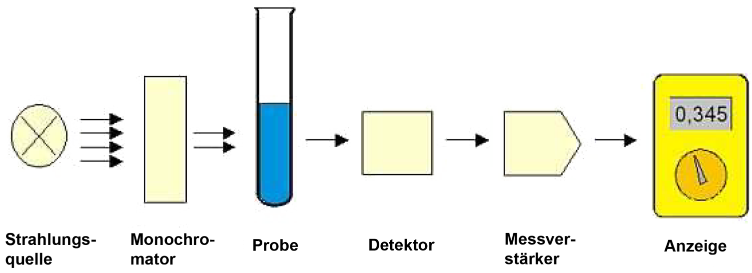
Mathematik 3: Übungsblatt - Statistik 2

1. Aufgabe:

Bestimmung von Stoffmengen in Lösungen wie z.B. Wasserproben durch Fotometrie.

Prinzip eines Einstrahl-Fotometers



Stoffe absorbieren eine bestimmte Wellenlänge des Lichtes. Die übrige "Intensität" kann gemessen werden. Als Messgröße dient eine Spannung U[mV], die von der Konzentration eines Stoffes S[mg/l] in der Lösung abhängig ist.

Um eine Zuordnung von Stoffmenge zur Spannung zu erhalten, muss das System kalibriert werden. Dies erfolgt z.B. mit 5 definierten Proben:

Stoffmenge [mg/l]	2	3	6	10	13
Spannung [mV]	6	21	42	83	93

a) Erstellen eines Diagramms (x=Stoffmenge, y=Spannung)

b) Eine Gerade durch möglichst viele Punkte zeichnen

c) Berechnung der Trendlinie (Regressionsgerade $f(x)$)

d) Einzeichnen der Regressionsgeraden in das Diagramm

e) Vergleich der geschätzten Linie zur berechneten Linie

f) Begriffsbestimmung, Berechnung und Interpretation des "Bestimmtheitsmaßes"

g) Ermittlung eines Probenwertes: Welche Stoffmenge befindet sich in der Probe bei einer Spannung von 35 mV

   i. Ablesen aus dem Diagramm

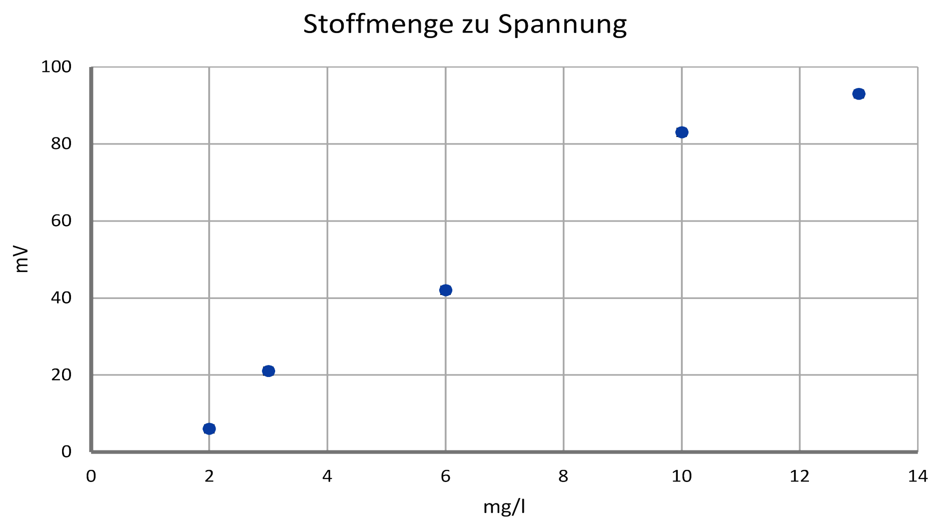
   ii. Berechnung nach $f(x)$

h) Welchen Schnittpunkt hat die Gerade mit der x-Achse und was bedeutet dieser Schnittpunkt?

i) PC-Übung: Ermittlung der Regressionsgeraden mittels Tabellenkalkulation oder Statistik Programm R und Vergleich der Ergebnisse

Lösung:

a)



b), c)

Formeln:

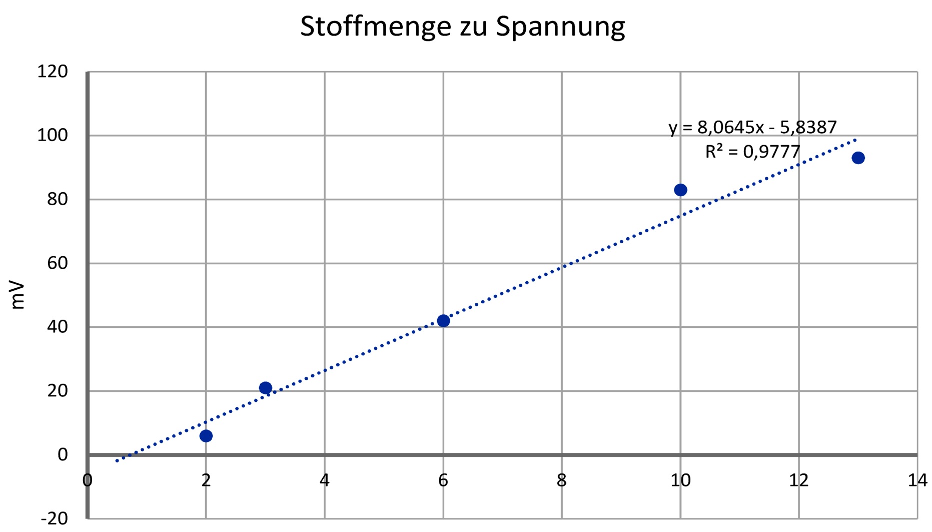
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

