

# Bericht Steinschlagrisiko

## Challenge CWM1

Livio Prosdocimo, Roberto Lorusso, Linus Ackermann

09.06.2024

## Inhaltsverzeichnis

<b>Problem</b>	<b>2</b>
<b>Wahrscheinlichkeitsmodellierung</b>	<b>3</b>
Zufallsvariablen . . . . .	3
Verteilungen . . . . .	3
Verteilungen Zone 1 . . . . .	3
Zone 1 Masse . . . . .	4
Zone 1 Geschwindigkeit . . . . .	8
Zone 1 Zeitabstände . . . . .	12
Verteilungen Zone 2 . . . . .	16
Zone 2 Masse . . . . .	16
Zone 2 Geschwindigkeit . . . . .	20
Zone 2 Zeitabstände . . . . .	24
<b>Simulation</b>	<b>29</b>
Generieren . . . . .	29
Wahrscheinlichkeitsberechnung . . . . .	30
Aussagekraft der Simulation . . . . .	31
<b>Ergebniss</b>	<b>32</b>
<b>Empfehlung</b>	<b>33</b>

## Problem

# Wahrscheinlichkeitsmodellierung

## Zufallsvariablen

## Verteilungen

Da Steinschläge über einen Zeitraum von drei Monaten aufgrund der begrenzten Anzahl an Ereignissen keine verlässlichen Aussagen über die aktuelle Gefahr ermöglichen, müssen Fälle nachgestellt werden. Um eine solche Simulation korrekt durchzuführen, werden die vorhandenen Daten der Steinschläge in die drei oben genannten Variablen Masse, Geschwindigkeit und zeitliche Abstände unterteilt.

Dann können diese Daten anhand ihren Eigenschaften simuliert werden. Dafür wird für jede Kategorie eine Wahrscheinlichkeitsverteilung zugeordnet. Durch die Zuordnung können die Eigenschaften der Daten anhand einem theoretischen Modell widerspiegelt werden und Fälle nachstellen, die der Vergangenheit am nächsten kommen.

## Verteilungen Zone 1

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v lubridate  1.9.3      v tibble     3.2.1
v purrr      1.0.2      v tidyr      1.3.1
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
Loading required package: MASS
```

Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

```
select
```

Loading required package: survival

## Zone 1 Masse

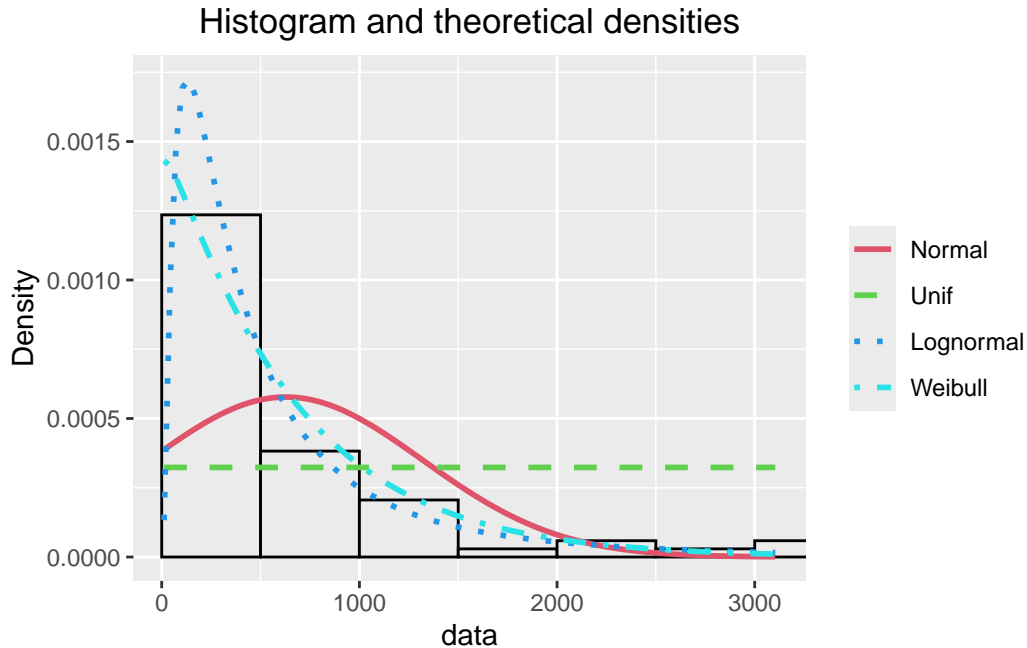
Zunächst wird der Hang 1 hinsichtlich der Masse untersucht.

Zuerst werden mit dem `fitdistr`-Package verschiedene Verteilungen auf den Datensatz der Masse angewendet um diese dann zu analysieren.

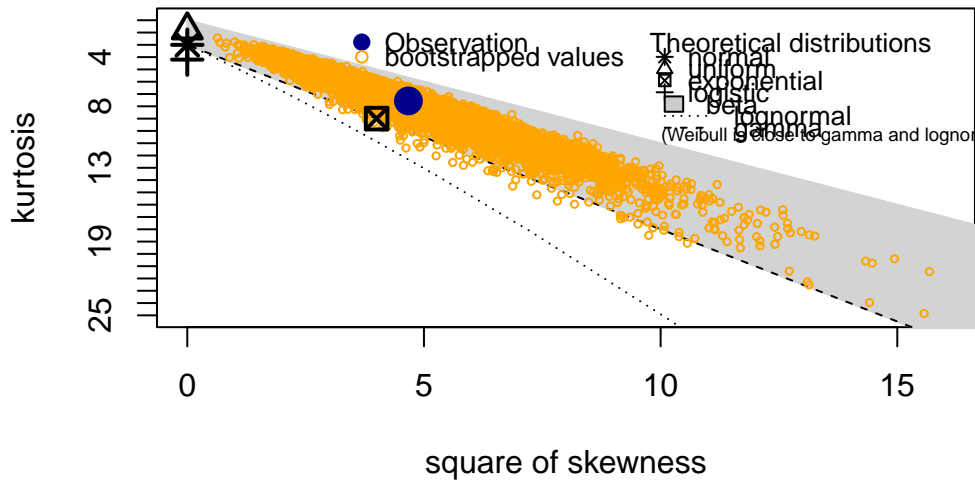
Alle Verteilungen werden dann mit einem Histogramm welches die Daten widerspiegelt angezeigt. In einem zweiten Plot, der ein Cullen-Frey-Diagramm darstellt, werden die Datenpunkte zusammen mit den Verteilungen verglichen. Die Verteilungen, die am nächsten an den Beobachtungspunkten liegen, sind tendenziell kompatibler.

Diese beiden Vergleiche allein erlauben noch keine endgültige Entscheidung darüber, welche Verteilung am besten geeignet ist. Sie dienen jedoch dazu, die Auswahl der möglichen Verteilungen einzuschränken und eine Orientierung zu geben, mit welchen Daten gearbeitet wird.

Dabei zeigt sich, dass nur die Weibull- und die Lognormalverteilung in Frage kommen könnten.



## Cullen and Frey graph



### summary statistics

-----

