Hypothesentest Pearson

Code

1) Grundlegende Konzepte: Was ist Pearson?

Die Korrelation, auch bivariate Korrelation oder Produkt-Moment-Korrelation genannt, beschreibt den Zusammenhang von zwei intervallskalierten Merkmalen/Variablen einer Zufallsstichprobe. Eine Möglichkeit, die Stärke des Zusammenhangs zu bestimmen, ist die Berechnung des Korrelationskoeffizienten r nach Bravais und Pearson. Voraussetzung ist hierbei, dass es sich um einen linearen Zusammenhang zwischen den analysierten Merkmalen handelt. Zusätzlich wird hier ein ungerichteter Zusammenhang untersucht, d.h. die Variablen sind unabhängig voneinander und folglich werden keine kausalen Aussagen gemacht.

Der Korrelationskoeffizient r kann Werte zwischen -1 und +1 annehmen und ist unabhängig von der Maßeinheit. Ein Wert von -1 beschreibt eine perfekt negative Korrelation und ein Wert von +1 eine perfekt positive Korrelation. Bei r = 0 liegt kein linearer Zusammenhang zwischen den Variablen vor.

Scheinkorrelation bezeichnet eine Korrelation (Übereinstimmung oder Entsprechung) zwischen zwei Größen, der kein Kausalzusammenhang, sondern nur eine zufällige oder indirekte Beziehung zu Grunde liegt.

2) Aufgabenstellung - Beschreibung

Im Rahmen einer Studie wurden Spatzen untersucht, dazu wurde das Gewicht der Spatzen, sowie deren Flügellänge erhoben. Es soll untersucht, ob es einen Zusammenhang zwischen das Gewicht und die Flügellänge des Spatzen besteht.

3) Hypothese

H1: Es gibt einen Zusammenhang zwischen Gewicht und Flügellänge eines Spatzen.

$$(r \neq 0)$$

H0: Es gibt keinen Zusammenhang zwischen Gewicht und Flügellänge eines Spatzen.

$$(r = 0)$$

Variable 1 = Weight -> Gewicht (g)

Variable 2 = WingLength -> Flügellänge (mm)

4) Voraussetzungen

Die Variablen sind mindestens intervallskaliert -> Erfüllt -> Gewicht und Flügellänge sind ratioskaliert.

Die Variablen sind normalverteilt (n=116>30) -> Erfüllt -> siehe Histogramm.

Der untersuchte Zusammenhang zwischen den Variablen muss linear sein -> Erfüllt -> siehe Streudiagramm.

Datensatz:

Hide

```
Pearson.Sparrows <- read.csv("C:/Data_Science_Projects/Repository_2/Aufgabe3/Pearson-Sparrows.csv")
View(Pearson.Sparrows)
```

