Blatt03

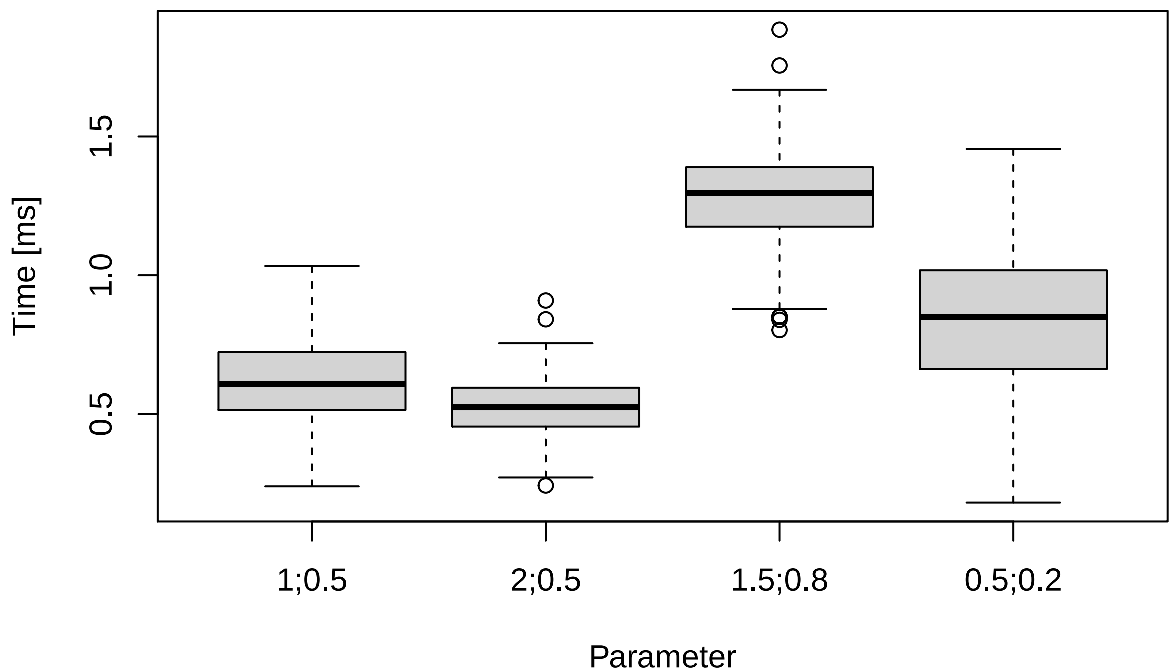
Henri Lübeck; Tobias Janßen; Johannes Burr

10 5 2020

R Markdown

```
source("po20_blatt03_code.R")

library("microbenchmark")
library("checkmate")


t1 = c(rep.int(0,500))
t2 = c(rep.int(0,500))
t3 = c(rep.int(0,500))
t4 = c(rep.int(0,500))

set.seed(123)
for (i in 1:500) {
  a = runif(1,min=-10,max=10)
  b = runif(1,min=-10,max=10)
  x_nul=c(a,b)

  res1 = microbenchmark( kompasssuche2(f,x0=x_nul,s=1,theta=0.5),times=100, unit = "ms")
  t1[i] = summary(res1)$median

  res2 = microbenchmark( kompasssuche2(f,x0=x_nul,s=2,theta=0.5),times=100, unit = "ms")
  t2[i] = summary(res2)$median

  res3 = microbenchmark( kompasssuche2(f,x0=x_nul,s=1.5,theta=0.8),times=100, unit = "ms")
  t3[i] = summary(res3)$median

  res4 = microbenchmark( kompasssuche2(f,x0=x_nul,s=0.5,theta=0.2),times=100, unit = "ms")
  t4[i] = summary(res4)$median


}



save(t1,file="t1.RData")
save(t2,file="t2.RData")
save(t3,file="t3.RData")
save(t4,file="t4.RData")
```

```r
boxplot(t1,t2,t3,t4,
        xlab="Parameter", ylab="Time [ms]",
        names=c("1;0.5","2;0.5","1.5;0.8","0.5;0.2"))
```



Die erste Parameter-Kombination funktioniert am zweit besten, mit einem leicht höheren Median und stärkerer Streuung als die zweite Variante.

