

# Handreichung zur Aufgabe „Regressionsgerade Schritt für Schritt“

Titel der Aufgabe: Regressionsgerade Schritt für Schritt

Screenshot der anfänglichen Aufgabe:

Ein Eisstand möchte den Zusammenhang zwischen der Nachfrage nach Eis und der Außentemperatur um 12:00 Uhr mittags untersuchen. Dazu werden an  $n = 5$  aufeinanderfolgenden Tagen im Sommer die Außentemperatur ( $x$  in Grad Celsius) sowie die verkaufte Eismenge ( $y$  in kg) notiert. Die Daten sind in der folgenden Tabelle zusammengefasst:

$i$	$x_i$	$y_i$
1	30	140
2	24	110
3	28	127
4	29	134
5	25	98

Das Ziel ist, den Korrelationskoeffizienten  $r_{x,y}$  und eine Regressionsgerade zu den Daten zu bestimmen.

(a) Bitte berechnen Sie dazu als erstes das arithmetische Mittel von  $x$  und  $y$ .

Antwort:

$\bar{x} =$

$\bar{y} =$

Autoren: Jonas Lache und Herold Dehling, Ruhr-Universität Bochum

Lizenz: CC BY-SA 4.0

Zielgruppe: Studierende der Mathematik und in Serviceveranstaltungen

Thema: Statistik

Tags: Stochastik, Wahrscheinlichkeitstheorie, Statistik, Regression, gemeinsame Verteilungen, Korrelation

Randomisierung: ja

Aufgabentyp: mehrteilige Aufgabe

Beschreibung: In der Aufgabenstellung wird die Situation vorgestellt, dass ein Eisstand den Zusammenhang zwischen der Nachfrage nach Eis und der Außentemperatur untersuchen möchte. Dazu ist eine Tabelle mit der Außentemperatur ( $x_i$  in °C) und der verkauften Eismenge ( $y_i$  in kg) an fünf Tagen gegeben. Das Ziel der Aufgabe ist, dass die Studierenden die Regressionsgerade und den Korrelationskoeffizienten zu den Daten bestimmen. Die Aufgabe ist in mehrere Teilaufgaben unterteilt, die die einzelnen Schritte dahin repräsentieren. In der ersten Teilaufgabe soll das arithmetische Mittel  $\bar{x}$  und  $\bar{y}$  beider Variablen bestimmt werden. Im Anschluss füllen die Studierenden sukzessive eine Tabelle aus, wobei die ursprüngliche Tabelle mit den Datenwerten in jeder Teilaufgabe um eine bis zwei Spalten erweitert wird. Es werden erst Spalten für  $x_i - \bar{x}$  und  $y_i - \bar{y}$ , dann für  $(x_i - \bar{x})^2$  und  $(y_i - \bar{y})^2$  und schließlich eine für  $(x_i - \bar{x})(y_i - \bar{y})$  ausgefüllt. In den übrigen Teilaufgaben geht es nun darum, die Tabelle für die Berechnung der Regressionsgeraden und des Korrelationskoeffizienten zu verwenden. Dazu wird in der fünften Teilaufgabe nach den Spaltensummen für  $(x_i - \bar{x})^2$ ,  $(y_i - \bar{y})^2$  und  $(x_i - \bar{x})(y_i - \bar{y})$  und in der sechsten Teilaufgabe nach dem Korrelationskoeffizienten  $r_{x,y}$  gefragt. In der siebten und achten Teilaufgabe sollen noch die Parameter der Regressionsgeraden (erst  $\beta$ , dann  $\alpha$ ) angegeben werden. Am Ende werden die Regressionsgerade und die Datenpunkte im abschließenden Feedback in einer Grafik veranschaulicht.

Didaktische Überlegungen: Diese Aufgabe ist wie eine Anleitung zur Bestimmung einer Regressionsgeraden zu einem Datensatz konzipiert. Die Studierenden lernen so die einzelnen Schritte kennen und führen sie direkt selbst aus. Dabei helfen ihnen Tipps im Feedback der Aufgabenteile. Die Grafik im Feedback der letzten Teilaufgabe veranschaulicht das Konzept der Regressionsgeraden visuell.

Enthaltene Fremdmaterialien: Diese Aufgabe bindet das Skript `stackselbstlern.js` von Michael Kallweit für die Aufgabennavigation ein..

Daten oder Links (evtl. aktualisieren): keine

**Lizenz:** „Handreichung zur Aufgabe ‚Regressionsgerade Schritt für Schritt‘“ wurde entwickelt von [Jonas Lache](#) an der Ruhr-Universität Bochum. Dieses Werk ist lizenziert unter der Lizenz „Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 International“: <http://creativecommons.org/licenses/by-sa/4.0/>.

## Screenshots aus der Aufgabe

a) Teilaufgabe – Arithmetisches Mittel bestimmen:

Ein Eisstand möchte den Zusammenhang zwischen der Nachfrage nach Eis und der Außentemperatur um 12:00 Uhr mittags untersuchen. Dazu werden an  $n = 5$  aufeinanderfolgenden Tagen im Sommer die Außentemperatur ( $x$  in Grad Celsius) sowie die verkaufte Eismenge ( $y$  in kg) notiert.

Die Daten sind in der folgenden Tabelle zusammengefasst:

$i$	$x_i$	$y_i$
1	30	140
2	24	110
3	28	127
4	29	134
5	25	98

Das Ziel ist, den Korrelationskoeffizienten  $r_{x,y}$  und eine Regressionsgerade zu den Daten zu bestimmen.

