Wirtschaftsstatistik

Übungsblatt Modul 6

Korrelation und Regression

# Aufgabe 1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Filiale** | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **H** | **I** | **J** |
| **Verkaufsfläche (1.000 qm)** | 7 | 5 | 6 | 3 | 8 | 2 | 4 | 6 | 4 | 7 |
| **Filialumsatz (Mio €)** | 35 | 22 | 41 | 15 | 38 | 12 | 34 | 28 | 25 | 52 |

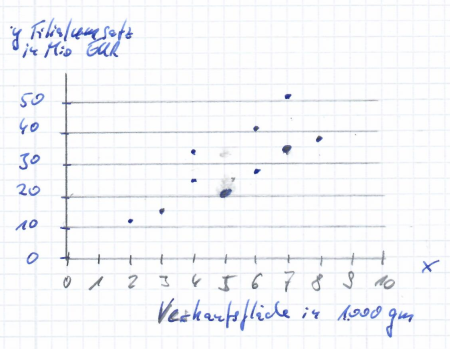
a)   
Visualisieren Sie für das obige Datenmaterial den Zusammenhang zwischen den beiden Merkmalen Verkaufsfläche und Filialumsatz in einem geeigneten Diagramm. Wie wird diese grafische Darstellung genannt?

b)

Welche Erkenntnisse liefert Ihnen eine erste qualitative Zusammenhangs- bzw. Abhängigkeitsanalyse auf der Basis des in a) erstellten Diagramms?

Lösung zu a)

Die grafische Lösung wird Streudiagramm genannt



Lösung zu b)

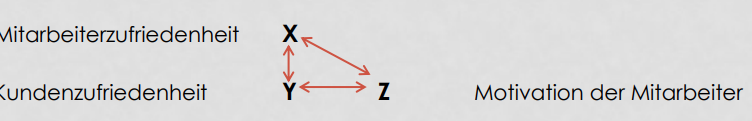
**Zusammenhangsanalyse (Interpendenzanalyse)**

Prüft,

* ob Zusammenhang zwischen Merkmalen (Variablen) besteht und, falls dem so ist,
* wie stark der Zusammenhang zwischen den Merkmalen (Variablen) ausgeprägt ist.
* Wie lässt sich die Stärke (der Grad, die Intensität) des Zusammenhangs messen
* Wie lässt sich der Zusammenhang (grafisch) darstellen?
* Lassen sich die Beobachtungswerte einer Variablen X durch mehrere andere Variablen (Y (Y1, Y2, …) näherungsweise bestimmen?
* Prüft die Wechselwirkung der Variablen untereinander
* Das Zusammenhangsmaß (Assoziationsmaß) bestimmt die Stärke und ggf. die Richtung eines Zusammenhangs

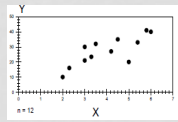
**KEIN gerichteter Zusammenhang**

* **zwischen den unabhängigen Variablen X, Y untereinander und zu einer unabhängiger Variablen Z besteht kein gerichteter Zusammenhang**
* der **Zusammenhang zwischen allen Variablen** (den unabhängigen Variablen X, Y untereinander und zwischen den unabhängigen Variablen X oder Y und der abhängigen Variablen Z) **ist bidirektional**  
  z. B. Zusammenhang zwischen den unabhängigen Variablen „Mitarbeiterzufriedenheit X“ und „Kundenzufriedenheit Y“ und der abhängigen Variable „Motivation der Mitarbeiter Z“

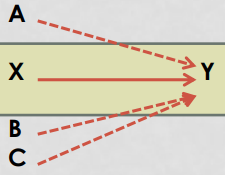


**Zusammenhangsanalyse zwischen metrischen Merkmalen X und Y heißt Korrelationsanalyse**

* Die **Korrelationsanalyse** prüft, ob ein **linearer Zusammenhang** zwischen den Variablen **besteht** und **wie stark der lineare Zusammenhang** ist.
* Grafische Darstellung **als Streudiagramm** (qualitative Zusammenhangsanalyse)



* **einfache Regressionsanalyse: Analyse des Zusammenhangs zwischen zwei metrischen Größen** 
  + **Eine Einflussgröße X (Regressor, unabhängiges Merkmal / unabhängige Variable)**
  + **Eine Zielgröße Y (Regressand, abhängiges Merkmal / abhängige Variable)**
  + **anhand von 2 Parametern wird eine Gerade durch die Punkwolke gezeichnet** und dadurch der lineare Zusammenhang gut veranschaulicht
* Es gibt **meist mehr als einen Einflussfaktor X (Regressor**, unabhängiges Merkmal / unabhängige Variable)
* **Probleme sind selten monokausal**
* **multiple Regressionsanalyse: Verallgemeinerung der einfachen linearen Regression** 
  + **mit k Regressoren (Einflussgrößen bzw. unabhängigen Merkmalen / Variablen)**, **welche den Regressand (abhängiges Merkmal / Variable) erklären sollen**
  + **mehrere Einflussgrößen** **X1, X2, …Xk** (Regressoren, unabhängige Merkmale / Variablen) und **eine Zielgröße Y** (Regressand / abhängiges Merkmal / Variable)  
    **X1, X2, …Xk → Y**



