Tut3

Aufgabe 1: Aktienrenditen und Value-at-Risk (VaR)

Im Finanzsektor ist der Value at Risk (VaR) ein Standardmaß, um das Risiko eines Portfolios bzw. einer Investition zu messen. Der Value at Risk gibt an, wie groß der Verlust einer Investition (mit einer bestimmten Wahrscheinlichkeit) in einem bestimmten Zeitraum, z.B. einem Tag, sein kann. Wenn z.B. eine Aktie einen eintägigen 95 %-VaR von -0.02 aufweist, bedeutet dies, dass die Aktie mit einer Wahrscheinlichkeit von 0.05 innerhalb eines Tages um mehr als 2 % an Wert verliert.

a)

Wir lesen zunächst Daten zum Schlusskurs der Amazon Aktie ein:

```
url <- paste0(
  "https://raw.githubusercontent.com/rfordatascience/",
  "tidytuesday/master/data/2023/2023-02-07/",
  "big_tech_stock_prices.csv"
)

stock_prices <- read.csv(url)

amazon_prices <- stock_prices[stock_prices$stock_symbol == "AMZN", ]
```

Geben Sie sich per view(amazon_prices) den Datensatz aus. Für welchen Tag liegt die erste Beobachtung vor? Und für welchen Tag die letzte?

```
View(amazon_prices)
```

Die erste Beobachtung liegt für den 4. Januar 2010 vor.

Die letzte Beobachtung liegt für den 29. Dezember 2022 vor.

b)

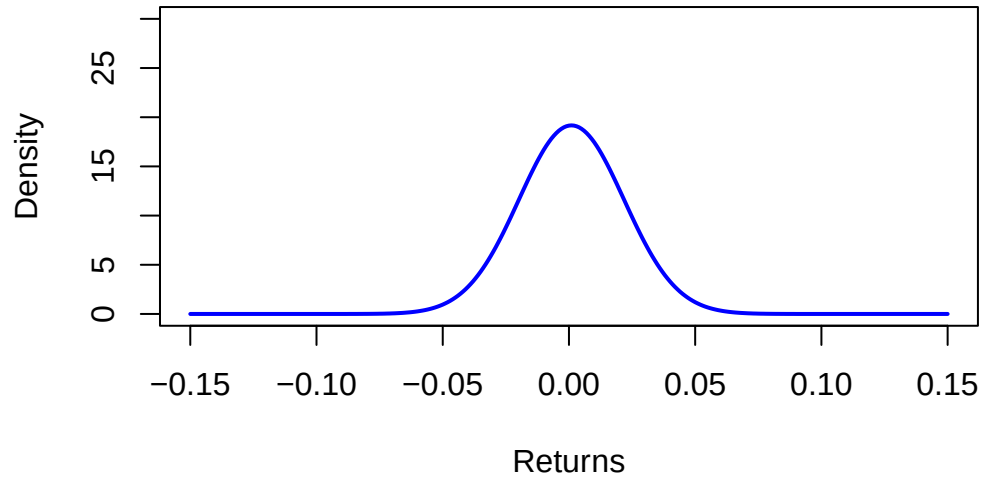
Um uns den VaR näher anzuschauen, berechnen wir zunächst die Tagesrenditen $y_t = \frac{p_t}{p_{t-1}} - 1$, $t = 2, \ldots, 3271$, wobei p_t den Schlusskurs an Tag t bezeichnet. In R können wir y_t wie folgt berechnen:

```
amazon_returns = (amazon_prices$adj_close[2:3271] /
                    amazon_prices$adj_close[1:3270]) - 1
```

c)

Um eine Verteilung für die Tagesrenditen der Amazon-Aktie zu schätzen, könnte man eine Normalverteilung verwenden. Schätzen Sie die Parameter μ und σ der Normalverteilung und plotten Sie die geschätzte Verteilung.

```
mu_hat <- mean(amazon_returns)
sigma_hat <- sd(amazon_returns)
x_seq <- seq(-0.15, 0.15, by = 0.001)
plot(x_seq, dnorm(x_seq, mu_hat, sigma_hat), col = "blue",
type = "l", ylim = c(0, 30), lwd = 2, xlab = "Returns", ylab = "Density")
```

