2.Übung Qualitätsmanagement

Henner Graubitz

31. Oktober 2006

Korrelation

Was ist eine Korrelation?

- Korrelation ist eine eindeutige Beziehung zwischen zwei Merkmalen

Korrelation

Was ist eine Korrelation?

- Korrelation ist eine eindeutige Beziehung zwischen zwei Merkmalen
- Sie gibt einen Trend über den Zusammenhang zweier Daten an

Korrelation

Was ist eine Korrelation?

- Korrelation ist eine eindeutige Beziehung zwischen zwei Merkmalen
- Sie gibt einen Trend über den Zusammenhang zweier Daten an
- Beispiel: Körpergröße vs. Körpergewicht

Korrelation

Was ist eine Korrelation?

- Korrelation ist eine eindeutige Beziehung zwischen zwei Merkmalen
- Sie gibt einen Trend über den Zusammenhang zweier Daten an
- Beispiel: Körpergröße vs. Körpergewicht
- Bei der Untersuchung von Ursache und Wirkung sind zwei Hilfsmittel von besonderer Bedeutung:

Korrelation

Was ist eine Korrelation?

- Korrelation ist eine eindeutige Beziehung zwischen zwei Merkmalen
- Sie gibt einen Trend über den Zusammenhang zweier Daten an
- Beispiel: Körpergröße vs. Körpergewicht
- Bei der Untersuchung von Ursache und Wirkung sind zwei Hilfsmittel von besonderer Bedeutung:
- **Korrelationsdiagramm** und die **Korrelationszahl**

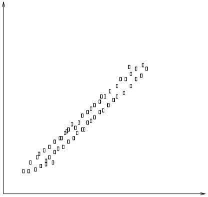
Korrelation

Was ist eine Korrelation?

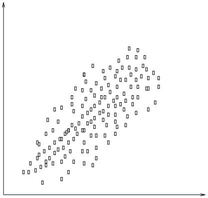
- Korrelation ist eine eindeutige Beziehung zwischen zwei Merkmalen
- Sie gibt einen Trend über den Zusammenhang zweier Daten an
- Beispiel: Körpergröße vs. Körpergewicht
- Bei der Untersuchung von Ursache und Wirkung sind zwei Hilfsmittel von besonderer Bedeutung:
- **Korrelationsdiagramm** und die **Korrelationszahl**

Was ist ein Korrelationsdiagramm

Starke Korrelation mit r=1

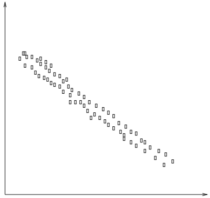


Schwache Korrelation mit r = 0.5

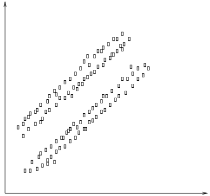


Was ist ein Korrelationsdiagramm

Schwache negative Korrelation mit r = -0.5



Keine Korrelation mit r = 0



Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen

Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen
- deshalb ist es wichtig, die Korrelation in einer Ziffer auszudrücken, der Korrelationszahl r

Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen
- deshalb ist es wichtig, die Korrelation in einer Ziffer auszudrücken, der Korrelationszahl r
- Für die Korrelationszahl r gilt: $-1 \leq r \leq +1$

Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen
- deshalb ist es wichtig, die Korrelation in einer Ziffer auszudrücken, der Korrelationszahl r
- Für die Korrelationszahl r gilt: $-1 \leq r \leq +1$
- wobei: r=1: perfekt positiver linearer Zusammenhang

Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen
- deshalb ist es wichtig, die Korrelation in einer Ziffer auszudrücken, der Korrelationszahl r
- Für die Korrelationszahl r gilt: $-1 \leq r \leq +1$
- wobei: r=1: perfekt positiver linearer Zusammenhang
- wobei: r=-1: perfekt negativer linearer Zusammenhang

Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen
- deshalb ist es wichtig, die Korrelation in einer Ziffer auszudrücken, der Korrelationszahl r
- Für die Korrelationszahl r gilt: $-1 \leq r \leq +1$
- wobei: r=1: perfekt positiver linearer Zusammenhang
- wobei: r=-1: perfekt negativer linearer Zusammenhang
- wobei: r=0: kein linearer Zusammenhang

