

Lösung zu Blatt 7 (RStudio)

Hausaufgabe 19

Erstellen der Vektoren

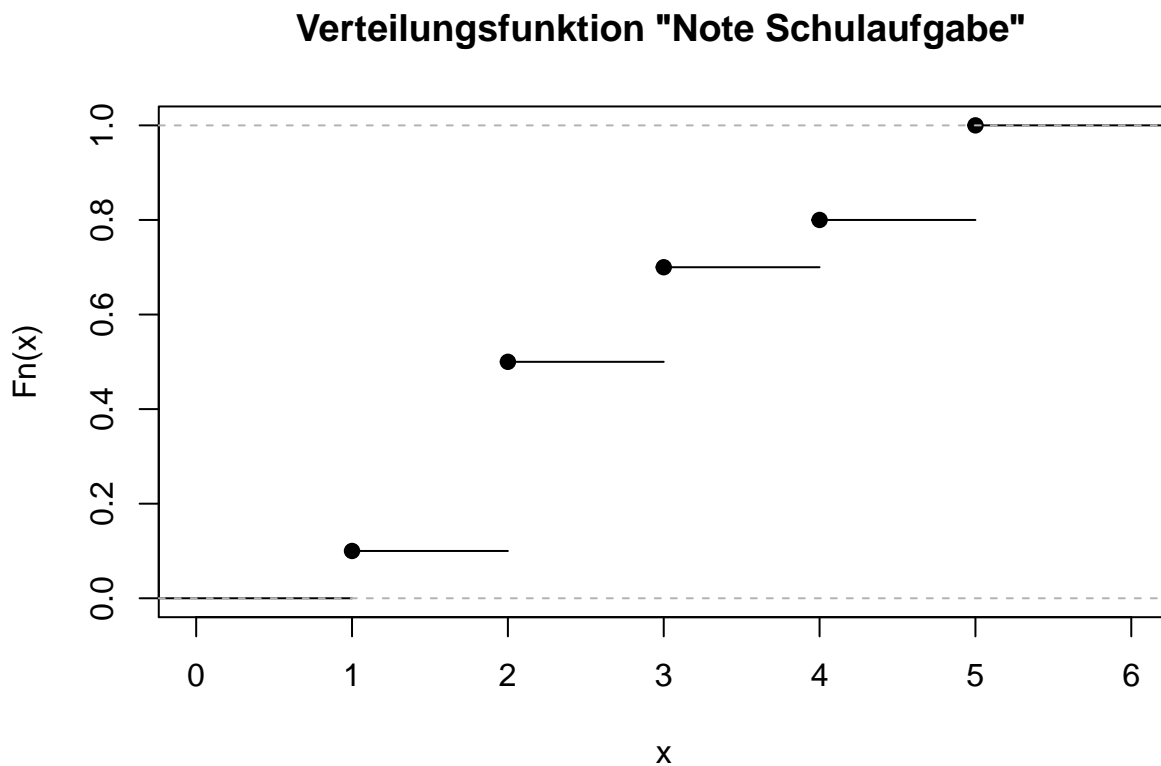
Im ersten Schritt erzeugen wir Vektoren x und y :

```
x = c(2,2,1,1,3,2,2,1,3,3)
y = c(2,3,1,2,5,3,2,2,4,5)
```

Verteilungsfunktion und Boxplot

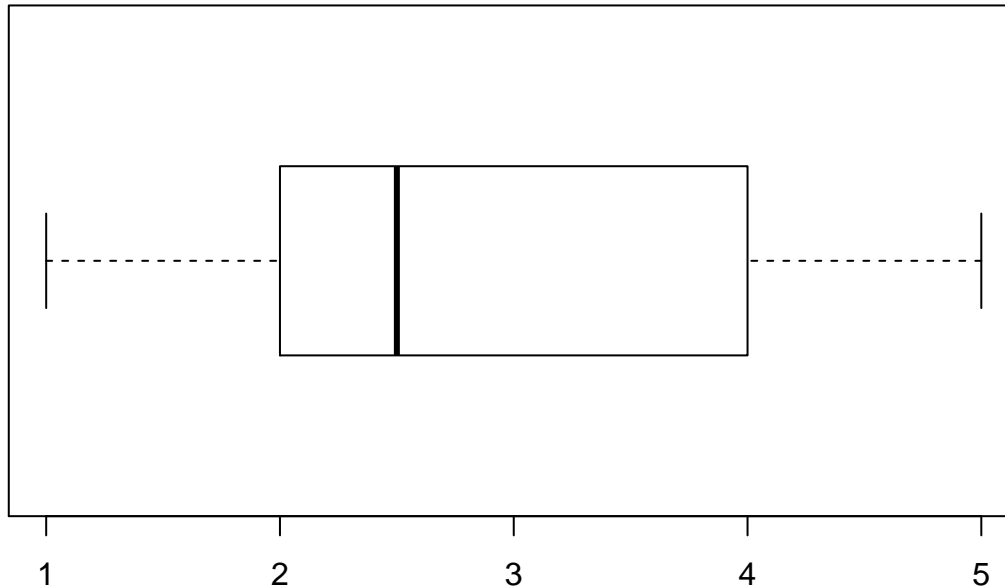
Nun zeichnen wir die empirische Verteilungsfunktion, sowie einen Boxplot für das Merkmal Y :

```
plot(ecdf(y), main = "Verteilungsfunktion 'Note Schulaufgabe'")
```



```
boxplot(y, horizontal = TRUE, range = 2, main = "Boxplot 'Note Schulaufgabe'")
```