```
min: 12    max: 3104
median: 402.5
mean: 628.6324
estimated sd: 695.8847
estimated skewness: 2.161327
estimated kurtosis: 7.560251
```

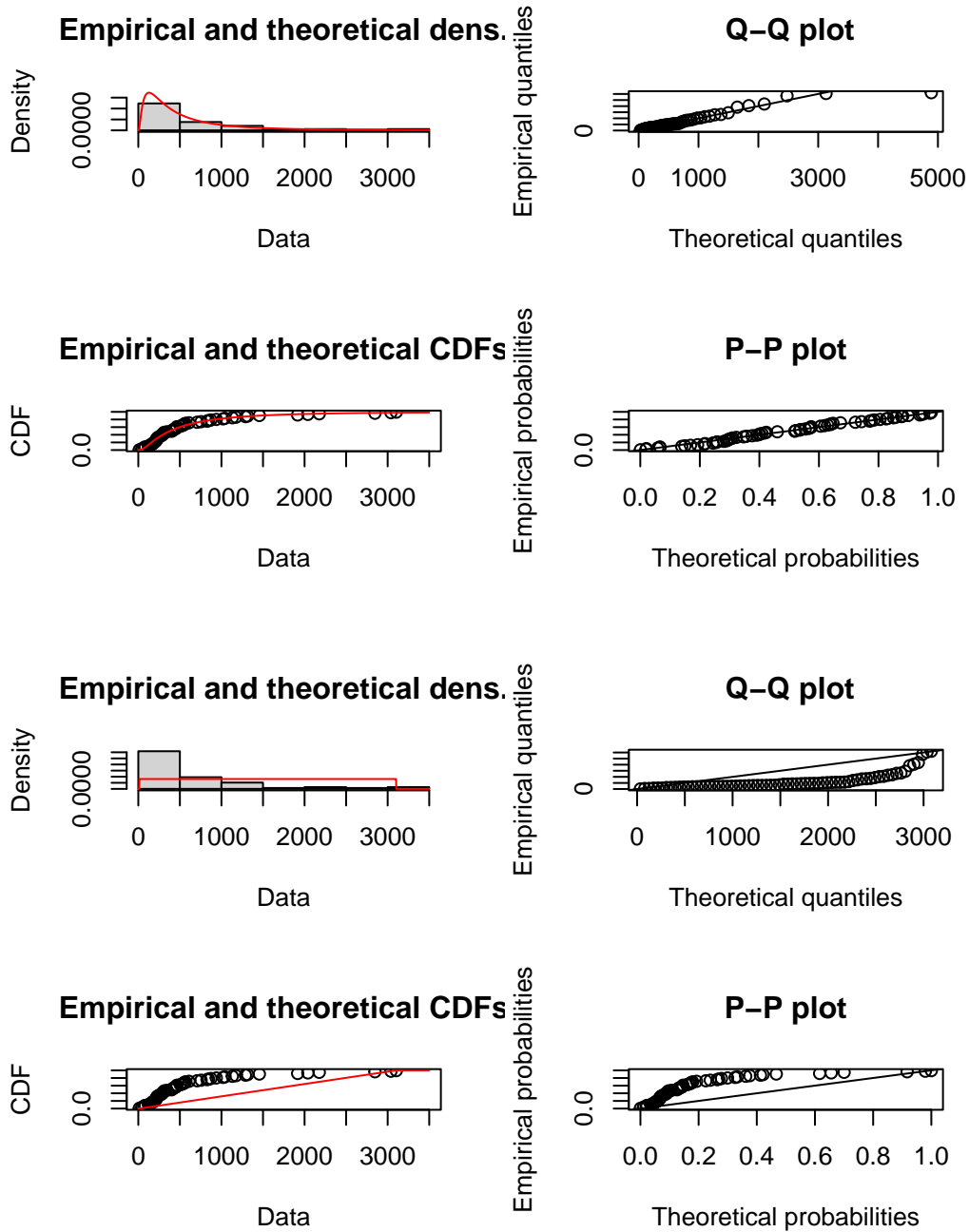
Diese werden in einem Goodness-of-Fit-Test verglichen. Dazu müssen die CDF-, QQ- und PP-Plots betrachtet werden. Die Punkte in den Plots stellen den Vergleich der vorhandenen Daten auf der Y-Achse dar, im Vergleich zu den theoretischen Werten auf der X Achse, welche die Verteilung erzeugt.

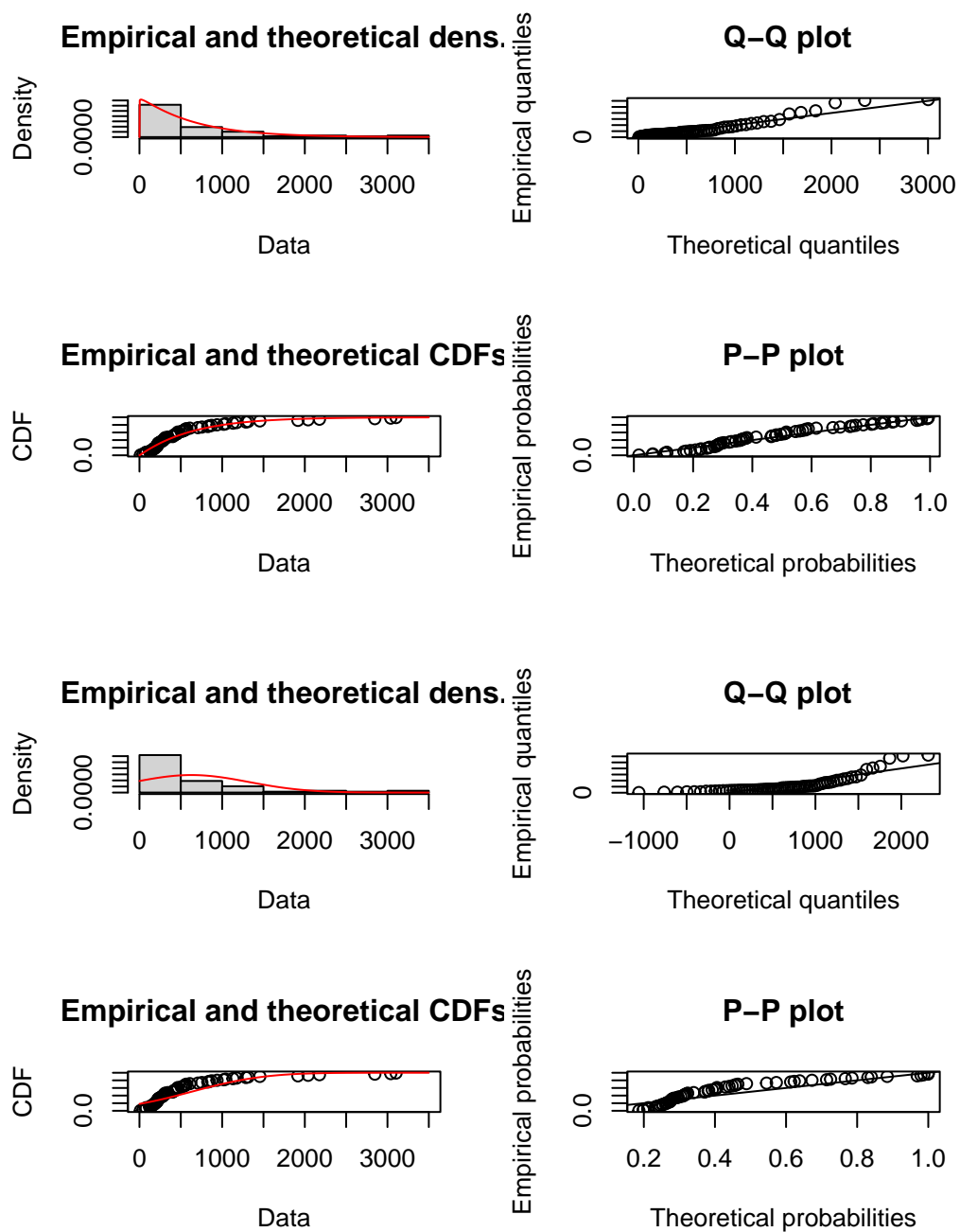
Der QQ-Plot vergleicht die Quantile, der PP-Plot die Perzentile und der CDF-Plot die kumulative Verteilungsfunktion. Wie gut eine Verteilung zu den Daten passt, erkennt man an den Linien. Wenn die Punkte der Gegenüberstellung auf der Linie liegen, passt die Verteilung gut zu den Daten.