Die zweite Parameter-Kombination benötigt im Mittel am wenigsten Zeit und hat die geringste Streung.

Die dritte Parameter-Kombination ist eindeutig die schlechteste mit einem Median von 1.49335, der über den oberen Quantilen der ersten beiden Varianten liegt.

Bei der vierten Parameter-Kombination ist die große Streuung hervorzuheben. Im Median liegt diese Variante zwischen 1&2 und 3, wobei einige Läufe sogar schneller als in den ersten beiden Varianten sind, andere Läufe aber auch schlechter als ein Großteil derer der dritten Variante.

Vorraussetzungen für die Tests

t-Test:

- unabhängige Stichproben: wegen der Pseudozufallszahlen garantiert
- stetig & normalverteilt: gemessene Einheit (Zeit) ist stetig Prüfung auf Normalverteilung durch QQ-Plots
- Homoskedastizität: F-Test auf gleiche Varianzen

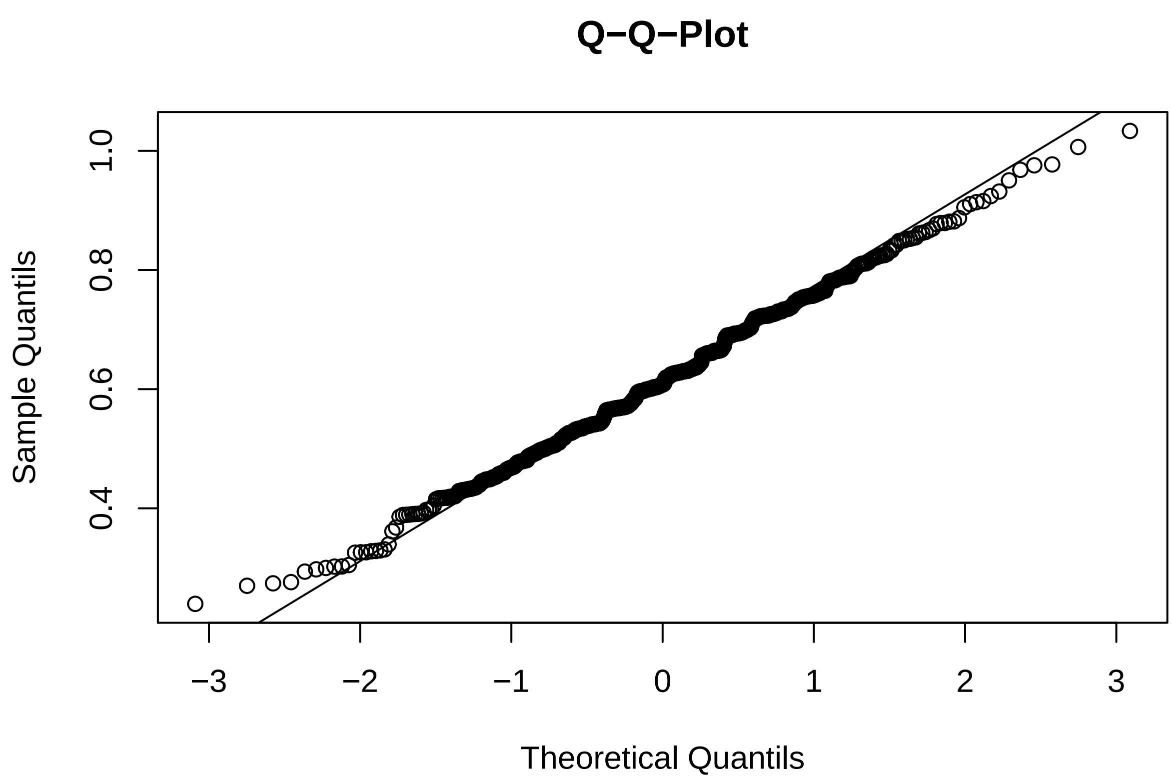
Wilcoxon-Rangsummen-Test:

- stetiges Merkmal: gemessene Einheit (Zeit) ist Stetig

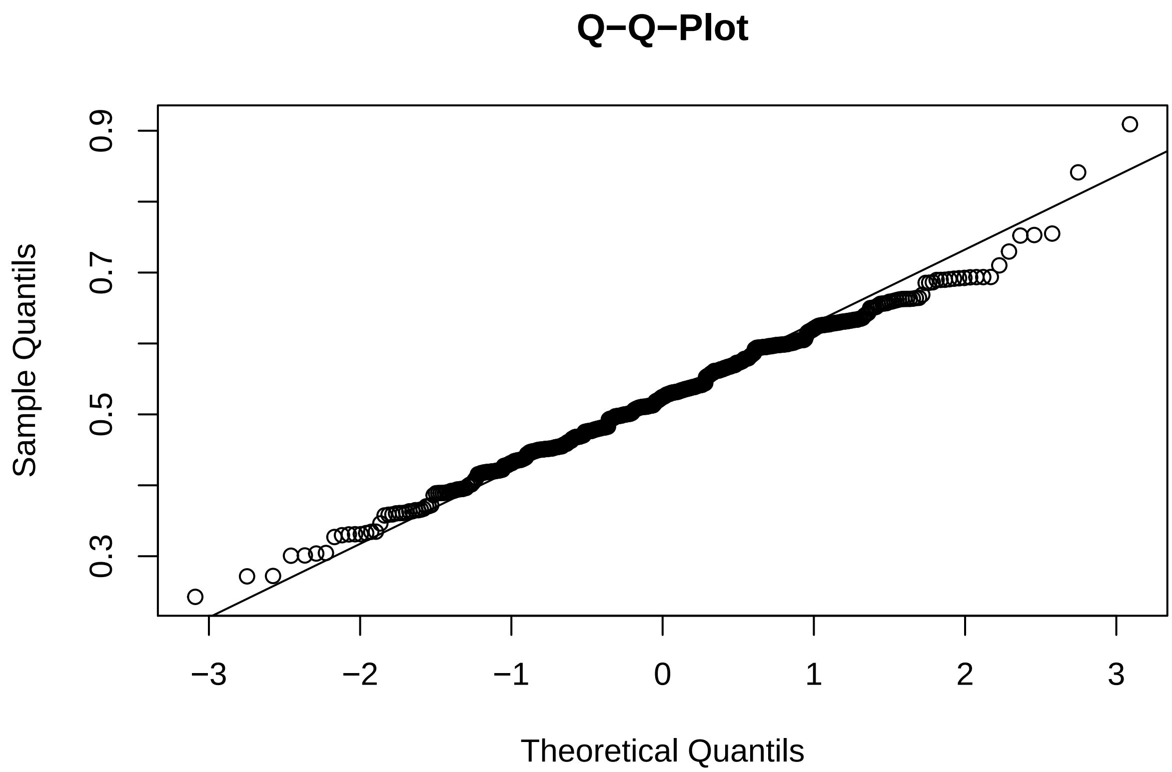
- gleiche Verteilungsform in beiden Stichproben: Durch QQ-Plots geprüft Wenn sich die Varianzen zu sehr unterscheiden, wird der Test zu liberal (Hayes, 2000)

Vorraussetzung nach: Eid, Gollwitzer, Schmitt (2010)

