

Übungszettel 7 - Lösung

Biostatistik - MoBi - SS2018 - Carl Herrmann

Aufgabe 1

Bei 10 männlichen Patienten wurden folgende Harnstoffwerte im Serum gemessen:

Alter	Harnstoff
70	4.1
76	5.6
81	8.8
85	9.7
44	2.9
58	3.1
51	3.5
74	6.8
45	2.8
67	3.8

- Bestimmen Sie aus den Daten den Pearson Korrelationskoeffizienten zwischen Alter und Harnstoffgehalt. Benutzen Sie dazu: $\sum_{i=1}^{10} (x_i - \bar{x})^2 = 1972.9$, $\sum_{i=1}^{10} (y_i - \bar{y})^2 = 57.369$, $\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 292.99$
- Bestimmen Sie die Steigung und der Schnittpunkt der Regressionsgeraden

```
b1 = cor(x[,1],x[,2])*sd(x[,2])/sd(x[,1])
b0 = mean(x[,2])-b1*mean(x[,1])
b0;b1
```

```
## [1] -4.557824
```

```
## [1] 0.1485073
```

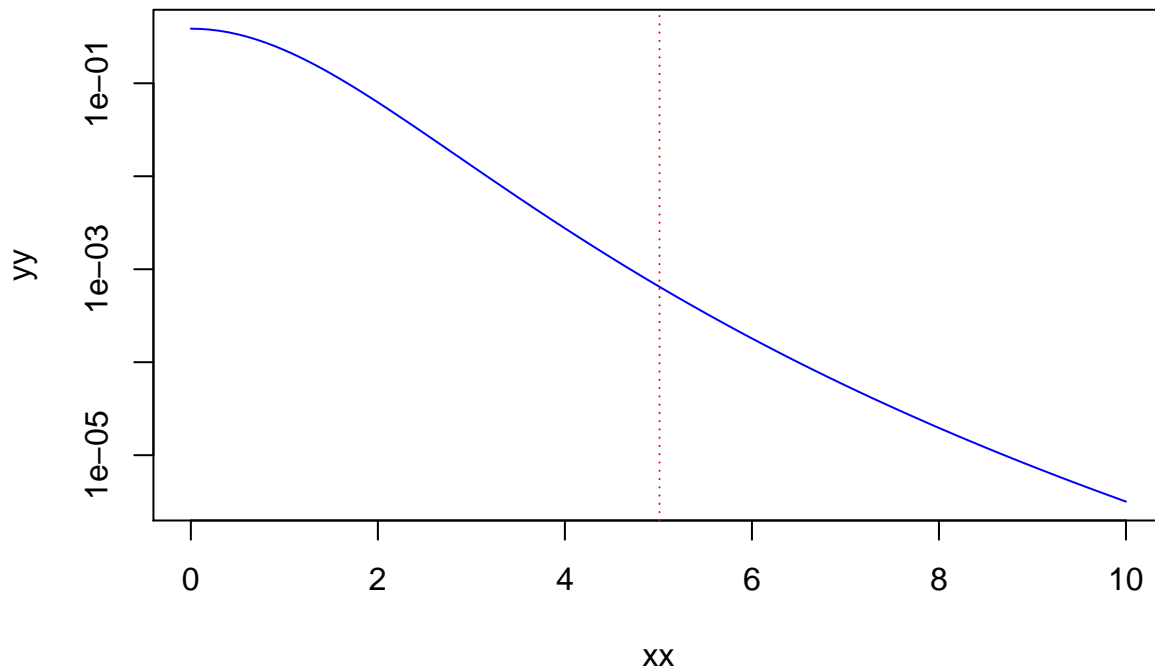
- Ist der Pearson Korrelationskoeffizient signifikant von 0 verschieden? Berechnen Sie die Teststatistik und der dazugehörigen P-Wert.

```
r = cor(x[,1],x[,2])
n = nrow(x)
t = r/sqrt((1-r^2)/(n-2))
r;t
```

```
## [1] 0.8708866
```

```
## [1] 5.011841
```

```
xx = seq(0,10,by=0.1);yy=dt(xx,n-2)
plot(xx,yy,type='l',col='blue',log='y');abline(v=t,lty=3,col='red')
```



```
p = pt(t,n-2,lower.tail=FALSE)
p
```

```
## [1] 0.0005186888
```

- Berechnen Sie den Spearman-Korrelationskoeffizienten und vergleichen Sie diesen mit dem Koeffizienten nach Pearson. Wie erklären Sie den Unterschied?

```
cor(x[,1],x[,2],method='spearman')
```

```
## [1] 0.9636364
```

Aufgabe 2

Von einer Firma sind über mehrere Jahre hinweg die Umsätze und die Beschäftigtenzahlen bekannt:

Jahr t	1	2	3	4	5
Umsatz $x(t)$	60	55	57	61	65
Anzahl der Beschäftigten $y(t)$	1000	1100	960	840	800