Hide

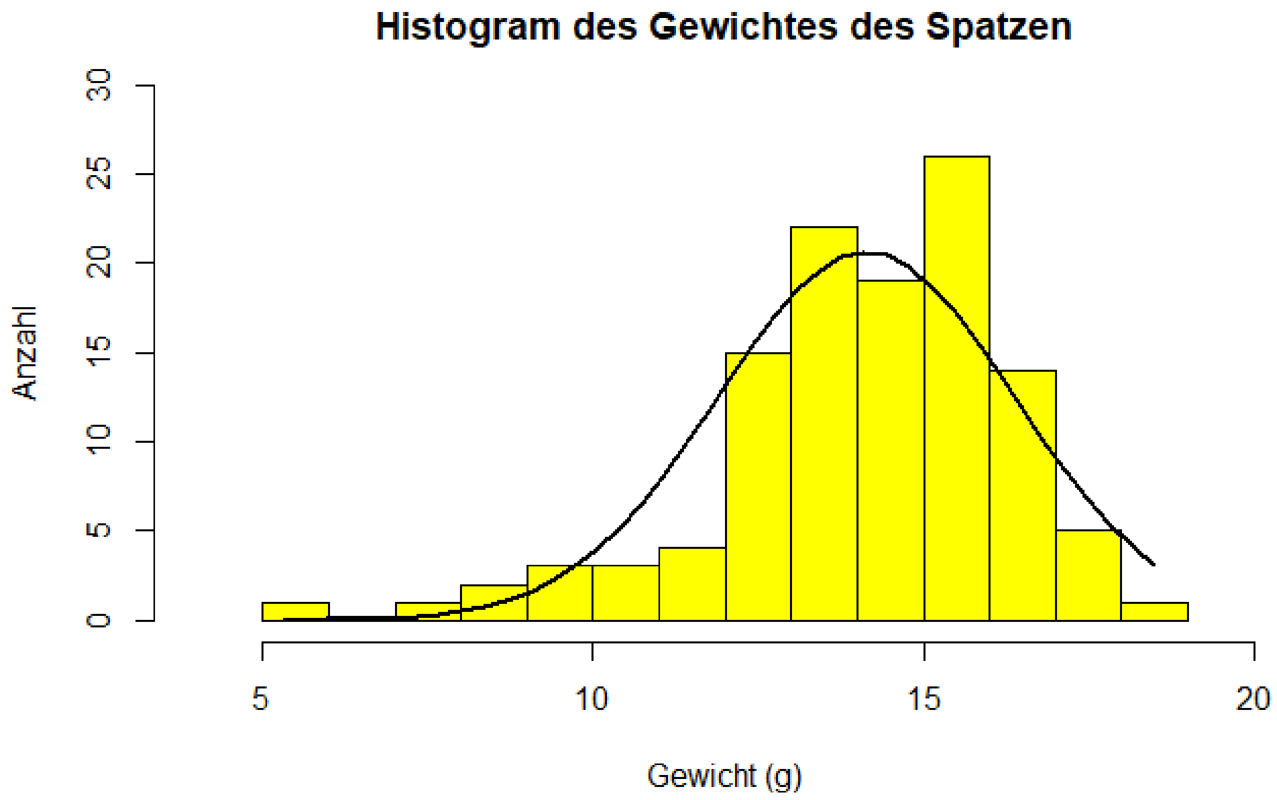
```
library(dplyr) #-> Rename
Pearson.Sparrows <- Pearson.Sparrows %>% #pipe
           rename(Gewicht = 'Weight',
                  Flügellänge  = 'WingLength'
                  )
```

Normalverteilung - Prüfung mittels Histogramm

Um einen ersten Überblick über die Daten zu gewinnen, empfiehlt es sich Histogrammm zu erstellen.