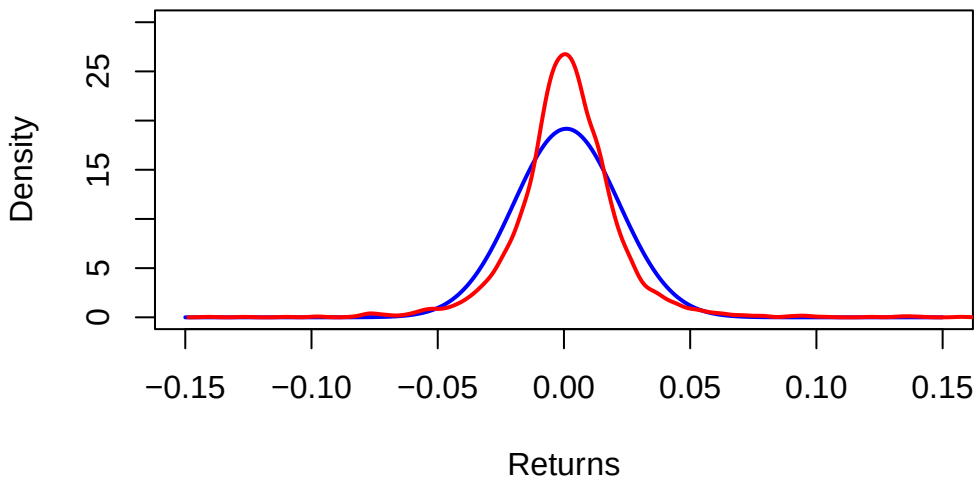


d)

Fügen Sie Ihrem Plot aus c) eine Kerndichteschätzung der Renditen hinzu. Halten Sie die Normalverteilungsannahme für plausibel?

```
plot(x_seq, dnorm(x_seq, mu_hat, sigma_hat), col = "blue",
type = "l", ylim = c(0, 30), lwd = 2, xlab = "Returns", ylab = "Density")

lines(density(amazon_returns), col = "red", lwd = 2)
```



Der Kerndichteschätzer zeigt eine höhere Dichte um den Nullpunkt. Das bedeutet, dass Renditen nahe 0 häufiger Auftreten als durch die zuvor bestimmte Normalverteilung erwartet. Der Kerndichteschätzer zeigt also eine positive Kurtosis (Wölbung). Der Kerndichteschätzer ist besonders am linken Rand fat tailed, was auf gelegentlich auftretenden negative Renditen hinweist.

e)

Für den VaR sind insbesondere die Werte sehr weit links auf der x-Achse relevant, d.h. die negativen Renditen. Sind Renditen von weniger als -5 % unter der Normalverteilung oder unter dem KDE wahrscheinlicher? Welche praktische Auswirkung hat die Verwendung der Normalverteilung hier hinsichtlich des VaR?

Siehe d): stark negative Renditen von weniger als -5% werden von der Normalverteilung als weniger wahrscheinlich eingeschätzt als vom KDE. Würde man in der Praxis auf die Normalverteilung vertrauen, würde man "zu selten" von negativen returns $<$ -5% ausgehen; diese kommen häufiger vor als von der Normalverteilung impliziert. Das heißt: Man würde das Risiko, dass die Aktie an einem Tag um mehr als -5% sinkt, unterschätzen.

Aufgabe 2: Poissonregression

In dieser Aufgabe beschäftigen wir uns mit dem Zusammenhang der Anzahl ausgeliehener Fahrräder an einer Verleihstation in Washington, D.C., USA und verschiedenen erklärenden Variablen. Der Datensatz enthält die Temperatur (temperature, in Grad Fahrenheit), die Windgeschwindigkeit (windspeed, in Meilen pro Stunde), ob der Tag auf ein Wochenende gefallen ist (weekend) und die Anzahl ausgeliehener Fahrräder (count). Wir lesen zunächst die Daten ein:

```
bikes = read.csv("https://tinyurl.com/5bkjupn2")
attach(bikes)
```

a)

Schätzen Sie mithilfe der glm()-Funktion das Poissonregressionsmodell

$$\text{count}_i \sim \text{Po}(\lambda_i), \quad \lambda_i = e^{\beta_0 + \beta_1 \,\text{temperature}_i + \beta_2 \,\text{windspeed}_i + \beta_3 \,\text{weekend}_i}.$$

