QM6:

Schätzgenauigkeit:

Punktschätzer, Standardfehler (SE) und Konfidenzintervall

Die Differenz als Schätzfehler

je größer die Stichprobe, desto näher sind die Stichprobenmittelwerte am wahren Populationsmittelwert

```
> favstats(flights_clean$arr_delay) # Population
 min  Q1 median Q3  max     mean       sd      n missing
 -86 -17     -5 14 1272 6.895377 44.63329 327346       0
```

<- normalerweise unbekannt

	min	Q1	median	Q3	max	mean	sd	n	missing	.row
...1	-41	-13.00	-1.5	28.25	146	22.3	60.26617	10	0	1
...2	-28	-6.75	16.5	43.75	87	21.5	36.05936	10	0	1
...3	-26	-7.75	10.5	33.25	123	19.3	42.52594	10	0	1
...4	-36	-20.00	-13.5	6.50	14	-10.5	18.00154	10	0	1
...5	-29	-14.25	-5.0	4.75	205	13.3	68.33911	10	0	1
...6	-35	-20.50	-8.0	7.00	83	3.3	37.13055	10	0	1

	min	Q1	median	Q3	max	mean	sd	n	missing
...1	-75	-17	-5	14	915	7.04148	44.61456	100000	0
...2	-86	-17	-5	14	1109	6.97302	44.68221	100000	0
...3	-86	-17	-5	14	1272	6.96174	44.85185	100000	0
...4	-74	-17	-5	14	1109	6.91156	44.84390	100000	0
...5	-86	-17	-5	14	1272	6.79285	44.78490	100000	0
...6	-86	-17	-5	14	1127	7.06829	44.74456	100000	0

Beobachtung:

- Kennzahlen ‚in der Mitte' unserer Stichprobe sind bei kleineren Stichproben bessere Schätzer (Median, iqr)
- Aggregierende Kennzahlen haben die Tendenz zur Mitte, der MW (mean) ist deshalb der häufigste Schätzer

Die Differenz als Schätzfehler

je größer die Stichprobe, desto näher sind die Stichprobenmittelwerte am wahren Populationsmittelwert

```
> favstats(flights_clean$arr_delay) # Population
```

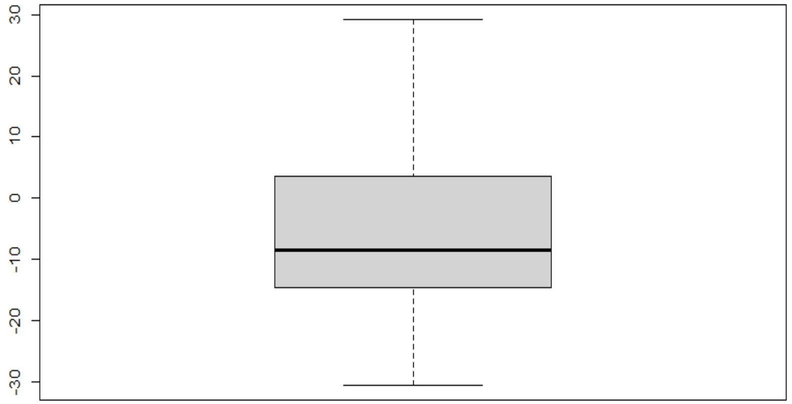
<- normalerweise unbekannt

min	Q1	median	Q3	max	mean	sd	n	missing
-86	-17	-5	14	1272	6.895377	44.63329	327346	0

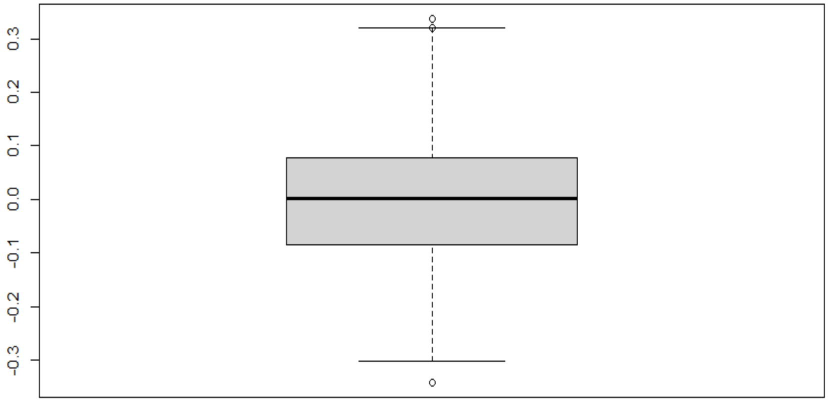
	min	Q1	median	Q3	max	mean	sd	n	missing	.row
...1	-41	-13.00	-1.5	28.25	146	22.3	60.26617	10	0	1
...2	-28	-6.75	16.5	43.75	87	21.5	36.05936	10	0	1
...3	-26	-7.75	10.5	33.25	123	19.3	42.52594	10	0	1
...4	-36	-20.00	-13.5	6.50	14	-10.5	18.00154	10	0	1
...5	-29	-14.25	-5.0	4.75	205	13.3	68.33911	10	0	1
...6	-35	-20.50	-8.0	7.00	83	3.3	37.13055	10	0	1

	min	Q1	median	Q3	max	mean	sd	n	missing
...1	-75	-17	-5	14	915	7.04148	44.61456	100000	0
...2	-86	-17	-5	14	1109	6.97302	44.68221	100000	0
...3	-86	-17	-5	14	1272	6.96174	44.85185	100000	0
...4	-74	-17	-5	14	1109	6.91156	44.84390	100000	0
...5	-86	-17	-5	14	1272	6.79285	44.78490	100000	0
...6	-86	-17	-5	14	1127	7.06829	44.74456	100000	0

starke Streuung (Bsp.)



uns interessiert das 95%- Intervall der Schwankung des means

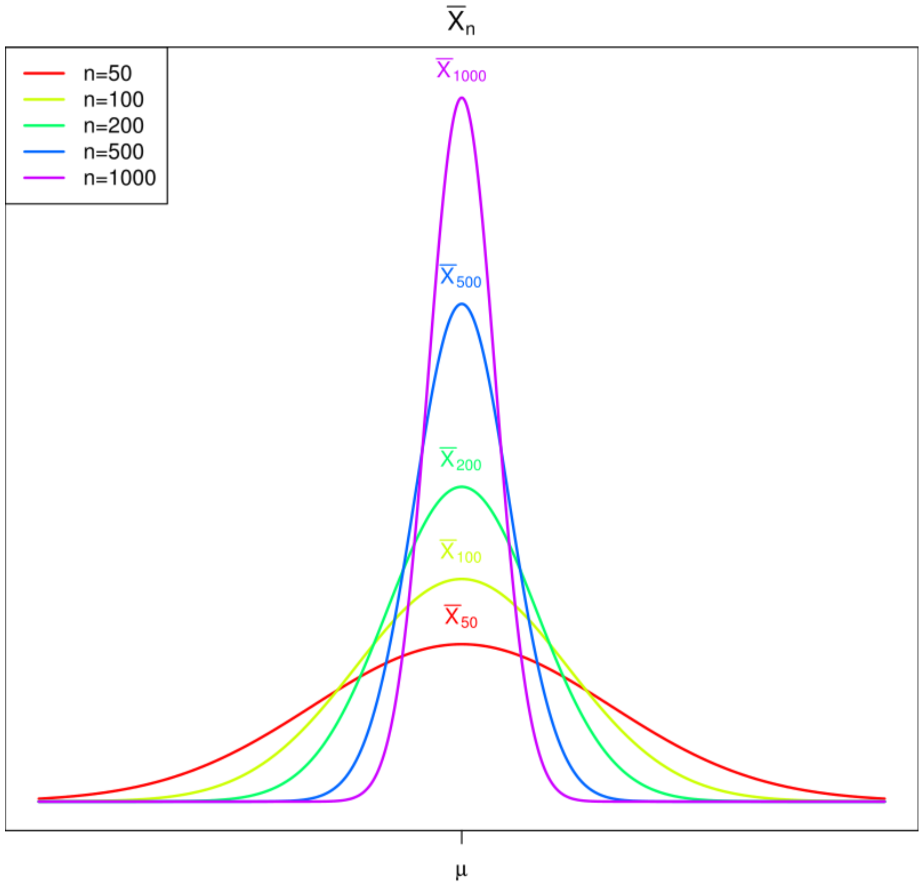


weniger Streuung

Fig.: Verteilung der Differenz zwischen population_mean und Stichprobenmittelwerte

Verteilung von Stichprobenmittelwerten

z.B. ‚flights'



μ = 6.98 min
(z.B. wie in flights)

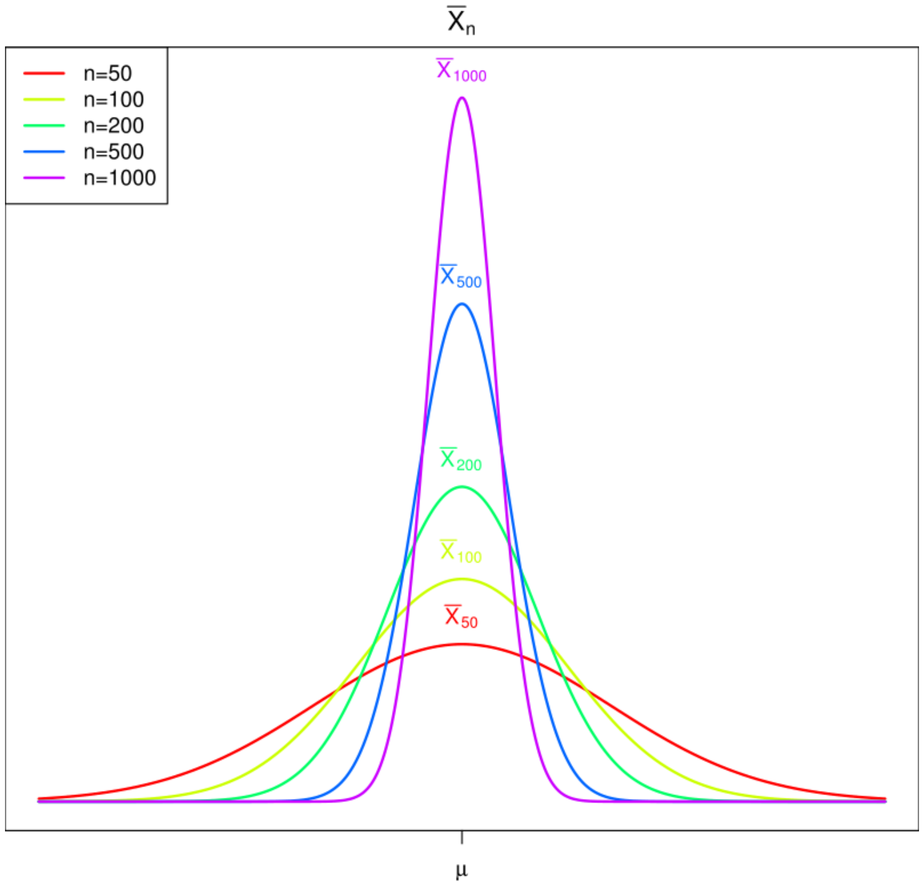
- Wir können nie wissen wie nah unser MW aus dem sample am wahren μ liegt, da μ in der Regel gesucht wird.

- Je größer die Stichprobe, desto schmaler ist die Verteilung der Stichprobenmittelwerte. Dadurch wird der Populationsmittelwert präziser geschätzt. Die Breite der Stichprobenverteilung wird durch die Streuung/ sd bestimmt.

Quelle: https://de.wikipedia.org/wiki/Erwartungstreue#/media/Datei: Distribution_of_mean_estimator.svg

Verteilung von Stichprobenmittelwerten

z.B. ‚flights'



μ = 6.98 min
(z.B. wie in flights)

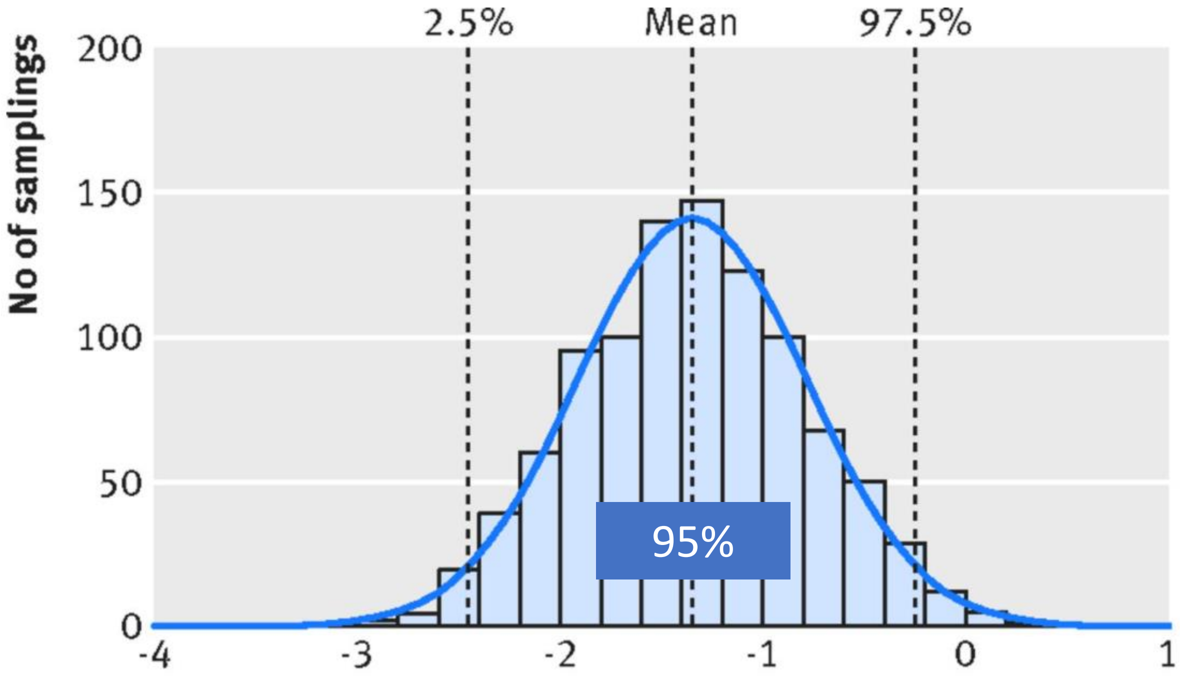
- **Wir benötigen demnach nur die sd der Stichprobenmittelwerte um die Genauigkeit unserer Punktschätzung anzugeben**

- Vorgehen:

1) Standardfehler berechnen (theoretisch oder praktisch)
2) 95%- Intervall berechnen

Quelle: https://de.wikipedia.org/wiki/Erwartungstreue#/media/Datei: Distribution_of_mean_estimator.svg

Punktschätzer und Konfidenzintervall

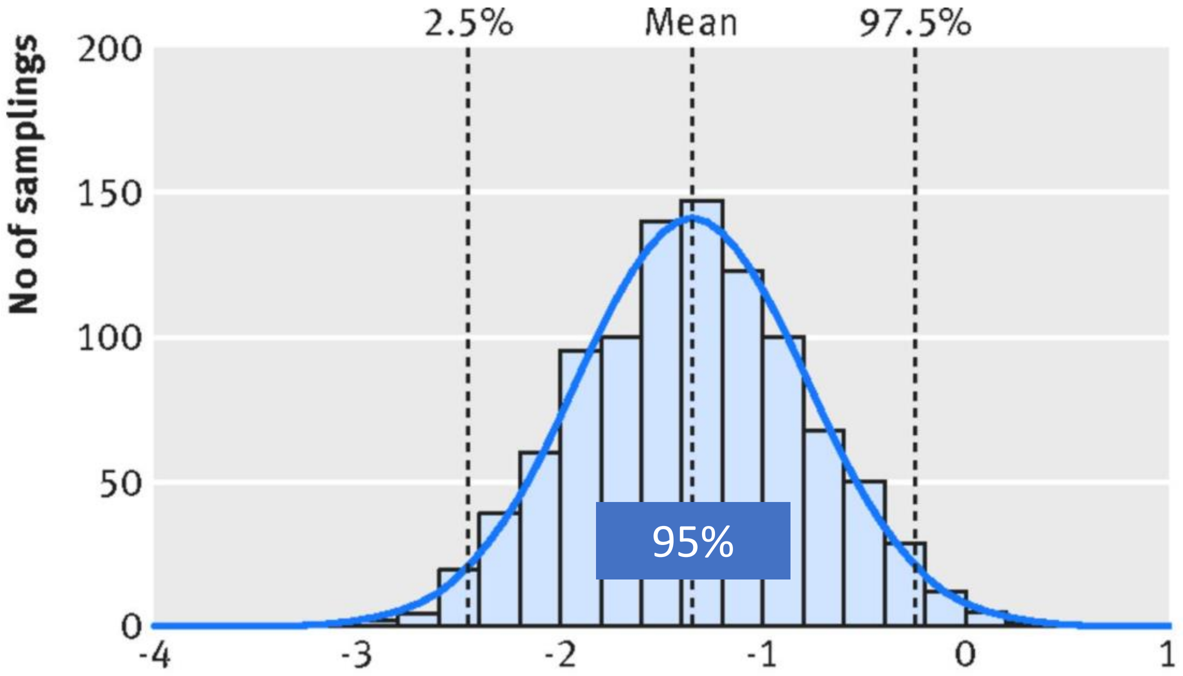


- Die Schätzgenauigkeit kann man oft besser durch eine untere und eine obere Grenze einordnen.

- Der Stichprobenmittelwert gilt dabei als **Punktschätzer.**
- Das Intervall nennt sich Vertrauensintervall, oder häufiger: **Konfidenzintervall**.