$$\underbrace{S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,}_{\text{Varianz der } x\text{-Werte}}$$

$$\hat{y}_i = \hat{a} + \hat{b}x_i \qquad \hat{b} = \frac{S_{XY}}{S_X^2}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}, \qquad \underbrace{S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\,.}_{\text{Kovarianz}}$$

			Xi-x̄	-4,8	-3,8	-0,8	3,2	6,2
x̄=	6,8		(Xi-x̄)²	23,04	14,44	0,64	10,24	38,44
ȳ=	49		Sx²	86,8				
			Yi-ȳ	-43	-28	-7	34	44
			(Xi-x̄)*(Yi -ȳ)	206,4	106,4	5,6	108,8	272,8
			Sxy	700				
b= Sxy/Sx²	8,064							
a= ȳ-b* x̄	-5,838							

d)



e) Wenn die Punkte gut liegen, kann auch eine geschätzte Linie eine gute Annäherung sein.

Variieren die Punkte zu viel, entsteht hier ein zu großer Freiraum für die Wahl der Geraden.

f) Das Bestimmtheitsmaß R^2 sagt aus, wie gut eine lineare Regression zu den Daten passt.

$$R^2 = \frac{S_{\hat{y}}^2}{S_y^2}, \quad S_{\hat{y}}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - \bar{\hat{y}}\right)^2 \quad \text{bzw.} \quad S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

R²	Summe (Yi-ȳ)²	5774,00					
0,9776864	ygi=a+bXi	10,29	18,35	42,55	74,81	99,00	
	(ygi-ȳ)²	1498,44	939,13	41,62	665,97	2500,00	
	Summe (ygi-ȳ)²	5645,16					

Je näher an 1, desto besser!

g) Gemäß Ablesen an der Geraden: Eine Spannung von 35 mV entspricht 5 mg/l des Stoffes.

Berechnung: $x = (y - a)/b$ D.h. $x =$ (35 mV - (-5,8 mV))/8,06 mV $\cdot$ l/mg = 5,06 mg/l

h) Schnittpunkt bei etwa 0,5 mg/l $\Rightarrow$ Nachweisgrenze.
Ab 0,5mg/l kann erst eine Spannung gemessen werden.

Berechnung: $Y = 0 \;\Rightarrow\; 0 = a + bx \;\Rightarrow\; x = (0 - a)/b =$ 5,8 mV/8,06 mV $\cdot$ l/mg = 0,7 mg/l

2. Aufgabe:

Führen Sie eine lineare Regressionsanalyse an dem aus dem Skript bekannten Sekt-Beispiel (Preis-Nachfrage-Problem) **mit R** durch.

Laden/Shop	i	1	2	3	4	5	6
Preis einer Flasche	x_i	20	16	15	16	13	10
verkaufte Menge	y_i	0	3	7	4	6	10

a) Definieren Sie zunächst zwei Variablen/Vektoren `X` und `Y` in R entsprechend der Stichprobentabelle.

b) Erstellen Sie einen Übersichtsplot (Streudiagramm) von `X` und `Y` in R.

c) Erstellen Sie eine KQ (Kleinste Quadrate)-Schätzung. Verwenden Sie dazu den Befehl `lm` ('linear Modell'). Machen Sie sich mit dem `lm` Befehl vertraut (s. z.B. help-Menü) und schreiben Sie das Modell in das Objekt `Modell` ($'\sim'$ befindet sich auf der Taste $'*, +'$ zusammen mit 'ALT').

d) Ermitteln Sie den Korrelationskoeffizienten b, Schnittpunkt (Intercept) a und somit die Gleichung der Regressionsgerade ($\hat{y} = a + bx$).

e) Wenden Sie den Befehl 'summary' auf das Modell an und interpretieren Sie die Ausgaben insbesondere: Multiple R-squared.

f) Zeichnen Sie die Regressionsgerade in das Streudiagramm mit ein.

g) Verwenden Sie die Funktion `predict` um eine Vorhersage für einen beliebigen Wert x, also den Preis auf die zu erwartende verkaufte Menge y zu erhalten. Schreiben Sie dazu die Funktion `predict` in das Objekt `Vorhersage` mit einem Konfidenzbereich von 95%.