* **Zusammenhangsmaß** heißt **Korrelationskoeffizient r**
* **Korrelationskoeffizienten r** ist die **Quantifizierung der Korrelation**
* **Korrelationskoeffizienten r** informiert über,
  + **Stärke (Grad) des linearen Zusammenhangs** (der Korrelation )
  + **Richtung des linearen Zusammenhangs** (der Korrelation )
* **Korrelationskoeffizient r kann Werte zwischen -1 und +1 haben**  
  **-1 ≤ rxy ≤ +1**
* **Korrelationskoeffizient ist dimensionsloses Maß für den linearen Zusammenhang**
* **Korrelationskoeffizient r > 0: Positiver Zusammenhang**  
  **hohe Werte in einem Merkmal (Variablen) X gemeinsam mit hohen Werten in einem anderen Merkmal (Variablen) Y**
* **Korrelationskoeffizient r < 0: Negativer Zusammenhang**  
  **hohe Werte in einem Merkmal (Variablen) X mit niedrigen Werten in einem anderen Merkmal (Variablen) Y**
* **Korrelationskoeffizient r = +1: extrem starker positiver Zusammenhang** (in der Praxis sehr selten)  
  Punktewolke liegt auf einer Geraden mit positiver Steigung (von links unten nach rechts oben)
* **Korrelationskoeffizient r = -1: extrem starker negativer Zusammenhang** (in der Praxis sehr selten)  
  Punktewolke liegt auf einer Geraden mit negativer Steigung (von links oben nach rechts unten)
* **Korrelationskoeffizient r = 0: kein Zusammenhang**

**Korrelation**

* Ist **linearer (gerichteter) Zusammenhang**
* Ist **zahlenmäßiger statistischer Zusammenhang zwischen zwei metrischen Merkmalen** X und Y
* **Aufgrund des zahlenmäßiger Zusammenhangs nur bei** **metrischen Merkmalen**
* **Korrelationsanalyse: Zusammenhangsanalyse zwischen metrischen Merkmalen X und Y**
* **Positive Korrelation: beide Merkmale X und Y entwickeln sich gleichförmig:**bei höheren Werten von X hat auch Y höhere Werte  
  wenn Werte von X steigen, steigen auch die Werte von Y
* **Negative Korrelation: beide Merkmale X und Y entwickeln sich gegenläufig**bei höheren Werten von X hat Y niedrigere Wertewenn Werte von X steigen, fallen die Werte von Y
* **Kausaler Zusammenhang**: **zwischen zwei Merkmalen X und Y** existiert ein **Ursache-Wirkungs-Beziehung**  
  die **Veränderung eines abhängigen Merkmals Y ist eindeutig auf Veränderung des unabhängigen Merkmals X zurückzuführen**  
  **Korrelation** sagt **nichts über den kausalen Zusammenhang** und **nichts über die Kausalitätsrichtung** aus.  
  **Korrelation ≠ Kausalität**
* **Korrelation ist für Kausalität notwendig**, **aber nicht ausreichende Voraussetzung**nicht jede Korrelation ist ein kausaler Zusammenhang, aber jeder kausale Zusammenhang basiert auf einer Korrelation

**Korrelation**:

statistischer **Zusammenhang**

**zwischen metrischen Merkmalen**

**Scheinkorrelation**:

**Kein kausaler Zusammenhang**

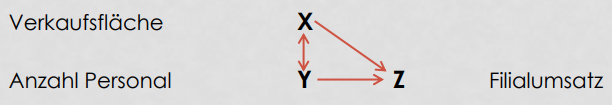
**Kausalität:**

**kausaler Zusammenhang**

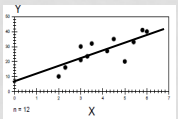
(Ursache-Wirkungs-Beziehung)

**Abhängigkeitsanalyse (Dependenzanalyse)**

* Wird **als Regressionsanalyse bezeichnet**
* **Unterscheidung zwischen abhängigen und unabhängigen Merkmalen**
* Unabhängige Merkmale beeinflussen abhängige Merkmale 🡆 **abhängige Variablen werden durch unabhängige Variablen beeinflusst**
* Es geht um **gerichteten Zusammenhang**
  + es besteht **vorab eine sachlogisch begründete Vorstellung bzw. Vermutung über den Zusammenhang zwischen den Merkmalen**
  + es i**st bekannt bzw. wird vermutet welche der Merkmale auf andere Merkmale einwirken**
  + **zwischen unabhängigen Variablen X, Y besteht ein gerichteter Zusammenhang zu einer unabhängiger Variablen Z** (Zusammenhang zwischen der unabhängigen Variablen X und der unabhängigen Variablen Z bzw. der unabhängigen Variablen Y und der unabhängigen Variablen Z ist unidirektional)  
    **zwischen den unabhängigen Variablen X, Y besteht kein gerichteter Zusammenhang** 🡆 der **Zusammenhang zwischen den unabhängigen Variablen X, Y ist bidirektional**  
    z. B. Zusammenhang zwischen den unabhängigen Variablen „Verkaufsfläche X“ und „Anzahl Personal Y“ und der abhängigen Variable „Umsatz Z“