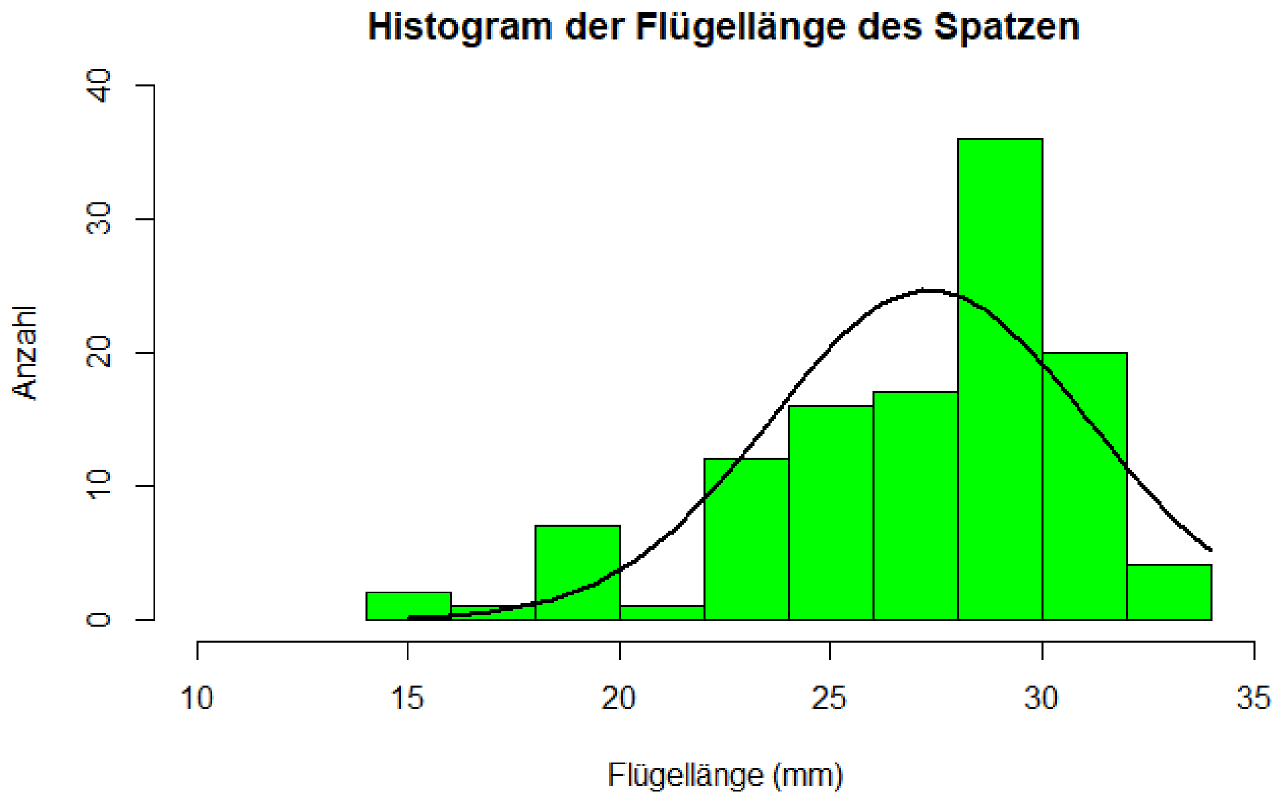
Hide

```
x <- Pearson.Sparrows$Gewicht

h <- hist(x,
          breaks=11,
          col="yellow",
          xlab="Gewicht (g)",
          xlim = c(4, 20),
          main="Histogram des Gewichtes des Spatzen",
          ylab= "Anzahl",
          ylim = c(0, 30))
xfit <- seq(min(x),
            max(x),
            length=40)
yfit <-dnorm(xfit,
             mean=mean(x),
             sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit,
      yfit,
      col="black",
      lwd=2)
```



Hide

```
x <- Pearson.Sparrows$Flügellänge

h <- hist(x,
          breaks=11,
          col="green",
          xlab="Flügellänge (mm)",
          xlim = c(10, 35),
          main="Histogram der Flügellänge des Spatzen",
          ylab= "Anzahl",
          ylim = c(0, 40))
xfit <- seq(min(x),
            max(x),
            length=40)
yfit <-dnorm(xfit,
             mean=mean(x),
             sd=sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit,
      yfit,
      col="black",
      lwd=2)
```

