

# Übung Wahrscheinlichkeitsrechnung und Statistik in den Ingenieurwissenschaften

## Übung 5

# Bivariate Daten

## grafische/ tabellarische Darstellung für bivariate Daten

diskrete Merkmale (qualitativ+ quantitativ diskret)	stetige Merkmale (quantitativ stetig)
Häufigkeitstabelle	Streudiagramm

## Zusammenhangsmaße

Hat das Merkmal  $X$  einen Einfluss auf das Merkmal  $Y$ ?

nominal	ordinal	quantitativ
Kontingenzkoeff. v. Pearson	Kontingenzkoeff. v. Pearson	Kontingenzkoeff. v. Pearson (dazu evtl. Daten klassieren)
	Rangkorrelations- koeff. v. Spearman	Rangkorrelations- koeff. v. Spearman
		Korrelationskoeff. von Pearson-Bravais

## Kontingenzkoeffizient von Pearson

- wird aus der Kontingenz-(Häufigkeits-)Tabelle berechnet

## Kontingenzkoeffizient von Pearson

- wird aus der Kontingenz-(Häufigkeits-)Tabelle berechnet
- nimmt Werte zwischen 0 und 1 an

## Kontingenzkoeffizient von Pearson

- wird aus der Kontingenz-(Häufigkeits-)Tabelle berechnet
- nimmt Werte zwischen 0 und 1 an
- dabei steht 0 für die Unabhängigkeit der Merkmale

## Kontingenzkoeffizient von Pearson

- wird aus der Kontingenz-(Häufigkeits-)Tabelle berechnet
- nimmt Werte zwischen 0 und 1 an
- dabei steht 0 für die Unabhängigkeit der Merkmale
- und 1 für die höchste Abhängigkeit



## Kontingenzkoeffizient von Pearson

- wird aus der Kontingenz-(Häufigkeits-)Tabelle berechnet
- nimmt Werte zwischen 0 und 1 an
- dabei steht 0 für die Unabhängigkeit der Merkmale
- und 1 für die höchste Abhängigkeit

Beispiel: Rostschutzmittel (Aufgabe 1), X: Typ (**nominales** Merkmal), Y: Wirksamkeit (ordinales Merkmal)

Wirksamkeit von Rostschutzmitteln:

Häufigkeitstabelle:

	gering	mittel	stark	
R1	65	103	106	274
R2	74	85	47	206
	139	188	153	480

Wirksamkeit von Rostschutzmitteln:

Häufigkeitstabelle:

	gering	mittel	stark	
R1	65	103	106	274
R2	74	85	47	206
	139	188	153	480

**Frage:** Haben die Rostschutzmittel eine unterschiedliche Wirkung?  
Also: Besteht ein Zusammenhang zwischen der Wirkung und dem Mittel?

Wirksamkeit von Rostschutzmitteln:

Häufigkeitstabelle:

	gering	mittel	stark	
R1	65	103	106	274
R2	74	85	47	206
	139	188	153	480

**Frage:** Haben die Rostschutzmittel eine unterschiedliche Wirkung?  
Also: Besteht ein Zusammenhang zwischen der Wirkung und dem Mittel?

Betrachte dazu den Kontingenzkoeffizient von Pearson.

# Aufgabe 1 in R

```
> Tabelle<-matrix(c(65,103,106,74,85,47),ncol=3,byrow=T)
> chi2A<-chisq.test(Tabelle)$statistic
> chi2A
X-squared
 15.74034
#Kontingenzkoeff (J=2, K=3,N=408):
> sqrt(chi2/(chi2+408)*2/1)
[1] 0.2725663
```

## Rangkorrelationskoeffizient von Spearman

- benutzt nur die Reihenfolge der Werte, Werte selbst nicht wichtig

## Rangkorrelationskoeffizient von Spearman

- benutzt nur die Reihenfolge der Werte, Werte selbst nicht wichtig
- nimmt Werte zwischen -1 und 1 an

## Rangkorrelationskoeffizient von Spearman

- benutzt nur die Reihenfolge der Werte, Werte selbst nicht wichtig
- nimmt Werte zwischen -1 und 1 an
- 1 genau dann, wenn aus  $x_n < x_m$  immer folgt  $y_n < y_m$



## Rangkorrelationskoeffizient von Spearman

- benutzt nur die Reihenfolge der Werte, Werte selbst nicht wichtig
- nimmt Werte zwischen -1 und 1 an
- 1 genau dann, wenn aus  $x_n < x_m$  immer folgt  $y_n < y_m$
- -1 genau dann, wenn aus  $x_n < x_m$  immer folgt  $y_n > y_m$