Figure: https://www.bmj.com/content/350/bmj.h2622

Punktschätzer und Konfidenzintervall



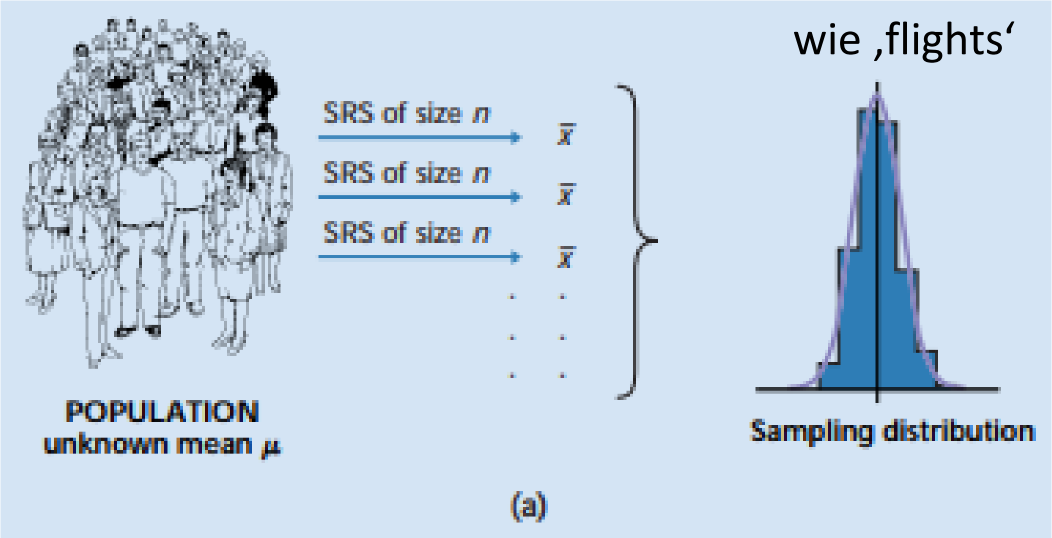
$$x_u = \bar{x} - z \cdot \frac{\sigma}{\sqrt{n}} \qquad x_o = \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}}$$

- Die Schätzgenauigkeit kann man oft besser durch eine untere und eine obere Grenze einordnen.

- Der Stichprobenmittelwert gilt dabei als **Punktschätzer.**
- Das Intervall nennt sich Vertrauensintervall, oder häufiger: **Konfidenzintervall**.

Mittels des Standardfehlers SE wird häufig die 95%-Umgebung um den Punktschätzer angegeben.

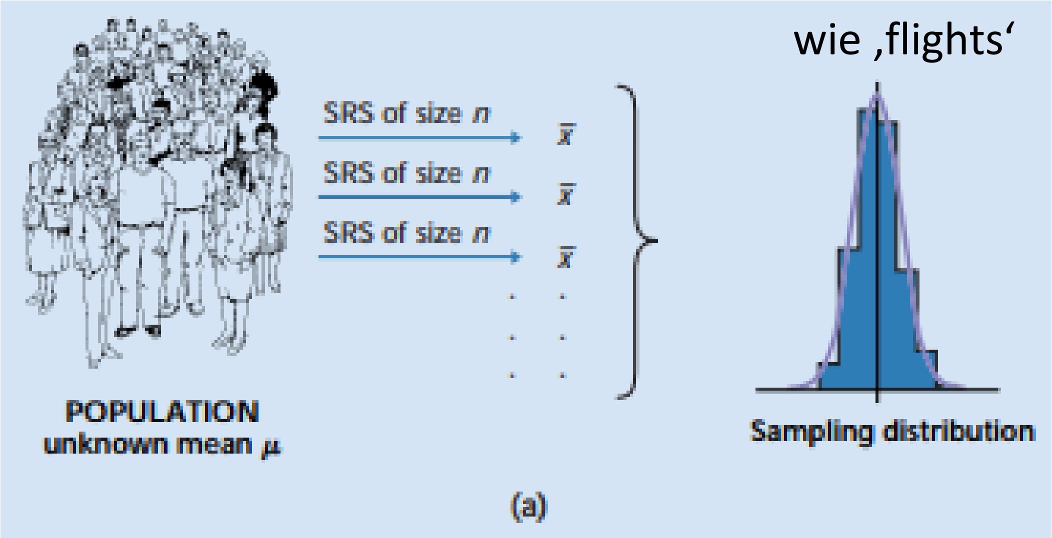
Herangehensweise: empirische oder theoretische **Berechnung des Standardfehlers** (standard error: SE)



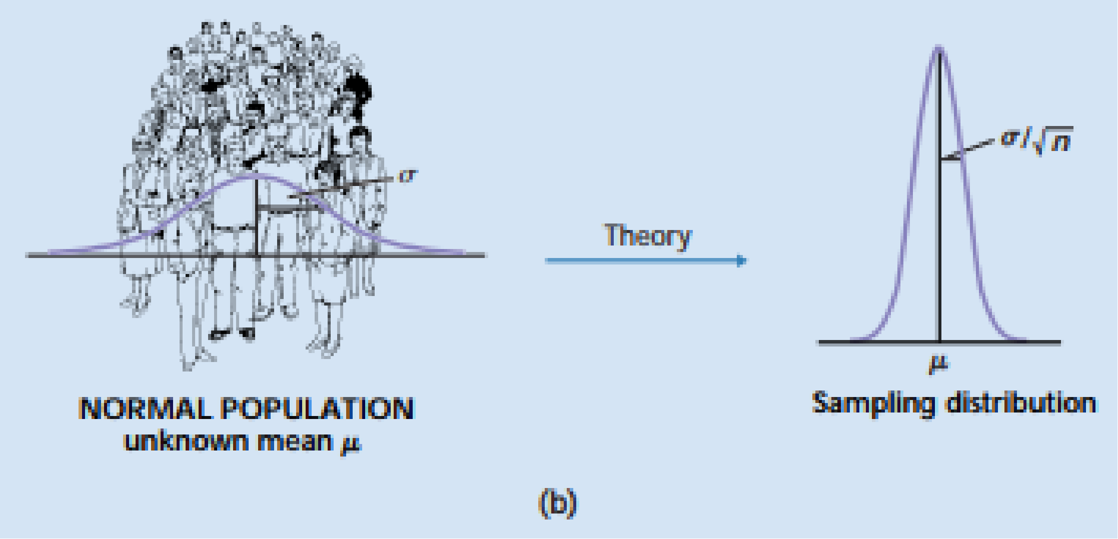
Praktische Berechnung:
ziehe samples
berechne means
berechne sd dieser means (=Standardfehler)

Quelle: http://statweb.stanford.edu/~tibs/stat315a/Supplements/bootstrap.pdf

Herangehensweise: empirische oder theoretische **Berechnung des Standardfehlers** (standard error: SE)



Praktische Berechnung:
ziehe samples
berechne means
berechne sd dieser means (=Standardfehler)



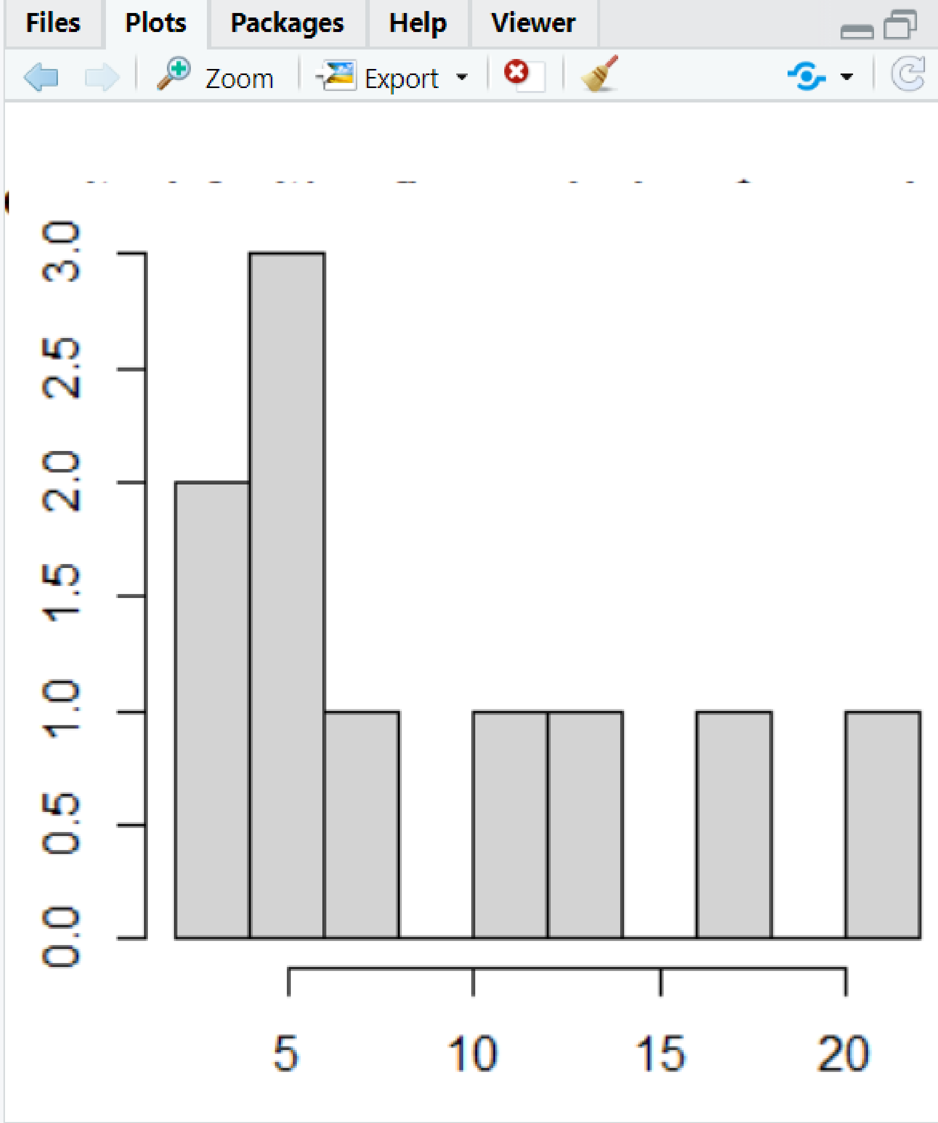
Theoretische Berechnung:
ziehe ein sample
berechne sd
teile sd durch Wurzel aus n (=Standardfehler)

$$\boxed{SE_{\bar{x}} = \frac{s}{\sqrt{n}}}$$

Quelle: http://statweb.stanford.edu/~tibs/stat315a/Supplements/bootstrap.pdf

Kleine(re) Stichprobe (n=100)

```
Console  Terminal  Jobs
~/P2/
> # erzeuge x samples der Große size und zeige favstats
> fl_sample_loop <- do(10) *
+   favstats(sample(flights_clean$arr_delay, size = 100))
>
> hist(fl_sample_loop$mean, breaks=10) #Histogramm
> plot(density(fl_sample_loop$mean,bw=1)) #Glättung Dichteplot
>
> mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
[1] 6.895377
> abline(v = mean(flights_clean$arr_delay), col = 4) #mean population
>
> m_<-mean(fl_sample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 9.044
> abline(v = mean(fl_sample_loop$mean), col = 2) #mean StichprobenMW
>
> ## 2.5% und 97.5% quantile (mittleren 95%)
> q<-quantile(fl_sample_loop$mean, prob=c(0.025,0.975));q
    2.5%    97.5%
 3.79700 19.64875
> abline(v = q, col = 2, lty = 2)
>
> sd(fl_sample_loop$mean) #Standardfehler, empirisch
[1] 5.991919
>
> # Berechnung des theoretischen SE = sd/Wurzel n
> se_t<-sd(sample(flights_clean$arr_delay, size = 100))/sqrt(100);se_t
[1] 3.106828
```



Kleine(re) Stichprobe (n=100)

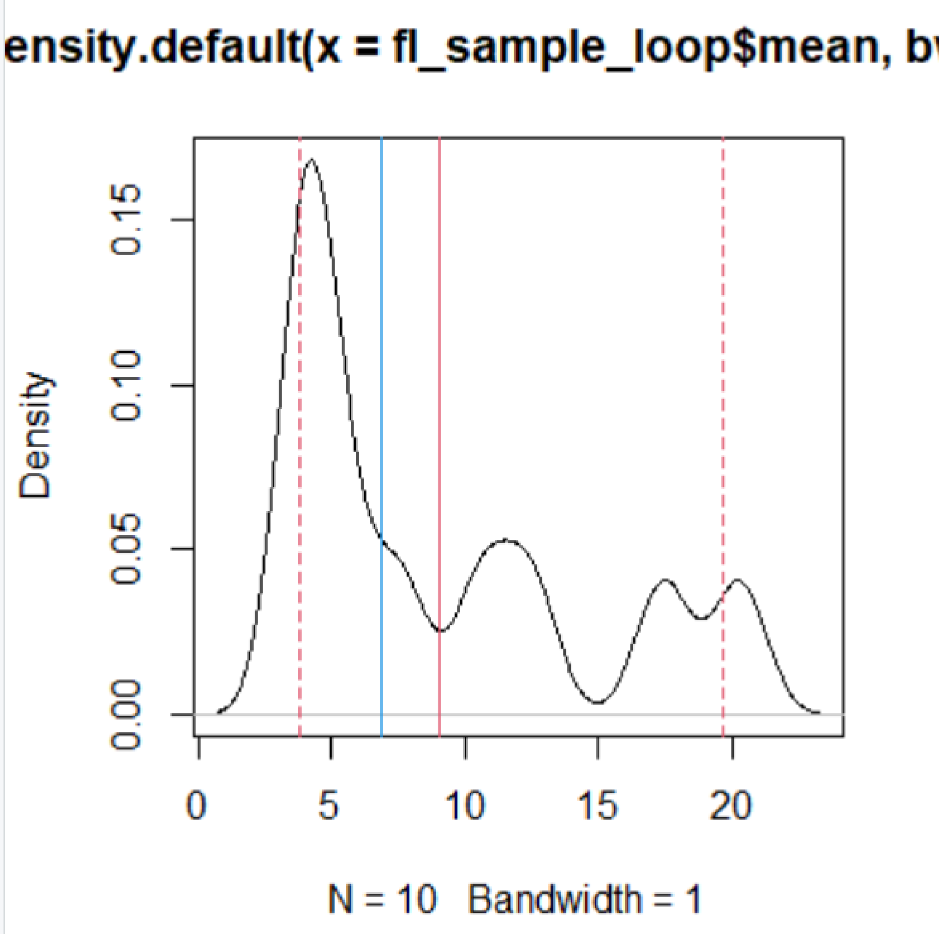
Console Terminal Jobs

~/P2/

```
> # erzeuge x samples der Größe size und zeige favstats
> fl_sample_loop <- do(10) *
+   favstats(sample(flights_clean$arr_delay, size = 100))
>
> hist(fl_sample_loop$mean, breaks=10) #Histogramm
> plot(density(fl_sample_loop$mean,bw=1)) #Glättung Dichteplot
>
> mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
[1] 6.895377
> abline(v = mean(flights_clean$arr_delay), col = 4) #mean population
>
> m_<-mean(fl_sample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 9.044
> abline(v = mean(fl_sample_loop$mean), col = 2) #mean StichprobenMW
>
> ## 2.5% und 97.5% quantile (mittleren 95%)
> q<-quantile(fl_sample_loop$mean, prob=c(0.025,0.975));q
    2.5%    97.5%
 3.79700 19.64875
> abline(v = q, col = 2, lty = 2)
>
> sd(fl_sample_loop$mean) #Standardfehler, empirisch
[1] 5.991919
>
> # Berechnung des theoretischen SE = sd/Wurzel n
> se_t<-sd(sample(flights_clean$arr_delay, size = 100))/sqrt(100);se_t
[1] 3.106828
```

Files Plots Packages Help Viewer

Zoom Export

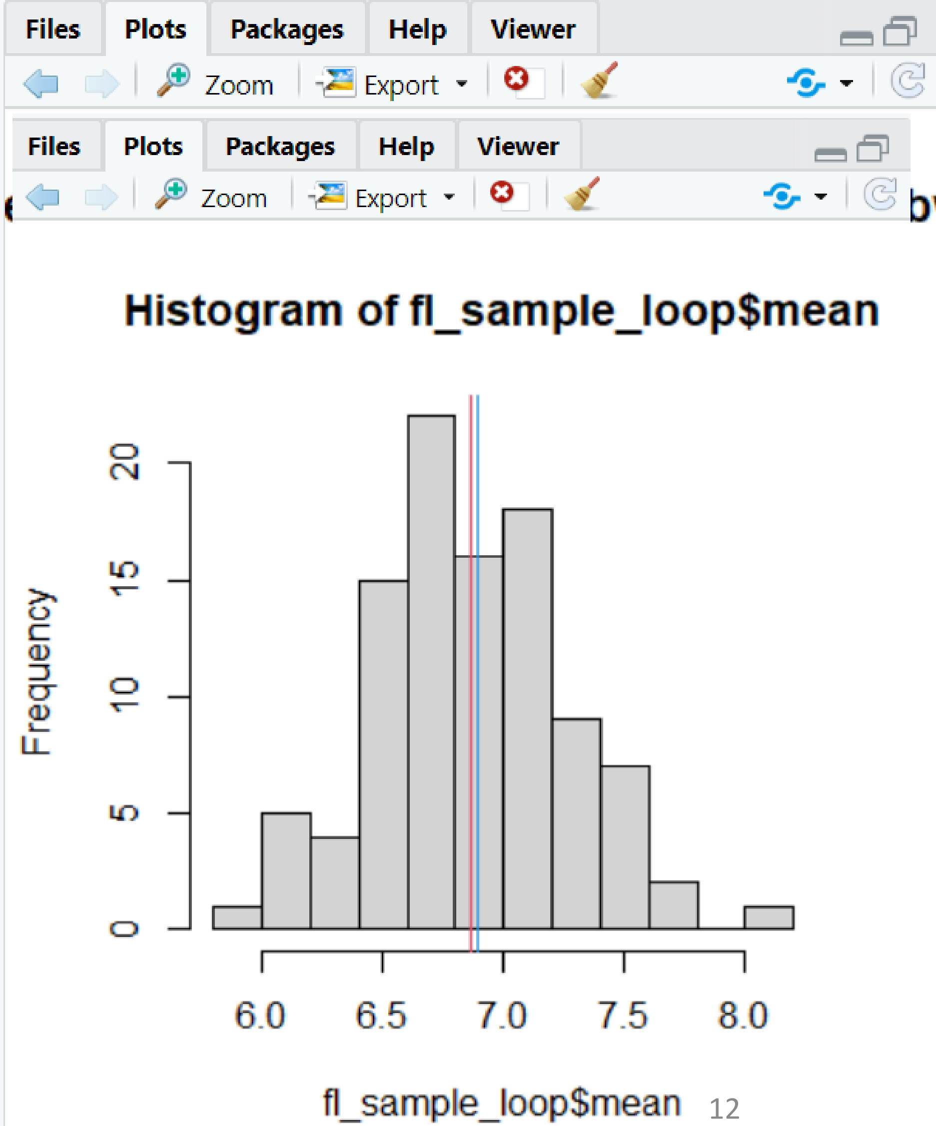


Große Stichprobe (n=10.000)