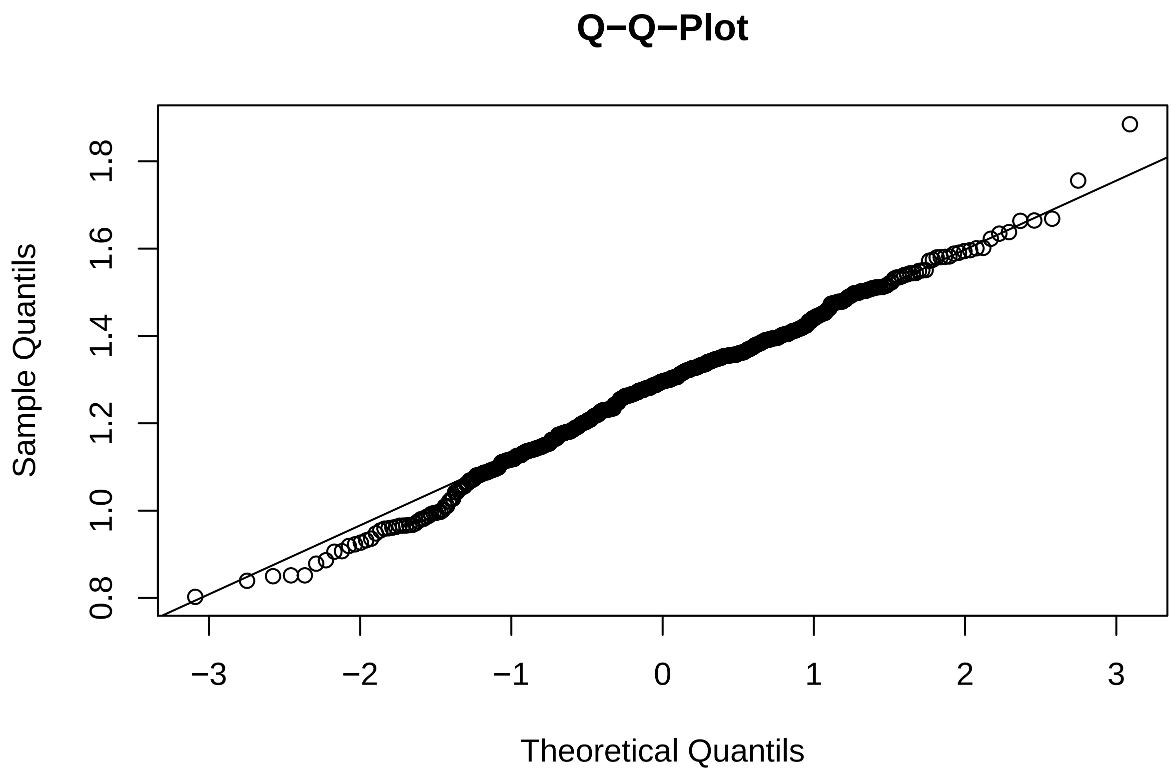
```
qqnorm(t1,
       xlab="Theoretical Quantils",
       ylab="Sample Quantils",
       main="Q-Q-Plot")
qqline(t1,distribution=qnorm)
```