**(a)** Bitte berechnen Sie dazu als erstes das arithmetische Mittel von  $x$  und  $y$ .

Antwort:

$$\bar{x} = \text{[ ]}$$

$$\bar{y} = \text{[ ]}$$

b) Teilaufgabe – Spalten  $x_i - \bar{x}$  und  $y_i - \bar{y}$  ausfüllen:

$i$	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$
1	30	140	[ ]	[ ]
2	24	110	[ ]	[ ]
3	28	127	[ ]	[ ]
4	29	134	[ ]	[ ]
5	25	98	[ ]	[ ]

**(b)** Füllen Sie nun die Tabelle aus und klicken Sie dann auf "Prüfen".

c) Teilaufgabe – Spalten  $(x_i - \bar{x})^2$  und  $(y_i - \bar{y})^2$  ausfüllen:

$i$	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	30	140	2.8	18.2	[ ]	[ ]
2	24	110	-3.2	-11.8	[ ]	[ ]
3	28	127	0.8	5.2	[ ]	[ ]
4	29	134	1.8	12.2	[ ]	[ ]
5	25	98	-2.2	-23.8	[ ]	[ ]

**(c)** Füllen Sie die Tabelle weiter aus und klicken Sie dann auf "Prüfen".

d) Teilaufgabe – Spalte  $(x_i - \bar{x})(y_i - \bar{y})$  ausfüllen:

$i$	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	30	140	2.8	18.2	7.84	331.24	<input type="text"/>
2	24	110	-3.2	-11.8	10.24	139.24	<input type="text"/>
3	28	127	0.8	5.2	0.64	27.04	<input type="text"/>
4	29	134	1.8	12.2	3.24	148.84	<input type="text"/>
5	25	98	-2.2	-23.8	4.84	566.44	<input type="text"/>

**(d)** Füllen Sie die Tabelle weiter aus und klicken Sie dann auf "Prüfen".

e) Teilaufgabe – Summen der Spalten angeben:

**(e)** Für die Berechnung der Regressionsgeraden benötigen Sie die folgenden Summen. Bitte geben Sie diese in die Eingabefelder ein und klicken Sie dann auf "Prüfen".

Antwort:

$\sum_{i=1}^5 (x_i - \bar{x})^2 =$

$\sum_{i=1}^5 (y_i - \bar{y})^2 =$

$\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) =$

f) Teilaufgabe – Korrelationskoeffizient angeben:

**(f)** Bestimmen Sie nun den Korrelationskoeffizienten von  $x$  und  $y$  und klicken Sie dann auf "Prüfen".

Bitte geben Sie den exakten Wert oder eine auf bis zu zwei Nachkommastellen gerundete Dezimalzahl ein.

Zur Erinnerung: Es gilt  $\bar{x} = 27.2$ ,  $\bar{y} = 121.8$ ,  $\sum_{i=1}^5 (x_i - \bar{x})^2 = 26.8$ ,  $\sum_{i=1}^5 (y_i - \bar{y})^2 = 1212.8$  und  $\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) = 167.2$ .

Antwort:  $r_{x,y} =$

g) Teilaufgabe – Parameter  $\beta$  der Regressionsgeraden angeben:

**(g)** Berechnen Sie den Parameter  $\beta$  der Regressionsgeraden  $y = \alpha + \beta x$  und klicken Sie nach der Eingabe auf "Prüfen".

Bitte geben Sie eine auf genau zwei Nachkommastellen gerundete Dezimalzahl ein.

$\beta =$

h) Teilaufgabe – Parameter  $\alpha$  der Regressionsgeraden angeben:

**(h)** Berechnen Sie jetzt den Parameter  $\alpha$  der Regressionsgeraden  $y = \alpha + \beta x$  und klicken nach der Eingabe auf "Prüfen".

Wenn Sie vorherige Ergebnisse verwenden, dann nutzen Sie bitte die auf zwei Nachkommastellen gerundeten Werte. Geben Sie dann das Ergebnis für  $\alpha$  als eine auf genau zwei Nachkommastellen gerundete Dezimalzahl ein.

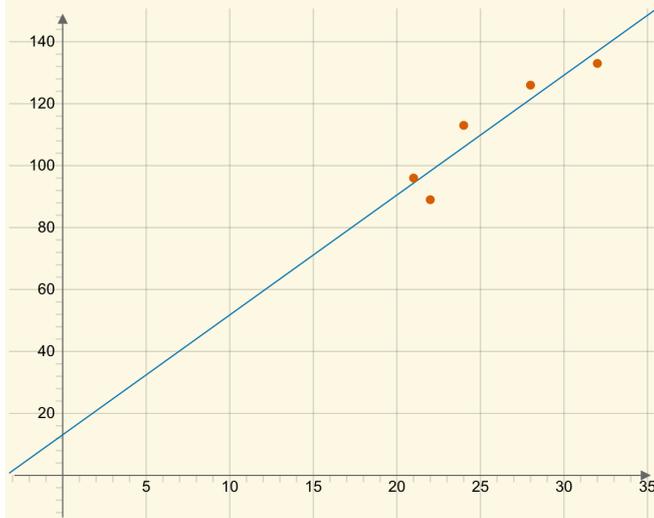
$\alpha =$

Grafik im Feedback nach der letzten Teilaufgabe:

Die Regressionsgerade lautet also wie folgt:

$$y = 3.87 \cdot x + 13.1$$

In der Abbildung sehen Sie einen Scatterplot der Daten und die von Ihnen berechnete Regressionsgerade:



Herzlichen Glückwunsch, Sie sind am Ende der Aufgabe angelangt! Nun sollten Sie aus einem Datensatz mit zwei Variablen eine Regressionsgerade und einen Korrelationskoeffizienten berechnen können.