BR1

Zur Erläuterung d. Methode und zu ihrer theoretischen Vertiefung, insbes. ihrer inneren Zusammenhänge im Detail, hat sich die Durchrechnung und Interpretation eines Beispiels meist als sehr hilfreich herausgestellt.

Für die $\boxed{\text{einfach lineare}}^{LR}$ sowie die $\boxed{\text{multiple lineare Regression}}^{MLR}$ soll ein Zahlenbeispiel analysiert werden, das in WONNACOTT/WONNACOTT aufgeführt ist.

(A) Beispiel für $\boxed{\text{LR (einfach lineare Regression)}}$:

a) Das Beispiel soll 2fach präsentiert werden, zunächst Schritt für Schritt durchgerechnet (so wie man die Rechnungen auch mittels Tabellenkalkulation oder per Hand durchführen könnte), zum anderen via Ausdruck eines Computerprogramms für Regressionsstatistik (wie es üblicherweise als Ausdruck d. Rechenergebnisse dargestellt wird)

b) Schritt für Schritt-Berechnung:

(1) Die Datentabelle samt Berechnungen:

Es handelt sich um ein wenig umfangreiches Datenmaterial mit $n = 7$. Es stammt aus der landwirtschaftlichen Praxis. Der Ertrag von Mais soll in einem Experiment in Abhängigkeit von d. Düngermenge untersucht werden.
Es wird zunächst der einfachste Fall eines linearen

Zusammenhänge beider Größen angenommen.

Es wird das sog. lineare Regressionsmodell-I angenommen, d.h. man geht davon aus, daß die Kriteriumsvariable Y (=Ertrag von Mais) durch die Prädiktorvariable X (=Dünger) insofern linear bedingt wird, als die Vorgaben d. X-Variable fixiert sind. Fixierung bedeutet hier, keine variable Vorgabe d. X-Werte (also z.B. 100 g Dünger, 200 g Dünger etc. und nicht 105,2 g oder 204,7 g). Als variierend wird ausschließlich der Ertrag von Mais, also Y angenommen.

Man kann auch dargestellt formulieren: die X-Variable geht fixiert, gewissermaßen mit Konstanten an bestimmten Punkten der X-Achse in die Berechnung ein.

Abweichungsquadrate ③, ④

Abweichungsprodukte ⑤

① a Y [mm] (Halmlänge)	① b X [g] (Düngermenge)	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})^2$	$(X_i-\bar{X})(Y_i-\bar{Y})$
40	100	90000	400	6000
50	200	40000	100	2000
50	300	10000	100	1000
70	400	0	100	0
65	500	10000	25	500
65	600	40000	25	1000
80	700	90000	400	6000
② Mittelwert 60	400			
Summe		⑥ 280000	⑦ 1150	⑧ 16500

Varianz für X: ⑨ $S_x^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} = 46666{,}6'$

und die Standardabweichung ⑩ $S_x = 216{,}025$

BR2

Varianz für Y: (11) $S_Y^2 = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{n-1} = 191{,}6$

und Standardabweichung f. Y (12) $S_Y = 13{,}844$

Kovarianz (XY): (13) $S_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1} = 2750$

Korrelation (XY): (14) $r_{XY} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2 \cdot \sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \cong 0{,}91951$

Bestimmtheitsmaß (XY) (15) $r_{XY}^2 \cong 0{,}845497$

Regressionskoeffizient (16) $b_1 = \frac{S_{XY}}{S_X^2} = 0{,}05893$

(2) Das Regressionsmodell:

geometrisch: Geradengleichung

(17)

MODELL
$$\begin{cases} \hat{Y}_i = b_0 + b_1 \cdot X_i \\ Y_i = \hat{Y}_i + \epsilon_i \end{cases}$$
(18)

b_0 ... Ordinatenabstand (intercept)

b_1 ... Steigung, Regressionskoeffizient (slope)

$\hat{Y}_i$ ... Schätzwert für d. Modell

ϵ_i ... Residualabweichung

Man bedenke, daß die Regressionsgerade stets durch das Centroid $(\bar{X}, \bar{Y})$ verläuft. Somit ist nach Kenntnis von b_1 b_0 leicht auszudrücken mit:

(19) $\bar{Y} = b_0 + b_1 \cdot \bar{X}$ und (20) $b_0 = \bar{Y} - b_1 \cdot \bar{X}$

Somit ist (21) $b_0 = 60 - 0{,}05893 \cdot 400 = 36{,}428$

3) Konkretes Modell:

22 $$\boxed{\hat{Y}_i = 36{,}428 + 0{,}0\,5893 \cdot X_i}$$

c) Ausdruck eines Computerprogramms zur Regressionsberechnung:
(siehe Anhang)

- Das hierfür verwendete Programm wurde auf d. Grundlage d. Matrizenkalküls programmiert.

Hervorhebenswert sind die 'fitted values' und 'residual values'.

