

Übung 5

-

Zeitreihen, Regression und Korrelation

Aktuelle Version: 19. Juli 2022

Hinweise:

- Übungen sind mit Vorteil alleine zu lösen.
- Benutzen Sie die Musterlösungen nur zur Korrektur.
- Die Übungen sind wichtige Vorbereitungen für die Prüfung. Lösen sie die Übungen sorgfältig und stellen Sie die Lösungswege übersichtlich dar.
- (Ergänzte) Vorlesungsunterlagen und Fachbücher helfen beim Lösen von Übungen und bringen gleichzeitig eine erweiterte Ansicht auf die Problemstellung.
- Wenn Sie die Übungen nicht verstehen, fragen Sie!

Übung 1. *Fragen*

1. Was ist eine Zeitreihe?
2. Welche Fragen interessieren bei der Untersuchung des Zusammenhangs zwischen zwei Merkmalen?
3. Wozu kann ein Streudiagramm in einem ersten Analyseschritt genutzt werden?
4. Wir haben die lineare Regression mit der Methode der kleinsten Quadrate kennengelernt. Welche weiteren, sinnvollen Formen können sie sich vorstellen?
5. Erläutern Sie den Begriff Kovarianz!
6. Erläutern Sie den Begriff Korrelationskoeffizient! Welchen Wertebereich kann der Korrelationskoeffizient annehmen?
7. Können Sie davon ausgehen, dass bei einem Korrelationskoeffizient von $|r| < 1$ *kein* perfekter Zusammenhang besteht?
8. Welcher Skala müssen Ihre Daten genügen, um eine Bravais-Pearson-Korrelation zu berechnen?
9. Welche Kausalität liegt bei einem Korrelationskoeffizient $|r| > 0$ der Merkmale x und y vor?
10. Erklären Sie den Unterschied zwischen formaler und sachlicher Abhängigkeit!

Übung 2. *Trend*

Die Daten der vierteljährlichen Anlieferungen in kg von Schlacke sind unten angegeben:

$$y_i = (841, 957, 999, 960, 894, 851, 863, 878, 792, 589, 657, 699, 675, 653, 628, 529, 480, 539, 557, 602, 659, 745, 845, 876, 808, 840, 893, 950, 884, 877, 905, 909)$$

1. Stellen Sie den Verlauf der Zeitreihe zeitlich dar.
2. Stellen Sie den Verlauf des gleitenden Mittels über 3 Quartale dar.
3. Stellen Sie den Verlauf des gleitenden Mittels über 4 Quartale dar.
4. Können sie einen Trend erkennen?

Übung 3. *Regression*

Der erforderliche Verbrauch an Grundmaterial eines Produktionsprozesses ist abhängig von der eingestellten Grösse des Produktes. Die Grösse kann jeden beliebigen reellen Wert zwischen 1 und 7 Einheiten einnehmen.

Um eine Formel zu entwickeln, die den Bedarf an Grundmaterial als Funktion der Grösse abschätzt, sind Experimente durchgeführt worden.

Die im Experiment gemessenen Daten sind jeweils für fünf Produktionstypen unten tabellarisch erfasst (Typ 1, Typ 2, Typ 3, Typ 4, Typ 5).

Bestimme Sie für jeden der Produktionstypen:

- Das entsprechende Streudiagramm und dadurch den vermuteten Zusammenhang.
- Eine Formel, die den Wert $f(x)$ bestimmt. Visualisieren Sie die Formel im Streudiagramm.
- Die Kovarianz und den Korrelationskoeffizienten, sofern sinnvoll.

