

---

**A1**

*UK Angewandte Statistik*

**Cordula Eggerth**

Wintersemester 2018

---

**Aufgabe 1:**

Gegeben ist eine Grundgesamtheit von 6 Elementen mit jeweils einem Merkmal y:

$U = \{1, 2, 3, 4, 5, 6\}$ ;  $y_1 = 2.6$ ,  $y_2 = 3.4$ ,  $y_3 = 5.7$ ,  $y_4 = 1.3$ ,  $y_5 = 0.9$ ,  $y_6 = 4.1$ .

Daraus sollen Stichproben im Umfang 3 ohne Zurücklegen (Anm.: Reihenfolge unwichtig) gezogen werden.

N ... Grundgesamtheit

y ... Vektor der Ausprägungen

n ... Stichprobenumfang

**a.) Wieviele Stichproben gibt es?**

Möglichkeit 1: Berechnung der Anzahl der möglichen Stichproben mittels der Formel

$$\frac{N!}{(N - n)! * n!}$$

Möglichkeit 2: Berechnung direkt durch den Binomialkoeffizienten („n über k“)

Beide Berechnungsmöglichkeiten ergeben 20 als die Anzahl der Stichproben.

Alle möglichen Stichproben aufgelistet:

```
> stichprobenmatrix <- (combn(y, m=3))
> stichprobenmatrix
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16] [,17] [,18] [,19]
[1,]  2.6  2.6  2.6  2.6  2.6  2.6  2.6  2.6  2.6  2.6  3.4  3.4  3.4  3.4  3.4  3.4  5.7  5.7  5.7
[2,]  3.4  3.4  3.4  3.4  5.7  5.7  5.7  1.3  1.3  0.9  5.7  5.7  5.7  1.3  1.3  0.9  1.3  1.3  0.9
[3,]  5.7  1.3  0.9  4.1  1.3  0.9  4.1  0.9  4.1  4.1  1.3  0.9  4.1  0.9  4.1  4.1  0.9  4.1  4.1
      [,20]
[1,]  1.3
[2,]  0.9
[3,]  4.1
```

**b.) Berechne für jede mögliche Stichprobe den Mittelwert und die Varianz. Bestimme die Verteilung der Mittelwerte und die Verteilung der Varianzen der geschätzten Mittelwerte von y.**

Mittelwerte aller möglichen Stichproben:

```
> mittelwerte_allerStichproben
[1] 3.900000 2.433333 2.300000 3.366667 3.200000 3.066667 4.133333 1.600000 2.666667 2.533333 3.466667
[12] 3.333333 4.400000 1.866667 2.933333 2.800000 2.633333 3.700000 3.566667 2.100000
```

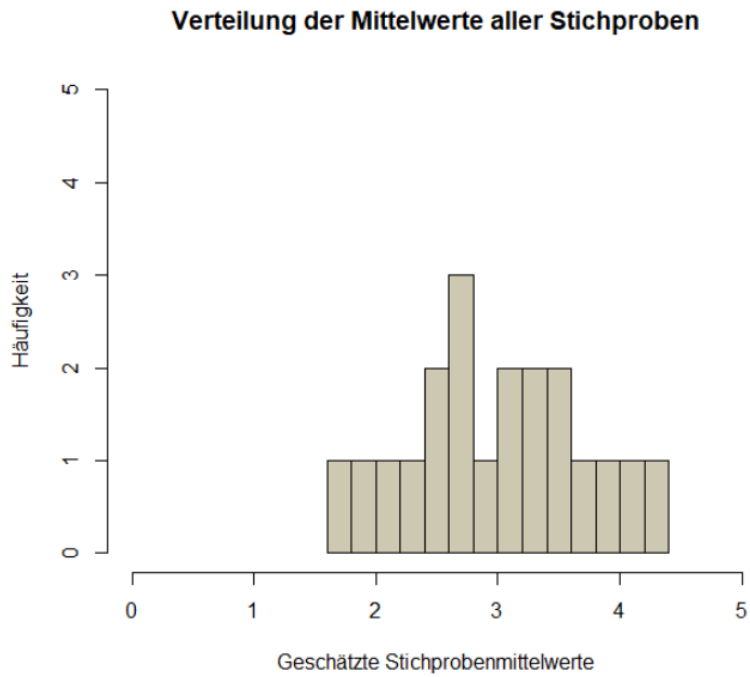
Varianz der obigen Mittelwerte:

```
> varianz_derMittelwerte
[1] 0.565614
```