## Rangkorrelationskoeffizient von Spearman

- benutzt nur die Reihenfolge der Werte, Werte selbst nicht wichtig
- nimmt Werte zwischen -1 und 1 an
- 1 genau dann, wenn aus  $x_n < x_m$  immer folgt  $y_n < y_m$
- -1 genau dann, wenn aus  $x_n < x_m$  immer folgt  $y_n > y_m$
- 0 bedeutet Unabhängigkeit der Daten

## Korrelationskoeffizient von Bravais-Pearson

- nimmt Werte zwischen -1 und 1 an

## Korrelationskoeffizient von Bravais-Pearson

- nimmt Werte zwischen -1 und 1 an
- 1 bedeutet, es gibt ein  $a > 0$  und ein  $b$  sodass  $y_n = a \cdot x_n + b$ , die Punkte liegen auf einer Geraden mit positiver Steigung  $a$

## Korrelationskoeffizient von Bravais-Pearson

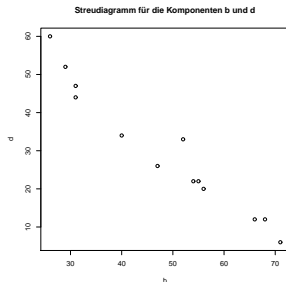
- nimmt Werte zwischen -1 und 1 an
- 1 bedeutet, es gibt ein  $a > 0$  und ein  $b$  sodass  $y_n = a \cdot x_n + b$ , die Punkte liegen auf einer Geraden mit positiver Steigung  $a$
- -1 bedeutet, es gibt ein  $a < 0$  und ein  $b$  sodass  $y_n = a \cdot x_n + b$ , die Punkte liegen auf einer Geraden mit negativer Steigung  $a$

Beispiel: Anteile der Zementkomponenten (quantitative Merkmale)  
Untersuche den Zusammenhang zwischen den Anteilen von  
Komponenten b und d.

# Aufgabe 2

Betrachte zunächst das Streudiagramm der Zementkomponenten b und d

```
> setting<-read.table("SETTING.DAT")  
> b<-setting[,2]  
> d<-setting[,4]  
> plot(b,d,main="Streudiagramm für Komponenten b und d")
```



Bestimme zunächst die Ränge:

- b: 26 29 56 31 52 55 71 31 54 47 40 66 68



Bestimme zunächst die Ränge:

- b: 26 29 56 31 52 55 71 31 54 47 40 66 68
- sortierter Datensatz:  
26 29 31 31 40 47 52 54 55 56 66 68 71

Bestimme zunächst die Ränge:

- b: 26 29 56 31 52 55 71 31 54 47 40 66 68

- sortierter Datensatz:

26 29 31 31 40 47 52 54 55 56 66 68 71

- Ränge:

26	29	56	31	52	55	71	31	54	47	40	66	68
1	2	10	3.5	7	9	13	3.5	8	6	5	11	12

Bestimme zunächst die Ränge:

- b: 26 29 56 31 52 55 71 31 54 47 40 66 68

- sortierter Datensatz:

26 29 31 31 40 47 52 54 55 56 66 68 71

- Ränge:

26	29	56	31	52	55	71	31	54	47	40	66	68
1	2	10	3.5	7	9	13	3.5	8	6	5	11	12

- der mittlere Rang ist  $\bar{R}(b) =$

$$\frac{1}{13}(1 + 2 + 10 + 3.5 + 7 + 9 + 13 + 3.5 + 8 + 6 + 5 + 11 + 12) = 7$$

Bestimme zunächst die Ränge:

- b: 26 29 56 31 52 55 71 31 54 47 40 66 68

- sortierter Datensatz:

26 29 31 31 40 47 52 54 55 56 66 68 71

- Ränge:

26	29	56	31	52	55	71	31	54	47	40	66	68
1	2	10	3.5	7	9	13	3.5	8	6	5	11	12

- der mittlere Rang ist  $\bar{R}(b) =$

$$\frac{1}{13}(1+2+10+3.5+7+9+13+3.5+8+6+5+11+12) = 7$$

- genauso be-

rechne die Ränge und den mittleren Rang für die Komponente d

60	52	20	47	33	22	6	44	22	26	34	12	12
13	12	4	11	8	5.5	1	10	5.5	7	9	2.5	2.5

Bestimme zunächst die Ränge:

- b: 26 29 56 31 52 55 71 31 54 47 40 66 68

- sortierter Datensatz:

26 29 31 31 40 47 52 54 55 56 66 68 71

- Ränge:

26	29	56	31	52	55	71	31	54	47	40	66	68
1	2	10	3.5	7	9	13	3.5	8	6	5	11	12

- der mittlere Rang ist  $\overline{R}(b) =$

$$\frac{1}{13}(1+2+10+3.5+7+9+13+3.5+8+6+5+11+12) = 7$$

- genauso be-

rechne die Ränge und den mittleren Rang für die Komponente d

60	52	20	47	33	22	6	44	22	26	34	12	12
13	12	4	11	8	5.5	1	10	5.5	7	9	2.5	2.5

- $\overline{R}(d) = 7$

Bestimme nun mögliche Zusammenhangsmaße:

- Spearmanscher Rangkorrelationskoeffizient

$$r_{bd}^{Sp} = \frac{\sum_{n=1}^N (R(b_n) - \bar{R}(b)) (R(d_n) - \bar{R}(d))}{\sqrt{\sum_{n=1}^N (R(b_n) - \bar{R}(b))^2 \sum_{n=1}^N (R(d_n) - \bar{R}(d))^2}}$$

Bestimme nun mögliche Zusammenhangsmaße:

- Spearmanscher Rangkorrelationskoeffizient

Bestimme nun mögliche Zusammenhangsmaße:

- Spearmanscher Rangkorrelationskoeffizient
- Korrelationskoeffizient von Bravais-Pearson

$$r_{bd} = \frac{s_{bd}}{s_b \cdot s_d},$$

wobei

$$s_{bd} = \frac{1}{N-1} \sum_{n=1}^N (b_n - \bar{b})(d_n - \bar{d})$$

und  $s_d, s_b$  die Standardabweichungen



Bestimme nun mögliche Zusammenhangsmaße:

- Spearmanscher Rangkorrelationskoeffizient

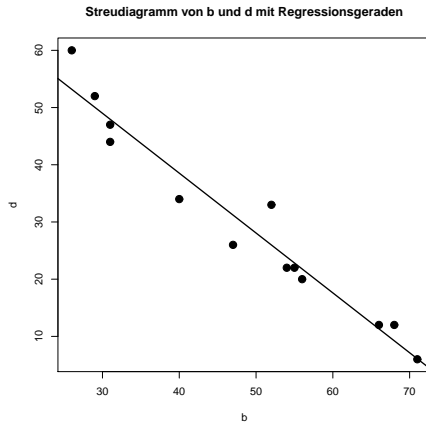
Bestimme nun mögliche Zusammenhangsmaße:

- Spearmanscher Rangkorrelationskoeffizient
- Korrelationskoeffizient von Bravais-Pearson

Bestimme nun mögliche Zusammenhangsmaße:

- Spearmanscher Rangkorrelationskoeffizient
- Korrelationskoeffizient von Bravais-Pearson
- Regressionsgerade, Steigung:  $\hat{m} = \frac{s_{bd}}{s_b^2}$ , y-Achsenabschnitt  
 $\hat{c} = \bar{d} - \hat{m}\bar{b}$

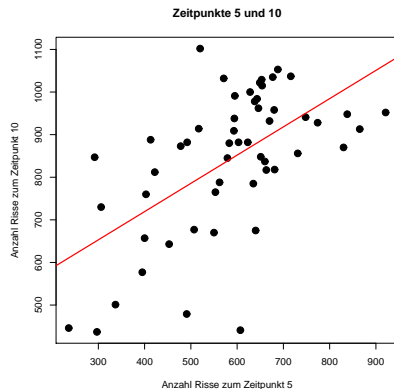
# Zeichne Regressionsgerade in Streudiagramm



Spearman'scher Rangkorrelationskoeffizient	<code>cor(rank(b),rank(d))</code>
Korrelationskoeffizient von Bravais-Pearson	<code>cor(b,d)</code>
Koeffizienten der Regressionsgeraden	<code>lsfit(b,d)\$coeff</code>

# Aufgabe 3

Rissanzen zum Zeitpunkt 5 und 10:



Korrelationskoeffizient von Bravais-Pearson: 0.582104

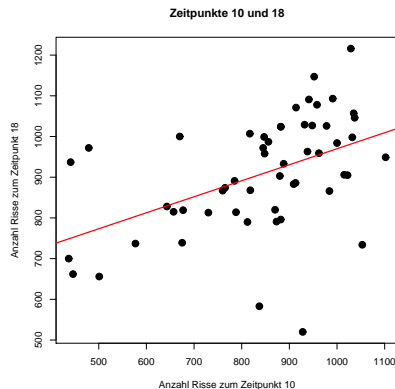
Spearman'scher Rangkorrelationskoeffizient: 0.562612

Koeffizienten der Regressionsgeraden: Achsenabschnitt: 454.172324

Steigung: 0.662917

# Aufgabe 3

Rissanzen zum Zeitpunkt 10 und 18:



Korrelationskoeffizient von Bravais-Pearson: 0.4692203

Spearman'scher Rangkorrelationskoeffizient: 0.562612

Koeffizienten der Regressionsgeraden: Achsenabschnitt:

576.7399626 Steigung: 0.3932361