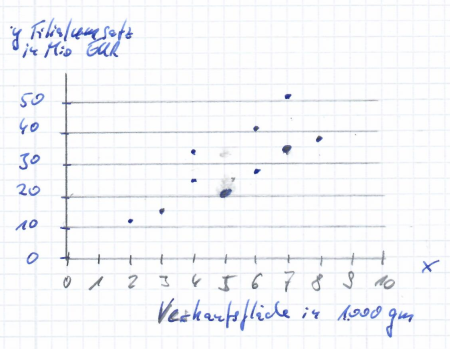


* **Abhängigkeitsanalyse muss sinnvoll** sein
* **Abhängigkeitsmaß ŷ**
* **Regressionsfunktion ŷ = a + b\*x**
* Grafische Darstellung **als Streudiagramm mit Regressionsgeraden** (qualitative Abhängigkeitsanalyse)



Lösung zu b)

Welche Erkenntnisse liefert Ihnen eine erste qualitative Zusammenhangs- bzw. Abhängigkeitsanalyse auf der Basis des in a) erstellten Diagramms?



**qualitative Zusammenhangsanalyse:**

Es besteht eine positive Korrelation (ein positiver Zusammenhang) zwischen der Verkaufsfläche und dem Filialumsatz. Eine größere Verkaufsfläche (höhere Werte zum Merkmal X „Verkaufsfläche in 1000 qm“) hat auch einen höheren Filialumsatz zur Folge (höhere Werte zum Merkmal Y „Filialumsatz in Mio EUR“).

Der lineare Zusammenhang ist „relativ“ stark.

Der Korrelationskoeffizient r muss > 0 sein.

**Qualitative Abhängigkeitsanalyse**

Es existiert ein kausaler Zusammenhang zwischen der Verkaufsfläche und dem Filialumsatz

Der Zusammenhang ist gerichtet und annähernd linear.

Die Größe der Verkaufsfläche hat Einfluss auf den Filialumsatz.

Das unabhängige Merkmal (die unabhängige Variable) beeinflusst das abhängige Merkmal (die abhängige Variable) Filialumsatz

# Aufgabe 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Monate** | **Januar** | **Februar** | **März** | **April** |
| **Produzierte Menge (in Mio. Stück)** | 5 | 3 | 2 | 4 |
| **Produktionskosten (Mio. €)** | 8 | 4 | 6 | 5 |

a)  
Berechnen Sie eine lineare Regressionsfunktion Ŷ = Ŷ(x) = a + b\*x, die den Zusammenhang zwischen produzierter Menge und Kosten möglichst gut charakterisiert   
(x produzierte Menge in Mio. Stück und Y Kosten in Mio. €)

Tipp: transponiere und erweitere obere Tabelle, um Zwischenberechnungen einzufügen

(s. z. B. Folien 20 und 39 Modul 6)

**Abhängigkeitsanalyse (lineare Regressionsfunktion Ŷ = a + b\*x): berechnet das Abhängigkeitsmaß**

**Voraussetzungen**:

* X und Y sind quantitative Merkmale
* gerichteter Zusammenhang zwischen X und Y (X -> Y)

**vorbereitende Arbeiten**

* prüfen, ob Abhängigkeitsanalyse sinnvoll ist
* Erhebung der Daten für X und Y

1. Visualisierung im Streudiagramm (qualitative Abhängigkeitsanalyse)
2. Auswahl des Funktionstyps (in diesem Seminar nur lineare Funktionen)
3. Berechnung der Regressionsfunktion (nach der Methode der kleinsten Quadrate)

**Regressionsfunktion Ŷ = a + b\*x nach der Methode der kleinsten Quadrate**

* **a und b** sind die **Regressionskoeffizienten (Kurvenparameter)**
* **Regressionskoeffizienten a und b werden so bestimmt, dass die Summe der quadratischen Abweichungen (Differenzen) von der Kurve der beobachteten Punkte minimal ist**

**∑(yi – Ŷ)²**

**∑(= (yi – (a + b\*x) )²**

**Vorgehen**

1. **REGRESSIONSRECHNUNG**
   1. Berechnungstabelle mit Hilfsgrößen **Xi \* Yi** und **Xi²** und Yi² erstellen

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Lfd. Nr | Monate | **Xi**  Produzierte Menge  (in Mio. Stück) | **Yi**  Produktionskosten  (in Mio. €) | **Xi \* Yi** | **Xi²** | Yi² |
| 1 | Januar | 5 | 8 | 40 | 25 | 64 |
| 2 | Februar | 3 | 4 | 12 | 9 | 16 |
| 3 | März | 2 | 6 | 12 | 4 | 36 |
| **4** | April | 4 | 5 | 20 | 16 | 25 |
| **Summe** |  | **14** | **23** | **84** | **54** | **141** |

* 1. **Berechnung der Regressionskoeffizienten a und b** mit Hilfe der Formeln

|  |  |
| --- | --- |
| **a =** ∑**Xi²** \* ∑**Yi** - ∑**Xi** \* ∑**Xi \* Yi**  **n** \* ∑**Xi²** - (∑**Xi**)² | **b = n** \* ∑**Xi \* Yi** - ∑**Xi** \* ∑**Yi**  **n** \* ∑**Xi²** - (∑**Xi**)² |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **a =** | ∑**Xi²** \* ∑**Yi** - ∑**Xi** \* ∑**Xi \* Yi**  **n** \* ∑**Xi²** - (∑**Xi**)² |  |  |  |
| **a =** | **54 \* 23 – 14 \* 84**  **4 \* 54 – 14²** | **1242 - 1176**  **216 - 196** | **66**  **20** | **3,3** |
|  | Hinweis:  Der Nenner muss nur 1 x berechnet werden, da er bei der Berechnung der Variablen b erneut verdendet wird. | | | |