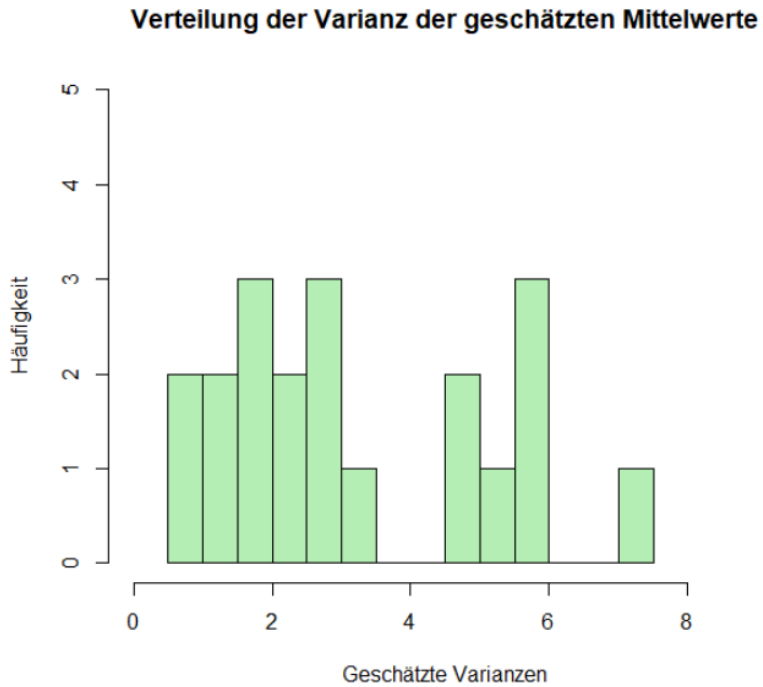
Varianz der einzelnen Stichproben:

```
> varianz_derStichproben
[1] 2.590000 1.123333 1.630000 0.563333 5.110000 5.923333 2.403333 0.790000 1.963333 2.563333
[11] 4.843333 5.763333 1.390000 1.803333 2.123333 2.830000 7.093333 4.960000 5.973333 3.040000
```

Verteilung der Mittelwerte aller möglichen Stichproben:



Verteilung der Varianz aller möglichen Stichproben:



*c.) Welche Relation besteht zwischen dem Mittelwert der Grundgesamtheit und dem Mittelwert der Verteilung der Schätzungen des Mittelwerts und der Verteilung der Schätzungen der Varianzen?*

Berechnung:

```
mittelwert_derGrundgesamtheit <- sum(y) / N
varianz_derGrundgesamtheit <- sd(y)^2
mittelwert_derSchaetzungen <- sum(mittelwerte_allerStichproben) /
                                length(mittelwerte_allerStichproben)
varianz_derSchaetzungen <- sd(varianz_derStichproben)^2
```

Ergebnisse:

```
> mittelwert_derGrundgesamtheit
[1] 3
> varianz_derGrundgesamtheit
[1] 3.224
> mittelwert_derSchaetzungen
[1] 3
> varianz_derSchaetzungen
[1] 3.963099
```

Die Mittelwerte sind also sowohl in der Grundgesamtheit als auch bei den Schätzungen gleich. Die Varianzen sind verschieden, wobei die Varianz der Schätzungen größer ist.

