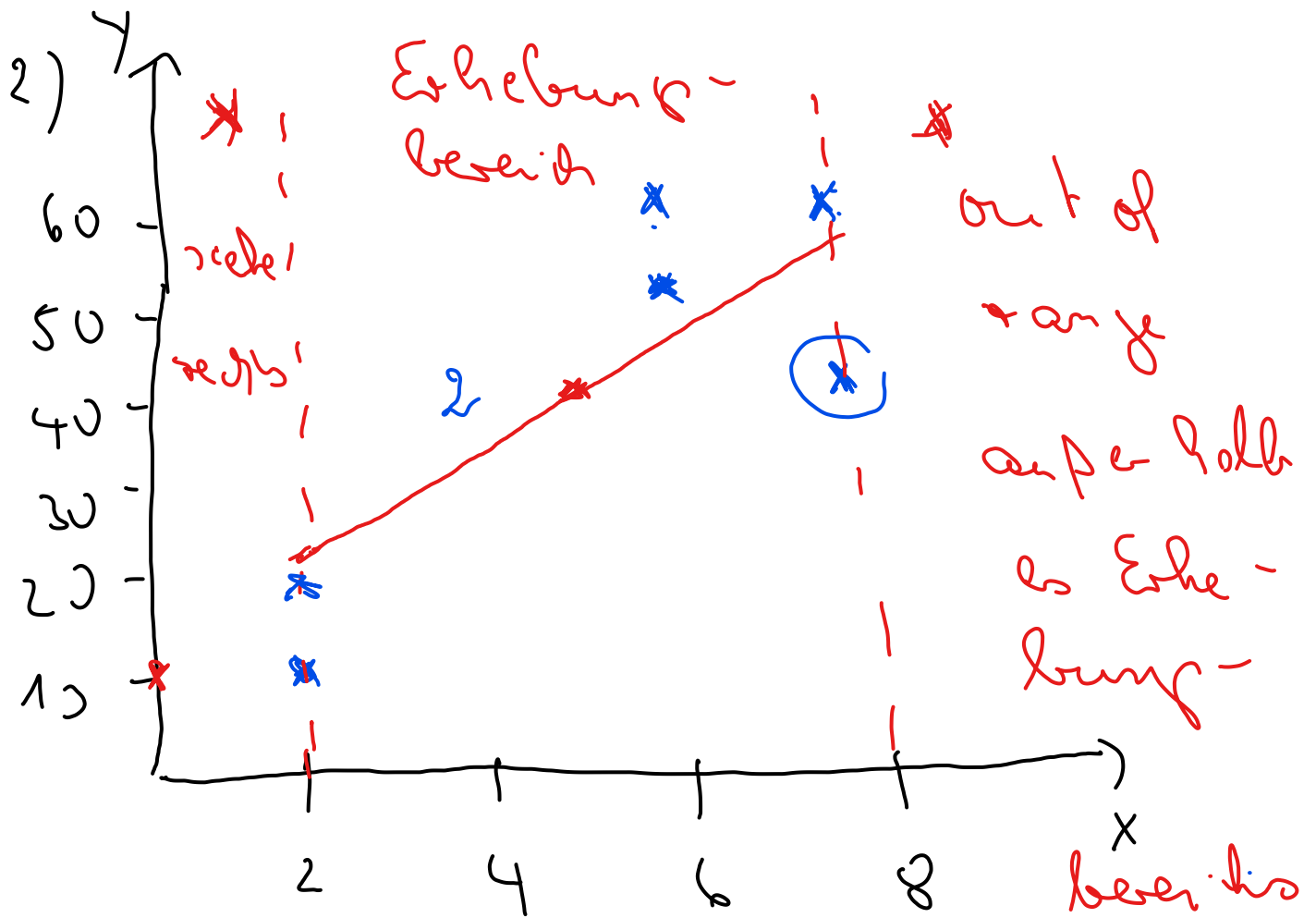
noch über Methode
der kleinste Quadrate.

1) X : Backpulvermenge (in g)
 Y : Höhe des Kuchens (in mm)

mit der Datenerhebung:

(- / $X = x_i$)

Die Höhe der Kuchen wurde für fest vorgegebene BP-Mengen ermittelt



lineare Ansatz ebenfalls
 okay (1) (8/40) Ansatzpar?

3.) $\hat{y}(x) = \hat{a} + \hat{b}x$

$$\hat{b} = \frac{\sum x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

$$\hat{a} = \bar{y} - \hat{b} \bar{x}$$

$$\hat{b} = \frac{1840 - 8 \cdot 5 \cdot 40}{240 - 8 \cdot 5^2} = 6 \quad \hat{a} = 40 - 6 \cdot 5 = 10$$

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\bar{y} = \frac{1}{n} \sum y_i$$

$$\bar{x} = \frac{1}{8} \cdot 40 = 5$$

$$\bar{y} = \frac{1}{8} \cdot 320 = 40$$

$$n = 8$$

x_i	y_i	$x_i y_i$	x_i^2
2	20	40	4
2	10	20	4
4	40	160	16
4	40	160	16
6	60	360	36
6	50	300	36
8	60	480	64
8	40	320	64
40	320	1840	240

$$\hat{y}(x) = 10 + 6x$$

4) jede so ermittelte Regressionsgerade verläuft stets durch:

a) $(\bar{x} | \bar{y})$ Schwerpunkt

hier

$(5 | 40)$

b) $(0 | \hat{a})$ Ordinatenabschnitt

hier:

$(0 | 10)$

Achtung:

Die Regressionsgerade gilt nur ein Erdbebenrisikoball.

5) \hat{b} : Steigung der Regressionsgerade.

inhaltlich. Wird die Bevölkerung

um 1% erhöht, erhöht sich die

Höhe des Kuchens um 6 mm.

$$\hat{b} = 6$$

b) Schönem

Erhebungspintervall $[2, 8]$

a) $x = 3 \quad 3 \in [2, 8]$

Schönem ist sinnvoll

$$\hat{y}(x) = 10 + 6 \cdot x$$

$$\hat{y}(3) = 10 + 6 \cdot 3 = 28$$

Bei 3 g BP wird ein Kuchen

28 mm hoch.

b) $x = 6 \quad 6 \in [2, 8]$

$$\hat{y}(6) = 10 + 6 \cdot 6 = 46$$

Bei 6 g BP wird ein

Kuchen 46 mm hoch

c) $10 \notin [2, 8]$

nicht
Element

out of
range

\Rightarrow ein Schöbun ist
nicht sinnvoll, da
ein Wert außerhalb
des Erkenntnisbereichs
gewählt wurde.

7) Stichprobenbestimmungsmaß R^2

Wie viel Prozent der Streuung
der y -Werte werden durch die
eingepasste Regressions-

gerade erklärt?

$$0 \leq R^2 \leq 1$$

$R^2 = 0,655$ 65,5% der Streuung
der y -Werte sind durch die
eingepasste Regressionsgerade
erklärt (multiplizierbarer Wert).

→ eventuell multiple Regres-
sion verwenden