Was ist die Korrelationszahl

- je stärker die Korrelation zweier Merkmal ist, desto eher kann auch ein ursächlicher Zusammenhang vorliegen
- deshalb ist es wichtig, die Korrelation in einer Ziffer auszudrücken, der Korrelationszahl r
- Für die Korrelationszahl r gilt: $-1 \leq r \leq +1$
- wobei: r=1: perfekt positiver linearer Zusammenhang
- wobei: r=-1: perfekt negativer linearer Zusammenhang
- wobei: r=0: kein linearer Zusammenhang

Beispiel:

Man läßt zwei Personen (X und Y) vier Konzepte beurteilen (nach den Noten von 1 - 4). Die Urteile X und Y sind ordinal skaliert. Dabei kommen beide Personen zu folgenden Ergebnis:

Konzept	a	b	c	d
Urteil X	3	4	2	1
Urteil Y	3	1	4	2

Frage: Sind die zwei Personen übereinstimmend in ihren Meinungen oder nicht?

Lösung:

Sortierung nach Urteil von X:

Konzept	d	c	a	b
Urteil X	1	2	3	4
Urteil Y	2	4	3	1

perfekt positiv hieße:

Konzept	d	c	a	b
Urteil X	1	2	3	4
Urteil Y	1	2	3	4

Lösung:

perfekt negativ hieße:

Konzept	d	c	a	b
Urteil X	1	2	3	4
Urteil Y	4	3	2	1

zur Auswertung benutzt Kendall's τ

Lösung nach Kendall's τ:

Möglichkeiten:

$$\binom{n}{2} = \frac{n(n-1)}{2} \quad (1)$$

.

Kendall's τ =

$$\frac{P-Q}{\frac{n(n-1)}{2}} \quad (2)$$

wobei:

P = Anzahl der Übereinstimmungen (als + gekennzeichnet)

Q = Anzahl der Nichtübereinstimmungen (als - gekennzeichnet)

Lösung:

Konzept	d	c	a	b
Urteil X	1	2	3	4
Urteil Y	2	4	3	1
		+	+	-
			-	-
				-

Somit ist

- P = 2
- Q = 4
- $\frac{n(n-1)}{2} = 6$
- $\tau = \frac{2-4}{6} = -\frac{1}{3}$

Aufgabe:

Zwei Weinkenner testen 8 Weine und geben folgende Benotung ab:

Weinsorte	C	G	D	B	F	E	H	A
Urteil von X	1	2	2	2	3	3	4	4
Urteil von Y	1	5	4	3	6	2	7	6

Lösung:

Weinsorte	C	G	D	B	F	E	H	A
Urteil von X	1	2	2	2	3	3	4	4
Urteil von Y	1	5	4	3	6	2	7	6
		+	+	+	+	+	+	+
			0	0	+	-	+	+
				0	+	-	+	+
					+	-	+	+
						0	+	0
							+	+
								0

Merke: ist bei einem Paarvergleich in mindestens einer Paarreihe kein Vergleich möglich wird der Vergleich ignoriert durch 0.

Lösung:

τ muß korrigiert werden, da durch die neutralen Entscheidungen 0 sieben Entscheidungen weggefallen sind.
Es gilt

$$\tau^* = \frac{P - Q}{\sqrt{\binom{n}{2} - \tau_x} \cdot \sqrt{\binom{n}{2} - \tau_y}} \tag{3}$$

wobei

$$\tau_x = \frac{1}{2} \cdot \sum_{i=1}^{k} g_i \cdot (g_i - 1) \tag{4}$$

und

$$\tau_y = \frac{1}{2} \cdot \sum_{j=1}^{l} h_j \cdot (h_j - 1) \tag{5}$$

Lösung:

k = Anzahl der Bindungen in der X-Reihe
l = Anzahl der Bindungen n der Y-Reihe
(Bindung bedeutet: wie oft wurde welcher Platz vergeben
Es gilt:

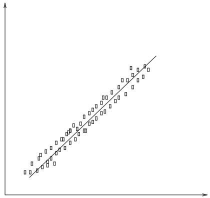
$$\tau_x = \frac{1}{2} \cdot (3 \cdot 2 + 2 \cdot 1 + 2 \cdot 1) = 5$$

$$\tau_y = \frac{1}{2} \cdot (2 \cdot 1) = 1$$

$$\tau^* = \frac{19 - 3}{\sqrt{(28-5)} \cdot \sqrt{(28-1)}} = +0,642$$

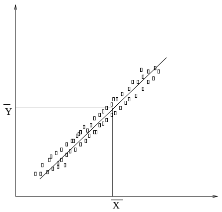
Interpretation: Die Urteile der beiden Weinkenner sind tendenziell recht gut übereinstimmend.