Beschreiben Sie stichwortartig die geschätzten Effekte der erklärenden Variablen (positiv/negativ, signifikant/insignifikant?).

```
mod = glm(count ~ temperature + windspeed + weekend, family = poisson,
          data = bikes)
summary(mod)
```

```
Call:
glm(formula = count ~ temperature + windspeed + weekend, family = poisson,
    data = bikes)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-13.0599   -2.8609   -0.4275    2.0934   10.6108
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.3782223  0.0253155  172.95   <2e-16 ***
temperature  0.0162645  0.0003077   52.86   <2e-16 ***
windspeed   -0.0132829  0.0012019  -11.05   <2e-16 ***
weekend     -0.3410637  0.0099013  -34.45   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 10853.3  on 364  degrees of freedom
Residual deviance:  5926.8  on 361  degrees of freedom
AIC: 8425.8

Number of Fisher Scoring iterations: 4
```

Alle Schätzer sind statistisch signifikant!

Die Temperatur hat einen positiven Effekt auf die Anzahl an geliehenen Fahrrädern. Die Windgeschwindigkeit und "Wochenende" haben einen negativen Effekt auf die Anzahl geliehener Fahrräder.

b)

Geben Sie für die Variablen temperature und windspeed eine quantitative Interpretation der geschätzten Effekte an.

```
exp(mod$coef[1])
```

```
(Intercept)
   79.69623
```

```
exp(mod$coef[2])
```

```
temperature
   1.016397
```

```r
exp(mod$coef[3])
```

```
windspeed
0.9868049
```

```r
exp(mod$coef[4])
```

```
  weekend
0.7110136
```

$count_i$ verändert sich multiplikativ um den Faktor e^{β_k}. Bei einem Temperaturanstieg von einem Grad Fahrenheit verändert sich die Schätzung der geliehenen Räder um $count_i \cdot e^{\beta_{\text{temperature}}}$.

Für GLMs gilt: $\beta > 0 \rightarrow y$ steigt
$\beta < 0 \rightarrow y$ sinkt.

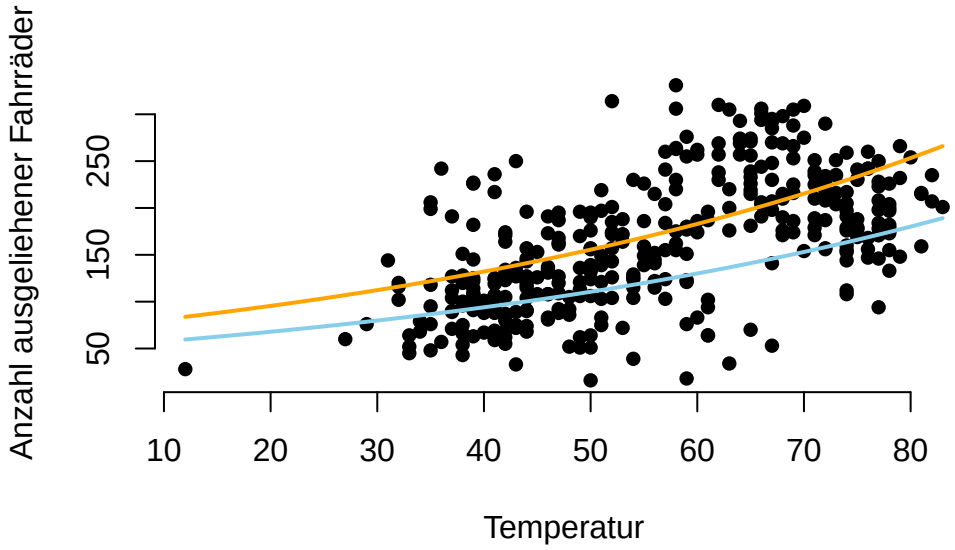
c)