**a = 3,3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **b =** | **n** \* ∑**Xi \* Yi** - ∑**Xi** \* ∑**Yi**  **n** \* ∑**Xi²** - (∑**Xi**)² |  |  |  |
| **b =** | **4 \* 84 – 14 \* 23**  **4 \* 54 – 14²** | **336 - 322**  **20** | **14**  **20** | **0,7** |
|  | Hinweis:  Der Nenner muss nur 1 x berechnet werden, da er bei der Berechnung der Variablen b erneut verdendet wird. | | | |

**b = 0,7**

**Regressionsfunktion nach der Methode der kleinsten Quadrate**

**Ŷ = Ŷ(x) = a + b\*x**

**Ŷ = Ŷ(x) = 3,3 + 0,7\*x (Mio €)**

b)

Zeichnen Sie das Streudiagramm und die Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!

* 1. **Zeichnen des Streudiagramms anhand der Wertepaare (Korodinaten) für Xi und Yi aus der in Schritt 1.1 erstellten Tabelle**
  2. **Ermitteln der Abweichungsquadrate anhand der Formel ∑(yi – Ŷ)²**Für **Ŷ** die Werte der Regressionskoeffizienten a und b und für Xi den gegebenen Wert aus der Tabelle einsetzen  
     **ui = ∑(yi – Ŷ)² = ∑(yi – (a + b \* Xi)²**

u ist das Residuum, welches für die Gerade der Regressionsfunktion benötigt wird.

a, b sind die Regressionskoeffizienten. Die Werte zu a, b wurden zuvor in der Nebenrechnung ermittelt.

Regressionskoeffizient **a = 3,3**

Regressionskoeffizient **b = 0,7**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lfd. Nr | **Xi**  Menge  (in Mio. Stück) | **Yi**  Kosten  (in Mio. €) | **a** | **b** | b \* Xi | Ŷ =  a + (b \* Xi) | Yi – Ŷ |
| 1 | **5** | 8 | **3,3** | **0,7** | **3,5** | 6,8 | 8 – 6,8 = **1,2** |
| 2 | **3** | 4 | **3,3** | **0,7** | **2,1** | 5,4 | 4 – 5,4 = **-1,4** |
| 3 | **2** | 6 | **3,3** | **0,7** | **1,4** | 4,7 | 6 – 4,7 = **1,3** |
| **4** | **4** | 5 | **3,3** | **0,7** | **2,8** | 6,1 | 5 – 6,1 = **-1,1** |
| **Summe** | **14** | **23** | **-** | **-** | - | - |  |

* 1. **Gerade für die Regressionsfunktion in das Streudiagramm einzeichnen**

Im Streudiagramm

* von den eingezeichneten Punkten jeweils um den in der letzten Spalte ermittelten Wert
  + **wenn der Wert > 0 nach unten** Eselsbrücke: es soll minimiert werden 🡆nach unten
  + **wenn der Wert < 0 nach oben**
* und die Punkte für die Gerade einzeichnen

Bsp: für das 1. Koordinatenpaar (X1 = 5, Y1 = 8)

u1 = (Y1 – Ŷ)² = (Y1 – (a + b \* X1))² | **a = 3,3, b = 0,7**

u1 = (8 – (3,3 + 0,7 \* 5))² = (8 – (3,3 + 3,5))² = (8 – 6,8)² = 1,2²

Zum Zeichnen der Geraden wird nur der Basiswert **1,2** benötigt (die **Potenzierung ist nicht erforderlich!**)

**u1 = Y1 – Ŷ = 1,2**

**🡆 gehe im Streudiagramm von Y1 = 8 🡆 1,2 nach unten (+1,2)**

Bsp: für das 2. Koordinatenpaar (X2 = 3, Y2 = 4)

u2 = (Y2 – Ŷ)² = (Y2 – (a + b \* X2))² | **a = 3,3, b = 0,7**

u2 = (4 – (3,3 + 0,7 \* 3))² = (4 – (3,3 + 2,1))² = (4 – 5,4)² = -1,4²

Zum Zeichnen der Geraden wird nur der Basiswert **-1,4** benötigt (die **Potenzierung ist nicht erforderlich!**)

**u2 = Y2 – Ŷ = -1,4**

**🡆 gehe im Streudiagramm von Y2 = 4 🡆 1,4 nach oben (-1,4)**

Bsp: für das 3. Koordinatenpaar (X3 = 2, Y3 = 6)

u3 = (Y3 – Ŷ)² = (Y3 – (a + b \* X3))² | **a = 3,3, b = 0,7**

u3 = (6 – (3,3 + 0,7 \* 2))² = (6 – (3,3 + 1,4))² = (6 – 4,7)² = 1,3²

Zum Zeichnen der Geraden wird nur der Basiswert **1,3** benötigt (die **Potenzierung ist nicht erforderlich!**)