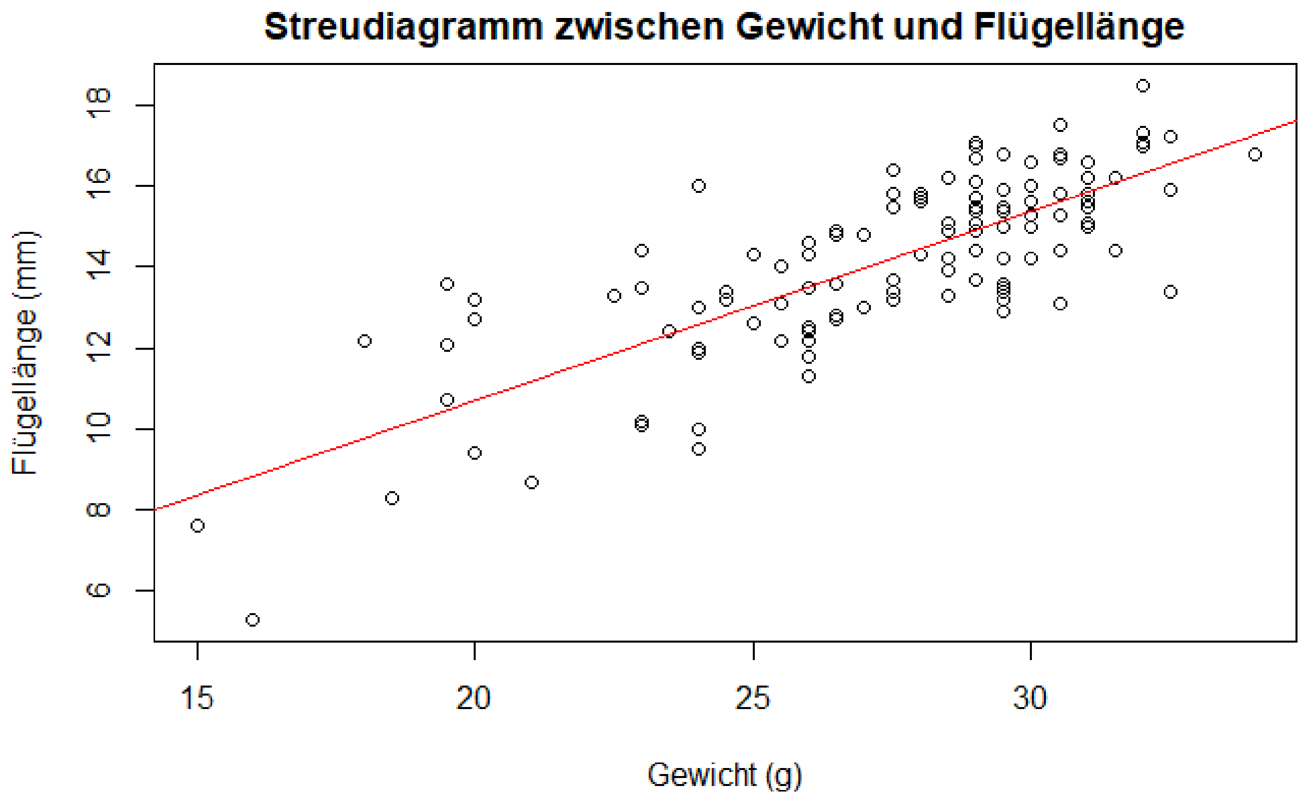


Die beiden Variablen scheinen normalverteilt zusein.Sie zeigt eine bauchige Mitte, insbesonders bei dem Gewicht. Darüber hinaus liegt die Strichprobengröße mit n=116 über der 30 Schwelle, was für eine Normalverteilung spricht.

5) Grafische Veranschaulichung des Zusammenhangs

Hide

```
plot(Pearson.Sparrows$Gewicht ~ Pearson.Sparrows$Flügellänge,
     main = "Streudiagramm zwischen Gewicht und Flügellänge",
     xlab = "Gewicht (g)",
     ylab= "Flügellänge (mm)")
abline(lm(Pearson.Sparrows$Gewicht ~ Pearson.Sparrows$Flügellänge, data = Pearson.Sparrows),
       col="red")
```



Das Streudiagramm in Abbildung zeigt eine tendenziell positive lineare Beziehung zwischen den beiden Variablen. Das heisst, die beiden Variablen korrelieren positiv vermutlich. Da die Korrelationsanalyse einen ungerichteten Zusammenhang untersucht, lässt er sich auf zwei Weisen ausformulieren: Je größer das Gewicht, desto länger ist die Flügellänge, oder je länger die Flügellänge, desto größer ist das Gewicht.

6) Deskriptive Statistik

Die Tabelle in Abbildung gibt die Mittelwerte, Standardabweichungen und Grössen der Variablen wieder. Diese Informationen werden für die Berichterstattung verwendet.

Hide

```
Sparrows <- Pearson.Sparrows[c(3,4)]
#head(Sparrows)
```

Hide

```
library(psych)
describe(Sparrows)
```

	vars <dbl>	n <dbl>	mean <dbl>	sd <dbl>	median <dbl>	trimmed <dbl>	mad <dbl>	min <dbl>	max <dbl>
Gewicht	1	116	14.13	2.24	14.4	14.36	2.00	5.3	18.5
Flügellänge	2	116	27.32	3.76	28.5	27.75	2.97	15.0	34.0

2 rows | 1-10 of 13 columns

In die Tabelle können die Mittelwerte und Standardabweichungen der Variablen Gewicht und Flügellänge abgelesen werden. Im Mittel liegt das Gewicht in g der Spatzen bei 14.13 (SD = 2.24,n=116)). Die Flügellänge gemessen in mm liegt durchschnittlich bei 27.32 (SD = 3.76, n=116) in der Einzelarbeit.