Erstellen Sie einen Scatterplot von count vs. temperature. Ergänzen Sie dann mithilfe der curve()-Funktion die in a) geschätzte Regressionsfunktion (als Funktion von temperature) i) für Wochentage und ii) für Wochenenden, wobei die beiden Regressionsfunktionen verschiedene Farben haben sollen. Für die Windgeschwindigkeit können Sie die mittlere Windgeschwindigkeit über alle Beobachtungen verwenden.

```r
plot(bikes$temperature, bikes$count, bty = "n", pch = 16, xlab = "Temperatur",
     ylab = "Anzahl ausgeliehener Fahrräder")

curve(exp(mod$coef[1] + mod$coef[2] * x + mod$coef[3] * mean(windspeed) +
            mod$coef[4] * 0), add = TRUE, lwd = 2, col = "orange") # Wochentage

curve(exp(mod$coef[1] + mod$coef[2] * x + mod$coef[3] * mean(windspeed) +
            mod$coef[4] * 1), add = TRUE, lwd = 2, col = "skyblue") # Wochenenden
```



d)

Für kommenden Samstag sind eine Temperatur von 40 Grad Fahrenheit und eine Windgeschwindigkeit von 20 Meilen pro Stunde vorhergesagt. Wie hoch ist die erwartete Anzahl ausgeliehener Fahrräder? Geben Sie die Verteilung der Anzahl ausgeliehener Fahrräder gemäß des Modells aus a) an.

```
exp(mod$coef[1] + mod$coef[2] * 40 + mod$coef[3] * 20 + mod$coef[4] * 1)
```

```
(Intercept)
   83.26928
```

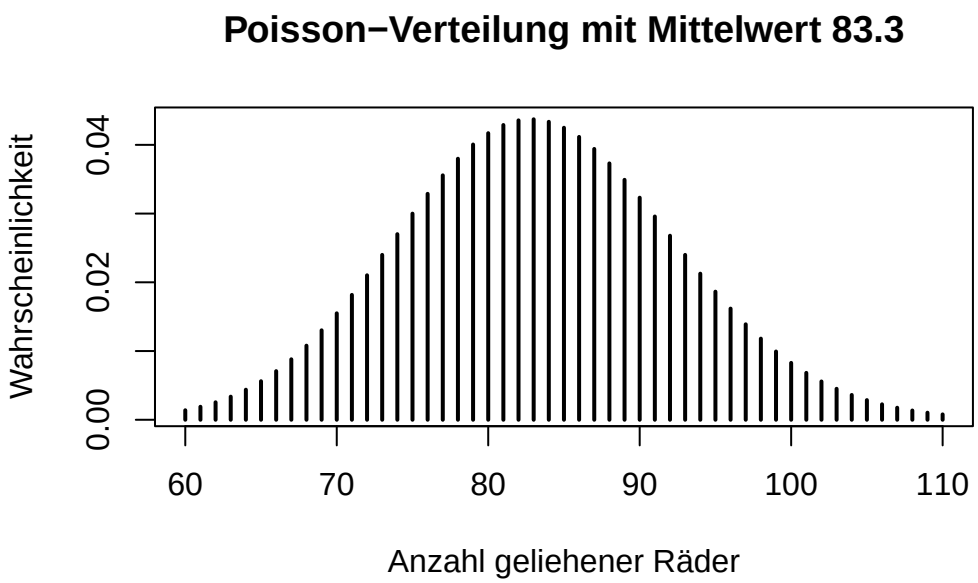
$$\text{count}_i \sim \text{Poisson}(\lambda), \quad \lambda = 83{,}3$$

```
x_vals <- 60:110

y_vals <- dpois(x_vals, lambda = 83.3)

plot(x_vals, y_vals, type = "h", lwd = 2,
```

```
     xlab = "Anzahl geliehener Räder",
     ylab = "Wahrscheinlichkeit",
     main = "Poisson-Verteilung mit Mittelwert 83.3")
```



e)

Bestimmen Sie die von dem Modell implizierte Wahrscheinlichkeit, dass kommenden Samstag mehr als 75 (mehr als 100) Fahrräder ausgeliehen werden. Die Wahrschein lichkeitsfunktion der Poissonverteilung ist in R unter ppois() verfügbar.

```
1 - ppois(q = 75, lambda = 83.3)
```

```
[1] 0.8023143
```

```
1 - ppois(q = 100, lambda = 83.3)
```

```
[1] 0.03270774
```

Die Wahrscheinlichkeit, dass mehr als 75 Räder geliehen werden liegt bei etwa 80%.

Die Wahrscheinlichkeit, dass mehr als 100 Räder geliehen werden liegt bei etwa 3%.