

Masterkurs Statistik – Aktualisierte Übungsbeispiele

In Klammern finden sie teilweise Lösungswerte (ohne Gewähr).

1. Gegeben ist die folgende Messung von zwei Merkmalen X und Y an n Werkstücken:

x_i	2	5	8	3	6
y_i	0.5	3	3.8	1.5	2.8

- (a) Bestimmen sie eine Regressionsgerade für die abhängige Variable Y .
 $[y = -0.2105 + 0.527x]$
- (b) Geben sie das R^2 an. $[R^2 = 0.923]$
- (c) Testen sie die Daten auf Korrelation und geben sie das Signifikanzniveau Ihrer Aussage an. $[t=5.983]$
- (d) Zeigen sie mit einem statistischen Test, dass der Anstieg der Regressionsgeraden größer als 0.25 ist (einseitig!, Fehlerwahrscheinlichkeit= 0.05).
 $[t = 3.146]$

2. Führen sie für die folgende Zeitreihe eine Trendrechnung mit einem loglinearen und einem halblogarithmischen Modell durch. Prognostizieren sie damit jeweils den Wert für die Periode $t = 9$. $[y = 66.81; y = 185.64]$

t_i	1	2	3	4	5	6
y_i	10	11	16	25	38	66

3. Bestimmen sie für die folgenden Daten eine quadratische Trendfunktion und prognostizieren sie den Wert für $t = 6$: $[y = 27.7]$

t_i	0	1	2	3	4
y_i	8	6	6.5	9	13

4. Für die folgenden Quartalsdaten Q_i der Jahre 2007, 2008 und 2009 soll unter Berücksichtigung der Saisonkomponenten eine lineare Trendrechnung durchgeführt werden. Prognostizieren sie damit den Wert für alle Quartale des Jahres 2011.

$[Q(t) = 3.888 + 0.513t; 14.13, 11.02, 10.66, 17.67]$

i	1	2	3	4	5	6	7	8	9	10	11
Q_i	6	3	2	9	8	5	5	12	10	7	6

5. Eine Buschenschank hat jeweils nur Freitag, Samstag und Sonntag geöffnet. In den Kalenderwochen 1 und 2 wurden die folgenden Umsatzzahlen beobachtet. Führen sie unter Berücksichtigung der Saisonkomponenten eine lineare Trendrechnung durch. Prognostizieren sie damit die Umsatzwerte für alle drei Öffnungstage der 4. Kalenderwoche. $[9.877, 14.948, 12.019]$

	KW 1			KW 2		
Tag i	1	2	3	4	5	6
Umsatz	6	9	6	6	12	9

6. Führen sie für die folgende Zeitreihe eine exponentielle Glättung mit der Methode von Holt-Winters mit den üblichen Startwerten durch. Verwenden sie die Parameter $\alpha = 0.4, \beta = 0.8$. $[g_{10}^H = 11.38, b_{10} = 1.6]$

t_i	1	2	3	4	5	6	7	8	9	10
y_i	2	6	2.5	7.5	3.5	10	4	13	6	16

7. Gegeben ist die lineare Funktion $z = 3 - 2.5x_1 + 3x_2 + 0.5x_3$ und folgende Daten einer Stichprobe mit $n = 4$:

i	x_1	x_2	x_3	y
1	1	1	1	1
2	2	0	3	0
3	2	1	1	0
4	3	1	0	0

- (a) Stellen sie eine logistische Regression auf und schätzen sie für alle Werte der Stichprobe das Auftreten des betrachteten Ereignisses Y .
[0.982, 0.377, 0.817, 0.182]
- (b) Bestimmen sie daraus die hit-ratio und das Press'sche Q . [$h = 0.75$]
- (c) Um welchen Faktor ändern sich die Odds, wenn eine Variable x_i um eine Einheit erhöht wird? Bestimmen sie den Faktor für x_1, x_2, x_3 . [0.082, 20.086, 1.6487]
- (d) Gegeben ist die Beobachtung $x_1 = 1, x_2 = -1, x_3 = 6$. Wie lauten die Odds für das Ereignis $y = 1$? [1.64] Wie lauten die Odds, wenn x_2 um eine Einheit erhöht wird? [33.12] Bestimmen sie diese neuen Odds sowohl auf direktem Weg als auch unter Verwendung der Ergebnisse von (c).
- (e) Bestimmen sie den Wert der LogLikelihoodfunktion für die Stichprobe.
[-2.395]
- (f) Testen sie mit Hilfe der Devianz die Güte des logistischen Regressionsmodells ($\alpha = 0.1$). (Verwenden Sie zur Bestimmung der Freiheitsgrade ausnahmsweise $n = 5$.)
8. Gegeben ist die lineare Funktion $z = 1 + 2x_1 - 0.5x_2 + 2.5x_3$ und folgende Daten einer Stichprobe mit $n = 5$:

i	x_1	x_2	x_3	y
1	1	2	1	1
2	-1	-1	0	0
3	0	2	1	1
4	2	4	-1	1
5	1.5	3	-0.5	1

- (a) Testen sie die Güte des mit z gebildeten logistischen Regressionsmodells über die Devianz ($\alpha = 0.05$). Geben sie genau die Nullhypothese und das Ergebnis des Tests an.
- (b) Bestimmen sie die hit-ratio und testen sie die Güte des Modells über das Press'sche Q (gleiche Details wie bei (a)).
- (c) Es soll der Einfluss der Variablen x_3 auf das Modell untersucht werden. Dazu wurde mit den Variablen x_1 und x_2 ein reduziertes Modell mit $z_R = 2 + 2.5x_1 - x_2$ berechnet. Testen sie den Einfluss von x_3 auf das Modell ($\alpha = 0.05$). Geben sie genau die Nullhypothese und das Ergebnis des Tests an.
9. Bestimmen sie für die folgenden beiden Personen A_1 und A_2 mit 11 binären Merkmalen die Werte von vier verschiedenen Ähnlichkeitsmaßen.

A_1	1	0	0	0	1	1	1	0	1	1	0
A_2	0	1	1	0	0	0	1	1	1	0	0

10. Bestimmen sie für die folgenden eindimensionalen Daten eine Clusterung unter Verwendung der L1-Norm. Verwenden sie dazu drei hierarchisch-agglomeratives Verfahren bis zum Ende, und zwar (i) single linkage, (ii) complete linkage, (iii) Centroid Clustering.

i	1	2	3	4	5	6	7	8
x_i	0,5	2,5	5,5	6,5	7,2	11	11,1	18

11. Gegeben sind 9 Messwerte, die durch ein Clusterungsverfahren bereits in 4 Cluster eingeteilt wurden:

$$C_1 = \{1, 3\} \quad C_2 = \{5, 6, 7\} \quad C_3 = \{10, 10\} \quad C_4 = \{14, 17\}$$

Führen sie das hierarchisch-agglomerative Clusterungsverfahren dreimal mit unterschiedlichen Cluster-Distanzmaßen wie in Bsp. 10 ausgehend von der gegebenen Clusterung jeweils bis zum Ende (nur mehr ein Cluster) durch.

12. Gegeben sind 6 Messwerte im \mathbb{R}^2 mit den Koordinaten (x_i, y_i) . Führen sie das hierarchisch-agglomerative Clusterungsverfahren mit single linkage einmal mit der L1-Norm und einmal mit der ∞ -Norm bis zum Ende (nur mehr ein Cluster) durch.

i	1	2	3	4	5	6
x_i	2	2	-3	4	2.5	1.5
y_i	3	-2	4	2	-2	-3

13. Gegeben sind 8 Messwerte im \mathbb{R}^2 mit den Koordinaten (x_i, y_i) .

i	1	2	3	4	5	6	7	8
x_i	-1	2	3	-2	3	5	8	11
y_i	-3	1	-4	3	-2	-3	7	12

- (a) Betrachten sie die Cluster $C_1 = \{2, 4, 6\}$ und $C_2 = \{1, 8\}$. Bestimmen sie die Cluster-Distanz mit average linkage und mit dem Centroid Verfahren unter Verwendung der L1-Norm.
- (b) Bestimmen sie den Abstand des Clusters $C_3 = \{7\}$ von C_1 mit average linkage unter Verwendung der ∞ -Norm.
- (c) Fusionieren sie C_2 und C_3 und bestimmen sie den Abstand des neuen Clusters von C_1 mit single und mit complete linkage bei Verwendung der L2-Norm.

14. Gegeben sind 5 Objekte mit den folgenden 8 binären Merkmalen:

O_1	1	0	0	1	1	1	1	1
O_2	1	1	1	0	0	0	1	0
O_3	0	0	1	1	1	0	1	1
O_4	1	1	0	1	0	0	0	1
O_5	0	0	1	0	1	0	0	0

- (a) Bestimmen sie eine Ähnlichkeitsmatrix unter Verwendung des SMC-Koeffizienten und wandeln sie diese in eine Distanzmatrix um.
- (b) Führen sie das hierarchisch-agglomerative Clusterungsverfahren mit complete linkage bis zum Ende durch.
- (c) Was ist die Distanz der Cluster $\{O_1, O_2\}$ und $\{O_3, O_5\}$ mit average linkage?