Mit der Funktion `UnivariateML` können verschiedene Verteilungen bereits im Vorhinein getestet werden, um herauszufinden, welche am besten zu den Daten passt. Diese Funktion vergleicht die Kompatibilität der Daten mit verschiedenen Verteilungen und bietet somit eine

Möglichkeit, eine mögliche Passform für die vorliegenden Daten zu ermitteln. Diese Funktion unterstützt die Auswahl von Verteilungen

In diesem Fall wird die Lognormalverteilung empfohlen.





Maximum likelihood estimates for the Loglogistic model

shape	rate
1.723909	0.002596

Vergleicht man die Masse am Hang 1 zwischen der Lognormalverteilung und der Weibullverteilung, so liegt die Lognormalverteilung nahe an der Linie, nur das letzte Quantil im QQ-Plot weicht deutlich ab. Bei der Weibullverteilung weichen die Punkte ebenfalls von der Linie ab, aber insbesondere das letzte Quantil liegt nahe an der Linie.

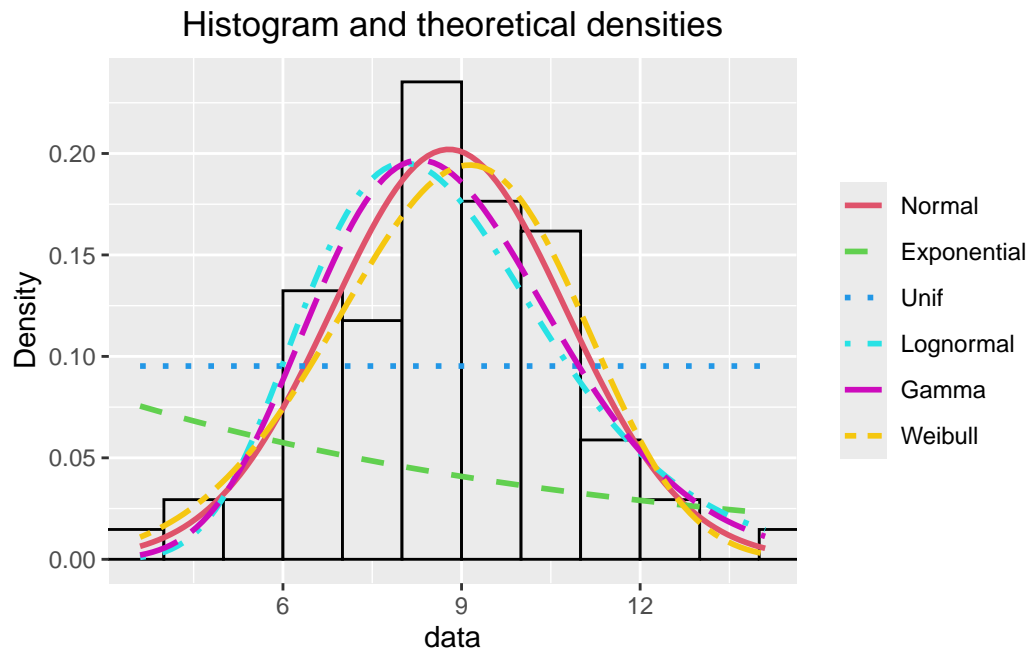
Trotz kleiner Abweichungen wird die Weibull-Verteilung gewählt. Größere Massen bedeuten ein höheres Risiko, daher ist es wichtiger, die schwersten Massen mit der Verteilung abzudecken.

Dieser Vorgang wird für die Auswahl bei allen Variablen wiederholt.

## Zone 1 Geschwindigkeit

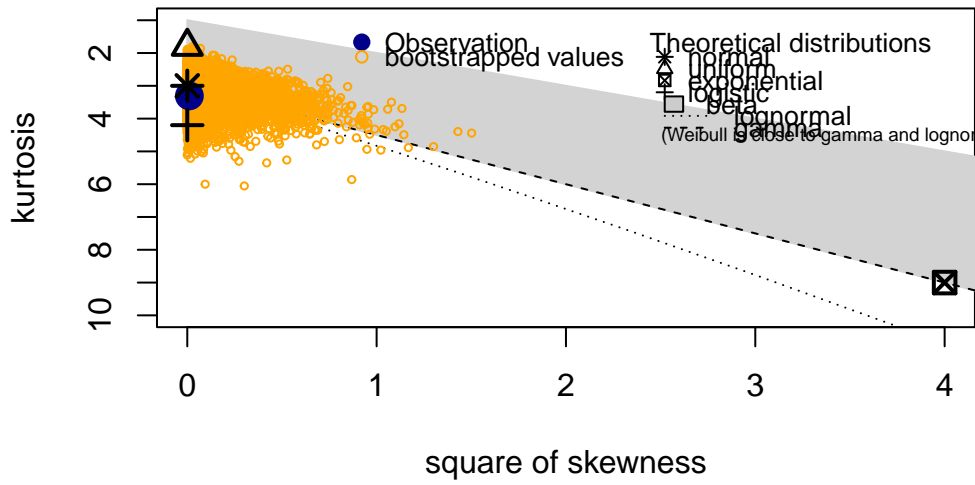
Durch die erste Analyse stellt sich heraus, dass viele Verteilungen passen könnten, Normalverteilung, Weibull, Exponential, Gamma oder Lognormal.

Der Vorschlag von UnivariateML ist die Normalverteilung.





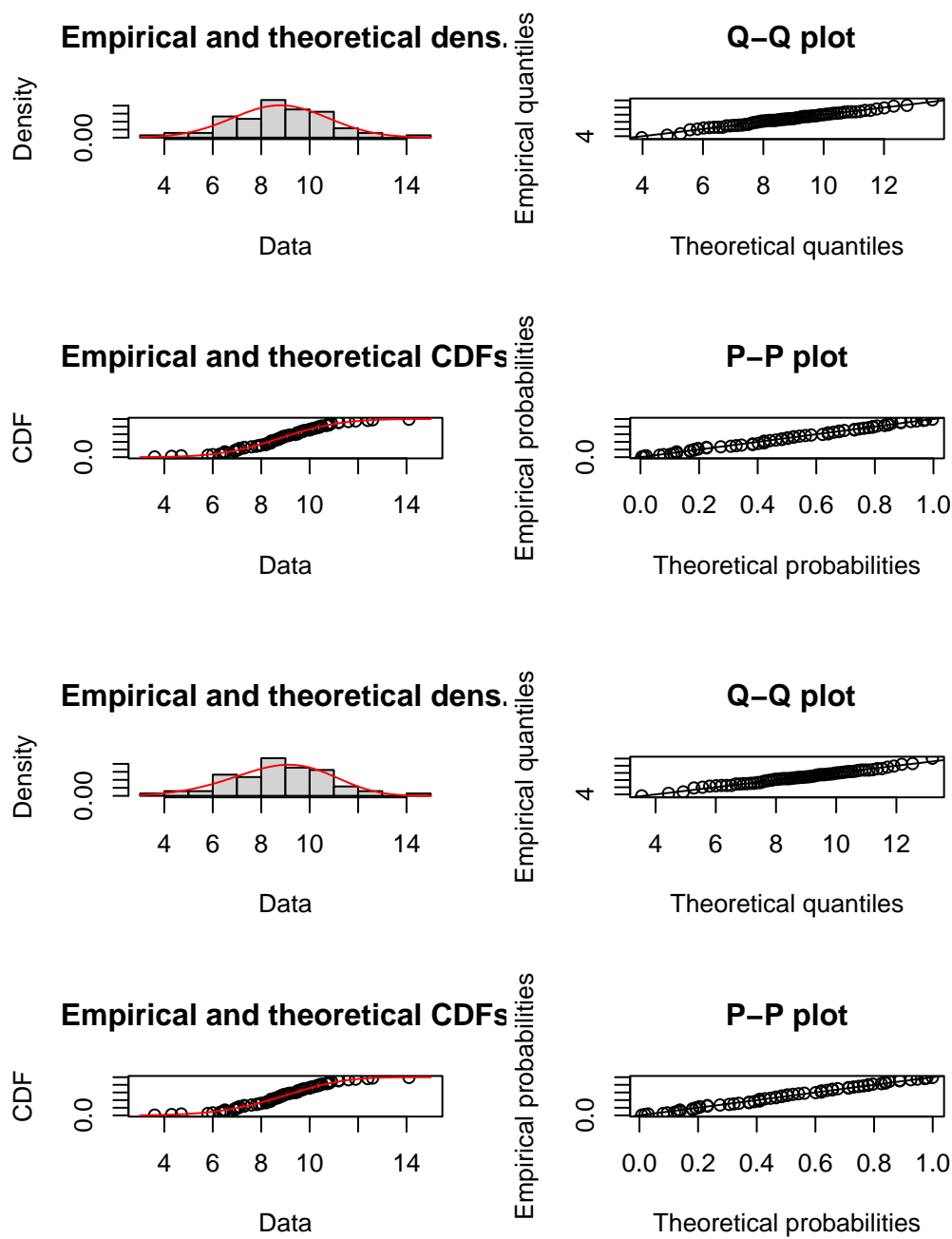
## Cullen and Frey graph

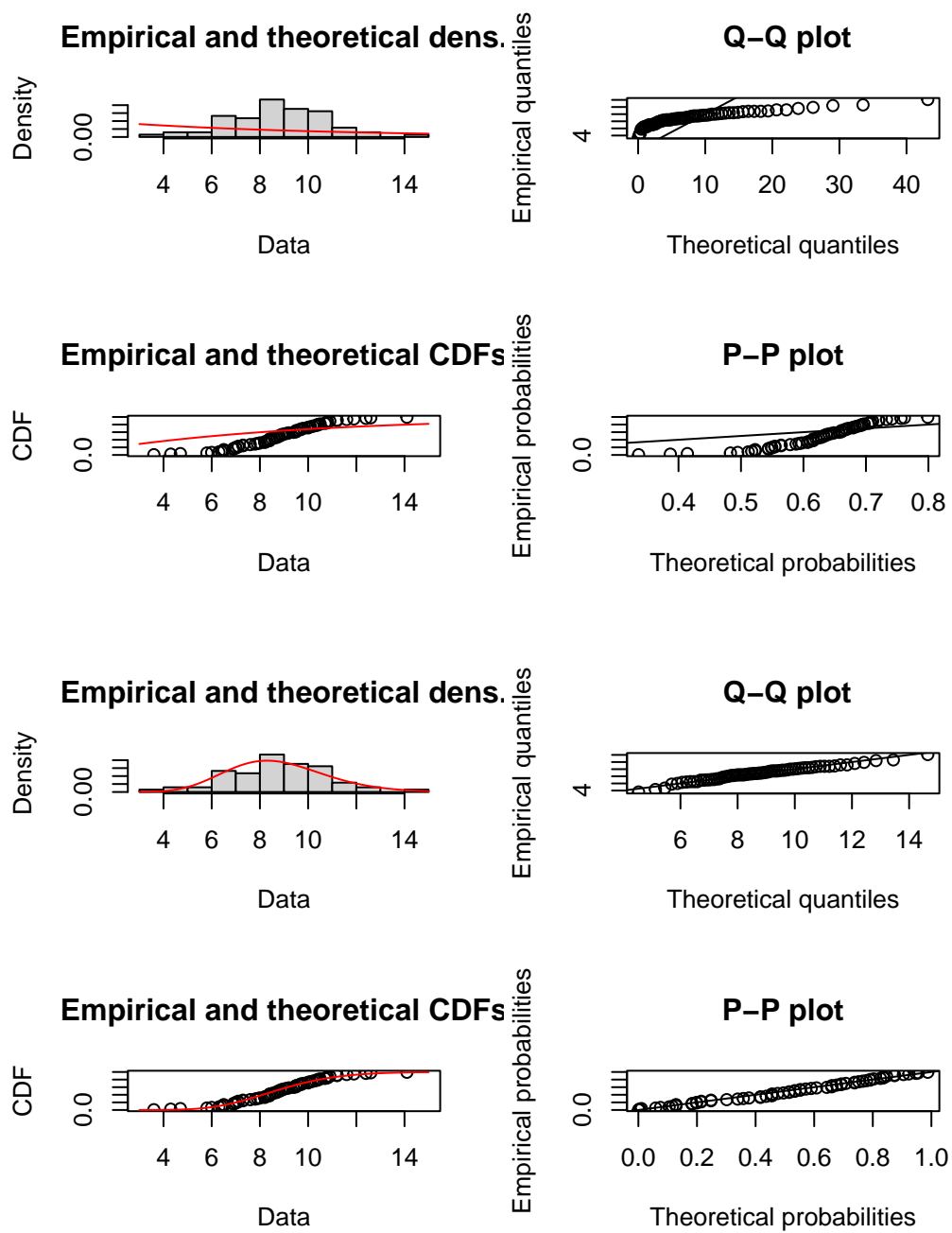


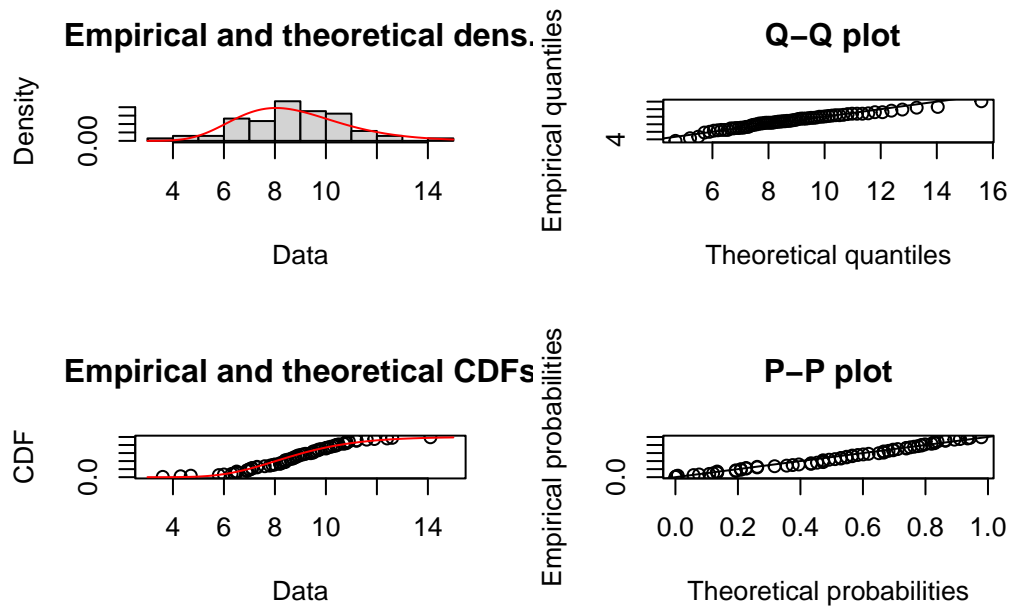
### summary statistics

-----