Console   Terminal   Jobs

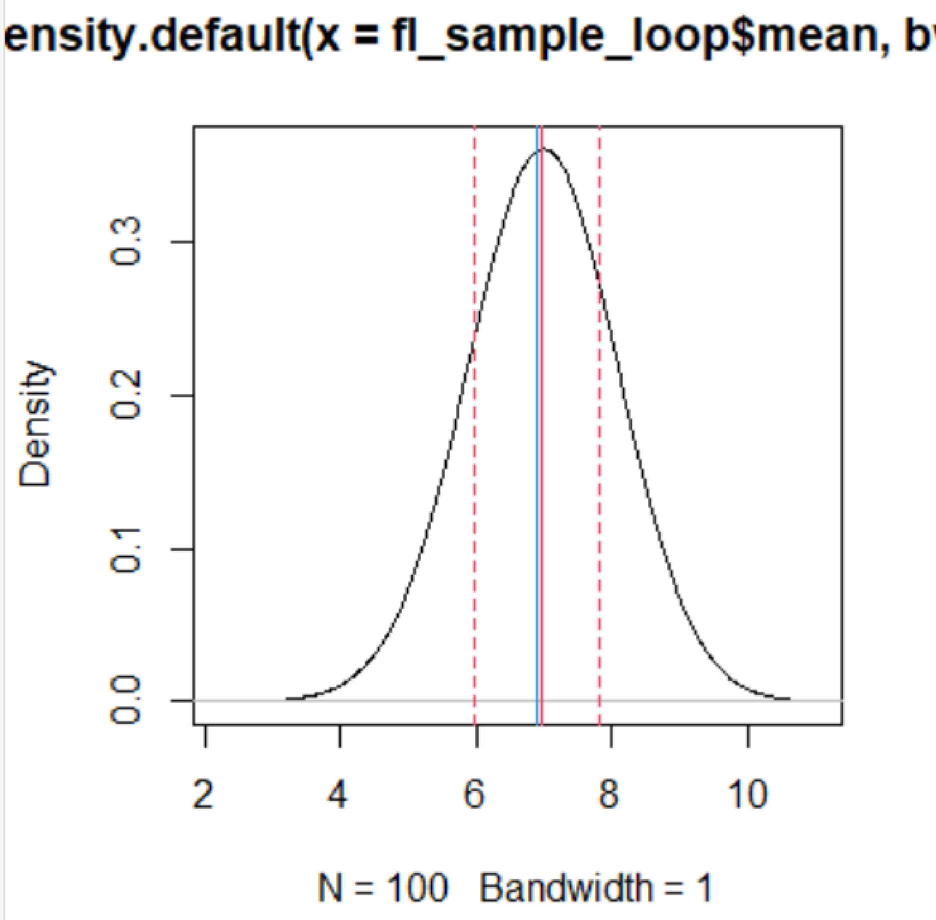
~/P2/

```
> # erzeuge x samples der Große size und zeige favstats
> fl_sample_loop <- do(100) *
+   favstats(sample(flights_clean$arr_delay, size = 10000))
>
> hist(fl_sample_loop$mean, breaks=10) #Histogramm
> plot(density(fl_sample_loop$mean,bw=1)) #Glättung Dichteplot
>
> mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
[1] 6.895377
> abline(v = mean(flights_clean$arr_delay), col = 4) #mean population
>
> m_<-mean(fl_sample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 6.967949
> abline(v = mean(fl_sample_loop$mean), col = 2) #mean StichprobenMW
>
> ## 2.5% und 97.5% quantile (mittleren 95%)
> q<-quantile(fl_sample_loop$mean, prob=c(0.025,0.975));q
    2.5%    97.5%
5.965695 7.818585
> abline(v = q, col = 2, lty = 2)
>
> sd(fl_sample_loop$mean) #Standardfehler, empirisch
[1] 0.4835039
>
> # Berechnung des theoretischen SE = sd/Wurzel n
> se_t<-sd(sample(flights_clean$arr_delay, size = 10000))/sqrt(10000);se_t
[1] 0.4513942
```



Große Stichprobe (n=10.000)

```
> # erzeuge x samples der Große size und zeige favstats
> fl_sample_loop <- do(100) *
+   favstats(sample(flights_clean$arr_delay, size = 10000))
>
> hist(fl_sample_loop$mean, breaks=10) #Histogramm
> plot(density(fl_sample_loop$mean,bw=1)) #Glättung Dichteplot
>
> mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
[1] 6.895377
> abline(v = mean(flights_clean$arr_delay), col = 4) #mean population
>
> m_<-mean(fl_sample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 6.967949
> abline(v = mean(fl_sample_loop$mean), col = 2) #mean StichprobenMW
>
> ## 2.5% und 97.5% quantile (mittleren 95%)
> q<-quantile(fl_sample_loop$mean, prob=c(0.025,0.975));q
    2.5%    97.5%
5.965695 7.818585
> abline(v = q, col = 2, lty = 2)
>
> sd(fl_sample_loop$mean) #Standardfehler, empirisch
[1] 0.4835039
>
> # Berechnung des theoretischen SE = sd/wurzel n
> se_t<-sd(sample(flights_clean$arr_delay, size = 10000))/sqrt(10000);se_t
[1] 0.4513942
```



QM7

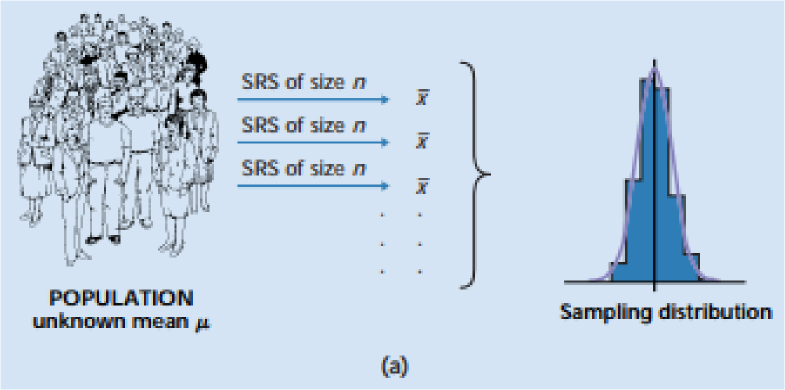
Einführung simulationsbasierte Inferenz (SBI):

Bootstrap (resample) Konfidenzintervall

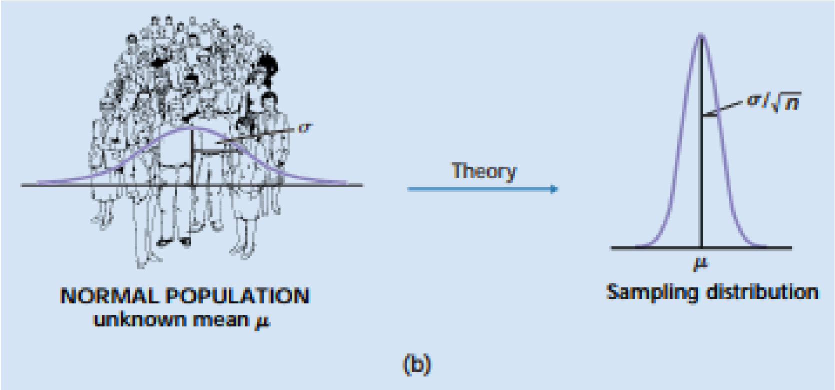
Schätzen vs. Hypothesen testen

Ein Ausreißertest

Herangehensweise: empirische oder theoretische **Berechnung des Standardfehlers** (standard error: SE)

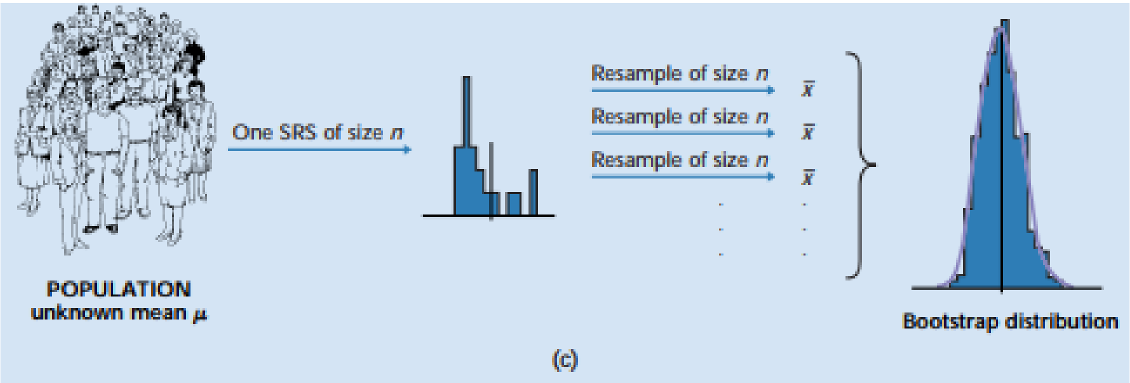


Praktische Berechnung:
ziehe samples
berechne means
berechne sd dieser means (=Standardfehler)



Theoretische Berechnung:
ziehe ein sample
berechne sd
teile sd durch Wurzel aus n (=Standardfehler)

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$



<- QM7: Resampling

Quelle: http://statweb.stanford.edu/~tibs/stat315a/Supplements/bootstrap.pdf

samplen vs. resamplen

```
### Resamplen: mehrfaches Ziehen aus EINER Stichprobe #######++++++++++

# Schritt 1: ziehe EIN sample

fl_sample<- sample(flights_clean$arr_delay, size = 3) #2500
fl_sample
mean(fl_sample)

# Schritt 2: resample: Ziehen mit Zurücklegen aus unserem (einzigen) sampl
fl_resample <- resample(fl_sample, size = 3)
fl_resample
```

```
> fl_sample<- sample(flights_clean$arr_delay, size = 3) #2500
> fl_sample
[1] -12 -29 128
> fl_resample <- resample(fl_sample, size = 3)
> fl_resample
[1] 128 -29 128
>
```

Mehrfache können gezogen werden

Ziehe ein sample der Größe = 3

Ziehe daraus ein sample der Größe 3. Zwischendurch zurücklegen

Ziehen ohne Zurücklegen (samplen)

Ziehen mit Zurücklegen (resamplen)

Wiederholtes resamplen (bootstrap)

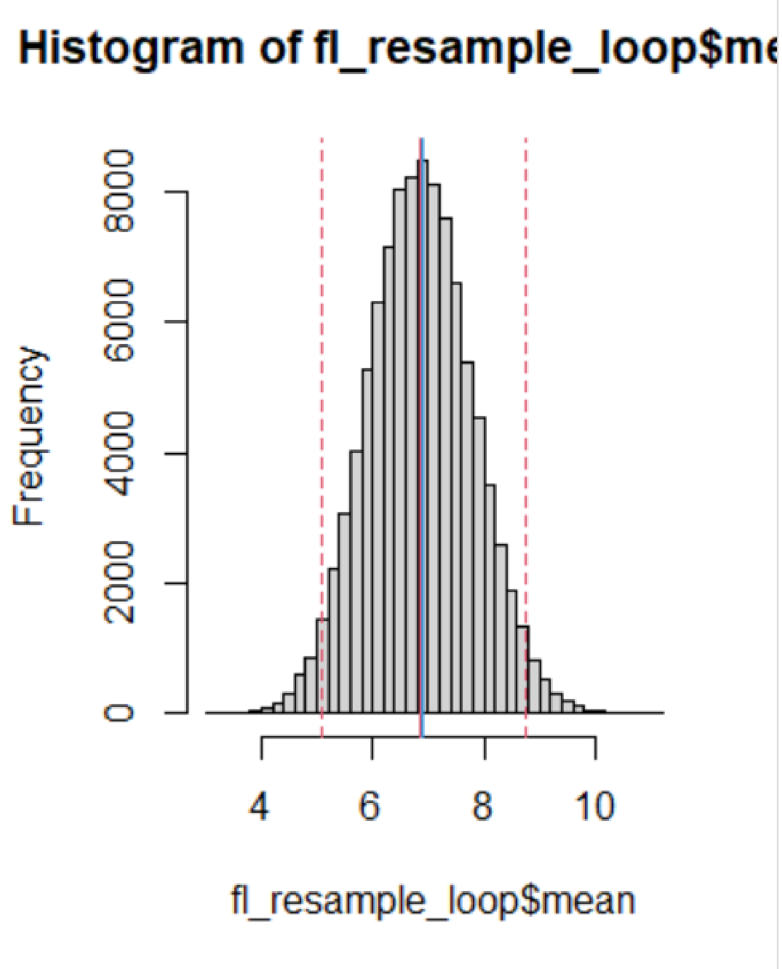
```r
############################################################
# als loop
# etwas Größeres sample ziehen
fl_sample<- sample(flights_clean$arr_delay, size = 2500)
mean(fl_sample) #an dieser Stelle läßt sich nur der mean des samples berechnen

#aus diesem sample resamplen um die Breite des Konf-intervalls zu schätzen
set.seed(2021)

fl_resample_loop <- do (100000)*mean(resample(fl_sample)) #for-loop
mean(fl_resample_loop$mean) #mean der means (grand mean)
sd(fl_resample_loop$mean) #Standardfehler
```

```
> mean(fl_sample)
[1] 6.8636
> set.seed(2021)
>
> fl_resample_loop <- do (100000)*mean(resample(fl_sample))
> mean(fl_resample_loop$mean)
[1] 6.862183
> sd(fl_resample_loop$mean)
[1] 0.9409522
>
```

Wiederholtes resamplen (bootstrap)

```r
hist(fl_resample_loop$mean, breaks=50) #Histogramm
plot(density(fl_resample_loop$mean,bw=1)) #Glättung Dichteplot

mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
abline(v = mean(flights_clean$arr_delay), col = 4) #mean population

m_<-mean(fl_resample_loop$mean);m_ # MW der Mittelwerte der samples
abline(v = mean(fl_resample_loop$mean), col = 2) #mean StichprobenMW

## 2.5% und 97.5% quantile (mittleren 95%)
q<-quantile(fl_resample_loop$mean, prob=c(0.025,0.975));q
abline(v = q, col = 2, lty = 2)

paste("Der Punktschätzer beträgt",mean(fl_resample_loop$mean),". Der wahre unbek
      in 95% der Stichproben zwischen ",q[1]," und",q[2])
```

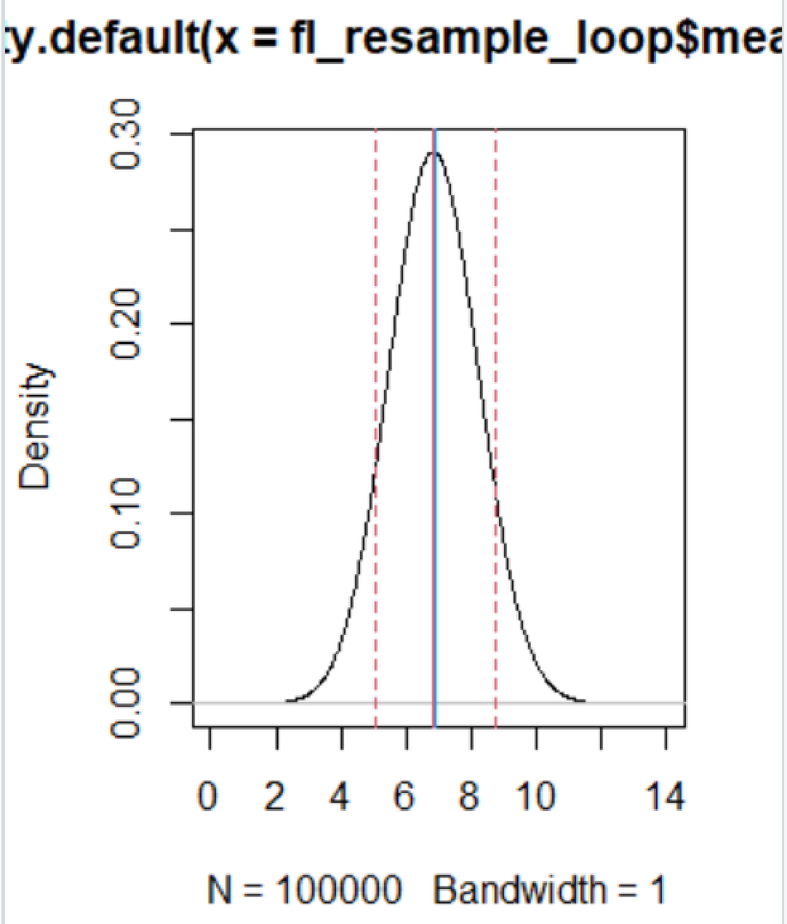
```
> mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
[1] 6.895377
> m_<-mean(fl_resample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 6.862183
> paste("Der Punktschätzer beträgt",mean(fl_resample_loop$mean),". Der wahre unbekannt
e Populationsparameter µ liegt
+       in 95% der Stichproben zwischen ",q[1]," und",q[2])
[1] "Der Punktschätzer beträgt 6.862183416 . Der wahre unbekannte Populationsparameter
 µ liegt\n       in 95% der Stichproben zwischen  5.06519  und 8.74161"
```



Wiederholtes resamplen (bootstrap)

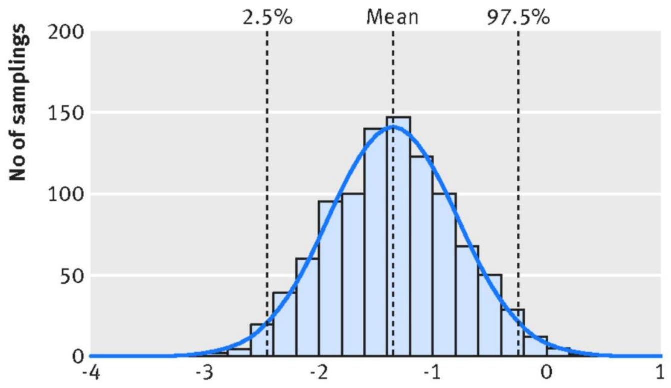
```r
hist(fl_resample_loop$mean, breaks=50) #Histogramm
plot(density(fl_resample_loop$mean,bw=1)) #Glättung Dichteplot

mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
abline(v = mean(flights_clean$arr_delay), col = 4) #mean population

m_<-mean(fl_resample_loop$mean);m_ # MW der Mittelwerte der samples
abline(v = mean(fl_resample_loop$mean), col = 2) #mean StichprobenMW

## 2.5% und 97.5% quantile (mittleren 95%)
q<-quantile(fl_resample_loop$mean, prob=c(0.025,0.975));q
abline(v = q, col = 2, lty = 2)

paste("Der Punktschätzer beträgt",mean(fl_resample_loop$mean),". Der wahre unbek
      in 95% der Stichproben zwischen ",q[1]," und",q[2])
```

```
> mu<-mean(flights_clean$arr_delay);mu # Mittelwert der population
[1] 6.895377
> m_<-mean(fl_resample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 6.862183
> paste("Der Punktschätzer beträgt",mean(fl_resample_loop$mean),". Der wahre unbekannt
e Populationsparameter µ liegt
+       in 95% der Stichproben zwischen ",q[1]," und",q[2])
[1] "Der Punktschätzer beträgt 6.862183416 . Der wahre unbekannte Populationsparameter
 µ liegt\n       in 95% der Stichproben zwischen  5.06519  und 8.74161"
```



Schätzen vs. Hypothesen prüfen

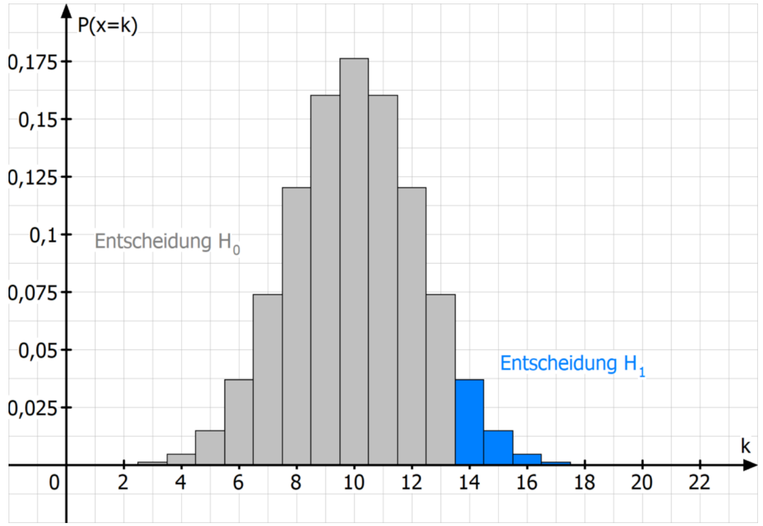
- Estimation (Schätzung)-> Confidence interval
  - In welchem Intervall liegt der unbekannte Wert?