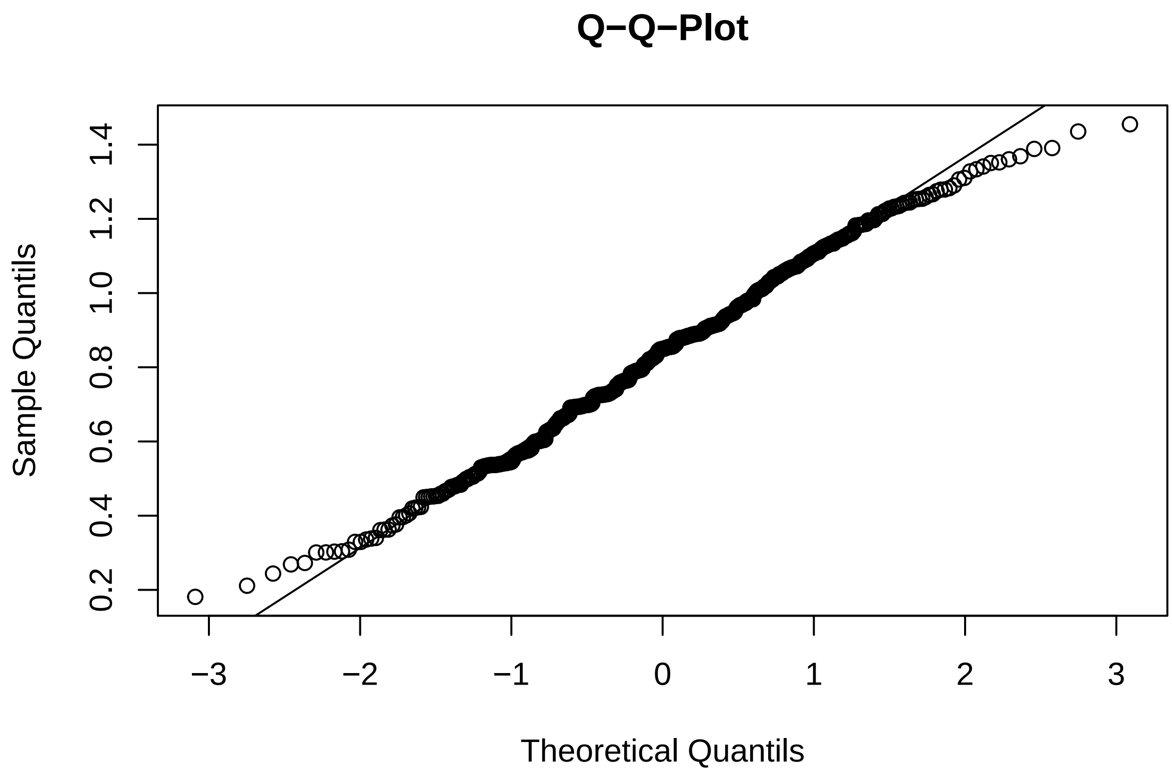

```
qqnorm(t2,
       xlab="Theoretical Quantils",
       ylab="Sample Quantils",
       main="Q-Q-Plot")
qqline(t2,distribution=qnorm)
```




```
qqnorm(t3,
       xlab="Theoretical Quantils",
       ylab="Sample Quantils",
       main="Q-Q-Plot")
qqline(t3,distribution=qnorm)
```




```r
qqnorm(t4,
       xlab="Theoretical Quantils",
       ylab="Sample Quantils",
       main="Q-Q-Plot")
qqline(t4,distribution=qnorm)
```



Im QQ-Plot werden die theoretischen Quantile gegenüber den empirischen Quantilen abgeglichen. Je näher die Punkte an der Einheitsgrade liegen, desto eher stimmen beide überein. Eine kleine Abweichung an den Extremwerten ist üblich.

Die Plots zeigen, das eine Normalverteilungsannahme gerechtfertig ist, wobei man beim Vierten aufpassen sollte. => die Ergebnisse sind alle normalverteilt und haben somit die selbe Verteilungsform

```r
var.test(t1, t3, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  t1 and t3
## F = 0.73807, num df = 499, denom df = 499, p-value = 0.000714
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6191526 0.8798164
## sample estimates:
## ratio of variances
##          0.7380655
```

```r
var.test(t3, t4, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  t3 and t4
## F = 0.4353, num df = 499, denom df = 499, p-value < 2.2e-16
```

```
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3651634 0.5188975
## sample estimates:
## ratio of variances
##          0.4352957
```

```r
var.test(t2, t4, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  t2 and t4
## F = 0.13637, num df = 499, denom df = 499, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1144016 0.1625648
## sample estimates:
## ratio of variances
##          0.1363733
```

```r
var.test(t1, t4, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  t1 and t4
## F = 0.32128, num df = 499, denom df = 499, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2695145 0.3829803
## sample estimates:
## ratio of variances
##          0.3212767
```