Boxplot "Note Schulaufgabe"



Korrelationskoeffizient

Für die Berechnung des Korrelationskoeffizienten benutzen wir folgende Formel:

$$r_{XY} = \frac{\langle X, Y \rangle - n\bar{x}\bar{y}}{\sqrt{(\|X\|^2 - n\bar{x}^2)(\|Y\|^2 - n\bar{y}^2)}}.$$

```
n = length(x)
m_x = mean(x)
m_y = mean(y)
n_x = sum(x^2)
n_y = sum(y^2)
(sum(x*y) - n*m_x*m_y)/sqrt((n_x-n*m_x^2)*(n_y-n*m_y^2))
```

```
## [1] 0.8937654
```

Trendgerade

Wir bestimmen die Gleichung der Trendgeraden mit dem Befehl `lm(y ~ x)`:

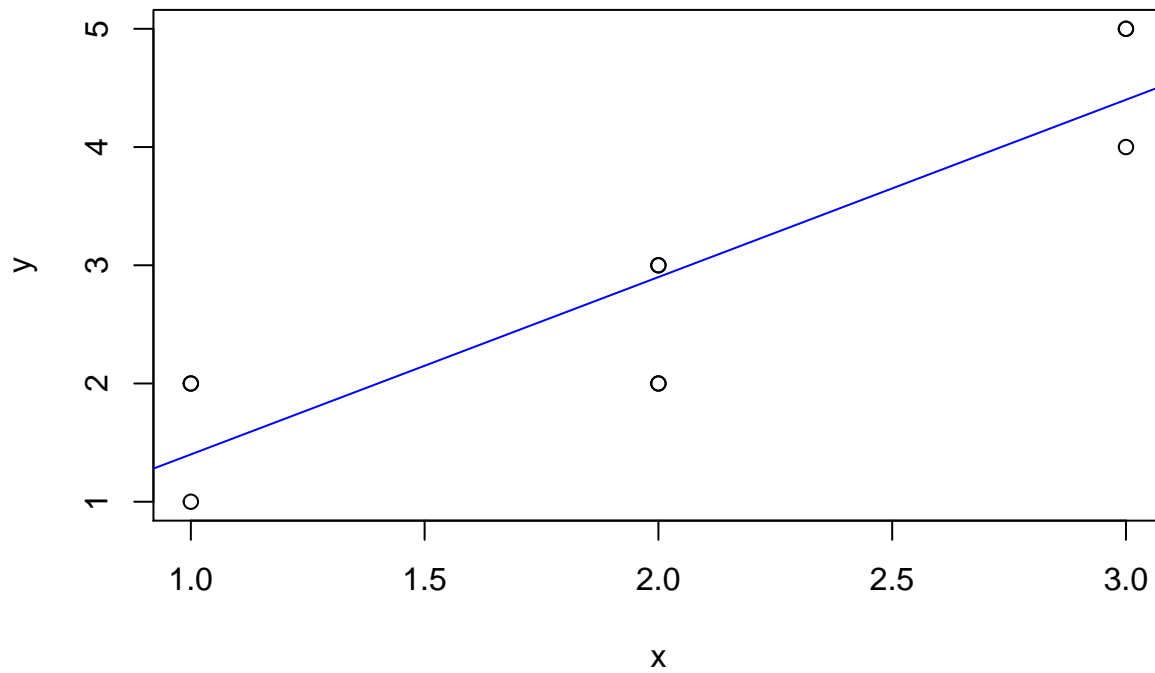
```
linear_model = lm(y ~ x)
linear_model
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)      x
##      -0.1         1.5
```

Somit lautet die Gleichung der Trendgeraden $y = 1.5x - 0.1$.

Nun zeichnen wir die entsprechende Punktwolke und fügen die Trendgerade hinzu:

```
plot(x, y)
abline(linear_model, col = "blue")
```



Interpretation

Da der Korrelationskoeffizient relativ nahe bei 1 liegt, kann von einer starken Korrelation gesprochen werden. Schüler, die gut mitarbeiten, haben auch eine gute Note in der Schulaufgabe.

Hausaufgabe 20

Die Matrix A hat folgende Struktur:

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}.$$

Wir erzeugen die Matrizen A und B , sowie den Vektor w :

```
v_1 = rep(1, n)
A = matrix(c(x, v_1), ncol = 2, byrow = FALSE)
A
```

```
##      [,1] [,2]
## [1,]    2    1
## [2,]    2    1
## [3,]    1    1
## [4,]    1    1
## [5,]    3    1
## [6,]    2    1
## [7,]    2    1
## [8,]    1    1
## [9,]    3    1
## [10,]   3    1
```

```
B = t(A)%*%A
B
```

```
##      [,1] [,2]
## [1,]   46   20
## [2,]   20   10
```

```
w = t(A)%*%y
w
```

```
##      [,1]
## [1,]    67
## [2,]    29
```

Die Spalten der Matrix B sind linear unabhängig. Daher folgt $\text{rang}(B) = 2$ und die Normalgleichung ist eindeutig lösbar. Die Lösung ist gegeben durch:

```
solve(B, w)
```

```
##      [,1]
## [1,]  1.5
## [2,] -0.1
```

Für die Trendgerade erhalten wir also $y = 1.5x - 0.1$. Dies bestätigt unser Ergebnis in Hausaufgabe 19.