Übung 2

Explorative Datenanalyse und Visualisierung

Wintersemester 2019
S. Döhler (FBMN, h_da)

Name:

Aufgabe 3. *Erzeugen Sie in R die Folge:* `1 2 3 1 2 3 1 2 3 1 2 3`

(Folgende R -Befehle können hilfreich sein: `rep, seq`*)*

Lösung

```
rep(seq(3),3)
#> [1] 1 2 3 1 2 3 1 2 3
```

Anmerkungen/Korrektur

Aufgabe 4. *Compute using R the 0.99 quantile of the standard normal distribution.*

Lösung

```
qnorm(0.99)
#> [1] 2.326348
```

Anmerkungen/Korrektur

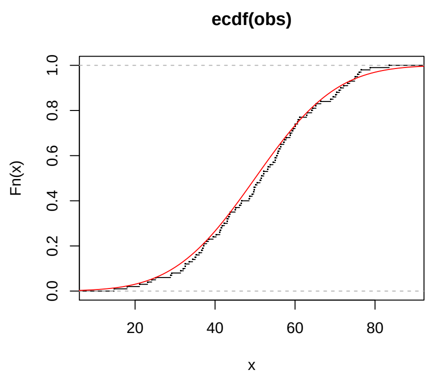
Aufgabe 5. *Simulate 100 observations from $N(50, 16)$. Plot the empirical distribution function for this sample and overlay the true distribution function F.*

(Folgender R -Befehle kann hilfreich sein: `ecdf`*)*

Lösung

```
obs <- rnorm(100, mean = 50, sd = 16)

#obs <- sort(obs)

fun <- ecdf(obs)
```

```
x <- seq(100)
f <- pnorm(x, mean = 50, sd = 16)

{plot(fun, verticals = FALSE, pch = ".")
lines(f, type = 'l', col = "red")}
```



Anmerkungen/Korrektur

Aufgabe 6. *Schreiben Sie eine R -Funktion* `diff.med`*, die zu einem gegebenen Vektor x die Differenz des median und des arithmetischen Mittels ausgibt.*

(a) Simulieren Sie 100 Beobachtungen von $N(50, 16)$ und wenden Sie `diff.med` *auf diesen Vektor an.*

(b) Simulieren Sie 100 Beobachtungen von $LN(1, 2)$ (log-normal) und wenden Sie `diff.med` *auf diesen Vektor an.*

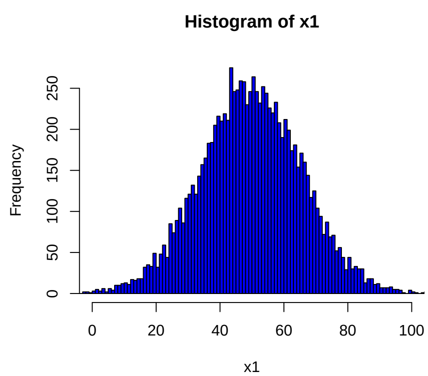
Vergleichen Sie die beiden Ergebnisse. Was würden Sie erwarten? Führen Sie ggf. mehr Simulationen durch und untersuchen Sie die Histogramme, die zu den Daten aus (a) und (b) gehören.
(Folgende R -Befehle können hilfreich sein: `mean, median, hist`*)*

Lösung

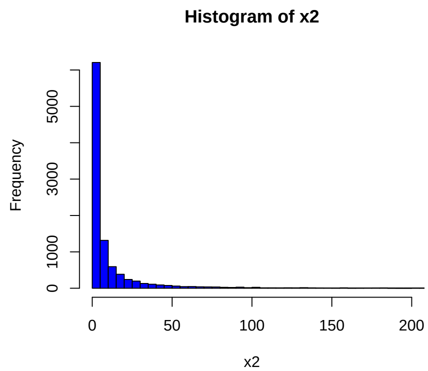
Mean und Median sollten bei einer Normalverteilung nahezu gleich sein (je mehr Observations wir haben, desto näher rücken sie zusammen).

Bei der Log Normal Distribution sollte der Mean über dem Median liegen. Dies kann man ganzt leicht am unten stehenden Histogramm erkennen. Die Log Normal Distribution hat noch einige "Ausreiser", die sich auf den Mittelwert auswirken.

```r
diff.med <- function(x){
  d <- median(x) - mean(x)
  return(d)
}

# part a
x1 <- rnorm(n = 10000, mean = 50, sd = 16)
loe <- diff.med(x1)
loe
#> [1] -0.05654697

# part b
x2 <- rlnorm(10000,1,2)
loeb <- diff.med(x2)
loeb
#> [1] -15.80039

hist(x1, breaks = 100, col = "blue", xlim = c(0,100))
```




```r
hist(x2, breaks = 1000, col = "blue", xlim = c(0,200))
```



Anmerkungen/Korrektur

Aufgabe 7. *Ein Risiko V hängt von einem anderen Risiko X folgendermaßen ab:*

$$V = e^X + 2X^2 \qquad mit \qquad X \sim N(0,1).$$

Schreiben Sie eine R -Funktion `Sim.V`, die zu einem gegebenen Niveau $\alpha \in (0,1)$ und gegebener Anzahl N_{Sim} von Simulationen von V das α-Quantil der Verteilung von V schätzt.

(Folgende R -Befehle können hilfreich sein: `quantile`)

Lösung

Anmerkungen/Korrektur