1. Typ 1

x_i	y_i
1	3.2
2	4.2
3	9
4	8
5	12
6	11.5
7	13

2. Typ 2

x_i	y_i
1	0.20
2	0.40
3	0.80
4	1.44
5	2.40
6	5.00
7	10.60

3. Typ 3

x_i	y_i
1	52.5
2	34.0
3	13.5
4	0.0
5	2.5
6	30.0
7	91.5

4. Typ 4

x_i	y_i
1	6
2	3.3
3	8.1
4	0.2
5	9
6	2.7
7	5.2

5. Typ 5

x_i	y_i
1	0.5
2	1.5
3	2
4	2.4
5	2.6
6	2.9
7	3

Zusatzaufgaben

Übung 4. Korrelations- und Determinationskoeffizienten

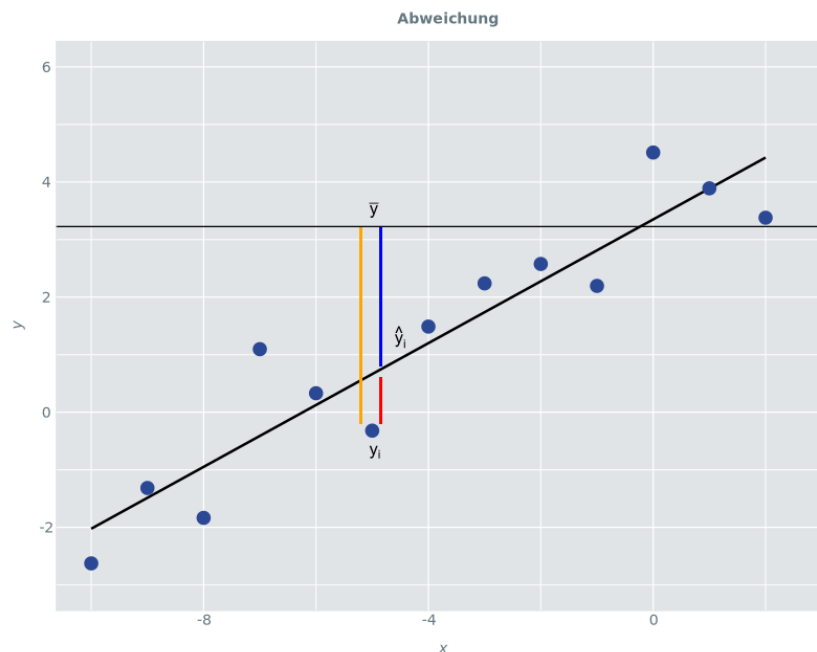
In drei Untersuchungen mit gleich grossen Stichproben wurde jeweils eine einfache lineare Regression durchgeführt. Es wurden folgende Korrelationskoeffizienten ermittelt: $r_1 = 0.75$, $r_2 = 0.49$ und $r_3 = 0.62$.

1. Welche lineare Regression würden Sie verwenden?

Hinweis: Der Determinationskoeffizient R^2 gibt an, wie viel der Gesamtvarianz durch die Regression erklärt wird¹. Bei einer einfachen linearen Regression berechnet sich dieser mittels:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum e^2}{\sum (y_i - \bar{y})^2} = r^2$$

Wobei es sich bei \hat{y} um die durch die Regression vorhergesagten Werte, bei y um die tatsächlich gemessenen Werten handelt und bei $e = y - \hat{y}$ (im Bild rot) um das Residuum handelt²:



2. Wie gross ist die Gesamtkorrelation? Hinweis: Korrelationskoeffizienten können nach *Fishers z-Transformation* $z = \frac{1}{2}(\ln(1 + r) - \ln(1 - r))$ gemittelt werden.

¹Mehr Informationen dazu findet sich zum Beispiel hier: https://www.inwt-statistics.de/blog-artikel-lesen/Bestimmtheitsmass_R2-Teil1.html

²Bildquelle https://www.inwt-statistics.de/blog-artikel-lesen/Bestimmtheitsmass_R2-Teil3.html

Übung 5. *Regression*

Sie haben die Anzahl Ebola-Neuansteckungen über einen Zeitraum in einer bestimmten Region erhoben:

x_i	y_i
1	5
2	7
3	20
4	34
5	45

Wie hoch schätzen Sie die Inzidenz für x_{30} ?