```
min:  3.6   max: 14.1
median: 8.8
mean:  8.788235
estimated sd:  1.989189
estimated skewness: -0.105241
estimated kurtosis:  3.30699
```







Maximum likelihood estimates for the Normal model

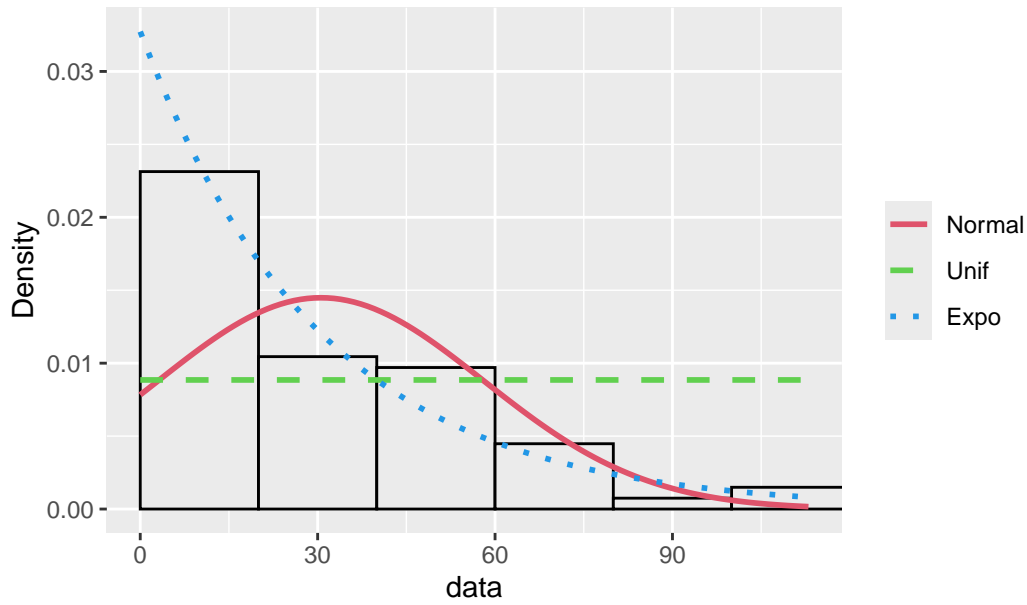
mean	sd
8.788	1.975

Nach dem Abgleich der Plots ist klar, dass die Normalverteilung am genauesten passt. Die Punkte verlaufen fast durchgehend entlang der Linie, was auf eine gute Übereinstimmung mit den Daten hinweist. Besonders in den oberen Quantilen ist diese Verteilung am nächsten bei der Linie.

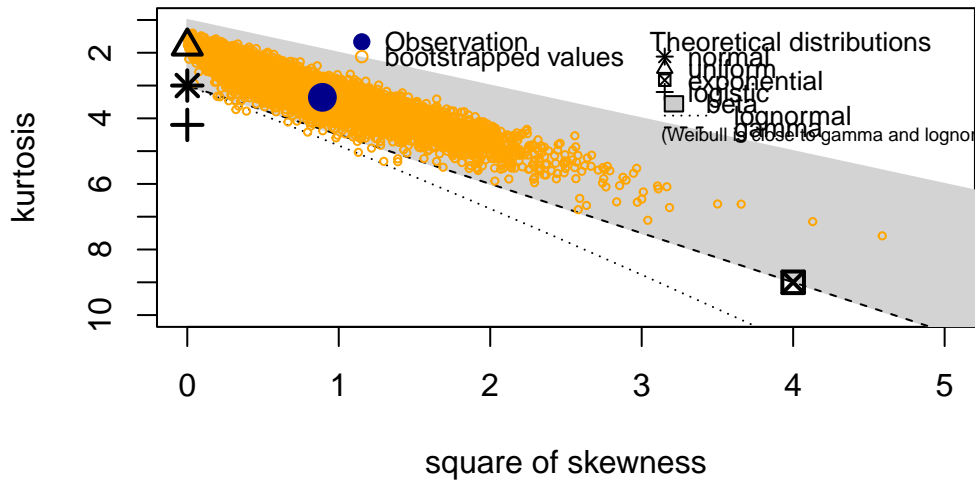
## Zone 1 Zeitabstände

Die erste Analyse deutet in diesem Fall klar darauf hin, dass nur eine Exponential- oder Normalverteilung passend sein könnten. Der Vorschlag von UnivariateML ist die Exponentialverteilung.

### Histogram and theoretical densities



### Cullen and Frey graph

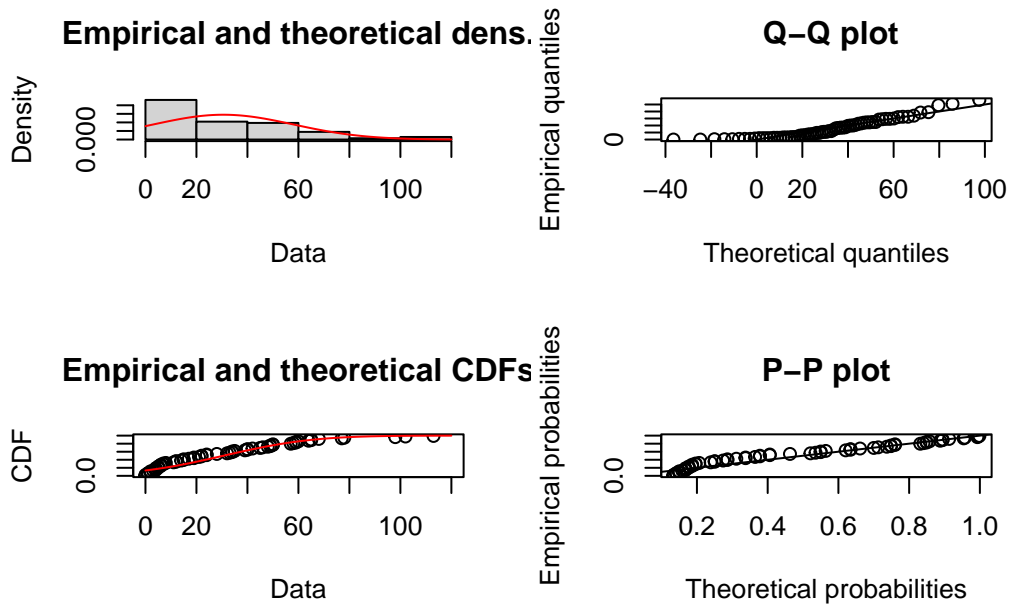


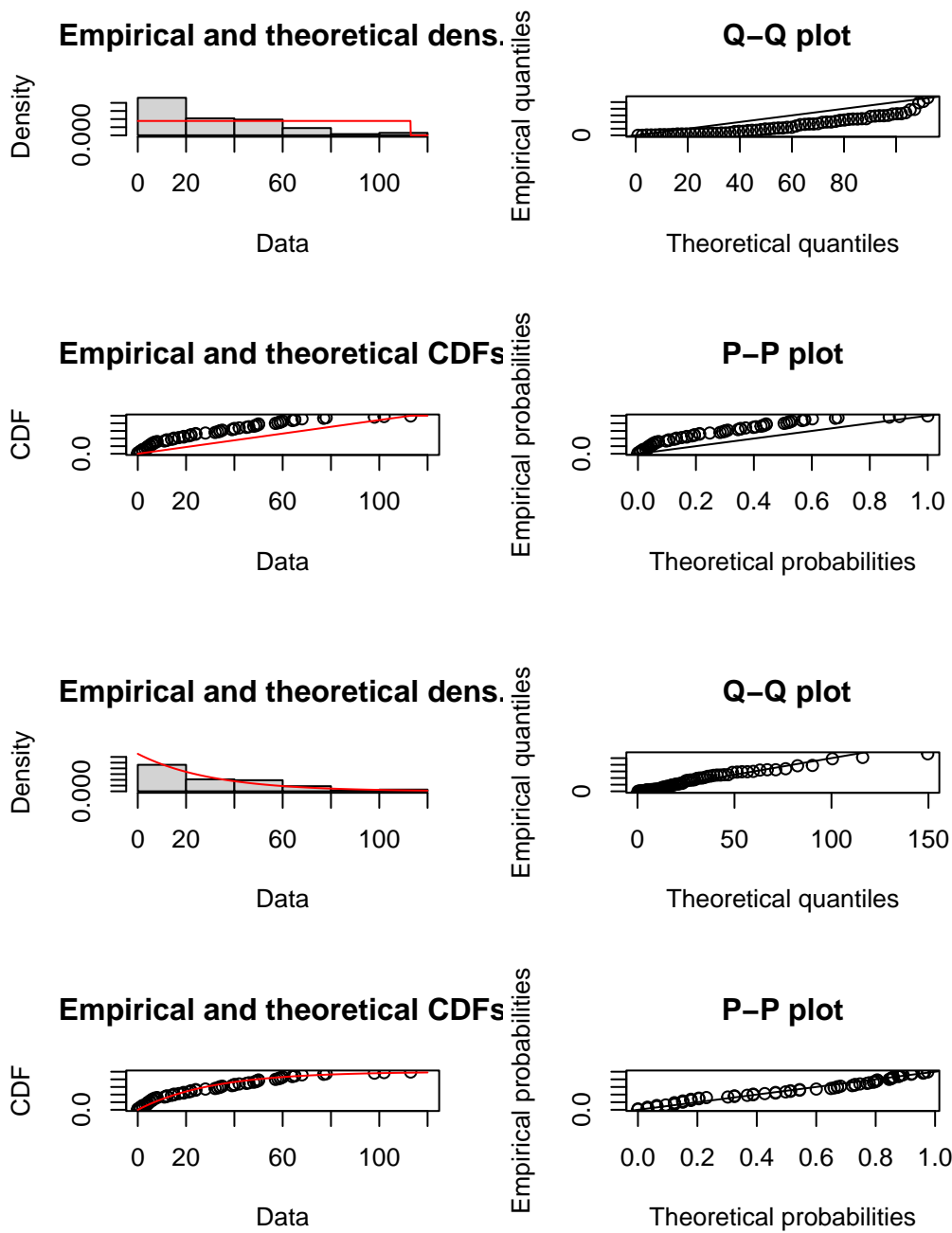
summary statistics

-----

