

7. Übung(sserie) Statistik für Ingenieure WiSe 18/19

Teil 2

Dieser Teil wird in der letzten Vorlesung am 07.02.2019 behandelt.

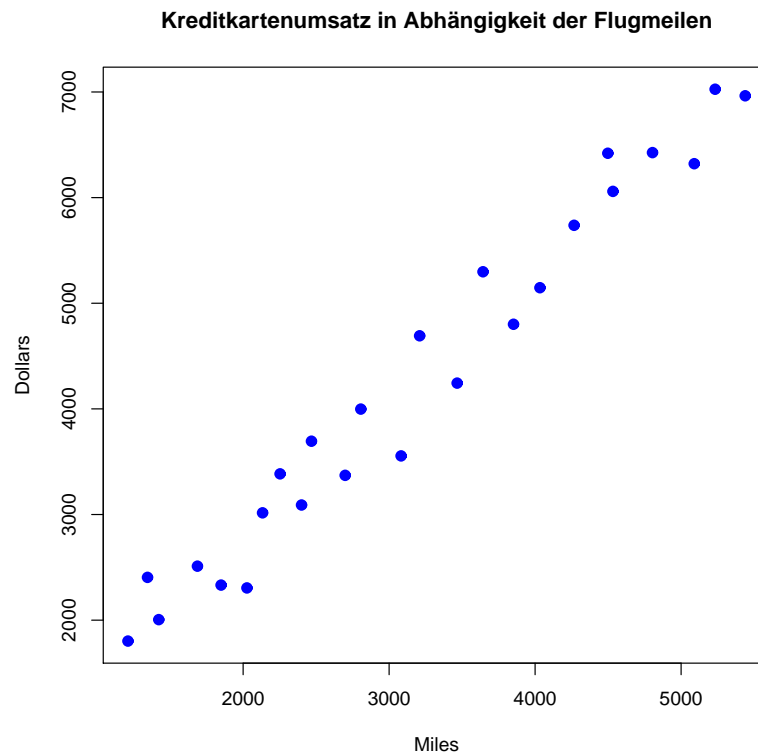
- Aufgabe:** American Express Company glaubte, dass Personen mit Kreditkarten intensiver reisen als Personen ohne. Die Frage die jetzt untersucht wird ist, ob es einen Zusammenhang zwischen der Reisedstrecke (Miles) und dem Kreditkartenumsatz (Dollars) gibt. Dazu wurden aus allen Karteninhabern 25 zufällig ausgewählt. Die Daten (in einem gewissen Zeitraum) sind in Miles.txt zu finden.

```
> miles<-read.table("D:/Miles.txt",header=T)
> miles
```

	Miles	Dollars
1	1211	1802
2	1345	2405
3	1422	2005
4	1687	2511
5	1849	2332
6	2026	2305
7	2133	3016
8	2253	3385
9	2400	3090
10	2468	3694
11	2699	3371
12	2806	3998
13	3082	3555
14	3209	4692
15	3466	4244
16	3643	5298
17	3852	4801
18	4033	5147
19	4267	5738
20	4498	6420
21	4533	6059
22	4804	6426
23	5090	6321
24	5233	7026
25	5439	6964

a) Was können Sie aus der folgenden Grafik ablesen?

```
> attach(miles)
> plot(miles,col="blue",lwd=6,cex=0.5,main="Kreditkartenumsatz
in Abhängigkeit der Flugmeilen")
```



b) Testen Sie zum Niveau $\alpha = 0,05$, ob sowohl der zufällige Kreditkartenumsatz, als auch die zufälligen Flugmeilen normalverteilt sind.

```
> shapiro.test(miles$Dollars)
```

Shapiro-Wilk normality test

```
data: miles$Dollars W = 0.936, p-value = 0.1195
```

```
> shapiro.test(miles$Miles)
```

Shapiro-Wilk normality test

```
data: miles$Miles W = 0.9518, p-value = 0.275
```

- c) Welchen der 2 folgenden Tests würden Sie verwenden, um zu testen, ob es eine signifikante Abhängigkeit des Kreditkartenumsatzes von der Reisetrecke gibt? Wie lauten die Hypothesen und wie die Testentscheidung bei $\alpha = 0,01$?

(1) `> cor.test(miles$Dollars, miles$Miles)`

Pearson's product-moment correlation

```
data: miles$Dollars and miles$Miles
t = 25.2482, df = 23, p-value <2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9599435 0.9923460
sample estimates:
      cor
0.9824339
```

(2) `> cor.test(miles$Dollars, miles$Miles,method="spearman")`

Spearman's rank correlation rho

```
data: miles$Dollars and miles$Miles S = 62, p-value = 6.869e-07
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9761538
```

- d) Wie muss man den R-Befehl in c) erweitern, um die Richtung des Zusammenhanges zwischen Kreditkartenumsatz und Reisetrecke signifikant zeigen zu können?

2. **Aufgabe:** In der Weinabteilung eines englischen Supermarktes, der sowohl deutsche als auch französische Weine feilbot, wurde an aufeinander folgenden Tagen entweder deutsche oder französische Musik gespielt. In der ersten Woche waren die deutschen Weine auf der linken Seite des Regals, die französischen rechts, in der zweiten Woche war es umgekehrt. Nationalflaggen an den Regalen machten den Herkunftsort der Weine unmissverständlich klar. Die Musik der ersten Woche war französische (im Wesentlichen Akkordeon-Musik) und in der zweiten Woche deutsche (im Wesentlichen Blaskapellen-Musik).

Herkunft des Weins	Musik	
	französische	deutsche
Frankreich	39	12
Deutschland	8	22

(nach M. Spitzer: *Musik, Wein und Bahnungseffekte*, Geist & Gehirn, dabei ist die Studie aus: North A, Hargreaves D, McKendrick J. *The influence of in-store music on wine selections*. Journal of Applied Psychology 1999; 84: 271-6.)

```

> Tabelle <- matrix(c(39,12,8,22), 2, 2, byrow=TRUE)
> rownames(Tabelle) <- c("Frankreich", "Deutschland")
> colnames(Tabelle) <- c("französische", "deutsche")
> Tabelle # Counts

```

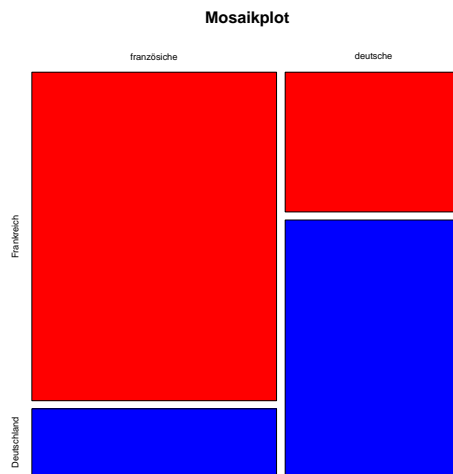
	französische	deutsche
Frankreich	39	12
Deutschland	8	22

a) Was können Sie aus folgender Grafik ablesen?

```

> mosaicplot(t(Tabelle), main="Mosaikplot", color=c(2,4))

```



b) Testen Sie zum Niveau $\alpha = 0,01$, ob die Wahl des Weines durch die Musik beeinflusst wird.

```

> chisq.test(Tabelle, correct=FALSE)

```

Pearson's Chi-squared test

data: Tabelle

X-squared = 19.2365, df = 1, p-value = 1.155e-05

```

> fisher.test(Tabelle)

```

Fisher's Exact Test for Count Data

data: Tabelle p-value = 1.806e-05

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

2.854093 28.966755

sample estimates:

odds ratio

8.64643