- Decision (Entscheidung) -> Hypothesis test

Forschungsfragen beantworten (z.B.):
Haben die Flüge am JFK eine größere Verspätung?
(angenommen die Gesamtpopulation würde nicht vorliegen)

Hypothesenpaar formulieren
H_1: Die Flüge am JFK haben eine größere Verspätung
H_0: Die Flüge am JFK haben **k**eine größere Verspätung
(in der Nullhypothese wird von Gleichheit ausgegangen)

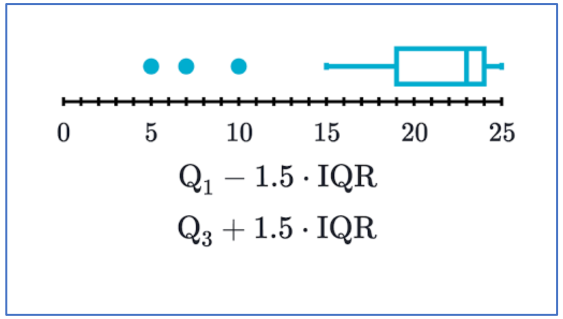


Ein Ausreißertest

Ausreißertest nach IQR

```r
# Ausreißertest: Es liegt eine Messung von Körpergrößen vor

height <- c(1.78,1.69,1.81,1.83,1.80,1.40) #c für Vektor
favstats(height) #Kennzahlen anzeigen
boxplot(height) #visualisieren
# der Wert 1.40 m wird hier als Ausreißer (outlier) gesehen, siehe
# Punkt unten unterhalb des boxplots
#iqr_test (nach boxplot), untere und obere Grenze
quantile(height) # wir brauchen Q1 und Q3

iqr_test_u <- function(x){
  quantile(x)[2]-1.5*(iqr(x))    #Q1-1.5*IQR
}
iqr_test_o <- function(x){
  quantile(x)[4]+1.5*(iqr(x))    #Q3+1.5*IQR
}
u<-iqr_test_u(height);u
o<-iqr_test_o(height);o
```

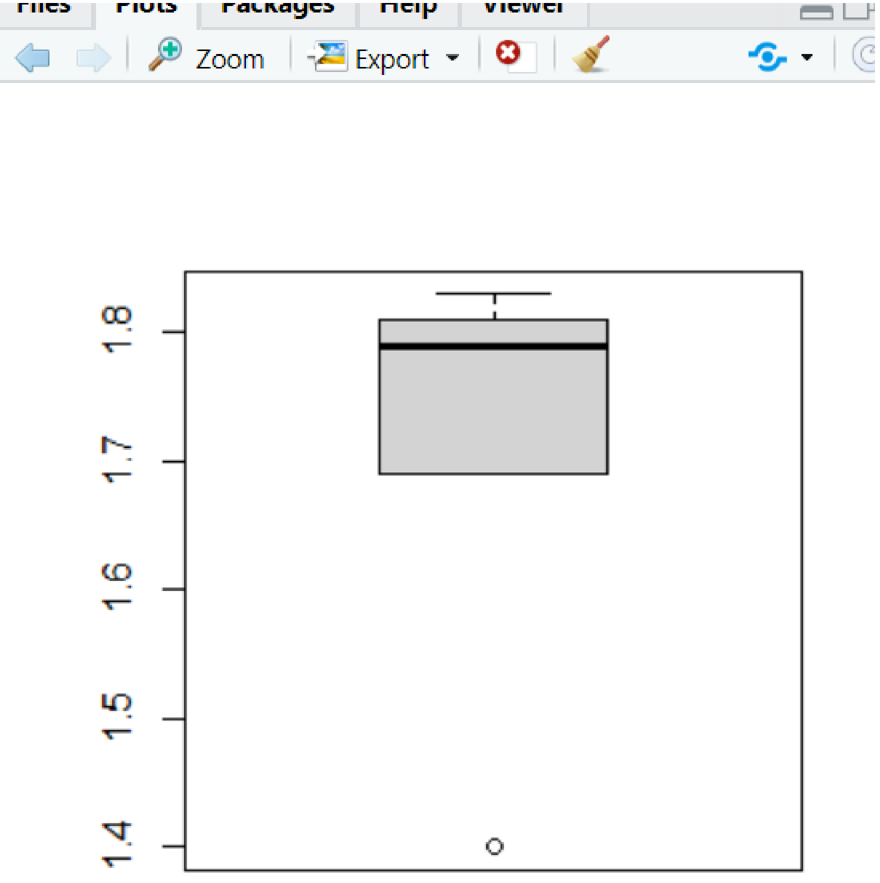
20:24 (Top Level) R Script



Console Terminal Jobs

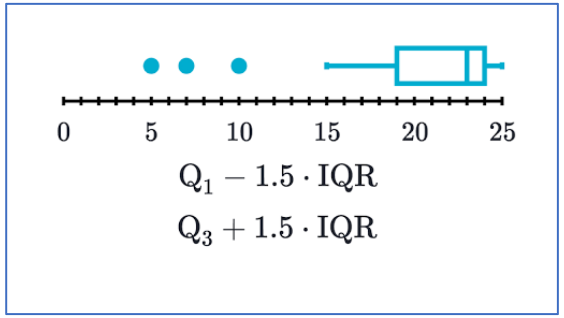
~/P2/

```
 min     Q1 median     Q3  max     mean        sd n missing
 1.4 1.7125   1.79 1.8075 1.83 1.718333 0.1633911 6       0
> quantile(height)
    0%    25%    50%    75%   100% 
1.4000 1.7125 1.7900 1.8075 1.8300 
```

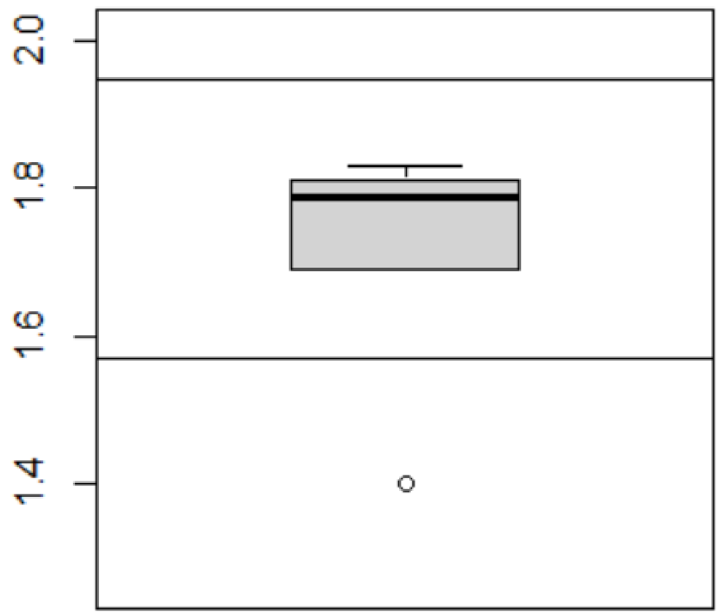


Ausreißertest nach IQR

```r
#iqr_test (nach boxplot), untere und obere Grenze
quantile(height) # wir brauchen Q1 und Q3

iqr_test_u <- function(x){
  quantile(x)[2]-1.5*(iqr(x))   #Q1-1.5*IQR
}
iqr_test_o <- function(x){
  quantile(x)[4]+1.5*(iqr(x))   #Q3+1.5*IQR
}
u<-iqr_test_u(height);u
o<-iqr_test_o(height);o

boxplot(height,ylim=c(0.9*min(height),1.1*max(height)))
abline(h=u); abline(h=o)
```



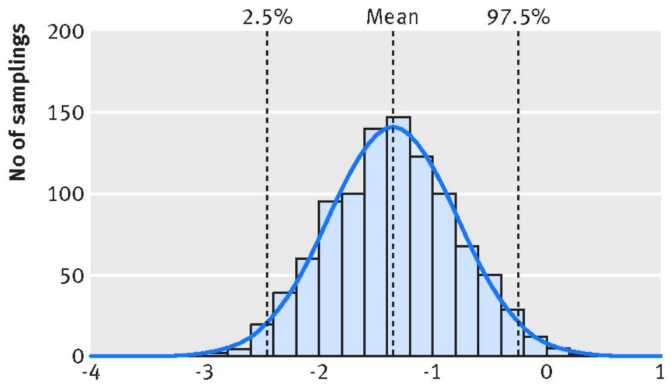

```
> u<-iqr_test_u(height);u
 25%
1.57
> o<-iqr_test_o(height);o
 75%
1.95
> boxplot(height,ylim=c(0.9*min(height),1.1*max(height)))
> abline(h=u); abline(h=o)
```



Simulationsbasierte Inferenz (SBI):
ein Hypothesentest

Schätzen vs. Hypothesen prüfen

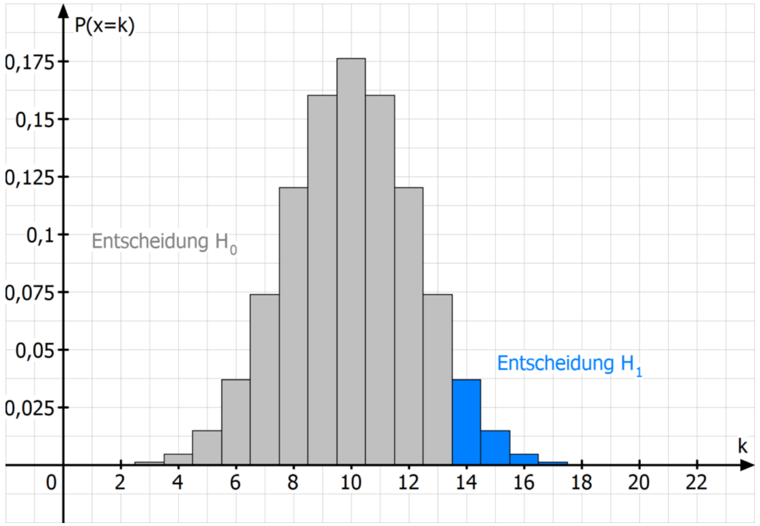
- Estimation (Schätzung)-> Confidence interval
  - In welchem Intervall liegt der unbekannte Wert?



- Decision (Entscheidung) -> Hypothesis test

Forschungsfragen beantworten (z.B.):
Haben die Flüge am JFK eine *kleinere* Verspätung?
(angenommen die Gesamtpopulation würde nicht vorliegen)

Hypothesenpaar formulieren
H_1: Die Flüge am JFK haben eine kleinere Verspätung
H_0: Die Flüge am JFK haben **k**eine kleinere Verspätung
(in der Nullhypothese wird von Gleichheit/ keine Auffälligkeit ausgegangen)



Hypothesen testen: formaler Aufbau

H_1: Es gibt einen Unterschied (etwas ist größer/kleiner etc.)

H_0: Es gibt **keinen** Unterschied

Beispiele:

Frage: Sind die Ankünfte am JFK pünktlicher?

H_1: Die Flüge am JFK haben weniger Verspätung

H_0: Die Flüge am JFK haben **nicht weniger/ gleiche** Verspätung

Frage: Hilft der Impfstoff?

H_1: Der Impfstoff erhöht die Immunität

H_0: Der Impfstoff erhöht **nicht** die Immunität

Hilft: Ist die Münze fair?

H_1: Die Münze ist gezinkt

H_0: Die Münze ist **nicht** gezinkt

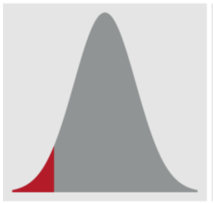
μ: unbekannter **Mittelwert** in der Gesamtheit(Population)

(π: unbekannter **Anteil** in der Gesamtheit(Population))

$$H_1: \mu_{JFK} < \mu_{non\_JFK}$$

$$H_0: \mu_{JFK} \geq \mu_{non\_JFK}$$

A



Gerichtet, "kleiner als"

$$H_1: \mu_{impf} > \mu_{non\_impf}$$

$$H_0: \mu_{impf} \leq \mu_{non\_impf}$$

B

Gerichtet, "größer als"

$$H_1: \mu_{Kopf} \neq \mu_{Zahl}$$

$$H_0: \mu_{Kopf} = \mu_{Zahl}$$

C

Ungerichte

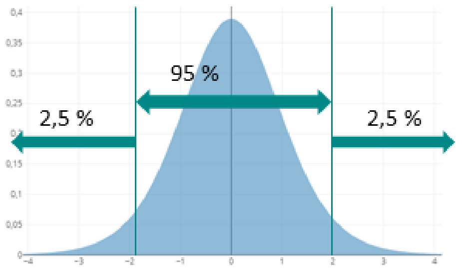
gerichtete (einseitige) vs. ungerichtete (zweiseitige) Hypothesen

Ein Hypothesentest hat das Ziel eine wissenschaftliche Forschungsfrage mit empirischen Methoden zu beantworten. Dazu benötigen wir Stichproben, die mit Unsicherheit behaftet sind (z.B. durch einen zu kleinen Stichprobenumfang oder fehlender Repräsentativität).

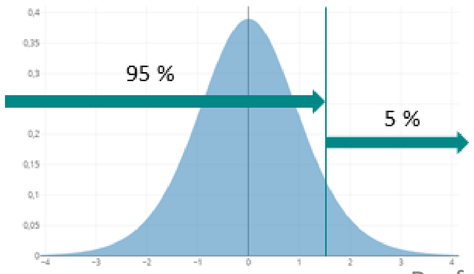
Wir prüfen z.B. ob ein Mittelwert noch mit der Schwankung des Zufallsprozesses (Streuung) zu erklären ist oder nicht. Die Verteilung und Streuung der Stichprobenmittelwerte (Standardfehler) zeigt welche Werte unter der Nullhypothese (kein Unterschied) am Wahrscheinlichsten sind (basierend auf den vorliegenden Daten).

Die Übergänge zu den weniger wahrscheinlichen Beobachtungen sind fließend sodass eine Regel für die Entscheidung festgelegt werden muss: typischerweise sagen wir dass wenn eine Beobachtung eine Wahrscheinlichkeit von < 5% hat, dieses Ergebnis *selten* ist. In diesem Fall wird die Grundannahme (dass es keine Auffälligkeit gibt: die Nullhypothese) verworfen (falsifiziert). Andernfalls, wenn das Ergebnis nicht selten ist unter der Grundannahme (> 5%), können wir die Nullhypothese nicht verwerfen. In diesem Fall wird sie bestätigt (verifiziert).

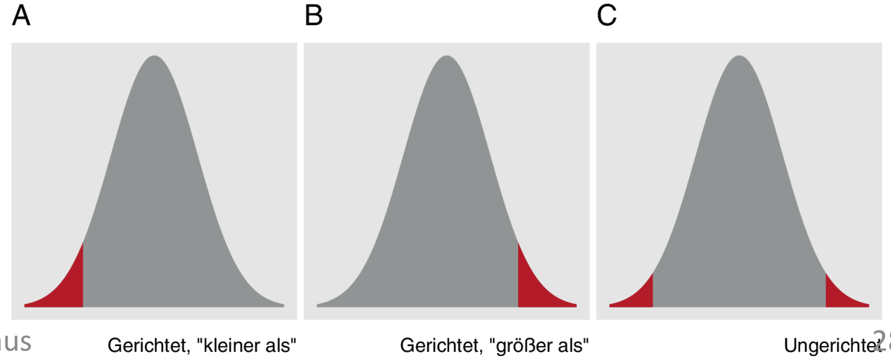
Zweiseitig / Ungerichtet



Einseitig / Gerichtet



Beispiel: Intelligenzquotient

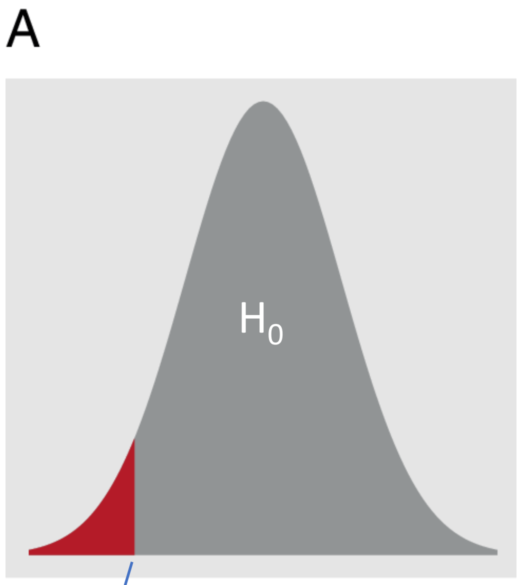


Beispiel Hypothesentest

Frage: Sind die Ankünfte am JFK weniger verspätet als an non_JFK?