min: 0 max: 113

median: 22  
mean: 30.55224  
estimated sd: 27.74903  
estimated skewness: 0.944373  
estimated kurtosis: 3.359452





Maximum likelihood estimates for the Exponential model  
rate  
0.03273

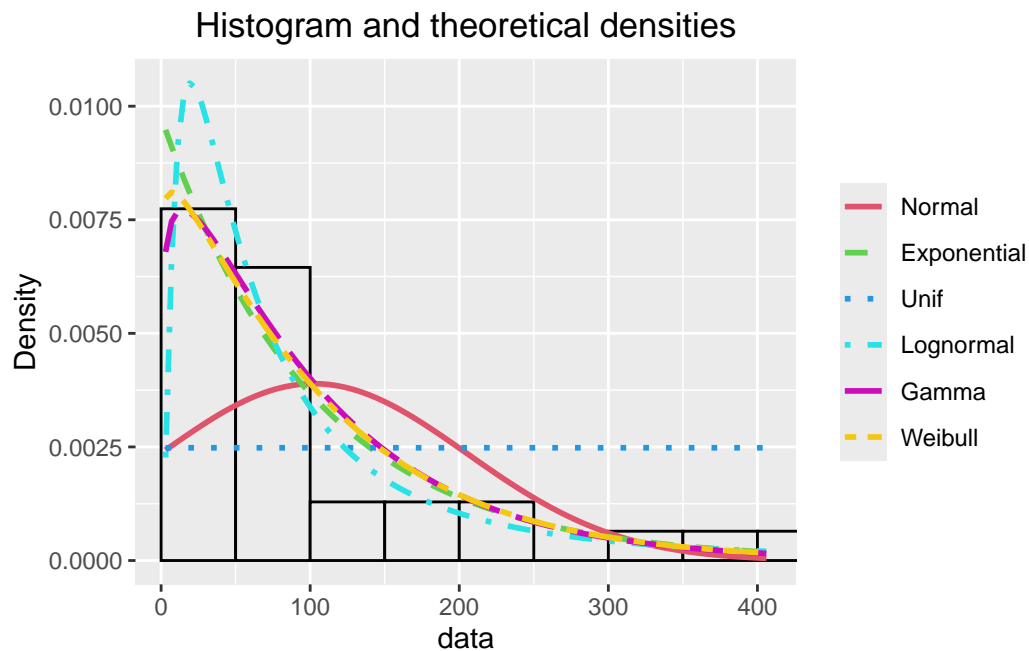
Die Auswertung der Plots zeigt deutlich, dass die Exponentialverteilung am besten passt.

Die kleineren Quantile werden bei der Exponentialverteilung besser gedeckt. Bei den Zeitlichen Abstände sind die kleineren Quantile wichtiger, da weniger Zeitintervalle mehrere Steinschläge bedeuten und somit eine größere Gefahr darstellen.

## Verteilungen Zone 2

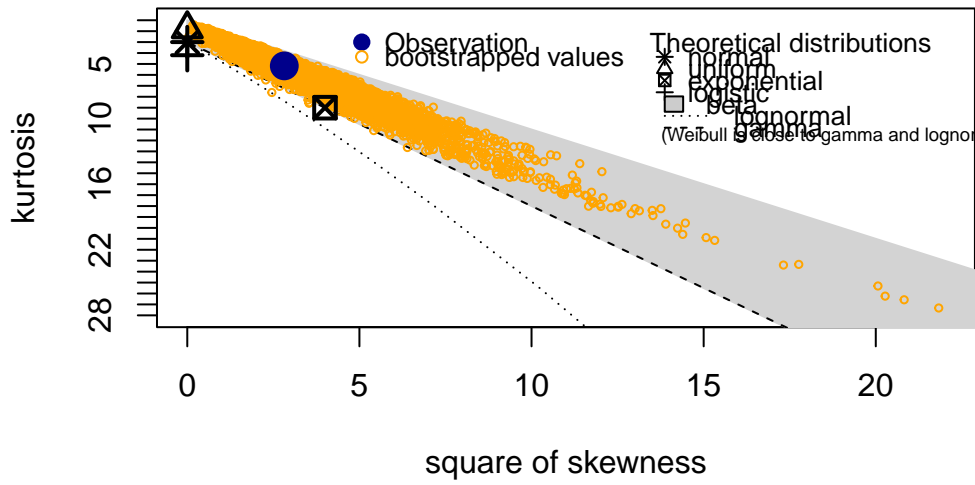
### Zone 2 Masse

Nach einer ersten Prüfung bleiben die Verteilungen Weibull, Gamma, Exponential und Lognormal zur Auswahl. Der Vorschlag von UnivariateML ist die Exponentialverteilung.





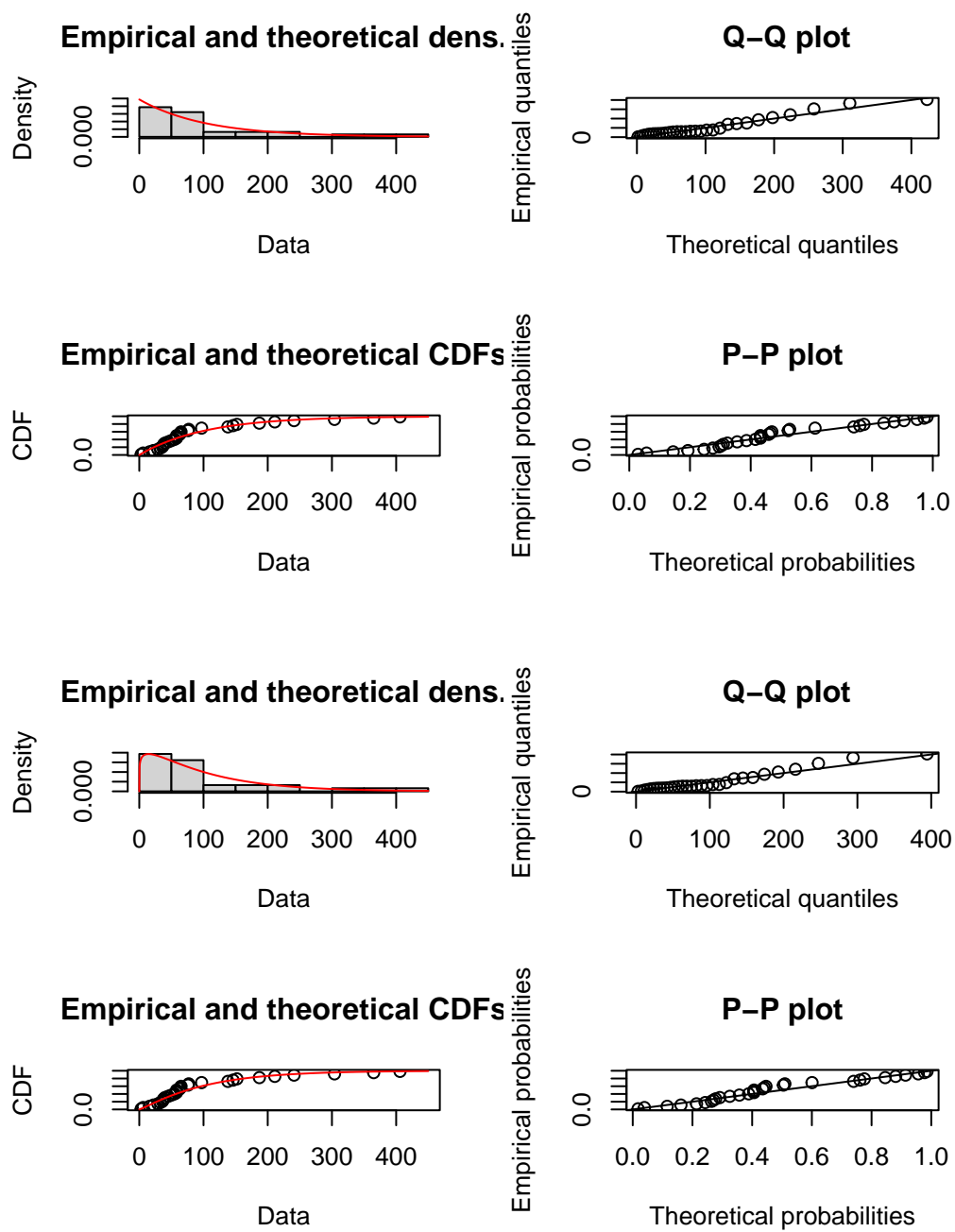
## Cullen and Frey graph

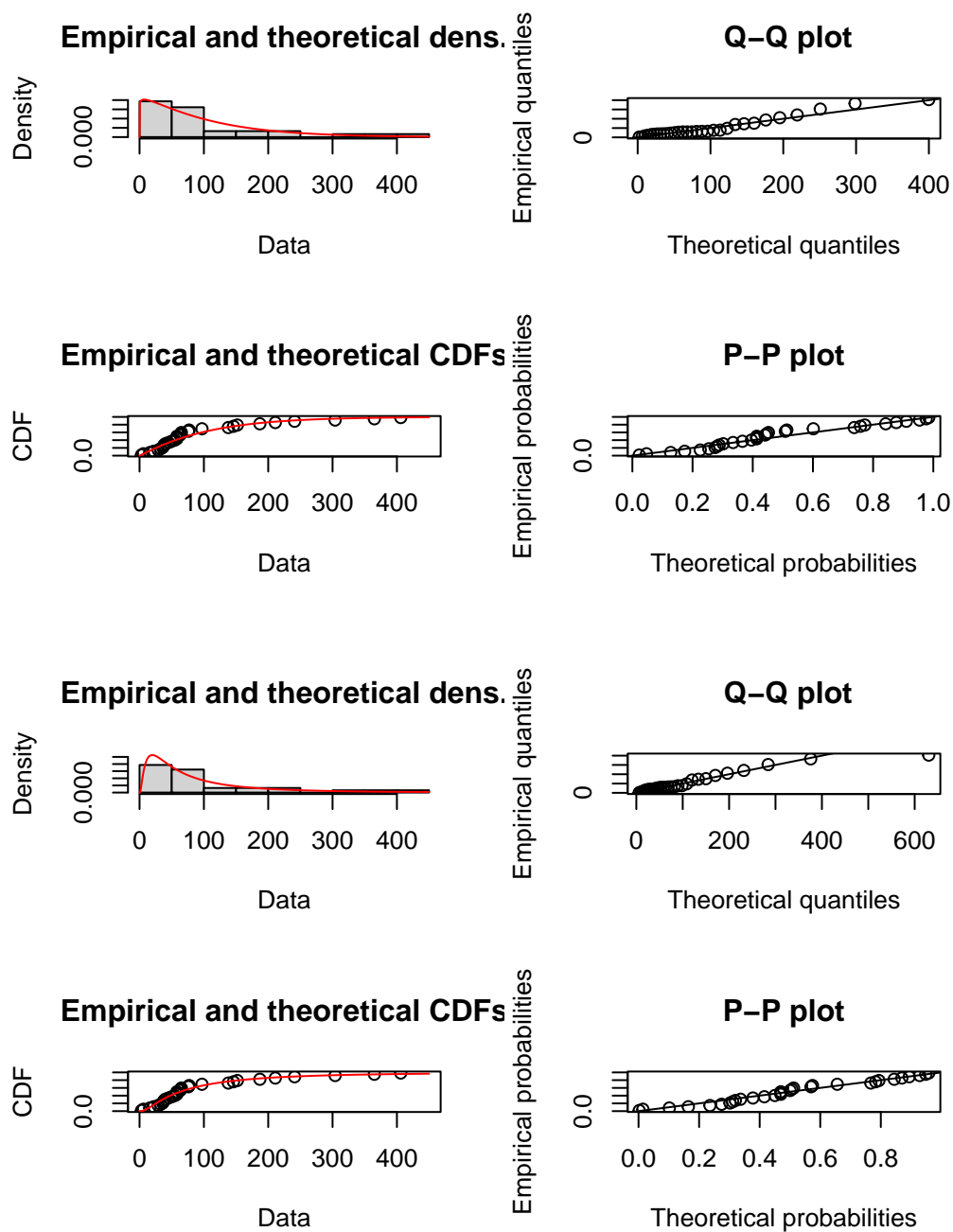


### summary statistics

-----