7) Ergebnisse der Korrelationsanalyse

Hide

```
SparrowTest <- cor.test(Pearson.Sparrows$Gewicht, Pearson.Sparrows$Flügellänge)
SparrowTest
```

```


	Pearson's product-moment correlation

data:  Pearson.Sparrows$Gewicht and Pearson.Sparrows$Flügellänge
t = 13.463, df = 114, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7013866 0.8451074
sample estimates:
     cor 
0.783512 
```

Hide

```
#View(test)
```

Der R-Output in Abbildung gibt den Korrelationskoeffizienten sowie den p-Wert (Signifikanz) wieder. Es wird ersichtlich, dass ein Zusammenhang vorliegt zwischen das Gewicht und die Flügellänge **(r = .783, p = .000, n = 116)**. Da r einen positiven Wert aufweist, kann von einem positiven linearen und signifikanter Zusammenhang zwischen das Gewicht und die Flügellänge ausgegangen werden. Das bedeutet: Je größer das Gewicht, desto länger ist die Flügellänge, oder je länger die Flügellänge, desto größer ist das Gewicht.

8) Berechnung des Bestimmtheitsmasses

$$Bestimmtheitsmass : r^2 \cdot 100 = 0.783512^2 \cdot 100 = 61.39$$

Hide

```
SparrowsRbestimmt <- SparrowTest$estimate^2*100
sprintf("Das Bestimmtheitsmaß liegt bei %.2f %%.", SparrowsRbestimmt)
```

```
[1] "Das Bestimmtheitsmaß liegt bei 61.39 %."
```

Wird dieser Wert mit 100 multipliziert, so ergibt sich ein Prozentwert. Dieser gibt an, welcher Anteil der Varianz in beiden Variablen durch gemeinsame Varianzanteile determiniert wird. Für das vorliegende Beispiel beträgt der Anteil der gemeinsamen Varianz 61.39%.

9) Berechnung der Effektstärke

Um die Bedeutsamkeit eines Ergebnisses zu beurteilen, werden Effektstärken berechnet. Im Beispiel ist die Korrelation der beiden Variablen signifikant, doch es stellt sich die Frage, ob der Zusammenhang gross genug ist, um ihn als bedeutend einzustufen. Der Korrelationskoeffizient r von Bravais-Pearson stellt selbst ein Mass für die Effektstärke dar.

Hide

```
sprintf("Die Effektstärke liegt bei %.4f.",SparrowTest$estimate)
```

```
[1] "Die Effektstärke liegt bei 0.7835."
```

Um zu bestimmen, wie gross der gefundene Zusammenhang ist, kann man sich an der Einteilung von Cohen (1992) orientieren:

$$\begin{aligned}
\text{Schwacher Effekt: } 0.10 &< ||r|| < 0.30 \\
\text{Schwacher bis mittlerer Effekt: } 0.30 &= ||r|| \\
\text{Mittlerer Effekt: } 0.30 &< ||r|| < 0.50 \\
\text{Mittlerer bis starker Effekt: } 0.50 &= ||r|| \\
\text{Starker Effekt: } 0.50 &< ||r||
\end{aligned}$$

Damit entspricht eine Effektstärke von 0.7835 einem starken Effekt.

10) Eine Aussage

Das Gewicht und die Flügellänge korrelieren signifikant **(r = .783, p = .000, n = 116)**. Je größer das Gewicht, desto länger ist die Flügellänge, oder je länger die Flügellänge, desto größer ist das Gewicht. 61.39% der Streuung der gemeinsamen Varianz kann durch das Gewicht und die Flügellänge erklärt werden. Dabei handelt es sich nach Cohen (1992) um einen starken Effekt. H0 kann verworfen werden.