- Bestimmen Sie den Korrelationskoeffizienten nach Spearman

```
## [1] -0.9
```

Jahr	Umsatz	Anzahl
1	60	1000
2	55	1100
3	57	960
4	61	840
5	65	800

- Bestimmen Sie die Parameter der Regressionsgeraden $y(t) = b_0 + b_1 t$.

```
b1 = cor(data$Jahr,data$Anzahl)*sd(data$Anzahl)/sd(data$Jahr)
b0 = mean(data$Anzahl)-b1*mean(data$Jahr)
b0;b1
```

```
## [1] 1138
```

```
## [1] -66
```

- Mit welcher Anzahl von Beschäftigten ist im Jahr 8 zu rechnen ?

```
b0+b1*8
```

```
## [1] 610
```

- Berechnen Sie den F-ratio und bestimmen Sie anhand dieser Tabelle (hier der Link) ob das Regressionsmodel signifikant besser als das Null-Model ist.

```
## Anzahl Beobachtungen
```

```
n = nrow(data)
```

```
## Vorhergesagte Anzahl von Beschäftigten
```

```
pred = b0+b1*data$Jahr
```

```
## Residuen
```

```
e=data$Anzahl-pred
```

```
ssr = sum(e^2)/(n-2)
```

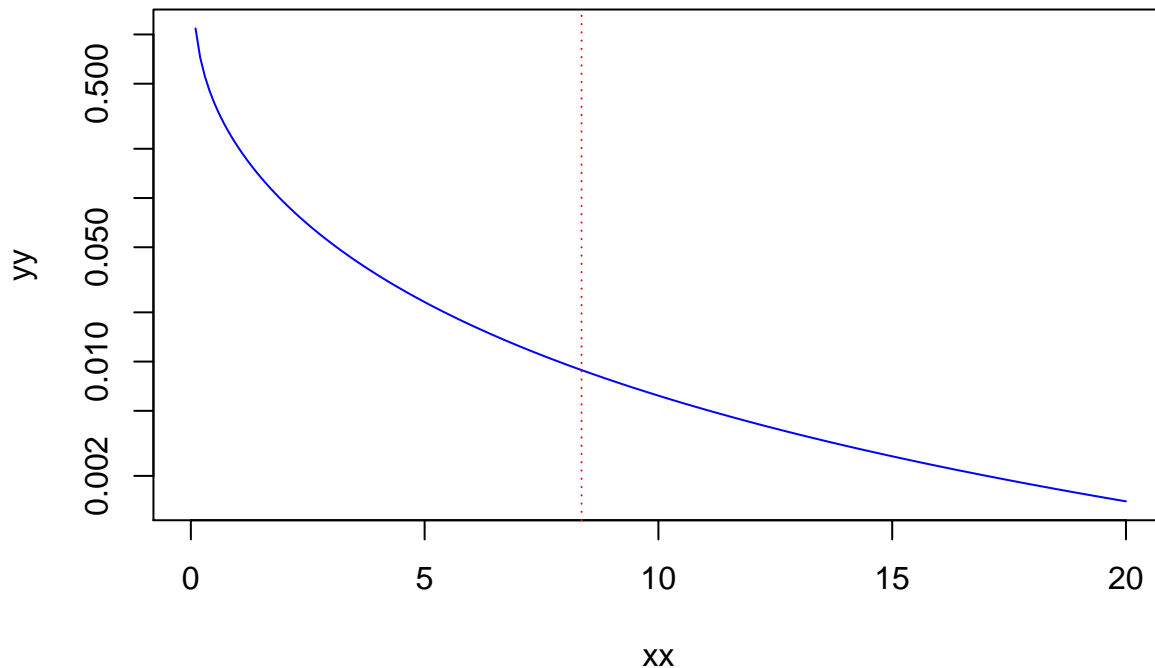
```
ssm = sum((pred-mean(data$Anzahl))^2)
```

```
f = ssm/ssr
```

```
## Plot der F=Verteilung
```

```
xx = seq(0,20,by=0.1); yy=df(xx,1,n-2)
```

```
plot(xx,yy,type='l',col='blue',log='y'); abline(v=f,lty=3,col='red')
```



```
pf(f,1,n-2,lower.tail=FALSE)
```

[1] 0.06298284

Aufgabe 3

In einem Versuch werden 10 Tests durchgeführt, die zu folgenden p-Werten führen. Bestimmen Sie, welche dieser signifikant sind unter Benutzung:

- eines Signifikanzwertes $\alpha = 0.05$
- Bonferroni-Korrektur mit $\alpha = 0.05$
- Benjamini-Hochberg FDR Korrektur mit $\text{FDR}=1\%$

PWert
0.3979419
0.0016410
0.0176781
0.9044153
0.4254804
0.0002708
0.0006478
0.2640552
0.8742515
0.0055376

p	noCorrection	Bonf	BH
0.0002708	TRUE	TRUE	TRUE
0.0006478	TRUE	TRUE	TRUE
0.0016410	TRUE	TRUE	TRUE
0.0055376	TRUE	FALSE	FALSE
0.0176781	TRUE	FALSE	FALSE
0.2640552	FALSE	FALSE	FALSE
0.3979419	FALSE	FALSE	FALSE
0.4254804	FALSE	FALSE	FALSE
0.8742515	FALSE	FALSE	FALSE
0.9044153	FALSE	FALSE	FALSE