H_1: Die Flüge am JFK haben weniger Verspätung (Ankunft) ← gilt wenn < 5%

H_0: Die Flüge am JFK haben **nicht weniger** Verspätung (Ankunft) ← bleibt bei ≥ 5%



Gerichtet, "kleiner als"

H_1: $\mu_{JFK} < \mu_{non\_JFK}$ ← gilt wenn < 5%

H_0: $\mu_{JFK} \geq \mu_{non\_JFK}$ ← bleibt bei ≥ 5%

$\alpha = 5\%$ (auch 1% oder 0,1%) Signifikanzniveau/ Irrtumswahrscheinlichkeit

Beispiel Hypothesentest

Frage: Sind die Ankünfte am JFK weniger verspätet als an non_JFK?

QM6c QM7 QM8.R | flights | Funktionen.R | jfk

```
########## QM8 ########################################

# arr_delay filtern nach 'JFK' und 'non_JFK'
#flights_clean <- drop_na(flights, arr_delay) #ohne NA

# nur jfk
jfk <- flights_clean %>% filter(origin == "JFK") %>% select(arr_delay)
str(jfk) #109.079

# nicht jfk
non_jfk <- flights_clean %>% filter(origin != "JFK") %>% select(arr_de
str(non_jfk) #218.267
# 109.079 + 218.267 = 327.346

# Anzahl prüfen mit Pupulation (ohne NA)
str(flights_clean) #327,346 x 19 --> stimmt
```

74:1 | QM8 | R Script

Console | Terminal | Jobs

```
tibble [109,079 x 1] (S3: tbl_df/tbl/data.frame)
 $ arr_delay: num [1:109079] 33 -18 -8 -2 -3 7 -4 -8 14 4 ...
> # nicht jfk
> non_jfk <- flights_clean %>% filter(origin != "JFK") %>% select(arr_delay)
> str(non_jfk) #218.267
tibble [218,267 x 1] (S3: tbl_df/tbl/data.frame)
```

JFK- Daten von den beiden anderen Flughäfen trennen (non_jfk)

kontrollieren

```r

##########################################################
### H1: Die Flüge am JFK haben weniger Verspätung (Ankunft)
### H0: Die Flüge am JFK haben nicht weniger Verspätung (Ankunft)


# ein JFK sample ziehen
jfk_sample<- sample(jfk$arr_delay, size = 10000)
x_<-mean(jfk_sample) #an dieser Stelle laesst sich nur der
# mean des samples berechnen
x_

non_jfk_sample<- sample(non_jfk$arr_delay, size = 10000)
mu<-mean(non_jfk_sample) #an dieser Stelle laesstt sich nur der
# mean des samples berechnen
mu
```

```
> x_
[1] 6.0769
> mu
[1] 7.2063
```

Ein sample aus jfk ziehen

Ein sample aus non_jfk ziehen

QM6c QM7 QM8.R | flights | Funktionen.R | jfk

```r
#aus diesem sample resamplen um die Breite des Konf-intervalls zu schÃ
set.seed(2021)

# resamplen mit der H_0: non-jfk
non_jfk_resample_loop <- do (10000)*mean(resample(non_jfk_sample)) #fo
mean(non_jfk_resample_loop$mean) #mean der means (grand mean)
sd(non_jfk_resample_loop$mean) #Standardfehler

hist(non_jfk_resample_loop$mean, breaks=50) #Histogramm
plot(density(non_jfk_resample_loop$mean,bw=1)) #GlÃ¤ttung Dichteplot

mu<-mean(non_jfk$arr_delay);mu # gesamte JFK
abline(v = mean(non_jfk$arr_delay), col = 4) #mean alle JFK

m_<-mean(non_jfk_resample_loop$mean);m_ # MW der Mittelwerte der sampl
abline(v = mean(non_jfk_resample_loop$mean), col = 2) #mean Stichprobe
```

Environment | History | Connections | Tutorial

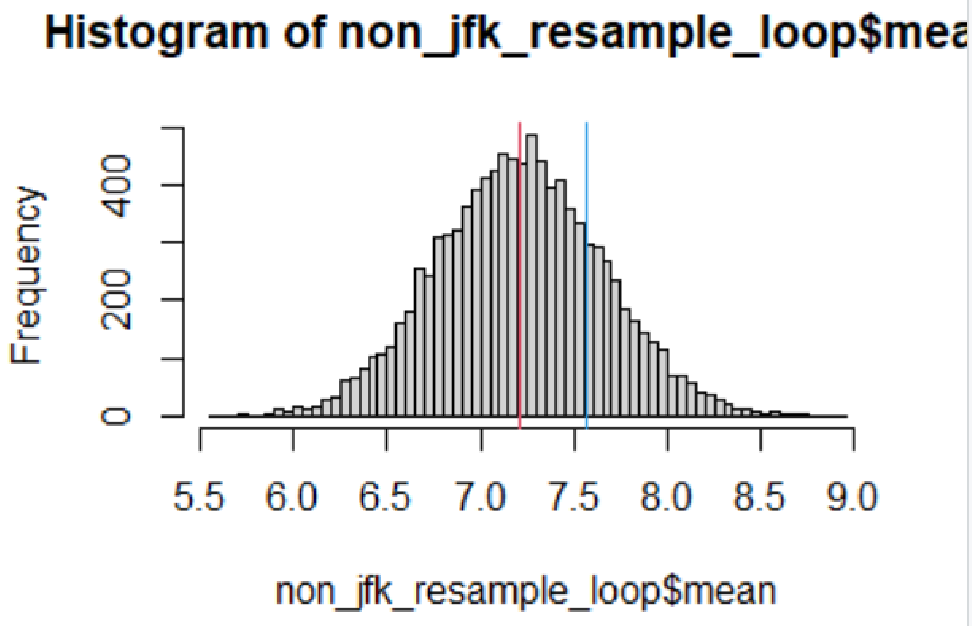
Global Environment

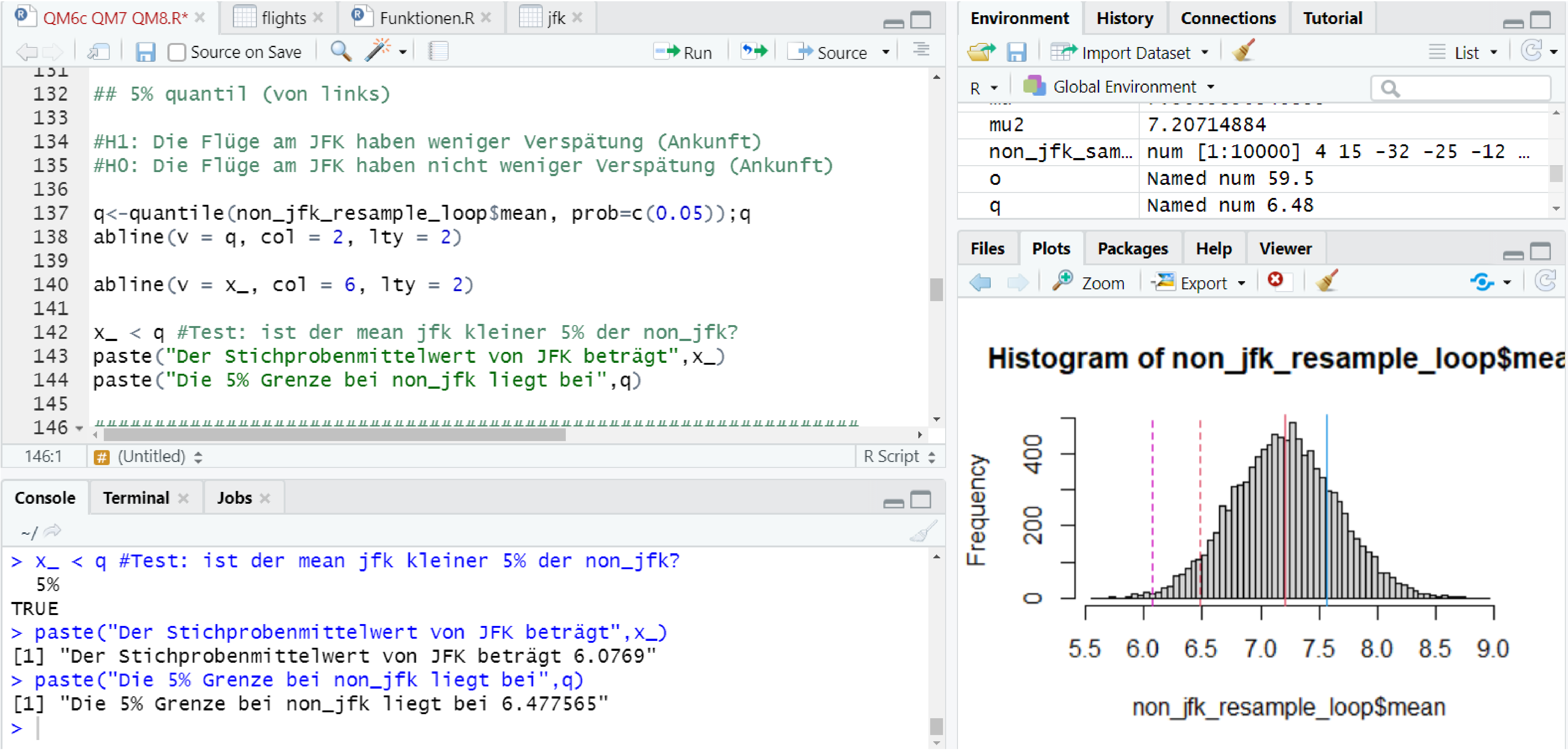
m_	7.20714884
mu	7.5669890546899
mu2	7.20714884
non_jfk_sam...	num [1:10000] 4 15 -32 -25 -12 ...

Console | Terminal | Jobs

```
> abline(v = mean(non_jfk$arr_delay), col = 4) #mean origins
> mu<-mean(non_jfk$arr_delay);mu # Mittelwert aller origins
[1] 7.566989
> hist(non_jfk_resample_loop$mean, breaks=50) #Histogramm
> m_<-mean(non_jfk_resample_loop$mean);m_ # MW der Mittelwerte der samples
[1] 7.207149
> abline(v = mean(non_jfk_resample_loop$mean), col = 2) #mean StichprobenMW
```

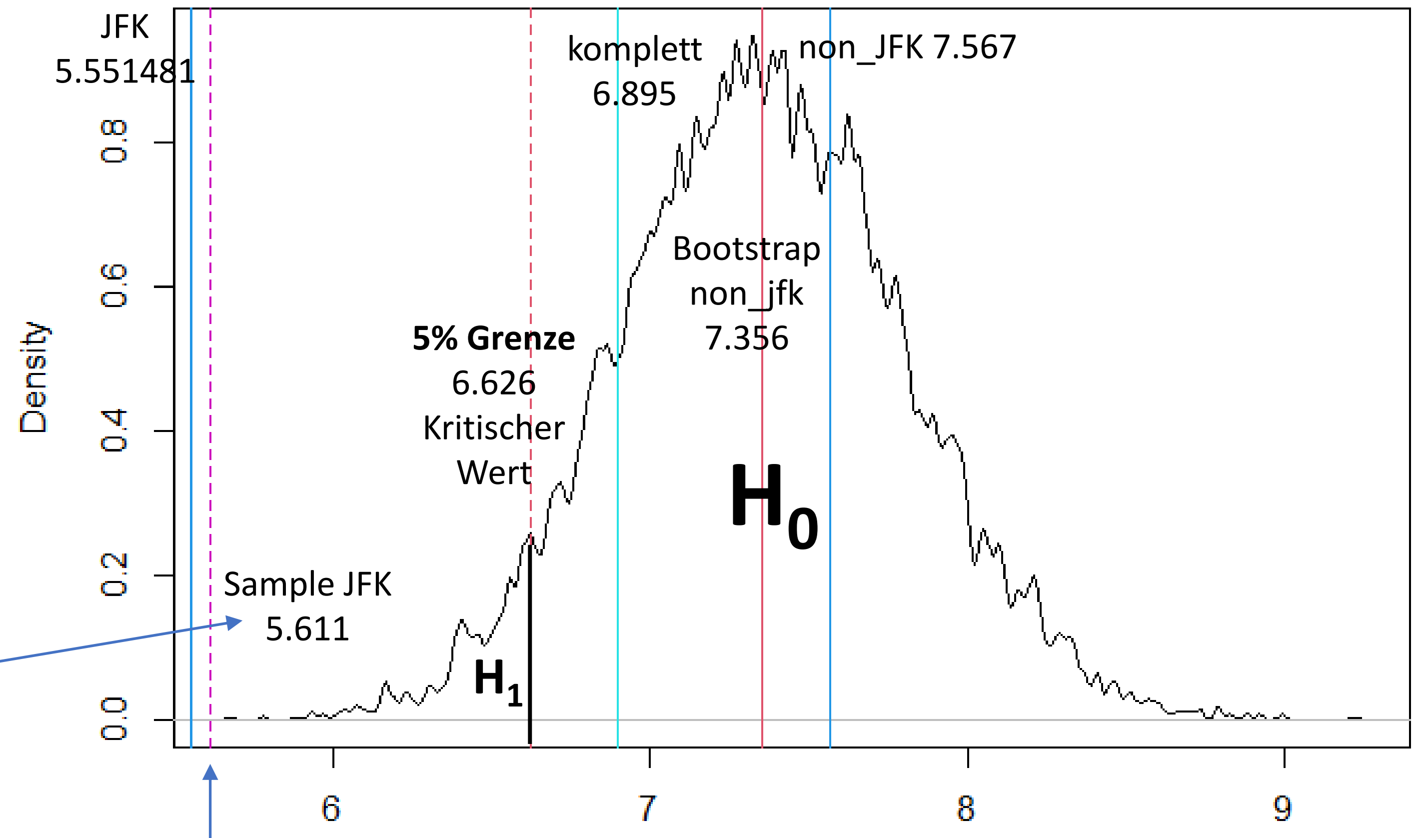
Files | Plots | Packages | Help | Viewer





QM6c QM7 QM8.R*
flights
Funktionen.R
jfk
Source on Save
Run
Source
5% quantil (von links)
#H1: Die Flüge am JFK haben weniger Verspätung (Ankunft)
#H0: Die Flüge am JFK haben nicht weniger Verspätung (Ankunft)
q<-quantile(non_jfk_resample_loop$mean, prob=c(0.05));q
abline(v = q, col = 2, lty = 2)
abline(v = x_, col = 6, lty = 2)
x_ < q #Test: ist der mean jfk kleiner 5% der non_jfk?
paste("Der Stichprobenmittelwert von JFK beträgt",x_)
paste("Die 5% Grenze bei non_jfk liegt bei",q)
146:1
(Untitled)
R Script
Console
Terminal
Jobs
~/
> x_ < q #Test: ist der mean jfk kleiner 5% der non_jfk?
5%
TRUE
> paste("Der Stichprobenmittelwert von JFK beträgt",x_)
[1] "Der Stichprobenmittelwert von JFK beträgt 6.0769"
> paste("Die 5% Grenze bei non_jfk liegt bei",q)
[1] "Die 5% Grenze bei non_jfk liegt bei 6.477565"
>
Environment
History
Connections
Tutorial
Import Dataset
List
R
Global Environment
mu2 7.20714884
non_jfk_sam… num [1:10000] 4 15 -32 -25 -12 …
o Named num 59.5
q Named num 6.48
Files
Plots
Packages
Help
Viewer
Zoom
Export
Histogram of non_jfk_resample_loop$mea
Frequency
0
200
400
5.5
6.0
6.5
7.0
7.5
8.0
8.5
9.0
non_jfk_resample_loop$mean

Beispiel Hypothesentest
Frage: Sind die Ankünfte am JFK weniger verspätet als an non_JFK?



<- unbekannt

Wie wahrscheinlich ist dieses Ergebnis von 5.6 min. unter der H_0 ?

Eher unwahrscheinlich!
Aber wie unwahrscheinlich?
→ p-Wert gibt genau diese Wahrscheinlichkeit an

Das Ergebnis ist < 0.05 und somit signifikant („überzufällig“)
-> die Nullhypothese wird abgelehnt. Es gilt die H_1

H_1: Die Flüge am JFK haben weniger Verspätung (Ankunft) ← gilt wenn < 5%
H_0: Die Flüge am JFK haben **nicht weniger** Verspätung (Ankunft) ← bleibt bei ≥ 5%

Korrigierte Stichprobenvarianz

Varianz = Streuung **um** den arithmetischen Mittelwert

Import „Schule“
Delimiter: tab
Decimal Mark: Comma
Data type: double

Erwartungstreue der Varianz

Jungen	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
	3,2	0,2	0,04
	3,5	0,5	0,25
	2,9	-0,1	0,01
	3,3	0,3	0,09
	3,4	0,4	0,16
	2,5	-0,5	0,25
	2,7	-0,3	0,09
	2,8	-0,2	0,04
	3,1	0,1	0,01
	2,6	-0,4	0,16
Summe	***30***	***0***	***1,1***

$$1{,}1 / 10 = 0{,}11$$

Mädchen	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
	1	-2	4
	1	-2	4
	2	-1	1
	2,5	-0,5	0,25
	3,2	0,2	0,04
	2,8	-0,2	0,04
	3,5	0,5	0,25
	2	-1	1
	6	3	9
	6	3	9
Summe	***30***	***0***	***28,58***

$$28{,}58 / 10 = 2{,}858$$

1,1/9 =
28,58/9 =

```
> var(Schule$boys)
[1] 0.1222222
> var(Schule$girls)
[1] 3.175556
```

default ist die sample variance, nicht die population variance

Sample Variance = $\frac{1}{n\boxed{-1}} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Sample Correction

```
> # function for population variance
> var_pop <- function(x){mean((x - mean(x))^2)}
> var_pop(Schule$boys)
[1] 0.11
> var_pop(Schule$girls)
[1] 2.858
```

1,1/10 =
28,58/10 =

Population Variance = $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

oder nur über die Nennen korrigieren mit *(n-1)/n

```
#von der built-in sample variance zur population variance:
var_pop <- function(x){var(x)*(length(x)-1) /length(x)}
```


https://statisticsglobe.com/variance-in-r-var-function

QM9

Testauswahl

Prüfen auf Normalverteilung

Ein Beispiel

Hypothesentest: Beispiel jfk vs non_jfk

1) H_1: Die Flüge am JFK haben weniger Verspätung (Ankunft)

   H_0: Die Flüge am JFK haben **nicht weniger** Verspätung (Ankunft)

2) Stichproben an den Flughäfen vornehmen

Simulation

Simulationsbasierte Verfahren

samplen
re-samplen

klassisch

NV?

ja

Parametrische Verfahren

nein

Nicht-parametrische Verfahren

U-Test
..