**R-Code zu Aufgabe 1:**

```
rm(list=ls())

### AUFGABE 1 ### -----
# 1. Gegeben ist eine Grundgesamtheit von 6 Elementen mit jeweils einem Merkmal y:
#   U={1,2,3,4,5,6}; y1=2.6, y2=3.4, y3=5.7, y4=1.3, y5=0.9, y6=4.1 .
#   Es werden Stichproben vom Umfang 3 ohne Zurücklegen (anm.: reihenfolge nicht wichtig)
#   gezogen.

# a) Wie viele Stichproben gibt es?

N <- 6
y <- c(2.6, 3.4, 5.7, 1.3, 0.9, 4.1)
n <- 3

# möglichkeit 1: formel  $N! / ((N-n)! * n!)$ 
anzahl_stichproben <- factorial(N) / (factorial(N-n) * factorial(n))
# möglichkeit 2: direkt binomialkoeffizient
anzahl_stichproben_2 <- choose(N, n)
# ANTWORT: 20 mögliche Stichproben (gemäß Daten aus der Angabe)

# b) Berechne für jede mögliche Stichprobe den Mittelwert und die Varianz der (Mittelwerte?)
#     und bestimme die Verteilung der Mittelwerte und die Verteilung der Varianzen der
#     geschätzten Mittelwerte von y.

stichprobenmatrix <- (combn(y, m=3))
mittelwerte_allerStichproben <- colSums(stichprobenmatrix) / n
varianz_derMittelwerte <- sd(mittelwerte_allerStichproben)^2

varianz_stichproben <- function(sample=stichprobenmatrix, means=mittelwerte_allerStichproben){
  varianzen <- rep(0, ncol(stichprobenmatrix))
  for(i in 1:ncol(stichprobenmatrix)){
    varianzen[i] <- sd(stichprobenmatrix[, i])^2
  }
  varianzen
}

varianz_derStichproben <- varianz_stichproben()

par(mfrow=c(1,1))
hist(mittelwerte_allerStichproben, breaks=10, col = "cornsilk3",
     main = "Verteilung der Mittelwerte aller Stichproben",
     xlim = c(0,5), ylim = c(0,5),
     xlab = "Geschätzte Stichprobenmittelwerte", ylab="Häufigkeit")

hist(varianz_derStichproben, breaks=10, col = "darkseagreen2",
     main = "Verteilung der Varianz der geschätzten Mittelwerte",
     xlim = c(0,9), ylim = c(0,5),
     xlab = "Geschätzte Varianzen", ylab="Häufigkeit")

# c) Welche Relation besteht zwischen dem Mittelwert der Grundgesamtheit und
#     dem Mittelwert der Verteilung der Schätzungen des Mittelwerts und der Verteilung der
#     Schätzungen der Varianzen.

mittelwert_derGrundgesamtheit <- sum(y)/N
varianz_derGrundgesamtheit <- sd(y)^2
mittelwert_derSchätzungen <- sum(mittelwerte_allerStichproben)/length(mittelwerte_allerStichproben)
varianz_derSchätzungen <- sd(varianz_derStichproben)^2
```