```
min: 3    max: 406
median: 58
mean: 102.4516
estimated sd: 104.1786
estimated skewness: 1.678832
estimated kurtosis: 5.174858
```





Maximum likelihood estimates for the Exponential model

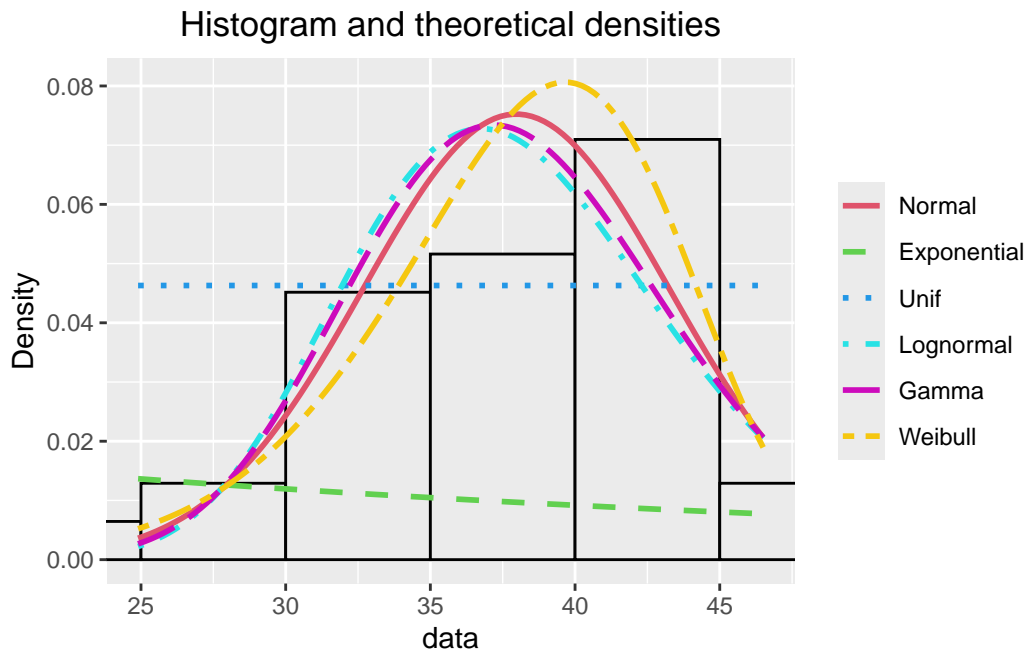
```
rate
0.009761
```

Nach Bewertung der Plots wird die Exponentialverteilung gewählt. Gamma, Weibull sind ähnlich wie die Exponentialverteilung, decken aber die oberen Quantile schlechter ab. Lognormal liegt im Allgemeinen näher an der Linie, lässt aber das letzte Quantil aus. Auch am zweiten Hang ist die größere Masse ausschlaggebender.

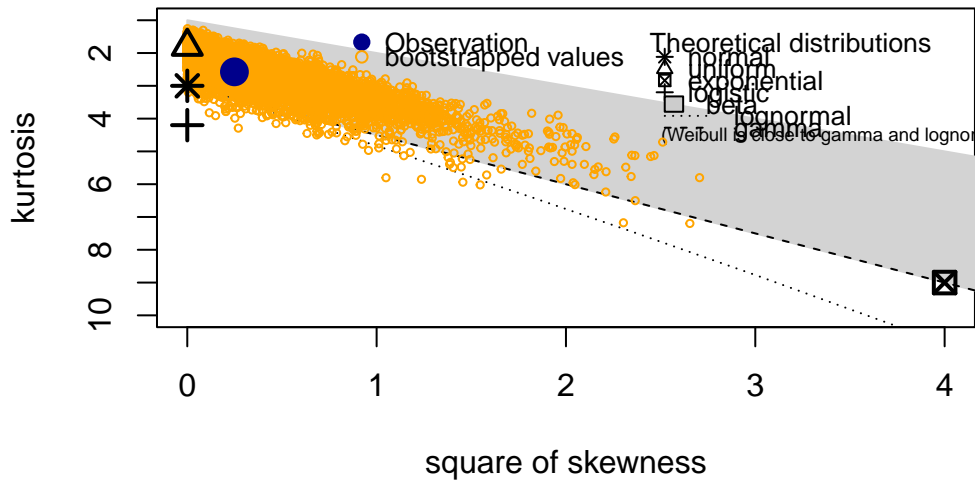
## Zone 2 Geschwindigkeit

Die Diagramme zeigen, dass viele Verteilungen ausgewählt werden können. Normal, Lognormal, Weibull und Gamma stehen zur Auswahl.

Der Vorschlag von UnivariateML ist die Weibullverteilung.



