

# W-Rechnung und Statistik für Ingenieure

## Übung 5

# Grafische/ tabellarische Darstellung für bivariate Daten

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

Stehen die Merkmale  $X$  und  $Y$  in einem Zusammenhang?

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

# Kontingenzkoeffizient von Pearson

- wird aus der Kontingenz-(Häufigkeits-)Tabelle berechnet
- nimmt Werte zwischen 0 und 1 an
- 0 bedeutet (empirische) Unabhängigkeit der Merkmale
- 1 bedeutet höchste Abhängigkeit

# Aufgabe 1 : 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?  
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)
> chi2<-chisq.test(Tabelle)$statistic
> chi2
X-squared
 15.74034
#Kontingenzkoeff (J=2, K=3,N=480):
> C<-sqrt((chi2*min(dim(Tabelle)))
/((chi2+sum(Tabelle))*(min(dim(Tabelle))-1)))
0.2519967
```

Es besteht also ein leichter Zusammenhang zwischen den Daten.

# 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  $y_n > y_m$
- 1 interpretierbar als gleichsinniges Verhalten der Ränge
- -1 interpretierbar als gegensinniges Verhalten
- 0 bedeutet (empirische) Unabhängigkeit der Daten

# Korrelationskoeffizient von Bravais-Pearson

- nimmt Werte zwischen -1 und 1 an
- 1 bedeutet, es gibt ein  $a > 0$  und ein  $b \in \mathbb{R}$  mit  $y_n = a \cdot x_n + b$   
d.h. die Punkte liegen auf einer Geraden mit positiver Steigung  $a$
- -1 bedeutet, es gibt ein  $a < 0$  und ein  $b \in \mathbb{R}$ , sodass  
 $y_n = a \cdot x_n + b$   
d.h. die Punkte liegen auf einer Geraden mit negativer Steigung  $a$
- 1 liefert einen gleichgerichteten Zusammenhang
- -1 liefert einen gegengerichteten Zusammenhang
- 0 bedeutet (empirische) Unabhängigkeit der Daten



# Aufgabe 2 : Zusammenhang von Zementkomponenten

Untersuche den Zusammenhang zwischen den Anteilen von Komponenten b und d der Daten.

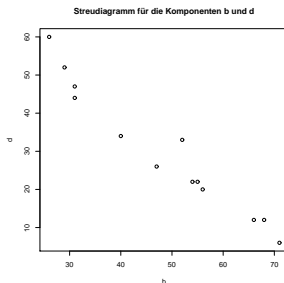
Es handelt sich um quantitative Merkmale.

Wir betrachten:

- Streudiagramm
- Ränge
- Spearmanns Rangkorrelationskoeffizienten
- Bravis-Pearson Korrelationskoeffizienten
- Regressionsgerade

## Aufgabe 2 : Streudiagramm

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



## Aufgabe 2 : Bestimmung der Ränge

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

- sortierter Datensatz mit `sort(b)`

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

- Ränge mit `rank(b)`

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 mit `mean(rank(b))`

$$\overline{R}(b) =$$

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

- genauso erhält man 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

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

- Spearmanscher Rangkorrelationskoeffizient

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

## Aufgabe 2 : 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

## Aufgabe 2 : Zusammenhangsmaße

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

## Aufgabe 2 in R

Spearman'scher Rangkorrelationskoeffizient

```
cor(rank(b),rank(d))
```

```
[1] -0.9903458
```

Korrelationskoeffizient von Bravais-Pearson

```
cor(b,d)
```

```
[1] -0.972955
```

Koeffizienten der Regressionsgeraden

```
lsfit(b,d)$coeff
```

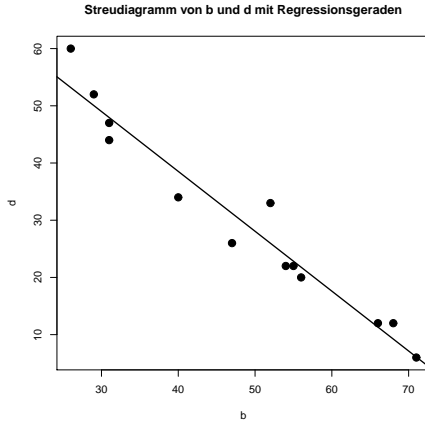
```
Intercept          X
```

```
80.396198 -1.046566
```

Es besteht ein deutlicher negativer Zusammenhang zwischen den Daten

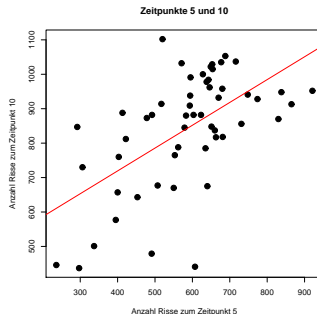
## Aufgabe 2 : Regressionsgrade

```
>abline(lsfit(b,d)$coeff)
```





# Aufgabe 3 : Risskenzahlen in Zeitpunkten 5 und 10



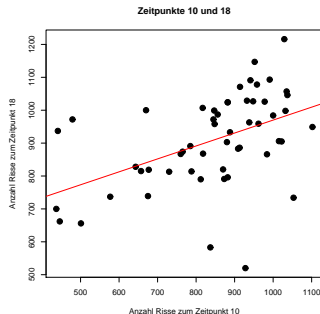
Korrelationskoeffizient von Bravais-Pearson: 0.582104

Spearman'scher Rangkorrelationskoeffizient: 0.562612

Koeffizienten der Regressionsgeraden:

Achsenabschnitt: 454.172324 Steigung: 0.662917

# Aufgabe 3 : Risskenzahlen in Zeitpunkten 10 und 18



Korrelationskoeffizient von Bravais-Pearson: 0.4692203

Spearman'scher Rangkorrelationskoeffizient: 0.5202806

Koeffizienten der Regressionsgeraden:

Achsenabschnitt: 576.7399626 Steigung: 0.3932361