Übersicht Hypothesentest

1) Hypothesenpaar aufstellen
2) Daten generieren
3) Explorative Analyse: z.B. Daten normalverteilt?
4) Testauswahl und -durchführung

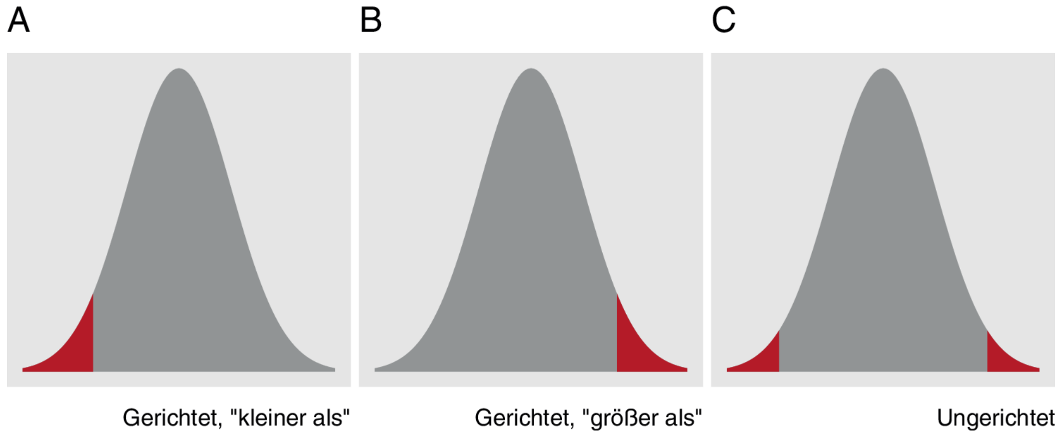
3) Prüfen ob die Daten normalverteilt sind: **QQ-plot und Shapiro-Wilk Test**

4) ..

Testauswahl

- Wenn Daten einer Normalverteilung (NV) folgen, wendet man in der Praxis sog. **parametrische** Hypothesentests an
- Parametrisch weil die NV durch 2 Parameter (Mittelwert und Standardabweichung) bestimmt wird

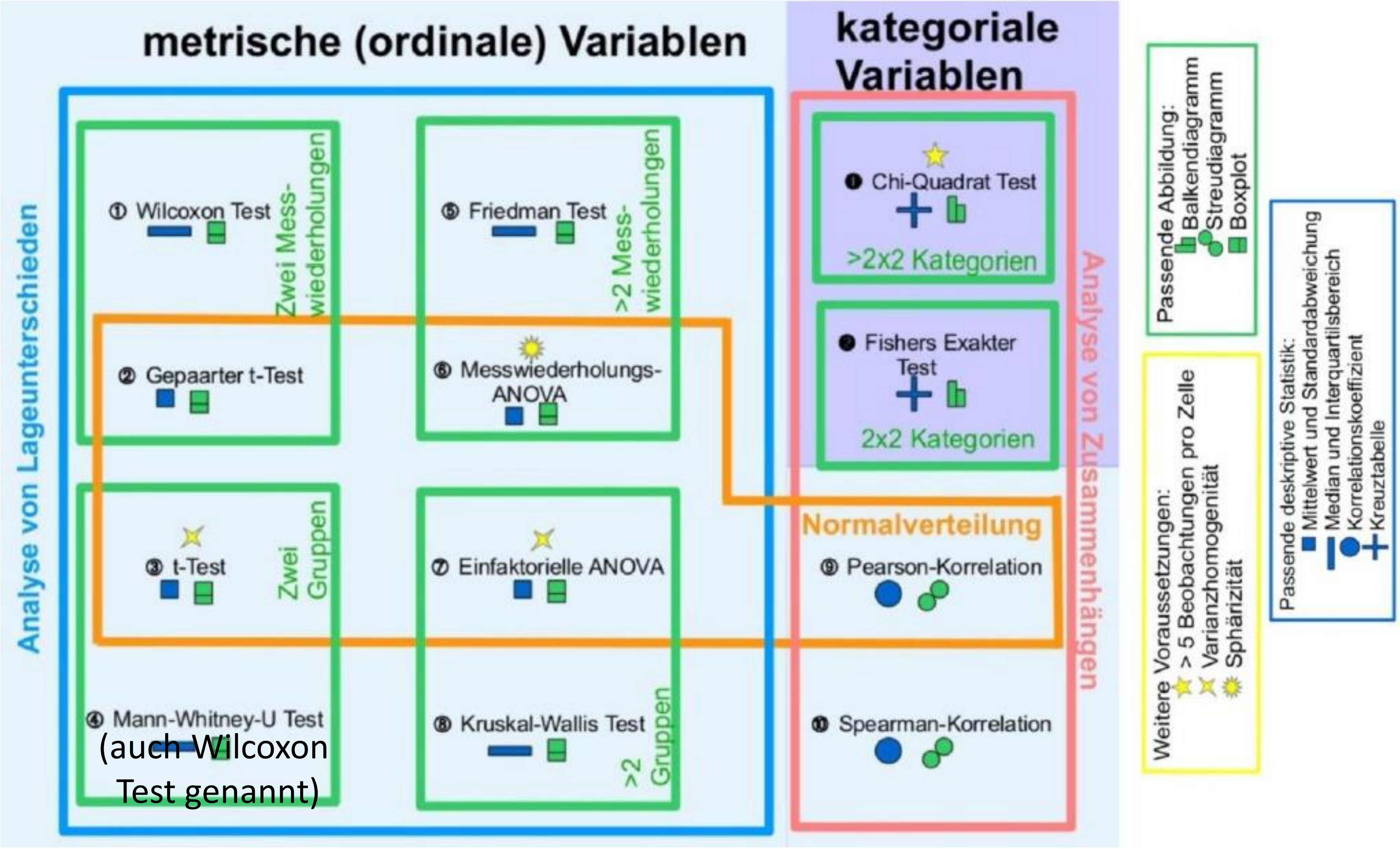
- Eine weit verbreitete Klasse von parametrischen Tests sind t-Tests (bei z.B. höchstens 2 Gruppen)
- Wir unterscheiden allgemein **einseitige** (gerichtete) von **zweiseitigen** (ungerichteten) Tests,



- sowie **abhängige** (vorher-nachher, dieselbe Gruppe) von **unabhängigen** (2 unterschiedliche Gruppen, z.B. geimpft vs. ungeimpft) Tests

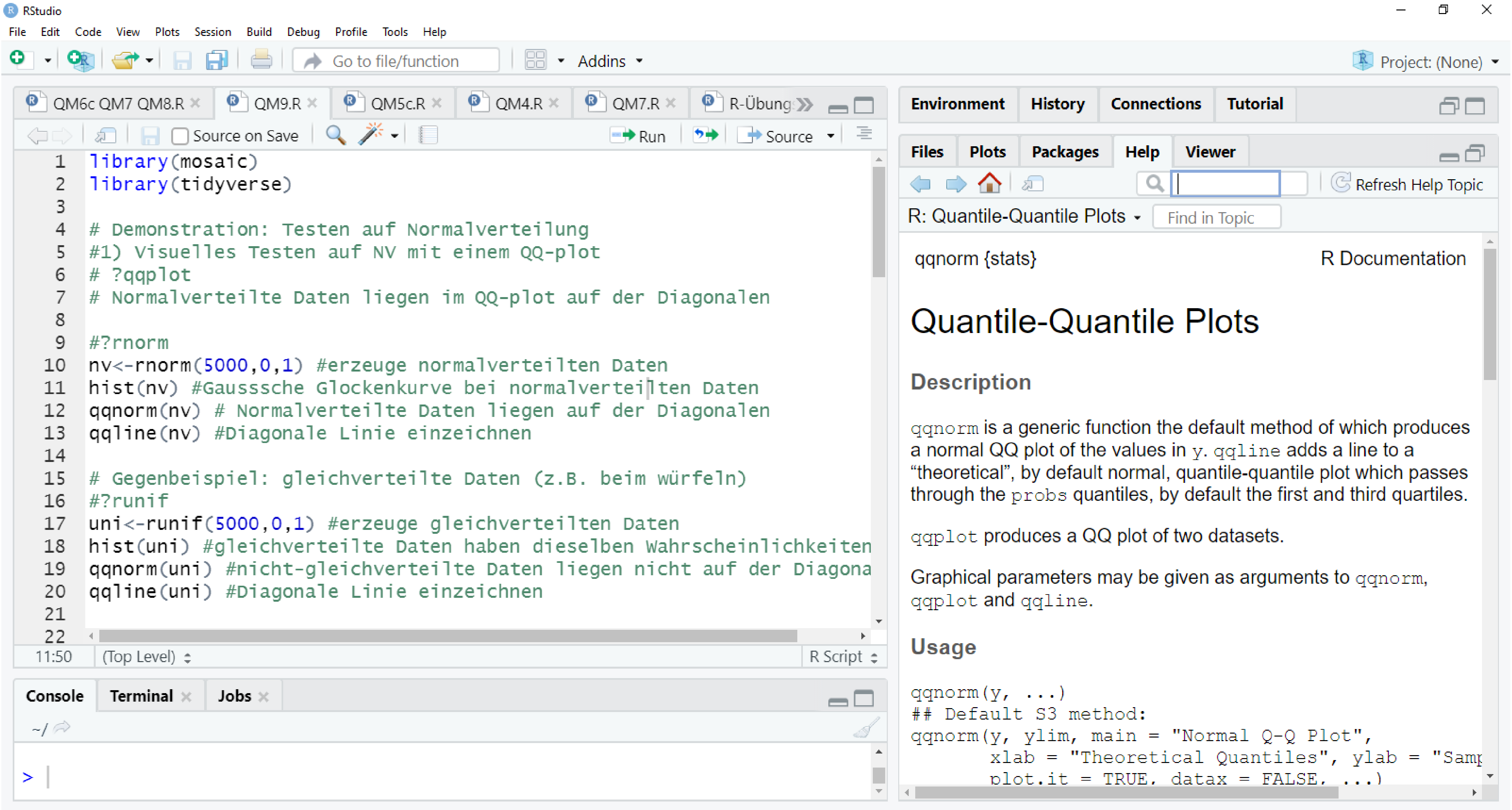
- Die Kenntnis ob Daten normalverteilt sind, ist wichtig für die richtige Testauswahl (siehe Schaubild nächste Folie)

Beispiele: Parametrische (NV) und nicht-parametrische Testverfahren



http://docplayer.org/72926040-Online-statistik-coaching.html

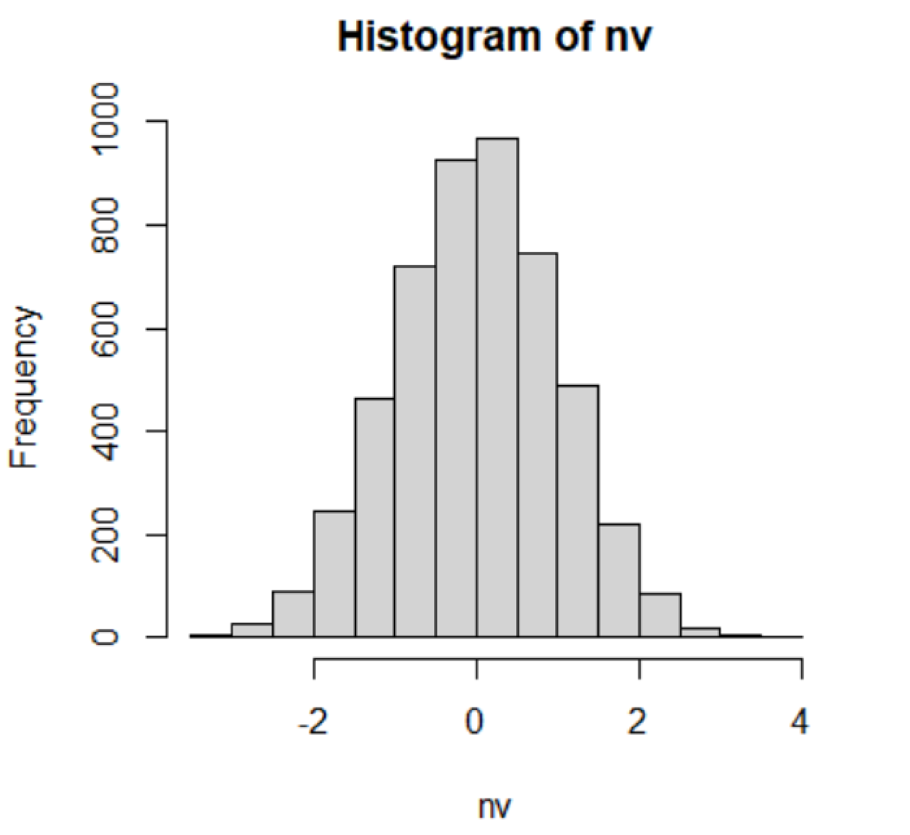
QQ-plot prüft visuell auf Normalverteilung



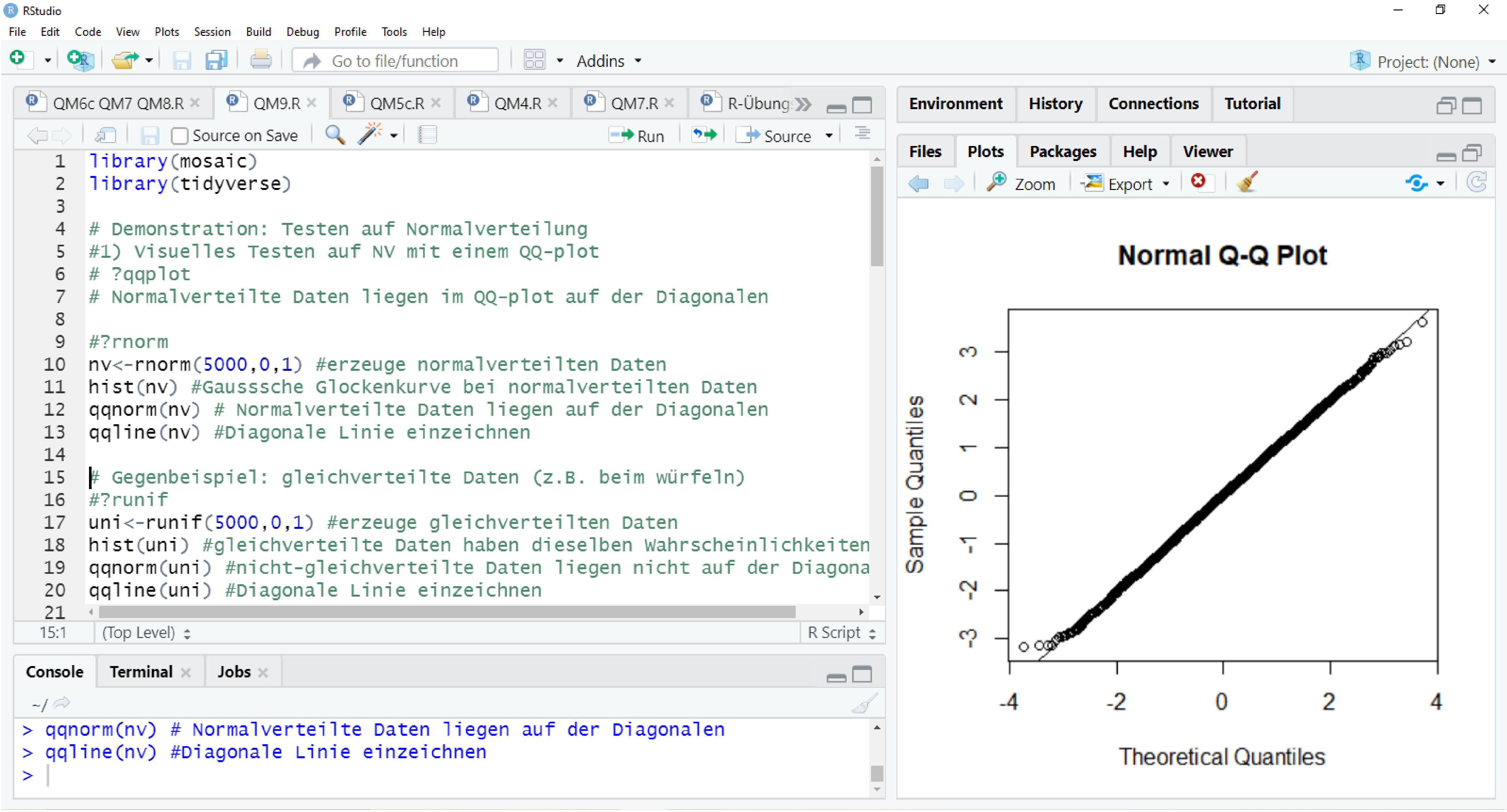
Beispiel: Gausssche Glockenkurve bei normalverteilten Daten

```r
library(mosaic)
library(tidyverse)

# Demonstration: Testen auf Normalverteilung
#1) Visuelles Testen auf NV mit einem QQ-plot
# ?qqplot
# Normalverteilte Daten liegen im QQ-plot auf der Diagonalen

#?rnorm
nv<-rnorm(5000,0,1) #erzeuge normalverteilten Daten
hist(nv) #Gausssche Glockenkurve bei normalverteilten Daten
qqnorm(nv) # Normalverteilte Daten liegen auf der Diagonalen
qqline(nv) #Diagonale Linie einzeichnen

# Gegenbeispiel: gleichverteilte Daten (z.B. beim würfeln)
#?runif
uni<-runif(5000,0,1) #erzeuge gleichverteilten Daten
hist(uni) #gleichverteilte Daten haben dieselben Wahrscheinlichkeiten
qqnorm(uni) #nicht-gleichverteilte Daten liegen nicht auf der Diagona
qqline(uni) #Diagonale Linie einzeichnen
```