**u3 = Y3 – Ŷ = 1,3**

**🡆 gehe im Streudiagramm von Y3 = 2 🡆 1,3 nach unten (+1,3)**

Bsp: für das 4. Koordinatenpaar (X4 = 4, Y4 = 5)

u4 = (Y4 – Ŷ)² = (Y4 – (a + b \* X4))² | **a = 3,3, b = 0,7**

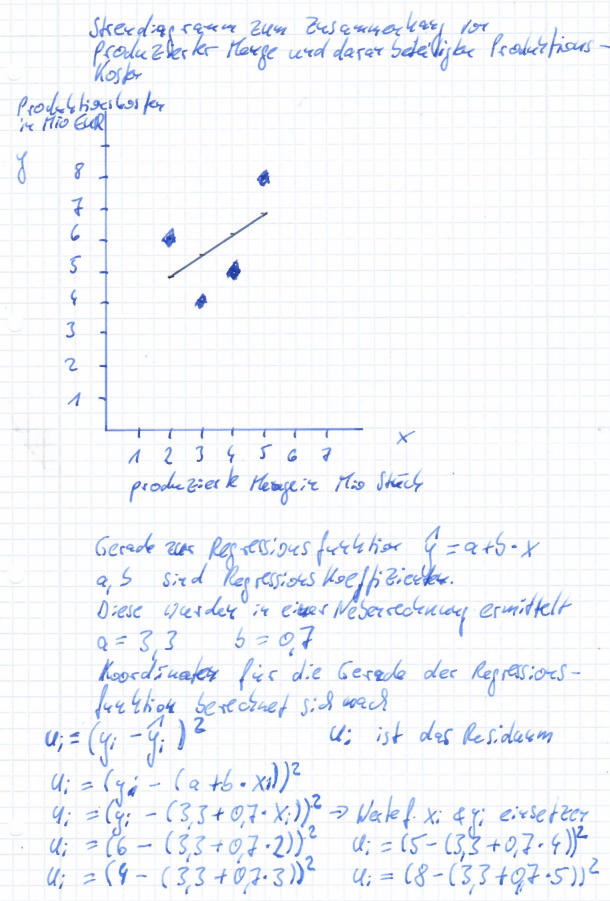
u4 = (5 – (3,3 + 0,7 \* 4))² = (5 – (3,3 + 2,8))² = (5 – 6,1)² = -1,1²

Zum Zeichnen der Geraden wird nur der Basiswert -**1,1** benötigt (die **Potenzierung ist nicht erforderlich!**)

**u4 = Y4 – Ŷ = -1,4**

**🡆 gehe im Streudiagramm von Y4 = 4 🡆 1,4 nach oben (-1,1)**

Streudiagramm zum Zusammenhang der produzierten Menge und der daran beteiligten Produktionskosten



c)

Interpretieren Sie die beiden Regressionskoeffizienten der berechneten Regressionsfunktion betriebswirtschaftlich.

Wie nennt man eine solche Funktion in der Betriebswirtschaftslehre?

Regressionskoeffizient **a = 3,3**

Regressionskoeffizient **b = 0,7**

**Produzierte Menge in Mio Stück: 2**

**Produktionskosten in Mio EUR**

3,3 + 0,7 \* **2** = 3,3 + 1,4 = **4,7** 4,7 : 2 = 2,35 Mio EUR / 1 Mio Stück produzierte Menge

**Produzierte Menge in Mio Stück: 3**

**Produktionskosten in Mio EUR**

3,3 + 0,7 \* **3** = 3,3 + 2,1 = **5,4** 5,4 : 3 = 1,8 Mio EUR / 1 Mio Stück produzierte Menge

**Produzierte Menge in Mio Stück: 4**

**Produktionskosten in Mio EUR**

3,3 + 0,7 \* **4** = 3,3 + 2,8 = **6,1** 6,1 : 4 = 1,525 Mio EUR / 1 Mio Stück produzierte Menge

**Produzierte Menge in Mio Stück: 5**

**Produktionskosten in Mio EUR**

3,3 + 0,7 \* **5** = 3,3 + 3,5 = 6,8 6,8 : 5 = 1,36 Mio EUR / 1 Mio Stück produzierte Menge

Interpretation der Regressionskoeffizienten:

**a = 3,3 Mio sind Fixkosten,**

**b = 0,7 sind variable (Stück)-Kosten**

**Die Stückkosten fallen bei zunehmender produzierte Menge.**

**In der Kostenrechnung/BWL wird diese Funktion lineare Kostenfunktion genannt.**

d)

Erstellen Sie auf der Basis der in a) ermittelten Regressionsfunktion eine

Kostenprognose für den Monat Mai, in dem eine Produktionsmenge von 6 Mio. Stück geplant ist

Kostenprognose für den Monat Mai mit 6 Mio Stück produzierte Menge

Ŷ(6) = 3,3 + 0,7 \* 6 = 3,3 + 4,2 = 7,5 Mio EUR

e)

Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß zur obigen

Regressionsrechnung

**Zusammenhangsanalyse zwischen zwei metrischen Merkmalen X und Y**

**Korrelationskoeffizient (Zusammenhangsmaß)** r

-1 ≤ r ≤ +1

* **ist dimensionsloses Maß für den linearen Zusammenhang**
* **Korrelationskoeffizient r > 0: Positiver Zusammenhang**  
  **hohe Werte in einem Merkmal (Variablen) X gemeinsam mit hohen Werten in einem anderen Merkmal (Variablen) Y**
* **Korrelationskoeffizient r < 0: Negativer Zusammenhang**  
  **hohe Werte in einem Merkmal (Variablen) X mit niedrigen Werten in einem anderen Merkmal (Variablen) Y**
* **Korrelationskoeffizient r = +1: extrem starker positiver Zusammenhang**   
  Punktewolke liegt auf einer Geraden mit positiver Steigung (von links unten nach rechts oben)
* **Korrelationskoeffizient r = -1: extrem starker negativer Zusammenhang**   
  Punktewolke liegt auf einer Geraden mit negativer Steigung (von links oben nach rechts unten)
* **Korrelationskoeffizient r = 0: kein Zusammenhang**

**Korrelationsrechnung ist die „Standardisierung“ der Kovarianz**

* **Zwei metrische Merkmale X, Y müssen** 
  + **mindestens intervallskaliert** sein (intervallskalierte, verhältnisskalierte oder absolut-skalierte Merkmale)
  + **jeweils eine positive Standardabweichung** haben und
  + **eine Kovarianz COV(X, Y)** haben

Für die Korrelationsrechnung (nach Pearson) werden für jedes Merkmal X, Y

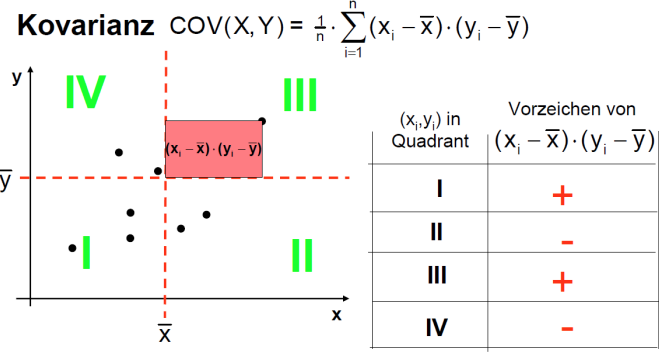
* die Kovarianz COV(X,Y) und
* die Standardabweichung Sx, Sy   
  (berechnet sich aus der Wurzel zur Summe der Quadrate aus der Differenz aus Merkmalswert und Mittelwert geteilt durch die Anzahl der Merkmale)  
  Sx = √∑((xi – x̅ )²) : n

benötigt

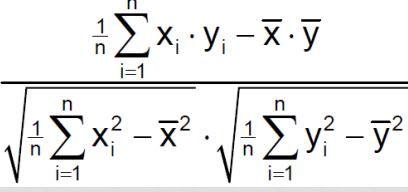
Korrelationskoeffizient 

**Kovarianz COV(X,Y)**

* informiert über gemeinsame Variabilität der Merkmale X und Y.
* Ist der **Zusammenhang (Korrelation) positiv, ist auch die Kovarianz positiv**
* ist der **Zusammenhang (Korrelation) negativ, ist auch die Kovarianz negativ**.
* **kein (linearer) Zusammenhang zwischen X und Y 🡆 Kovarianz liegt in der Nähe von 0**.
* **Bei positiver Kovarianz (= positivem Zusammenhang (Korrelation)) 🡆 steigende Gerade im Streudiagramm** (von links unten nach rechts oben)
* **Bei negativer Kovarianz (= positivem Zusammenhang (Korrelation)) 🡆 fallende Gerade im Streudiagramm** (von links oben nach rechts unten)



**Kovarianz COV(X,Y) = (xi – x̅) \* (yi – y̅) : n**



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lfd. Nr | Monate | **Xi**  Produzierte Menge  (in Mio. Stück) | **Yi**  Produktionskosten  (in Mio. €) | **xi \* yi** | **x²** | **y²** |  |
| 1 | Januar | 5 | 8 | 40 | 25 | 64 |  |
| 2 | Februar | 3 | 4 | 12 | 9 | 16 |  |
| 3 | März | 2 | 6 | 12 | 4 | 36 |  |
| **4** | April | 4 | 5 | 20 | 16 | 25 |  |
| **Summe** |  | **14** | **23** | **84** | **54** | **141** |  |

**x̅ = 14 : 4 = 3,50 x̅² = 12,25**

**y̅ = 23 : 4 = 5,75 y̅² = 33,06**

**x̅ \* y̅ = 3,5 \* 5,75 = 20,12**

rx,y =

¼ \* 84 – 3,5 \* 5,75 21 – 20,125 0,875

0,875 0,875 = 0,52914

1,118 \* 1,479 1,6536

Korrelationskoeffizient rx,y = 0,529

**Bestimmtheitsmaß 𝑹²**

Bei einfacher linearen Regression (nur eine unabhängige Variable) entspricht das Bestimmtheitsmaß dem Quadrat des Korrelationskoeffizienten nach Pearson rXY

𝑹² = (𝑟𝑋𝑌)²

𝑹² = 0,529² = 0,28

f)

Interpretieren Sie das Bestimmtheitsmaß

g)

Berechnen Sie die Varianz der Produktionsmengen, der Produktionskosten und der Regressionswerte (Summen aus der Berechnungstabelle verwenden).