Die Residualwerte sind die Grundlage für eine optische Sichtung bezüglich der Tauglichkeit des Modells.

d) $\boxed{\text{RESIDUALPLOT}}$ – Analyse d. Plots

Im Anhang finden Sie den Residualplot.
Es sind zueinander (Abszisse) Y-Werte gegen (Ordinate) Residualwerte aufgetragen.

Das Bild spricht dafür, daß die Modellannahme passt.
Die Residualwerte (das Scatter) erscheinen als zufällig gestreut, ohne einen sichtbaren Trend im Bild.

```
**********************ANFANG***************************
       Multipel-Lineare Regression
(Inverse Matrix via Neumann- und Schulz-Iterationsverfahren)


Programm: REGRMUQB.BAS, version 17.01.2019

Datum: 04-02-2020
Uhrzeit: 20:12:37

ASCII-Daten-File: ZEA2.TXT
Anzahl der OTUs: 7
Anzahl der Merkmale: 2

Die abhaengige Variable (usrpuengl. Nummer): 1

************************************
Die Rohdaten (OTUs zeilenweise und Variablen spaltenweise):
  Halmlänge        Dünger
    40.000000   100.000000
    50.000000   200.000000
    50.000000   300.000000
    70.000000   400.000000
    65.000000   500.000000
    65.000000   600.000000
    80.000000   700.000000

************************************

Die invertierte Matrix in single-precision:
 .7142857 -1.428571E-03
-1.428571E-03  3.571429E-06

**************************************************

****************************************

fitted values:

    42.321430
    48.214287
    54.107143
    60.000000
    65.892860
    71.785713
    77.678574

residual values:

    -2.321429
     1.785714
    -4.107143
    10.000000
    -0.892857
    -6.785714
     2.321429

******************************
Standardschaetzfehler:  5.961184 mit  5 Freiheitsgraden

(partieller) Regressionskoeffizient, SE, t-value, two-sided probability:

b0 const :    36.428570     5.038120     7.230588     0.000791
b  2  :     0.058929     0.011266     5.230852     0.003381
```

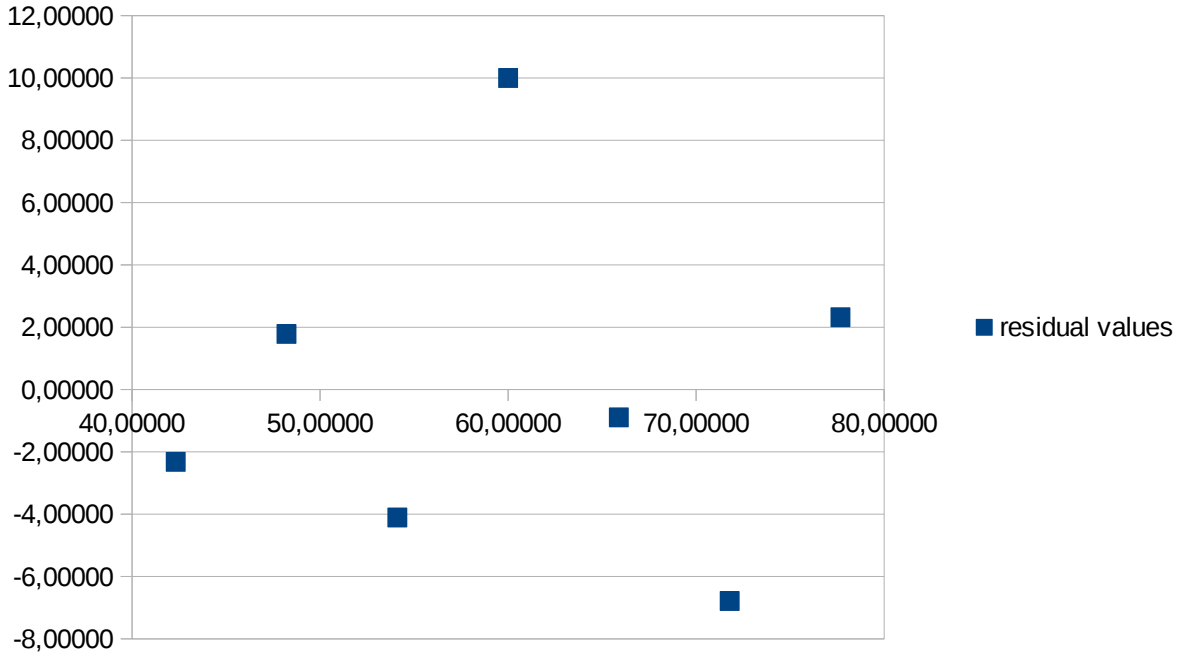
OTU... operational taxonomic unit. (hier ist jede Einheit ein Pflanzenindividuum)

↗ hier als Regressionskoeffizient mit Var-2 als Prädiktor

b_1

Multiples Bestimmheitsmass: .8454969

**********ENDE**********

Tabelle1

OTU	fitted values	residual values
1	42,32143	-2,32143
2	48,21429	1,78571
3	54,10714	-4,10714
4	60,00000	10,00000
5	65,89286	-0,89286
6	71,78571	-6,78571
7	77,67857	2,32143