Bei allen F-Tests wird die Nullhypothese bei einem Signifikanzniveau von 5% verworfen. Damit sind die Varianzen zur Irrtumswahrscheinlichkeit von 5% verschieden. Damit ist die Homoskedastizitätsverraussetzung des t-Tests nicht erfüllt.

```r
#Tests zu t1 und t3
t.test(t1,t3,alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  t1 and t3
## t = -67.518, df = 975.84, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.6501093
## sample estimates:
## mean of x mean of y
## 0.6153385 1.2816970
```

```r
wilcox.test(t1,t3,alternative="less")
```

```
##
##  Wilcoxon rank sum test with continuity correction
```

```
##
## data:  t1 and t3
## W = 340, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
```

```r
#Tests zu t3 und t4
t.test(t3,t4,alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  t3 and t4
## t = 32.992, df = 864.22, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf 0.4708682
## sample estimates:
## mean of x mean of y
## 1.2816970 0.8332124
```

```r
wilcox.test(t3,t4,alternative="less")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  t3 and t4
## W = 232156, p-value = 1
## alternative hypothesis: true location shift is less than 0
```

```r
#Tests zu t2 und t4
t.test(t2,t4,alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  t2 and t4
## t = -25.665, df = 632.62, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.2905102
## sample estimates:
## mean of x mean of y
## 0.5227776 0.8332124
```

```r
wilcox.test(t2,t4,alternative="less")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  t2 and t4
## W = 33399, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
```

```r
#Tests zu t1 und t4
t.test(t1,t4,alternative="less")
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  t1 and t4
## t = -16.705, df = 789.64, p-value < 2.2e-16
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.1963956
## sample estimates:
## mean of x mean of y
## 0.6153385 0.8332124
```

```r
wilcox.test(t1,t4,alternative="less")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  t1 and t4
## W = 58634, p-value < 2.2e-16
## alternative hypothesis: true location shift is less than 0
```

Der t-Test und der Wilcoxon-Rangsummen-Test kommen bei allen 4 Fällen zur gleichen Testentscheidung

Fall1: Die erste Parameterkombination ist im Mittel signifikant schneller als die dritte.

Fall2: Die dritte Parameterkombination ist nicht signifikant schneller als die vierte und somit wird die Nullhypothese nicht verworfen.

Fall3: Die zweite Parameterkombination ist signifikant schneller als die vierte.

Fall4: Die erste Parameterkombination ist signifikant schneller als die vierte.

Aufgabe 2

Φ sei ein Test für das beschriebene Testproblem. Weil es ein Test zum Niveau α ist, muss die Gütefunktion auf ganz $\Theta \setminus \theta_0 < \alpha$ sein. Sollte der Test für H1, also an der Stelle θ_0, eine Güte$> \alpha$ haben, müsste die Gütefunktion hier sprunghaft an- und direkt wieder absteigen. Dann kann die Gütefunktion aber nicht stetig sein, was dem Hinweis widerspricht.

Inhaltlich bedeutet das zuerst einmal, dass ein sinnvoller Test auf Gleichheit nicht möglich ist. Allgemein ist es nicht möglich eine H_1 mit nur genau einem Punkt zu formulieren. Erweiternd könnte man den Bezug zu Popper herstellenm Falszifikationismus nach Popper ziehen, wonach positivistisches Testen grundsätzlich unmögich ist.

Aufgabe 3

a) Nein, die nachzuweisende Aussage sollte als H_1 formuliert werden. Wird H_0 tatsächlich verworfen, kann man mit der Irrtumswahrscheinlichkeit alpha annehmen, dass H_1 stimmt (Weil der Test entweder die korrekte Entscheidung trifft oder mit alpha den Fehler 1. Art begeht). Andersrum gilt das nicht. Wenn H_0 nicht verworfen werden kann, kann man im Grunde keine Aussage machen. Das wahre theta kann sowohl aus Θ_1 wie auch Θ_0 kommen.Die Irrtumswahrscheinlichkeit ist nicht bezifferbar, weil man den Fehler 2. Art nicht kontrollieren kann.

b) Nein, sie kann nicht sicher sein, dass eine Normalverteilung vorliegt, was an oben beschriebenem Fehler 2. Art liegt, der nicht kontrolliert werden kann. Das gilt ebenso bei p=0.8. Der p-value darf (streng genommen) nicht als Indikator der Evidenz interpretiert werden.

c) Nein. Signifikanz und Relevanz sind zwei verschiedene Dinge. Wissenschaftliche Relevanz basiert zwar häufig auf statistischer Signifikanz, andersrum ist dies aber nicht der Fall. Es dürfte sich leicht zeigen lassen, dass das Coronavirus signifikant kleiner als zB eine Biene ist. Publiziert wird diese Erkenntnis aber bestimmt nicht, weil es keinen Beitrag dazu leistet, die Pandemie nachzuverfolgen. Was wissenschaftliche Relevanz besitzt, muss immer wieder abgewogen werden und recht sicher nicht bezifferbar.