## Aufgabe 2:

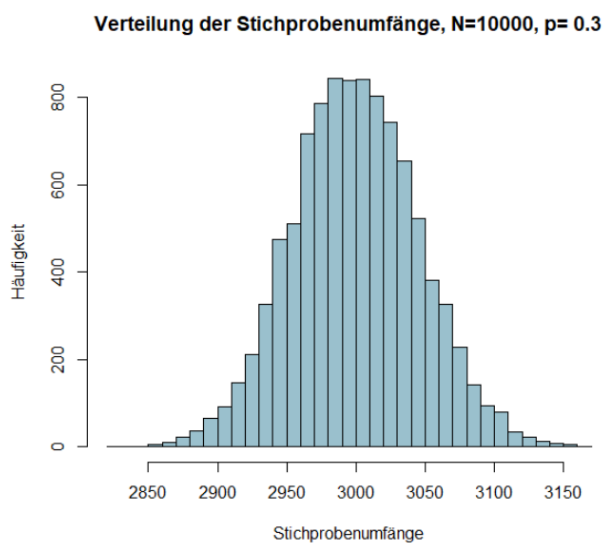
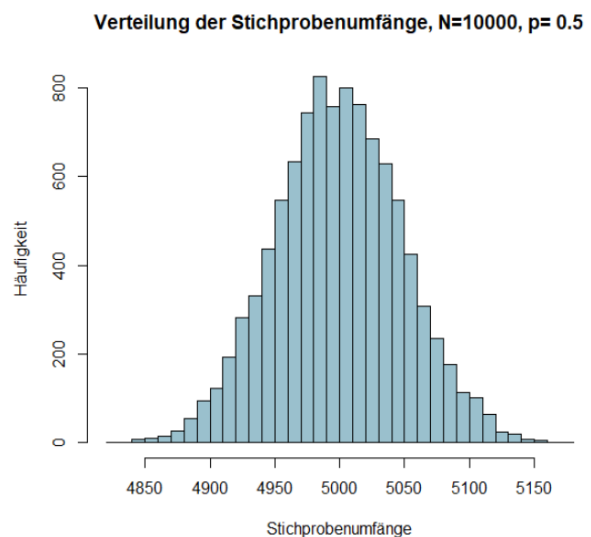
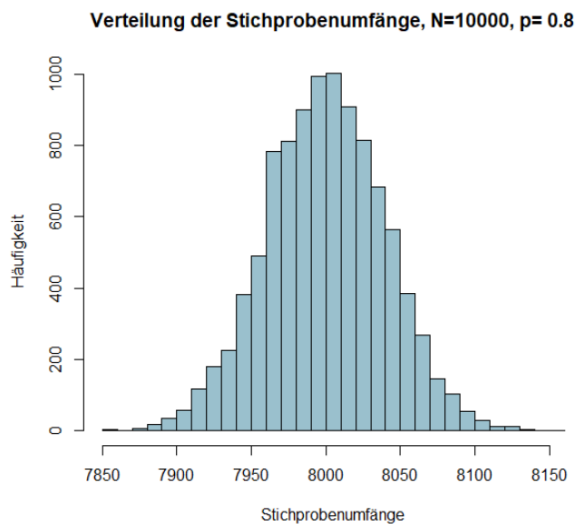
Bernoulli Sampling:

*Man zeige, dass die möglichen Umfänge der Stichprobe bei dieser Methode nach einer Binomialverteilung  $B(N,p)$  verteilt sind.*

Zunächst wird eine Stichprobe gleichverteilter Zufallszahlen vom Umfang  $N$  zwischen 0 und 1 mittels runif-Funktion im R generiert.  $\pi$  (hier als Parameter  $p$ ) ist der Anteil bzw. der Wert, mit dem die Zufallszahlen jeweils verglichen werden. Wenn die Zufallszahl  $u_i$  kleiner als  $p$ , dann wird  $u_i$  die Stichprobe hineingenommen.

Dann werden hier (beispielsweise) 10000 Stichproben mit den Werte  $p=0.8$ ,  $p=0.5$  und  $p=0.3$  gezogen. Für jede dieser Stichproben wird der Umfang ermittelt.

Danach wird die Verteilung der Stichprobenumfänge für die verschiedenen  $p$ -Werte geplottet und man sieht, dass das Histogramm eine Binomialverteilung anmuten lässt.