Zeigen Sie an diesem Beispiel, dass das Bestimmtheitsmaß angibt, welcher Anteil der Varianz der Produktionskosten erklärt wird durch die Regressionsfunktion bzw. die Varianz der Produktionsmengen Tipp: Varianzzerlegung s. Modul 6

# Aufgabe 3

Die Marktforschungsabteilung eines Unternehmens will für ein neues Produkt die Abhängigkeit der Absatzmenge x vom Preis Y empirisch untersuchen. Dazu wird in 4

vergleichbar großen Testmärkten unter sonst annähernd gleichen Rahmenbedingungen 8 Wochen lang jeweils ein unterschiedlicher Preis gefordert. Die Ergebnisse dieses Tests stehen in der folgenden Tabelle:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Testmärkte** | **A** | **B** | **C** | **D** |
| **x: Preis pro Stück (€)** | 1,90 | 2,10 | 1,50 | 2,50 |
| **Y: Absatzmenge in 8 Wochen (Stück)** | 5.000 | 4.000 | 8.000 | 3.000 |

1. Berechnen Sie eine lineare Regressionsfunktion Ŷ = Ŷ(x) = a + b\*x, die die

Abhängigkeit zwischen Preis und Absatzmenge möglichst gut charakterisiert. (x Preis in € und Y Absatzmenge in Stück)

1. Zeichnen Sie das Streudiagramm und die Regressionsfunktion in das Koordinatensystem ein. Beschriftung der Achsen nicht vergessen!
2. Interpretieren Sie die beiden Regressionskoeffizienten der in a) berechneten

Regressionsfunktion betriebswirtschaftlich

1. Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß zur obigen

Regressionsrechnung

1. Interpretieren Sie das Bestimmtheitsmaß für diesen Fall
2. Erstellen Sie auf der Basis der in a) ermittelten Regressionsfunktion eine

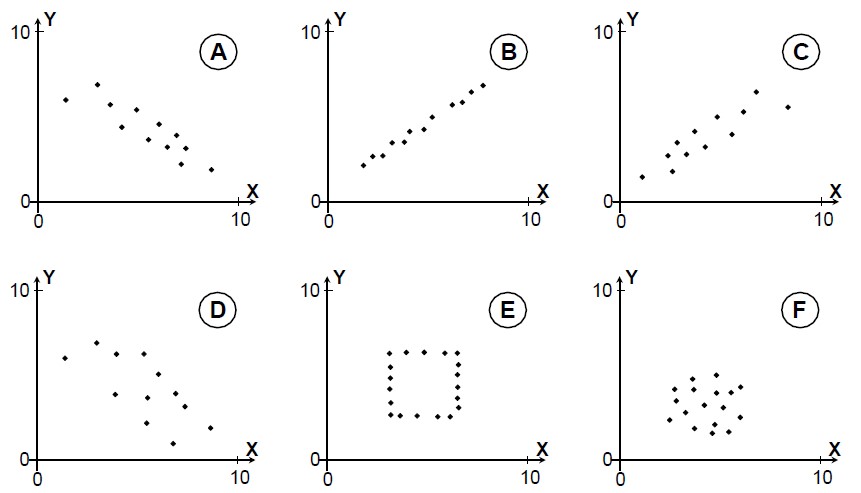
Absatzmengen-Prognose für einen Testmarkt M, der den Testmärkten A, B, C und D entspricht. Welche Absatzmenge ist dort in 8 Wochen zu erwarten bei einem Preis von € 1,70

1. Berechnen Sie für die Daten die Varianz der empirisch ermittelten Absatzmengen und die Varianz der dazugehörigen Regressionswerte (= entsprechende Absatzmengen auf der Regressionsfunktion). Was ergibt der Quotient „Varianz der Regressionswerte zur Varianz der Absatzmengen“?

# Aufgabe 4

In den folgenden Abbildungen sind 6 Streudiagramme dargestellt, die den Zusammenhang zwischen den zwei Merkmalen X und Y visualisieren. Bringen Sie die Korrelationskoeffizienten rA, rB, rC, rD, rE, rF nach ihren Werten in eine aufsteigende Rangfolge.

(rA: Korrelationskoeffizient für das Streudiagramm A, usw.)



**Interpretation der Korrelationskoeffizienten r**

* Korrelationskoeffizienten r < 0 🡆 negativer Zusammenhang (negative Korrelation)
  + Punkte sind im Punktdiagramm fallend (von links oben nach rechts unten)
* Korrelationskoeffizienten r = -1 🡆 starker negativer Zusammenhang
  + Punkte liegen dicht und linear beieinander, fast wie in einer Geraden   
    (ähnlich Diagramm A)
* Korrelationskoeffizienten r > 0 🡆 positiver Zusammenhang (positive Korrelation)
  + Punkte sind im Punktdiagramm steigend (von links unten nach rechts oben)
* Korrelationskoeffizienten r = +1 🡆 starker positiver Zusammenhang
  + Punkte liegen dicht und linear beieinander, fast wie in einer Geraden   
    (ähnlich Diagramm B)

**Je dichter die Punkte zusammenliegen, desto stärker ist der (negative oder positive) Zusammenhang**

**Lösung:**

-1 < rA < rD < rF  ≈ 0 ≈ rE < rC < rB < +1

# Aufgabe 5

Es soll den Zusammenhang zwischen Preis P und Absatzmenge X empirisch untersucht werden. Dazu werden in 20 ausgewählten Testmärkten unterschiedliche Preise eingesetzt und Absatzmengen zu diesen Preisen dokumentiert. Die Regressionsrechnung liefert die folgende lineare Regressionsfunktion:

Ŷ = Ŷ(p) = a + b\*p = 2.000 – 500\*p (p in €, Y in Tsd. Stück)

Der Korrelationskoeffizient beträgt r = - 0,75.