Korrelationskoeffizient nach Bravais:



- wir gehen davon aus, daß die Gerade durch den Schwerpunkt aller Punkte verläuft
- der Schwerpunkt wird berechnet, indem der Mittelwert der X und der Y-Werte aller Punkte bestimmt wird

Korrelationskoeffizient nach Bravais:



Wobei gilt:

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$$\overline{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i$$

Korrelationskoeffizient nach Bravais:

Für den Korrelationskoeffizient nach Bravais gilt dann:

$$r = \frac{\sum(x_i \cdot y_i) - n \cdot \overline{xy}}{\sqrt{[\sum x_i^2 - n \cdot \overline{x^2}] \cdot [\sum y_i^2 - n \cdot \overline{y^2}]}} \tag{6}$$

Beispiel:

In einer Obstplantage wird die Korrelation zwischen der Baumhöhe und dem Ernteertrag untersucht. Dabei werden die Bäume einer bestimmten Höhe x_i und der dazugehörige Ernteertrag (y_i) in kg gemessen:

x_i	y_i
1	1
2	1,5
3	4
4	4,5

Lösung:

x_i	y_i	$x_i \cdot y_i$	x^2	y^2
1	1	1	1	1
2	1.5	3	4	2.25
3	4	12	9	16
4	4.5	18	16	20.25
10	11	34	30	39.5

$\overline{x} = 2,5$ sowie $\overline{y} = 2,75$

Lösung:

Einsetzen in:

$$r = \frac{\sum(x_i \cdot y_i) - n \cdot \overline{xy}}{\sqrt{[\sum x_i^2 - n \cdot \overline{x^2}] \cdot [\sum y_i^2 - n \cdot \overline{y^2}]}}$$

$$= \frac{34 - 4 \cdot 2,5 \cdot 2,75}{\sqrt{(30 - 4 \cdot 2,5^2) \cdot (39,5 - 4 \cdot 2,75^2)}}$$

$$= \frac{6,5}{\sqrt{5 \cdot 9,25}}$$

$$= 0,955$$

Aufgabe 1:

Ein Psychologe untersucht in einem Betrieb verschiedene Arbeitsgruppen danach, ob ein Zusammenhang zwischen der Attraktivität einer Arbeitsgruppe (Zusammenhalt und Wohlbefinden ihrer Mitglieder) und der Streuung ihrer Arbeitsleistung besteht. Die Gruppenattraktivität wurde mit Hilfe einer Ordinalskala gemessen (von $1 =$ sehr hoch bis $6 =$ sehr niedrig), die Streuung der Leistung bezieht sich auf die Zahlen der von einer Arbeitsgruppe an verschiedenen Tagen fertiggestellten Produkte.

Aufgabe 1:

Die Untersuchung brachte folgendes Ergebnis:

Arbeitsgruppe	A	B	C	D	E	F	G
Gruppenattraktivität	1	2	2	3	4	5	6
Leistungsvarianz	32	20	40	32	64	68	60

Berechnen Sie einen geeigneten Korrelationskoeffizienten.

Aufgabe 2:

Die folgende Tabelle gibt für die Monate des ersten Halbjahres 20.. jeweils die Durchschnittszahlen der täglich in Magdeburg fahrenden Pkw's und die Zahl der Verkehrsunfälle, an denen Pkw's beteiligt waren, an:

Monat	Jan	Feb	Mär	Apr	Mai	Jun
Pkw pro Tag in Mio.	1,4	1,4	1,6	1,8	1,8	1,6
Unfälle (in Tsd.)	16	16	15	14	15	14

Berechnen Sie den Korrelationskoeffizienten nach Bravais-Pearson ! Hinweis Wenn Sie die Daten linear auf einstellige Zahlen transformieren läßt es sich besser rechnen.