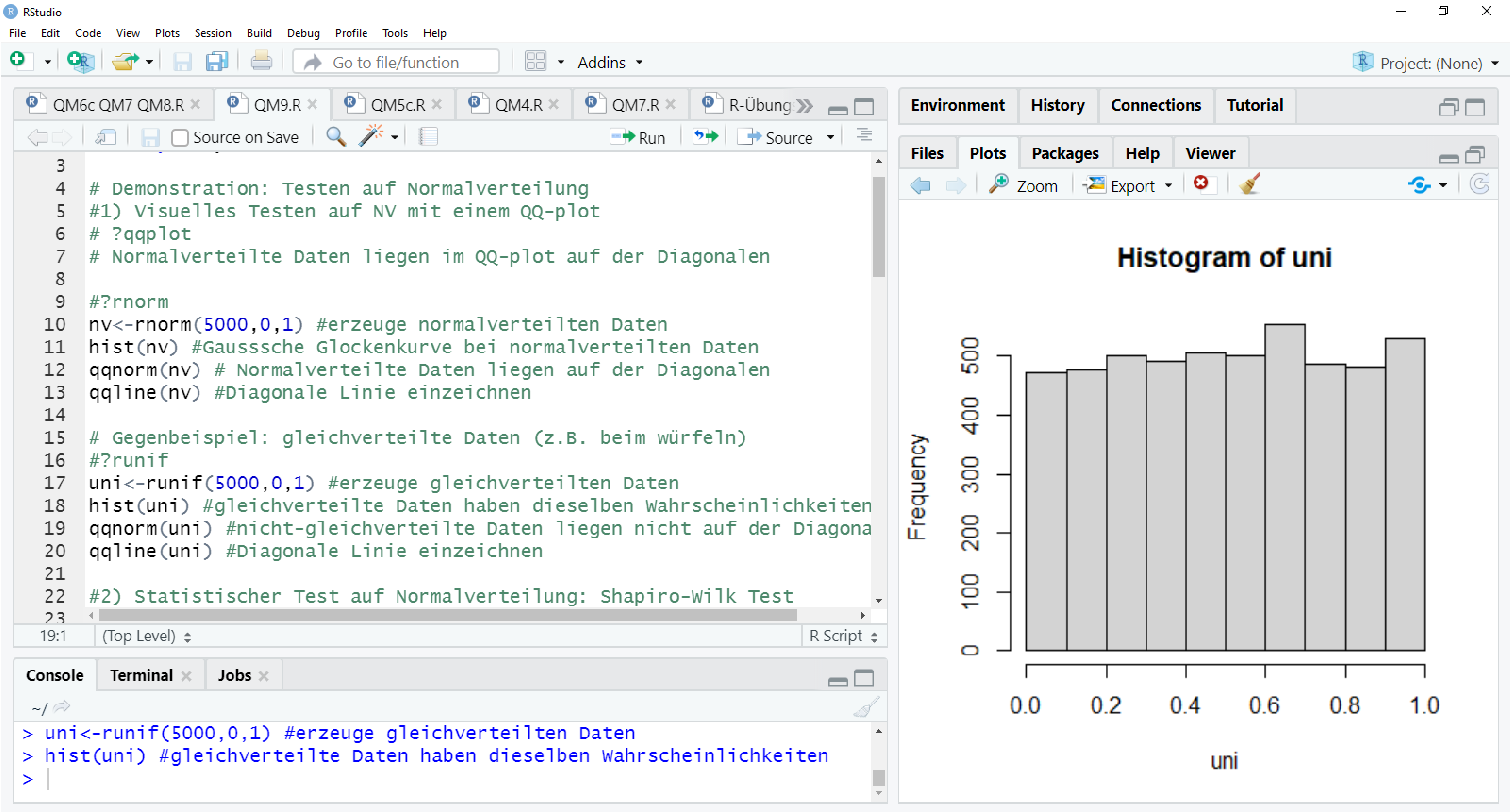
```
> hist(nv) #Gausssche Glockenkurve bei normalverteilten Daten
>
```



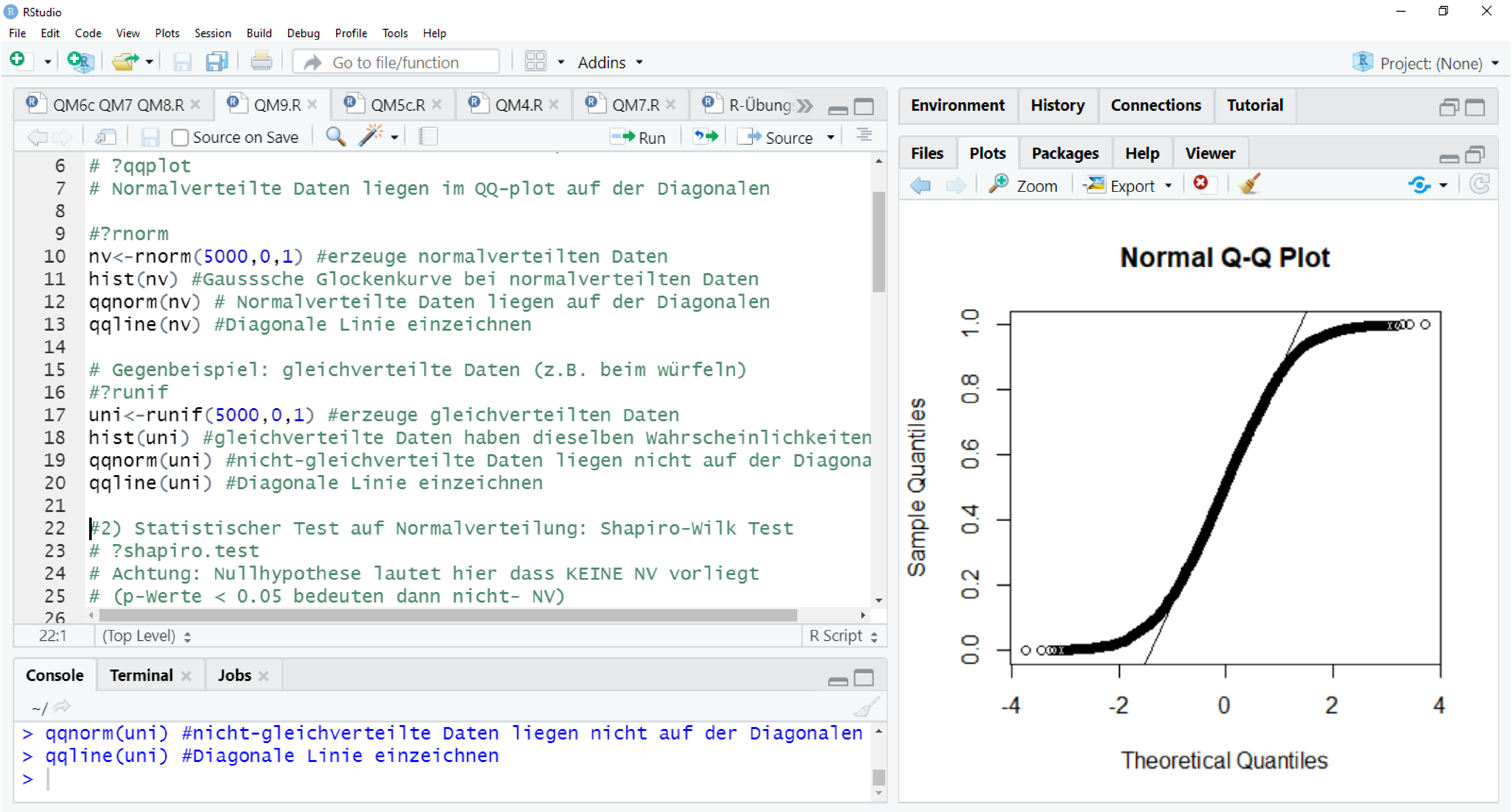
Beispiel: Diagonale im QQ-plot bei normalverteilten Daten



Beispiel: keine Gausssche Glockenkurve bei nicht-normalverteilten Daten



Beispiel: keine Diagonale im QQ-plot bei nicht-normalverteilten Daten



Beispiel: Shapiro- Wilk Test testet auf Normalverteilung

```r
#2) Statistischer Test auf Normalverteilung: Shapiro-Wilk Test
# ?shapiro.test
# Achtung: Nullhypothese lautet hier dass EINE NV vorliegt
# (p-Werte < 0.05 bedeuten dann nicht- NV)
# p-Werte > 0.05 bestätigen dann die H_0: Daten sind dann NV
shapiro.test(nv) #testen auf NV
shapiro.test(uni) #testen auf NV
```

Kurzform Veranschaulichung Shapiro- Wilk Test

```
        Shapiro-Wilk normality test

data:  nv
W = 0.99955, p-value = 0.2959
```

H_0 gilt: NV weil > 0.05

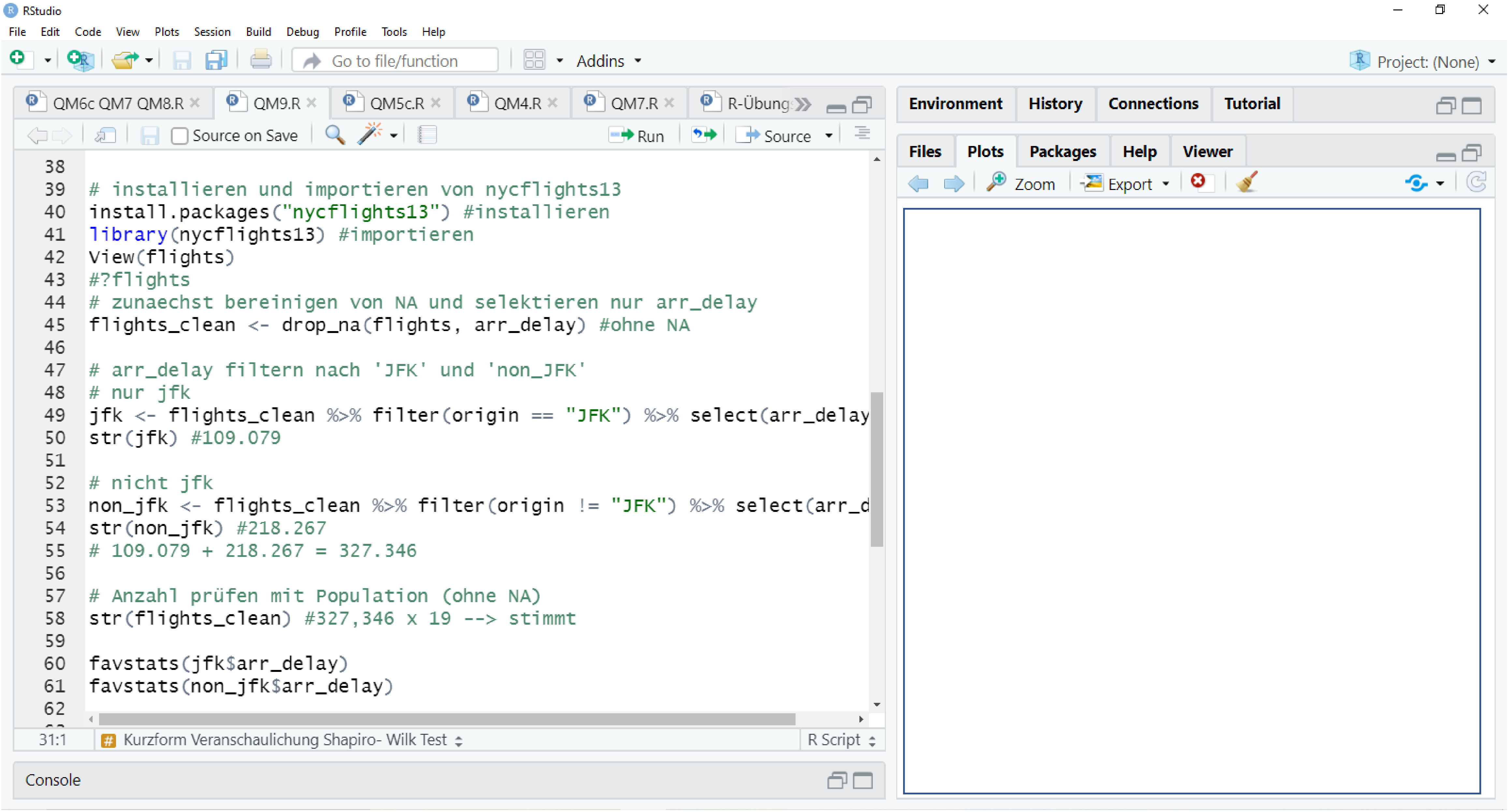
```
> shapiro.test(uni) #testen auf NV

        Shapiro-Wilk normality test

data:  uni
W = 0.95663, p-value < 2.2e-16
```

H_1 gilt: keine NV weil < 0.05

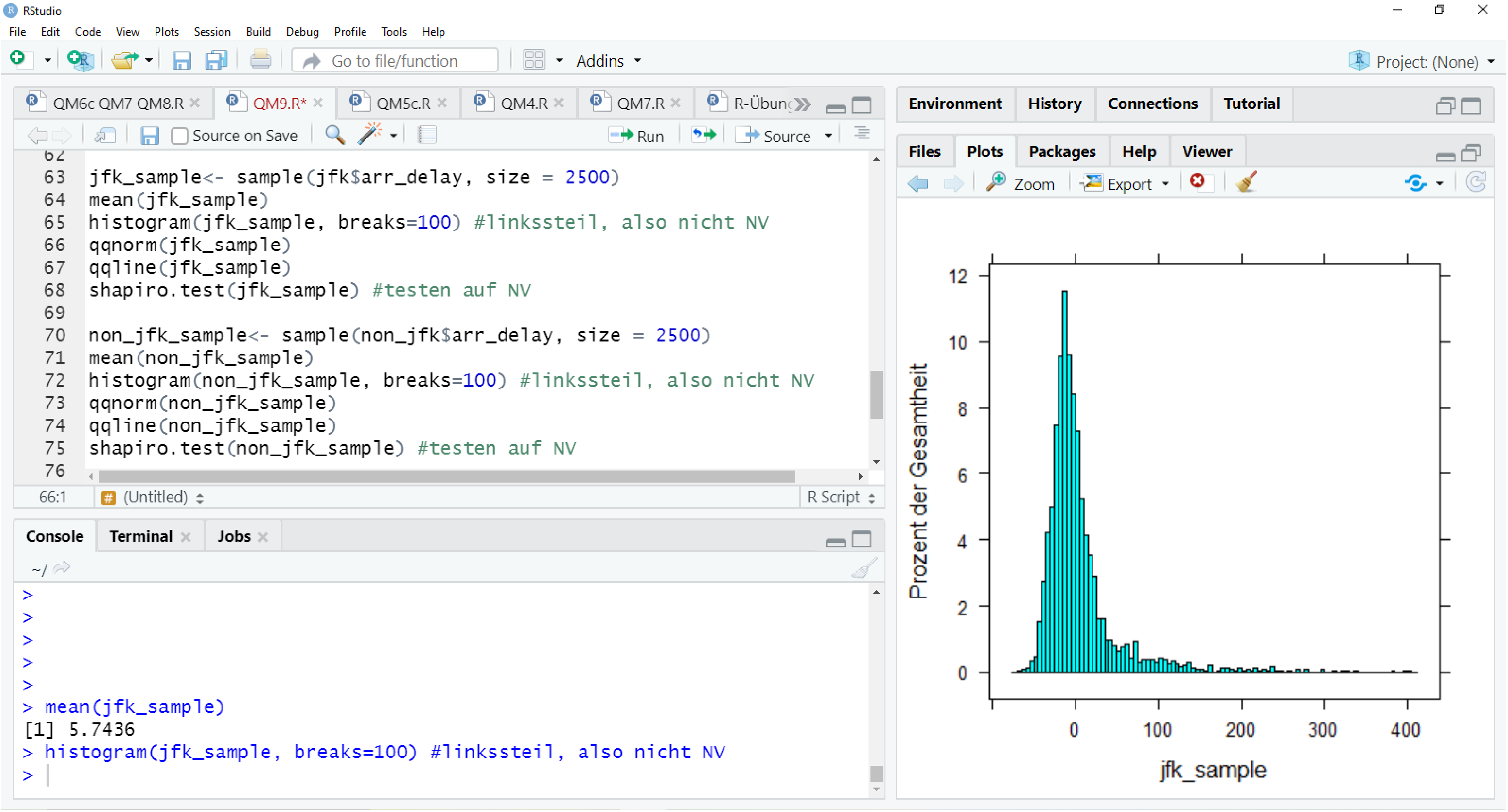
Shapiro- Wilk Test

Nullhypothese: Daten **sind** NV (p>0.05)
Alternativhypothese: Daten **sind nicht** NV (p>0.05)