**R-Code zu Aufgabe 2:**

```
# 2. Eine oft verwendete Möglichkeit zum Ziehen einer Stichprobe ist das sogenannte Bernoulli
# Sampling:
# Eine Grundgesamtheit besteht aus insgesamt N Elementen.
# Will man eine Stichprobe realisieren, die einen Anteil p der Elemente der Grundgesamtheit
# hat, so erzeugt man N gleichverteilte Zufallszahlen u_i und entscheidet nach folgender
# Regel, ob das Element zur Stichprobe gehört:
# Falls (u_i < p), gehört das Element i zur Stichprobe, ansonsten nicht.
# Man zeige, dass die möglichen Umfänge der Stichprobe bei dieser Methode nach einer
# Binomialverteilung B(N,p) verteilt sind.

stichprobe_generieren <- function(N=10000, p=0.3){

  u <- runif(N, min=0, max=1)

  bool_u_kleiner_p <- u<p

  stichprobe <- u[bool_u_kleiner_p]

  stichprobe
}

# stichproben damit generieren:
anzahl_zuGenerierendeStichproben <- 10000
stichprobenumfaenge_p0.8 <- rep(0, anzahl_zuGenerierendeStichproben)
stichprobenumfaenge_p0.5 <- rep(0, anzahl_zuGenerierendeStichproben)
stichprobenumfaenge_p0.3 <- rep(0, anzahl_zuGenerierendeStichproben)

for(j in 1:anzahl_zuGenerierendeStichproben){
  stichprobenumfaenge_p0.8[j] <- length(stichprobe_generieren(10000,0.8))
  stichprobenumfaenge_p0.5[j] <- length(stichprobe_generieren(10000,0.5))
  stichprobenumfaenge_p0.3[j] <- length(stichprobe_generieren())
}

liste_stichprobenumfaenge <- list(stichprobenumfaenge_p0.8,
                                stichprobenumfaenge_p0.5,
                                stichprobenumfaenge_p0.3)

# verteilung der moeglichen stichprobenumfaenge
p_vec <- c(0.8, 0.5, 0.3)
for(i in 1:length(liste_stichprobenumfaenge)){
  hist(liste_stichprobenumfaenge[[i]], breaks=40, col = "lightblue3",
       main = paste("Verteilung der Stichprobenumfänge, N=10000, p=", p_vec[i]),
       xlab = "Stichprobenumfänge", ylab="Häufigkeit")
}
```

### Aufgabe 3:

Dichte der Weibullverteilung:

*Sie hat für die verschiedene Wahl der Parameter unterschiedliche Formen. Visualisieren Sie die Dichten der Weibullverteilung für die angegebenen Parameter.*

Anmerkung:

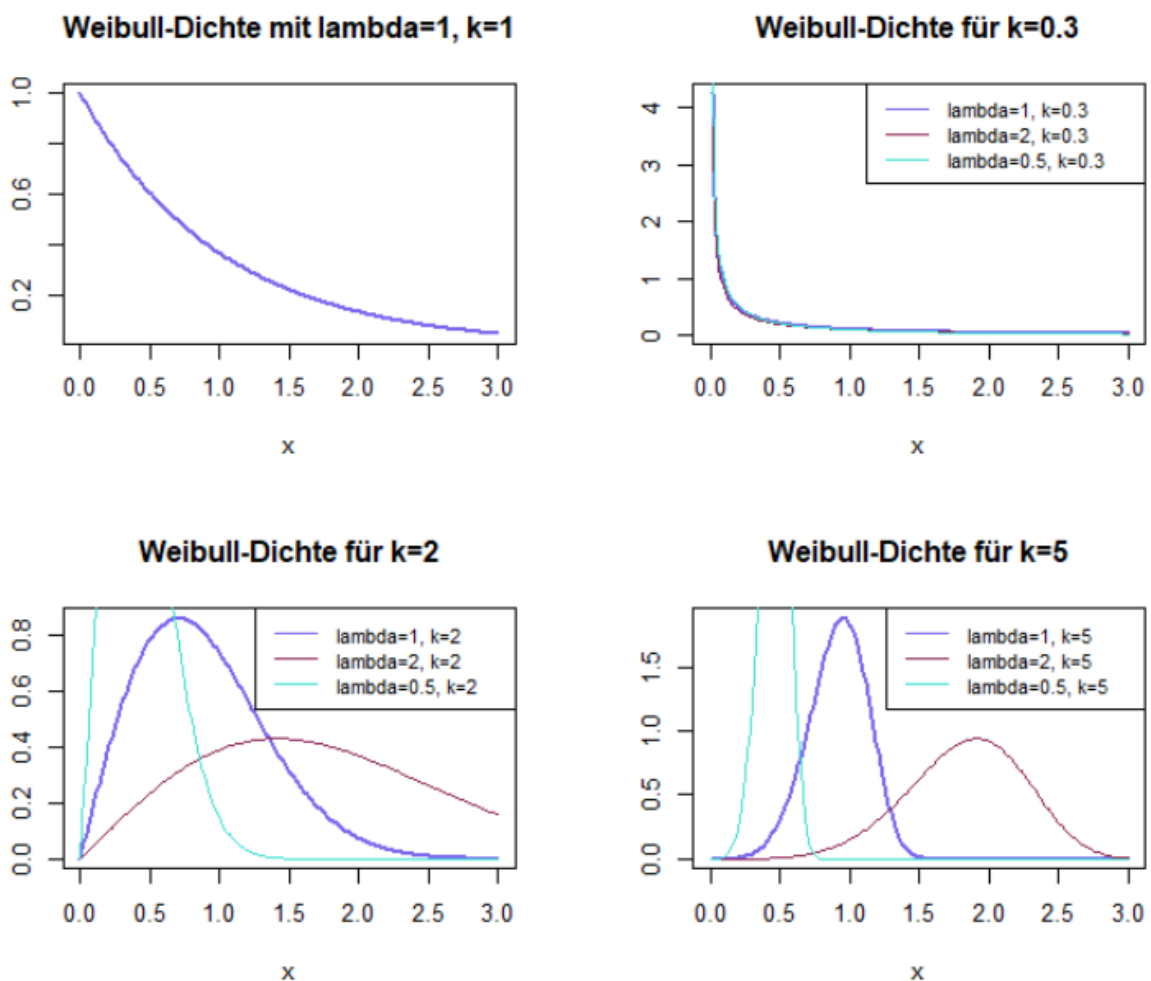
k ... shape

lambda,  $\lambda$  ... scale

Dichte: 
$$f(x; \lambda, k) = \frac{k}{\lambda} \left( \frac{x}{\lambda} \right)^{k-1} e^{-(x/\lambda)^k}, x \geq 0$$

Die **Dichte mit  $\lambda = 1$  sowie  $k = 1$  entspricht der Dichte der Exponentialverteilung.**<sup>1</sup> Im Plot wird dieser Fall oben links dargestellt.

Darstellung der verschiedenen Fälle für die angegebenen Parameterwerte:



<sup>1</sup> Quelle: [http://math.bme.hu/~nandori/Virtual\\_lab/stat/special/Weibull.pdf](http://math.bme.hu/~nandori/Virtual_lab/stat/special/Weibull.pdf).



***Was bedeuten die Parameter  $\lambda$  (bzw. Scale) und  $k$  (bzw. Shape) inhaltlich?<sup>2</sup>***

**$\lambda$  (bzw. Scale):**

Der Scale-Parameter bestimmt, wie weit die Dichte nach oben gezogen bzw. wie weit die Dichtefunktion auf der x-Achse verschoben wird. Je höher Scale, desto höher die Dichte. Die Änderung des Scale-Parameters entspricht einer Verschiebung der x-Koordinate. Bei konstantem  $k$  wird bei zunehmendem  $\lambda$  die Dichte in die Höhe gezogen.

**$k$  (bzw. Shape):**

Der Shape-Parameter bestimmt die Form der Weibulldichte. Der Parameter wird auch „Steigung“ genannt. Dies entspricht der Regressionsgerade im Wahrscheinlichkeitsplot. Wie auch aus den obigen angefertigten Plots ersichtlich, hat der Parameter einen starken Einfluss auf die Form. Beispielsweise für  $k=0.3$  ist keine „Glockenform“ ersichtlich, für  $k=2$  ist die Linie glockenförmiger, aber stark nach links verschoben, für  $k=5$  ist die Glockenform enger und im Vergleich zum vorigen  $k$ -Wert weiter nach rechts gerückt.