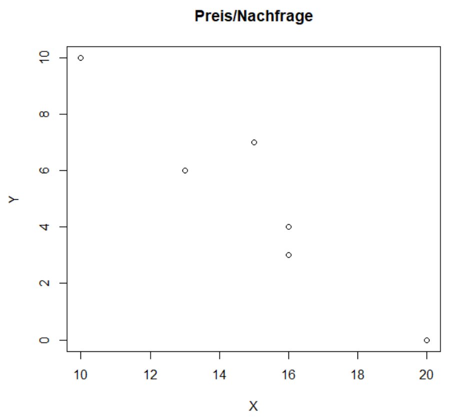
h) (Zusatzaufgabe) Geben Sie das Vorraussage-Modell mit ihrem Vertrauensbereich als Grafik aus mit dem Befehl `matplot`.

Lösung:

a)
- Mit Eingabeaufforderung: `> X <- scan()` bzw. `> Y <- scan()`
- Direkt:

```
> X <- c(20, 16, 15, 16, 13, 10)
> Y <- c(0, 3, 7, 4, 6, 10)
```

b) `> plot(X,Y, main="Preis/Nachfrage")`



c)

```
> Modell <- lm(Y ~ X)
> Modell
```

```
Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)       X
19.7321     -0.9821
```

d) $\hat{y} = 19.7321 - 0.9821 \cdot x$

```
> summary(Modell)

Call:
lm(formula = Y ~ X)

Residuals:
       1        2        3        4        5        6 
-0.08929 -1.01786  2.00000 -0.01786 -0.96429  0.08929 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  19.7321     2.5016   7.888  0.00140 **
X            -0.9821     0.1634  -6.010  0.00386 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.223 on 4 degrees of freedom
Multiple R-squared:  0.9003,    Adjusted R-squared:  0.8754 
F-statistic: 36.12 on 1 and 4 DF,  p-value: 0.003859

> 
```

e)

Interpretation der Werte:

Residuals:
Abweichung zwischen den Y-Werten zur Modellschätzung und den über das Modell geschätzten Y-Werten.

Coefficients:
Angaben zur Signifikanz der geschätzten Parameter.

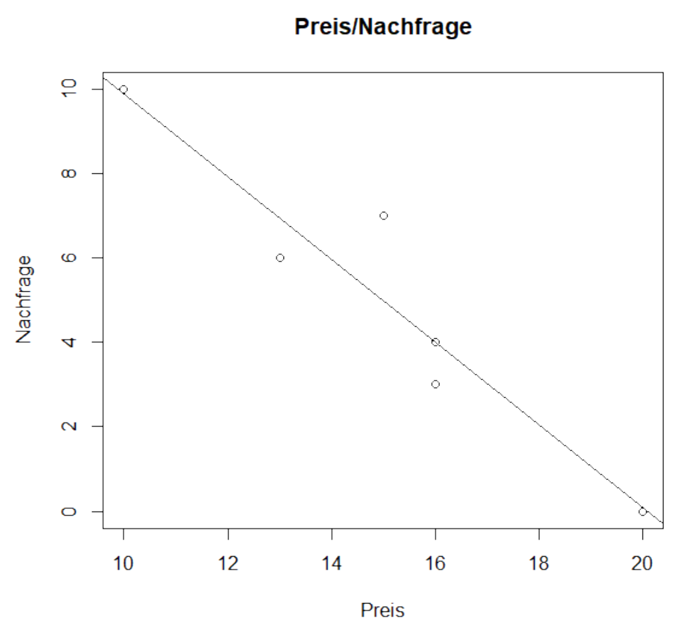
Zusätzlich:

Korrelationskoeffizient (r) (auch: `> r <- cor(X,Y)`)
Maß dafür, ob negative oder positive Korrelation vorliegt (-1.. 0 ..1).

Multiple R-squared, Bestimmtheitsmaß r^2: `> r^2`
der eigentliche lineare Übereinstimmungsfaktor (0..1).

f) `> plot(X,Y, main="Preis/Nachfrage", xlab="Preis", ylab="Nachfrage")`

`> abline(Modell)`



f)
```
>  neu <- data.frame(X=sort(X))
```

```
> X
[1] 20 16 15 16 13 10
> neu
   X
1 10
2 13
3 15
4 16
5 16
6 20
> Vorhersage <- predict(Modell, neu, interval = "c", level = 0.95)
> Vorhersage
         fit       lwr       upr
1 9.91071429  7.252124 12.569305
2 6.96428571  5.307513  8.621058
3 5.00000000  3.613845  6.386155
4 4.01785714  2.559333  5.476382
5 4.01785714  2.559333  5.476382
6 0.08928571 -2.569305  2.747876
```

g)
```
>  matplot(neu, Vorhersage, lty=c(1,2,2), type = "l", main ="Preis/Nachfrage",
+ xlab="Preis in Euro")
```