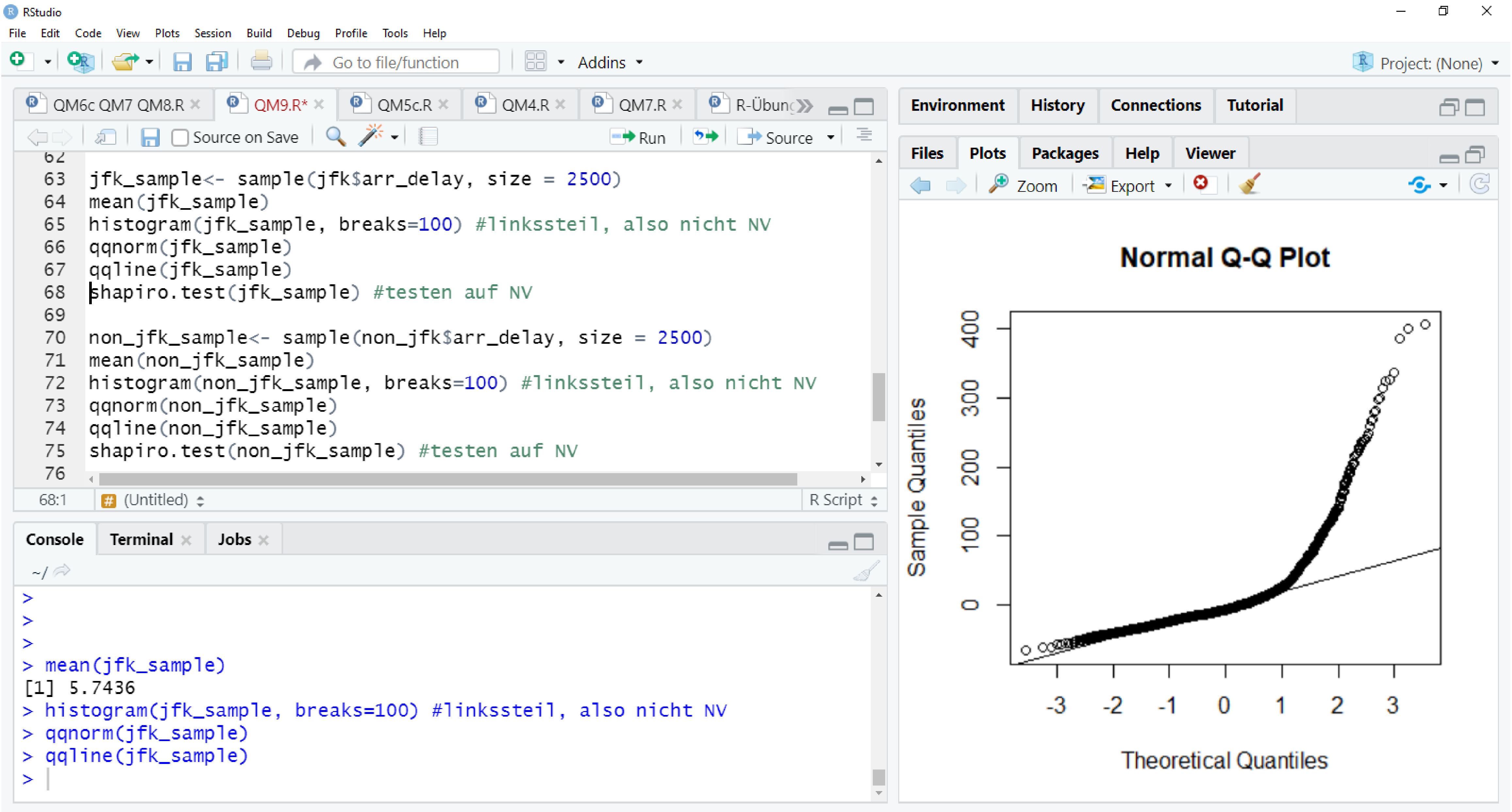
Beispiel: jfk und non_jfk auf Normalverteilung prüfen



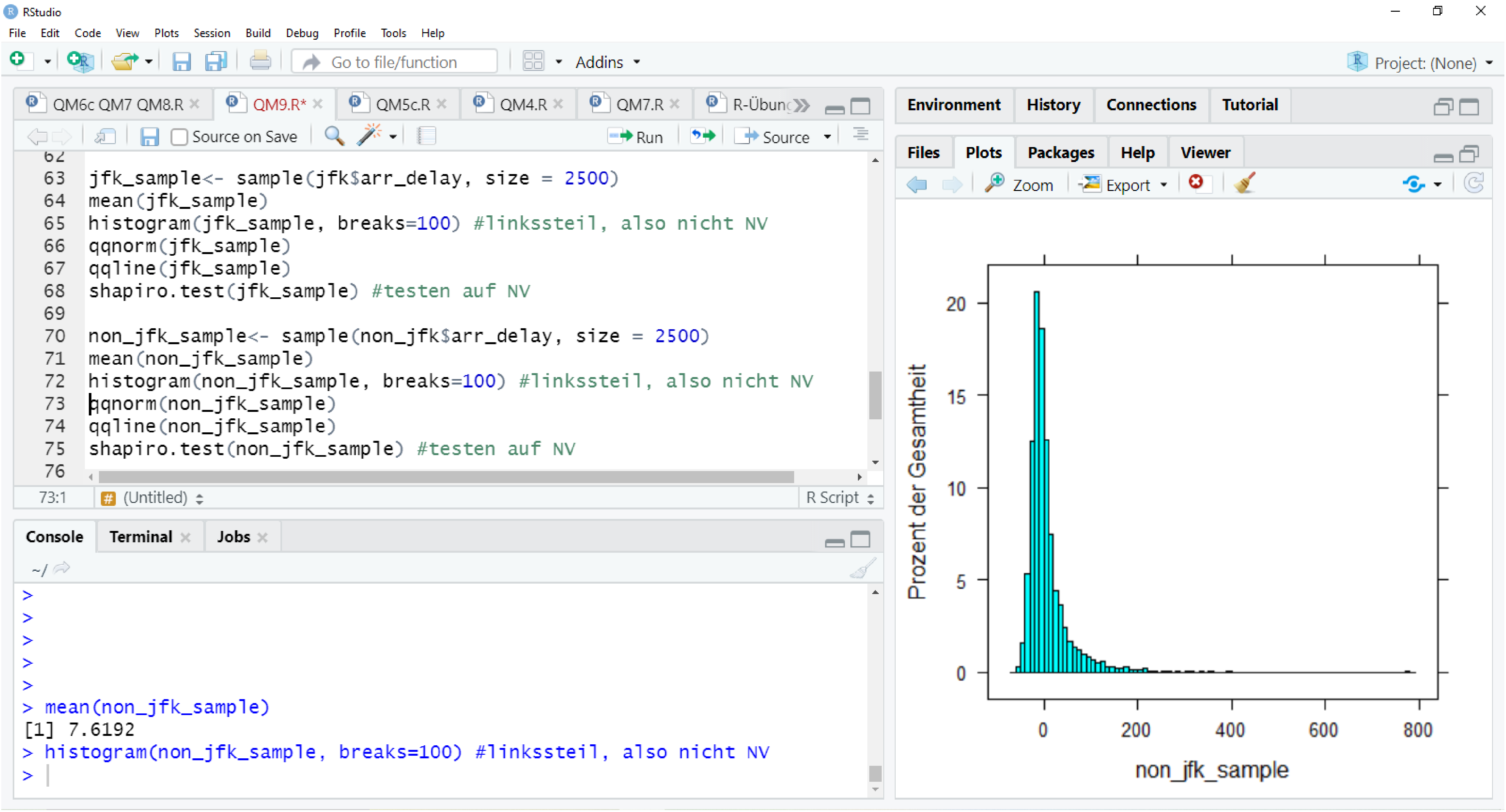
Beispiel: jfk ist linkssteil



Beispiel: jfk ist eher nicht NV



Beispiel: non-jfk ist linkssteil



Beispiel: non_jfk ist eher nicht NV

Beispiel: Shapiro- Wilk Test lehnt NV ab

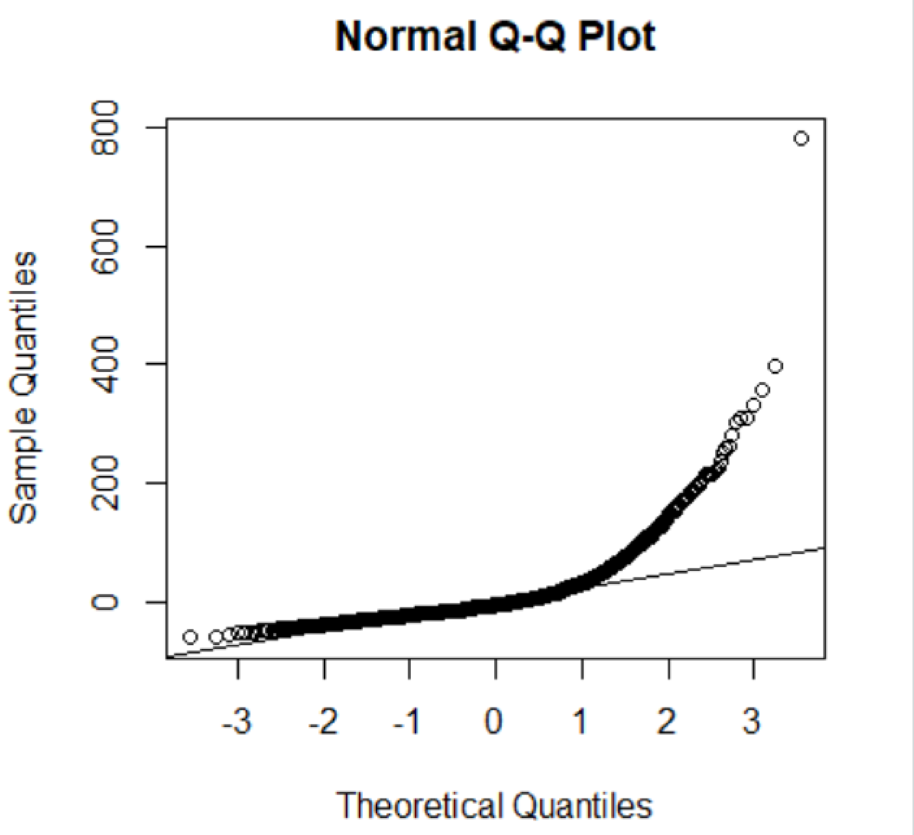
```
shapiro.test(jfk_sample) #testen auf NV
shapiro.test(non_jfk_sample) #testen auf NV

#FAZIT: beide Datensätze sind nicht normalverteilt.
# Es können deshalb keine parametrischen Verfahren angewandt werden.
# Es liegen 2 unabhängige Gruppen vor (kein vorher-nachher)
```

```
> shapiro.test(jfk_sample) #testen auf NV

        Shapiro-Wilk normality test

data:  jfk_sample
W = 0.68492, p-value < 2.2e-16
```

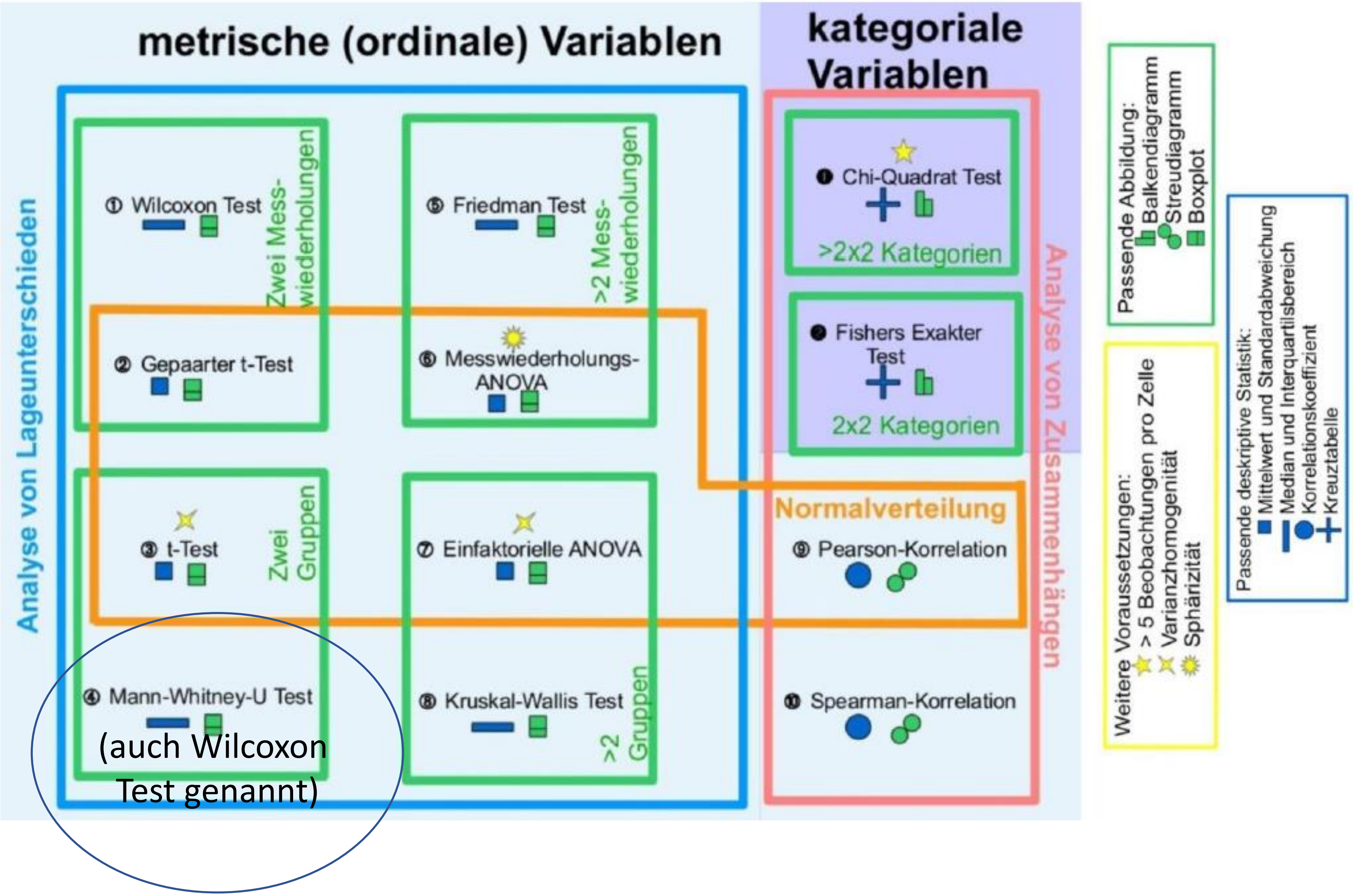
→ H_1 gilt: keine NV weil < 0.05

```
> shapiro.test(non_jfk_sample) #testen auf NV

        Shapiro-Wilk normality test

data:  non_jfk_sample
W = 0.67879, p-value < 2.2e-16
```

→ H_1 gilt: keine NV weil < 0.05



Hilfestellung: Parametrische (NV) und nicht-parametrische Testverfahren



Ablauf Hypothesentest

1) Hypothesenpaar aufstellen
2) Daten generieren
3) Explorative Analyse: z.B. Daten normalverteilt?
4) **Testauswahl und -durchführung**

http://docplayer.org/72926040-Online-statistik-coaching.html

Hypothesentest: Beispiel jfk vs non_jfk

1) H_1: Die Flüge am JFK haben weniger Verspätung (Ankunft)

H_0: Die Flüge am JFK haben **nicht weniger** Verspätung (Ankunft)

$$H_1: \mu_{\mathrm{JFK}} < \mu_{\mathrm{non\_JFK}}$$

$$H_0: \mu_{\mathrm{JFK}} \geq \mu_{\mathrm{non\_JFK}}$$

2) Stichproben an den Flughäfen vornehmen

3) Prüfen ob die Daten normalverteilt sind: QQ-plot und Shapiro-Wilk Test

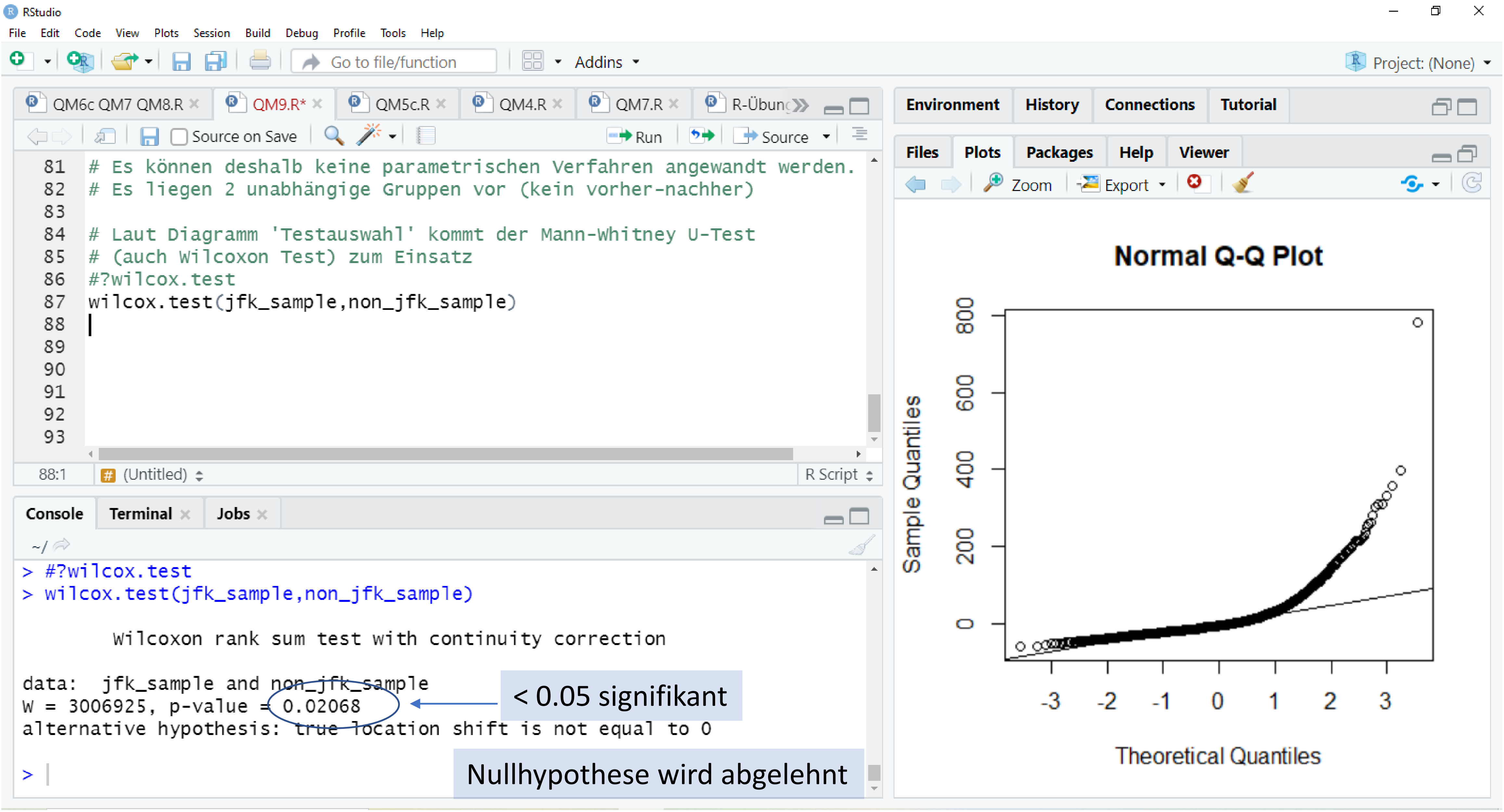
4) 2 unabhängige nicht-normalverteilte Gruppen:

Mann-Whitney U – Test (auch Wilcoxon Test genannt)

Ablauf Hypothesentest

1) Hypothesenpaar aufstellen
2) Daten generieren
3) Explorative Analyse: z.B. Daten normalverteilt?
4) Testentscheidung und -durchführung

Beispiel: Wilcoxon Test ist signifikant. Die Nullhypothese (nicht weniger Verspätung) wird abgelehnt



Ende