---

<sup>2</sup> Quellen für die Beantwortung der Frage: [http://reliawiki.org/index.php/The\\_Weibull\\_Distribution](http://reliawiki.org/index.php/The_Weibull_Distribution);  
<https://www.weibull.com/hotwire/issue14/relbasics14.htm>.

**R-Code zu Aufgabe 3:**

```
par(mfrow=c(2,2))

x <- seq(0, 3, length=200)

# ad 2.: lambda=1, k=1
dichte_weibull <- dweibull(x, shape=1, scale=1, log=FALSE)
plot(x, dichte_weibull, type = "l", col="mediumslateblue", lwd=2,
     main="Weibull-Dichte mit lambda=1, k=1", xlab="x", ylab="")

# ad 1.: lambda=1,k=0.3; lambda=2,k=0.3; lambda=0.5,k=0.3
dichte_weibull_l1 <- dweibull(x, shape=0.3, scale=1, log=FALSE)
dichte_weibull_l2 <- dweibull(x, shape=0.3, scale=2, log=FALSE)
dichte_weibull_l3 <- dweibull(x, shape=0.3, scale=0.5, log=FALSE)

plot(x=x, y=dichte_weibull_l1, type = "l",
     col=c("mediumslateblue"), lwd=2,
     main="Weibull-Dichte für k=0.3", xlab="x", ylab="")
lines(x, dichte_weibull_l2, type = "l", lty = 1, col = "violetred4")
lines(x, dichte_weibull_l3, type = "l", lty = 1, col = "turquoise")

legend("topright", legend=c("lambda=1, k=0.3", "lambda=2, k=0.3", "lambda=0.5, k=0.3"),
     col = c("mediumslateblue", "violetred4", "turquoise"),
     border = "black", lwd=1, cex = 0.8)

# ad 3.: lambda=1,k=2; lambda=2,k=2; lambda=0.5,k=2
dichte_weibull_31 <- dweibull(x, shape=2, scale=1, log=FALSE)
dichte_weibull_32 <- dweibull(x, shape=2, scale=2, log=FALSE)
dichte_weibull_33 <- dweibull(x, shape=2, scale=0.5, log=FALSE)

plot(x=x, y=dichte_weibull_31, type = "l",
     col=c("mediumslateblue"), lwd=2,
     main="Weibull-Dichte für k=2", xlab="x", ylab="")
lines(x, dichte_weibull_32, type = "l", lty = 1, col = "violetred4")
lines(x, dichte_weibull_33, type = "l", lty = 1, col = "turquoise")

legend("topright", legend=c("lambda=1, k=2", "lambda=2, k=2", "lambda=0.5, k=2"),
     col = c("mediumslateblue", "violetred4", "turquoise"),
     border = "black", lwd=1, cex = 0.8)

# ad 4.: lambda=1,k=5; lambda=2,k=5; lambda=0.5,k=5
dichte_weibull_41 <- dweibull(x, shape=5, scale=1, log=FALSE)
dichte_weibull_42 <- dweibull(x, shape=5, scale=2, log=FALSE)
dichte_weibull_43 <- dweibull(x, shape=5, scale=0.5, log=FALSE)

plot(x=x, y=dichte_weibull_41, type = "l",
     col=c("mediumslateblue"), lwd=2,
     main="Weibull-Dichte für k=5", xlab="x", ylab="")
lines(x, dichte_weibull_42, type = "l", lty = 1, col = "violetred4")
lines(x, dichte_weibull_43, type = "l", lty = 1, col = "turquoise")

legend("topright", legend=c("lambda=1, k=5", "lambda=2, k=5", "lambda=0.5, k=5"),
     col = c("mediumslateblue", "violetred4", "turquoise"),
     border = "black", lwd=1, cex = 0.8)
```

---