Statistik für Ingenieure
Übung 7, 2.Teil

Prof. Dr. Hans-Jörg Starkloff

TU Bergakademie Freiberg
Institut für Stochastik

3. Februar 2020

1. 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 Reisestrecke (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
```

Daten

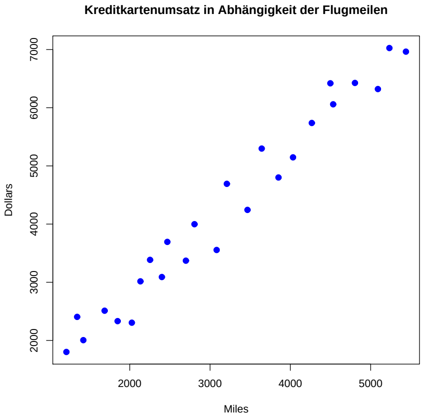
	Miles	Dollars
1	1211	1802
2	1345	2405
3	1422	2005
4	1687	2511
5	1849	2332
6	2026	2305
7	2133	3016
8	2253	3385
9	2400	3090
10	2468	3694
11	2699	3371
12	2806	3998
13	3082	3555
14	3209	4692
15	3466	4244

	Miles	Dollars
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

Teilaufgabe a)

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")
```



Antwort Teilaufgabe a)

Die Datenpunkte im Streudiagramm streuen nicht sehr stark um eine Gerade mit positivem Anstieg. Also scheint es einen Zusammenhang zwischen der Reisestrecke umd dem Kreditkartenumsatz zu geben, der auch gut durch eine lineare Funktion beschrieben werden kann.

Teilaufgabe b)

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
```

Antwort Teilaufgabe b)

Bez.: X – Reisestrecke ("Miles"), Y – Kreditkartenumsatz ("Dollars").

H_0 : $Y \sim \mathbf{N}(\mu_Y, \sigma_Y^2)$ Kreditkartenumsatz ist normalverteilt.
H_A : $Y \not\sim \mathbf{N}(\mu_Y, \sigma_Y^2)$ Kreditkartenumsatz ist nicht normalverteilt.
$p = 0.1195 > 0.05 = \alpha \;\Rightarrow\; H_0$ wird nicht abgelehnt.

H_0 : $X \sim \mathbf{N}(\mu_X, \sigma_X^2)$ Reisestrecke ist normalverteilt.
H_A : $X \not\sim \mathbf{N}(\mu, \sigma_Y^2)$ Reisestrecke ist nicht normalverteilt.
$p = 0.275 > 0.05 = \alpha \;\Rightarrow\; H_0$ wird nicht abgelehnt.

Sowohl der Kreditkartenumsatz als auch die Reisestrecke unterscheiden sich nicht signifikant von einer Normalverteilung.

Unter Berücksichtigung der Bonferroni-Korrektur:
H_0 : X und Y sind beide normalverteilt.
H_A : mindestens eine der Zufallsgrößen X und Y ist nicht normalverteilt.

$p_Y = 0.1195 > 0.025 = \alpha/2$, $p_X = 0.275 > 0.025 = \alpha/2$,
folglich wird H_0 nicht abgelehnt.

Teilaufgabe c)

c) Welchen der 2 folgenden Tests würden Sie verwenden, um zu testen, ob es eine signifikante Abhängigkeit des Kreditkartenumsatzes von der Reisestrecke 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
```

Fortsetzung Teilaufgabe c)

(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
```

Antwort Teilaufgabe c)

Falls X, Y normalverteilt $\Rightarrow$ Pearson; sonst Spearman.

Hier also Test (1) aufgrund der Ergebnisse von b).

$H_0 : \rho_{X,Y} = 0 \quad (X, Y \text{ unabhängig})$
$H_A : \rho_{X,Y} \neq 0 \quad (X, Y \text{ abhängig})$
$p = 2.2 \cdot 10^{-16} < 0.01 = \alpha$
H_0 wird abgelehnt, d.h. es gibt eine signifikante Abhängigkeit zwischen der Reisestrecke und dem Kreditkartenumsatz.

Teilaufgabe d)

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

Antwort Teilaufgabe d)

Einseitiger Test, so dass man bei Ablehnung der Nullhypothese auf eine signifikante positive Korrelation schließen kann.

```
> cor.test(miles$Dollars, miles$Miles,
+ alternative="greater")
```

2. Aufgabe

In der Weinabteilung eines englischen Supermarktes, der sowohl deutsche als auch französische Weine feilbot, wurde an aufeinanderfolgenden 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. Nationalfahnen 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).

	Musik	
Herkunft des Weins	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: 271-276.)

R-Code

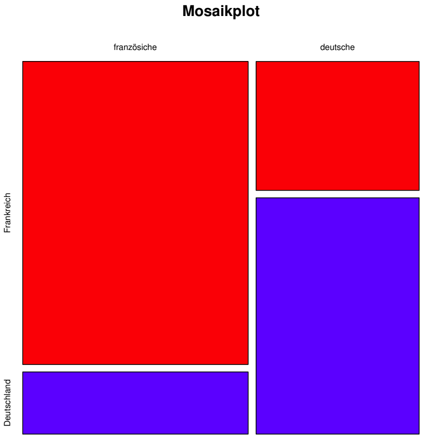
```
> Tabelle <- matrix(c(39,12,8,22), 2, 2, byrow=TRUE)
> rownames(Tabelle) <- c("Frankreich","Deutschland")
> colnames(Tabelle) <- c("französische","deutsche")
> Tabelle # Counts
```

```
            französische deutsche
Frankreich            39       12
Deutschland            8       22
```

Teilaufgabe a)

a) Was können Sie aus folgender Grafik ablesen?

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



Antwort Teilaufgabe a)

Erkennbar sind bedingte Verteilungen in der Stichprobe:

unter der Bedingung, dass französische Musik gespielt wird:
großer Anteil von Wein aus Frankreich,
kleiner Anteil von Wein aus Deutschland;

unter der Bedingung, dass deutsche Musik gespielt wird:
großer Anteil von Wein aus Deutschland,
kleiner Anteil von Wein aus Frankreich.

Teilaufgabe b)

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
```

Antwort Teilaufgabe b)

H_0 : Die Wahl des Weines ist von der gespielten Musik unabhängig.
H_A : Die Wahl des Weines ist von der gespielten Musik abhängig.

`fisher.test`: Der exakte Test von Fisher geht nur bei $2 \times 2-$Tafeln (Vierfeldertafeln).

$p = 1.806 \cdot 10^{-5} < 0.01 = \alpha \quad \Rightarrow \quad H_0$ wird abgelehnt.

`chisq.test`: Der χ^2-Test ist ein asymptotischer Test, geht bei „großen" Stichproben immer und nicht nur bei $2 \times 2-$Tafeln (Vierfeldertafeln).

$p = 1.155 \cdot 10^{-5} < 0.01 = \alpha \quad \Rightarrow \quad H_0$ wird abgelehnt.

Die Wahl des Weines und die gespielte Musik sind abhängig. Es gibt eine signifikante Abhängigkeit der Wahl des Weines (nach Herkunftsland) von der gespielten Musik.