Nehmen Sie kurz Stellung zu den beiden folgenden Aussagen a) und b):

1. Aus der Regressionsfunktion kann man erkennen, dass der Zusammenhang zwischen Preis und Absatzmenge sehr stark ist.
2. Die Absatzmengenunterschiede in den Testmärkten lassen sich zu 75% durch die unterschiedliche Preispolitik erklären, d.h. durch die Preisunterschiede in den Testmärkten.

# Aufgabe 6

Für ein neues Produkt wird untersucht, wie die Käufer auf unterschiedliche Preise reagieren. In 4 vergleichbar großen Testgeschäften in verschiedenen Gegenden wird das neue Produkt einen Monat lang zu jeweils unterschiedlichen Preisen angeboten. In der folgenden Tabelle sind für die Testphase die jeweils geforderten Preise (**p**) und die Reaktion der Käufer ausgedrückt in Absatzmengen des Produktes (**Y**) zusammengestellt.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Testgeschäft Nr.** | **1** | **2** | **3** | **4** |
| **x: Preis pro Stück (€)** | 3,00 | 1,50 | 1,00 | 2,50 |
| **Y: Absatzmenge (in Tausend Stück)** | 2 | 6 | 14 | 10 |

Nach der Methode der kleinsten Quadrate wurde mit Hilfe der Berechnungstabelle und den entsprechenden Formeln für die beiden Regressionskoeffizienten die folgende lineare PreisAbsatz-Funktion ermittelt:

Ŷ = Ŷ(x) = a + b\*x = 16 – 4\*x

1. Erstellen Sie eine Berechnungstabelle. Berechnen Sie die Regressionswerte und Residualwerte und tragen Sie diese in die Tabelle ein. Zeichnen Sie in das Koordinatensystem ein:
   1. die Beobachtungswertepaare durch ∎
   2. die berechnete Regressionsfunktion,
   3. die Regressionswerte Ŷ**i** durch •,
   4. die Residualwerte (wenn farbiger Stift vorhanden, in Farbe)
2. Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß für die

Daten in a) und interpretieren Sie die beiden Ergebnisse bezogen auf unser Beispiel

1. Prognostizieren Sie auf der Basis der Regressionsfunktion die monatliche Absatzmenge für einen den 4 Testgeschäften vergleichbaren Einzelhandelsbetrieb, wenn dort ein Preis von € 2,00 pro Stück verlangt wird
2. Was halten Sie von der Güte der obigen Prognose (unter c))?
3. Berechnen Sie mit Hilfe der Berechnungstabelle aus a) die Varianz 𝑠𝑌2 der Absatzmengen Yi und die Varianz 𝑠𝑌̂2 der entsprechenden Regressionswerte Ŷi. Zeigen Sie, dass hier die folgende Beziehung gilt:

Bestimmtheitsmaß = 𝑅2 = 𝑠𝑠𝑌𝑌̂22

1. Berechnen Sie außerdem mit Hilfe der Berechnungstabelle aus a) die Varianz 𝑠𝑥2 der Preise xi. Zeigen Sie, dass die Varianz der Regressionswerte bestimmt wird durch die Varianz des unabhängigen Merkmals, d.h., dass hier folgende Beziehung gilt:

𝑠𝑌̂2 =𝑏²∗𝑠𝑥2 (b  Regressionskoeffizient)

# Aufgabe 7

Für 4 Monate liegen die Daten über den Hypothekenzinssatz **x** sowie über den saisonbereinigten monatlichen Auftragseingang **Y** im Bauhauptgewerbe vor, der auf den privaten Wohnungsbau entfällt.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Testgeschäft Nr.** | **1** | **2** | **3** | **4** |
| **x: Hypothekenzinssatz in %** | 6 | 7 | 5,5 | 7,5 |
| **Y: Auftragseingang (in Mrd. €)** | 30 | 40 | 50 | 20 |

1. Berechnen Sie nach der Methode der kleinsten Quadrate die lineare

Regressionsfunktion, die die „mittlere“ Abhängigkeit des Auftragseingangs im

Bauhauptgewerbe vom Hypothekenzinssatz möglichst gut beschreibt

1. Interpretieren Sie für diesen Fall die beiden Regressionskoeffizienten
2. Zeichnen Sie in das Koordinatensystem die Wertepaare des Streudiagramms und die in a) berechnete Regressionsfunktion. Markieren Sie die Regressionswerte Ŷi
3. Berechnen Sie den Korrelationskoeffizienten und das Bestimmtheitsmaß zur obigen Regressionsrechnung
4. Interpretieren Sie das Bestimmtheitsmaß für diesen Fall
5. Prognostizieren Sie auf der Basis der Regressionsanalyse den (saisonbereinigten) monatlichen Auftragseingang im Bauhauptgewerbe, der zu erwarten ist bei einem

Hypothekenzinssatz von 6,6% und von 7,2%

1. Welche Kennzahl kann man zur Beurteilung der Güte der obigen Prognose nutzen? Was halten Sie von der in f) erstellten Prognose?