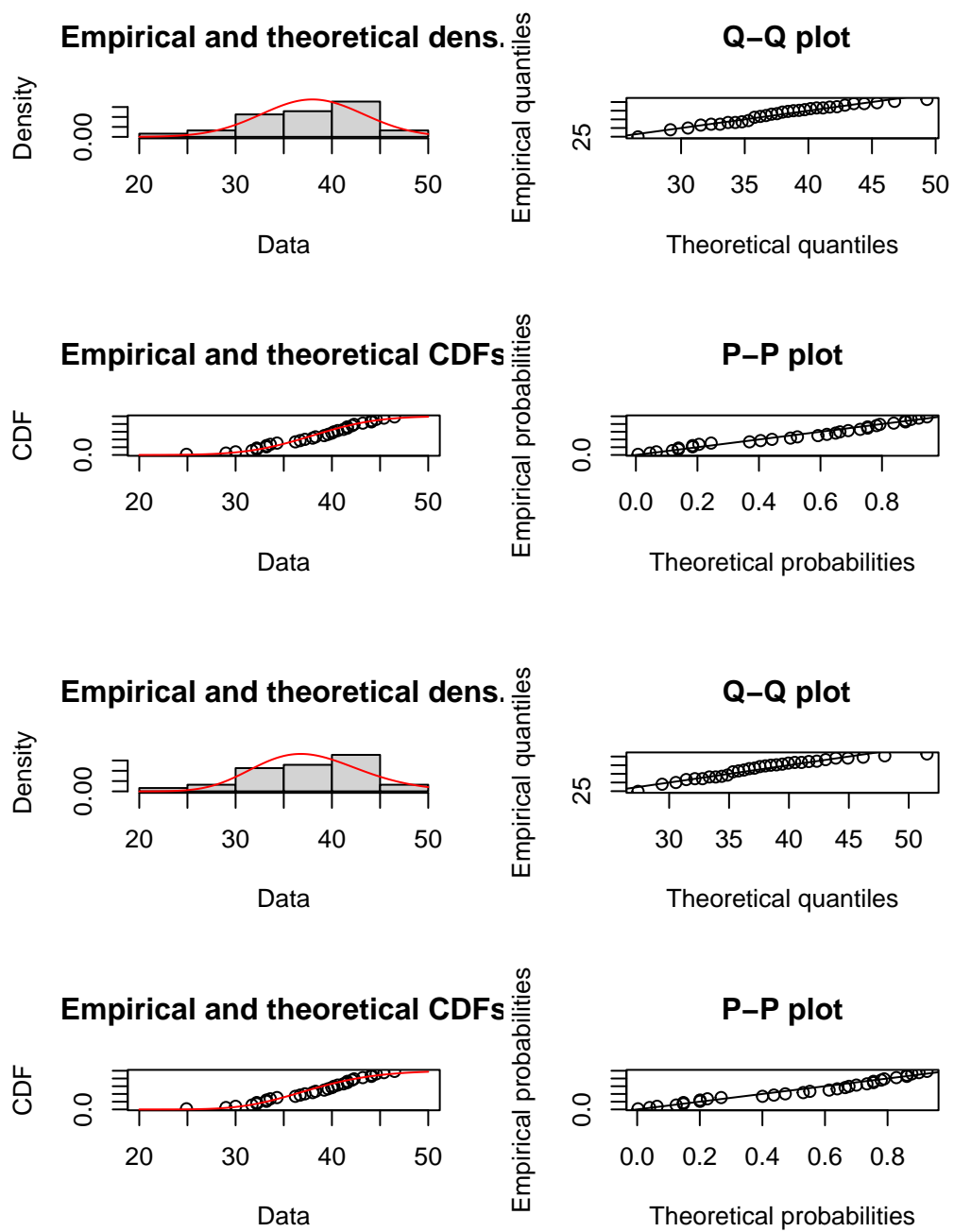
## Cullen and Frey graph

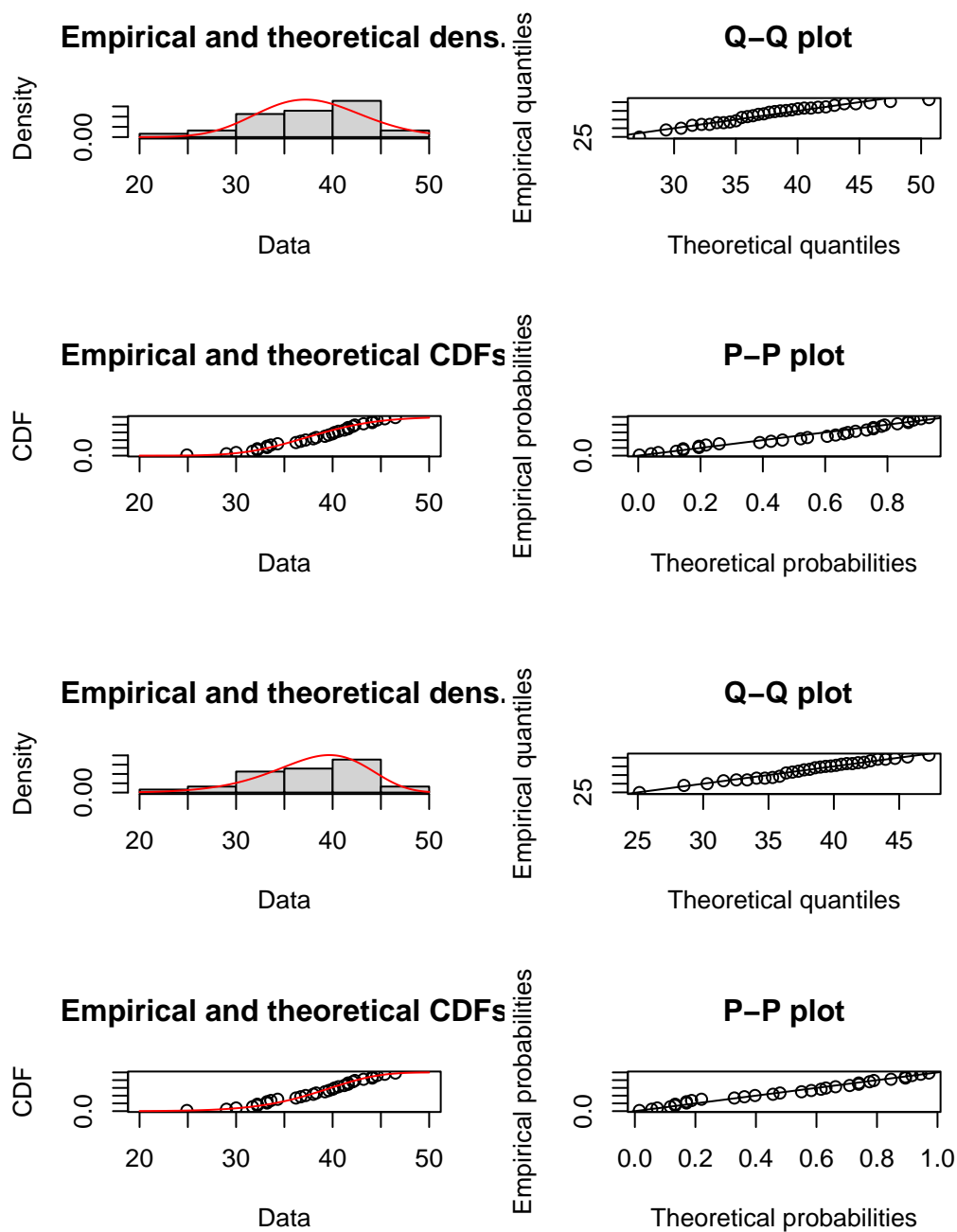


### summary statistics

-----

min: 24.9    max: 46.5  
 median: 39.2  
 mean: 37.96774  
 estimated sd: 5.389582  
 estimated skewness: -0.49904  
 estimated kurtosis: 2.574054





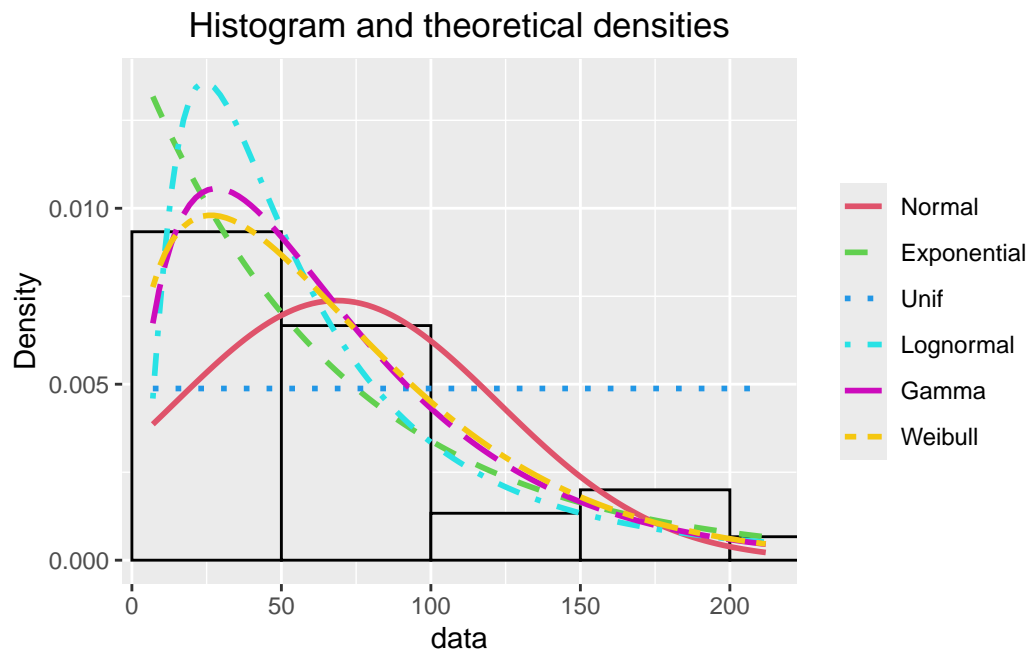
Maximum likelihood estimates for the Weibull model

shape	scale
8.755	40.212

Die Auswertung zeigt, dass nur die Weibullverteilung die Daten umfangreich decken kann. Daher wird diese Verteilung gewählt.

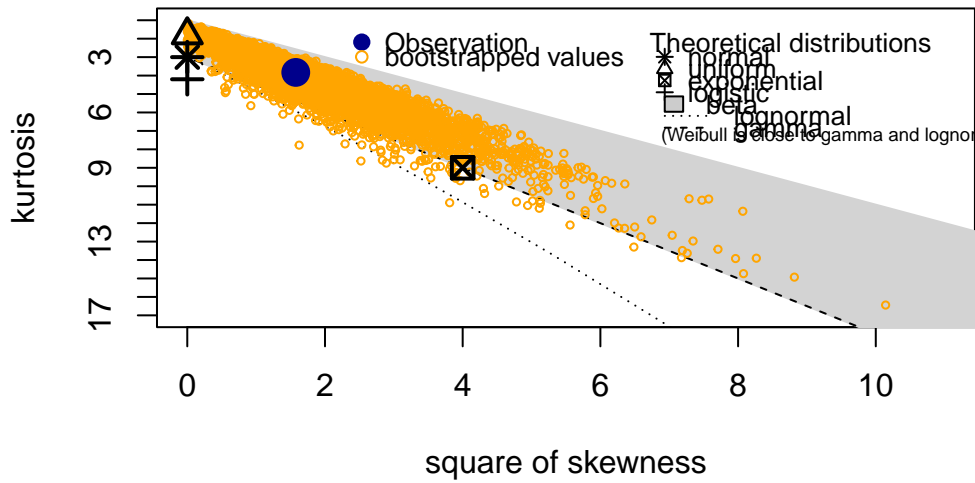
## Zone 2 Zeitabstände

In diesem Fall erkennt man ebenfalls, dass die Verteilungen Weibull, Gamma, Exponential und Lognormal zur Auswahl stehen. Der Vorschlag von UnivariateML ist die Gammaverteilung





## Cullen and Frey graph



summary statistics

-----

min: 7 max: 212

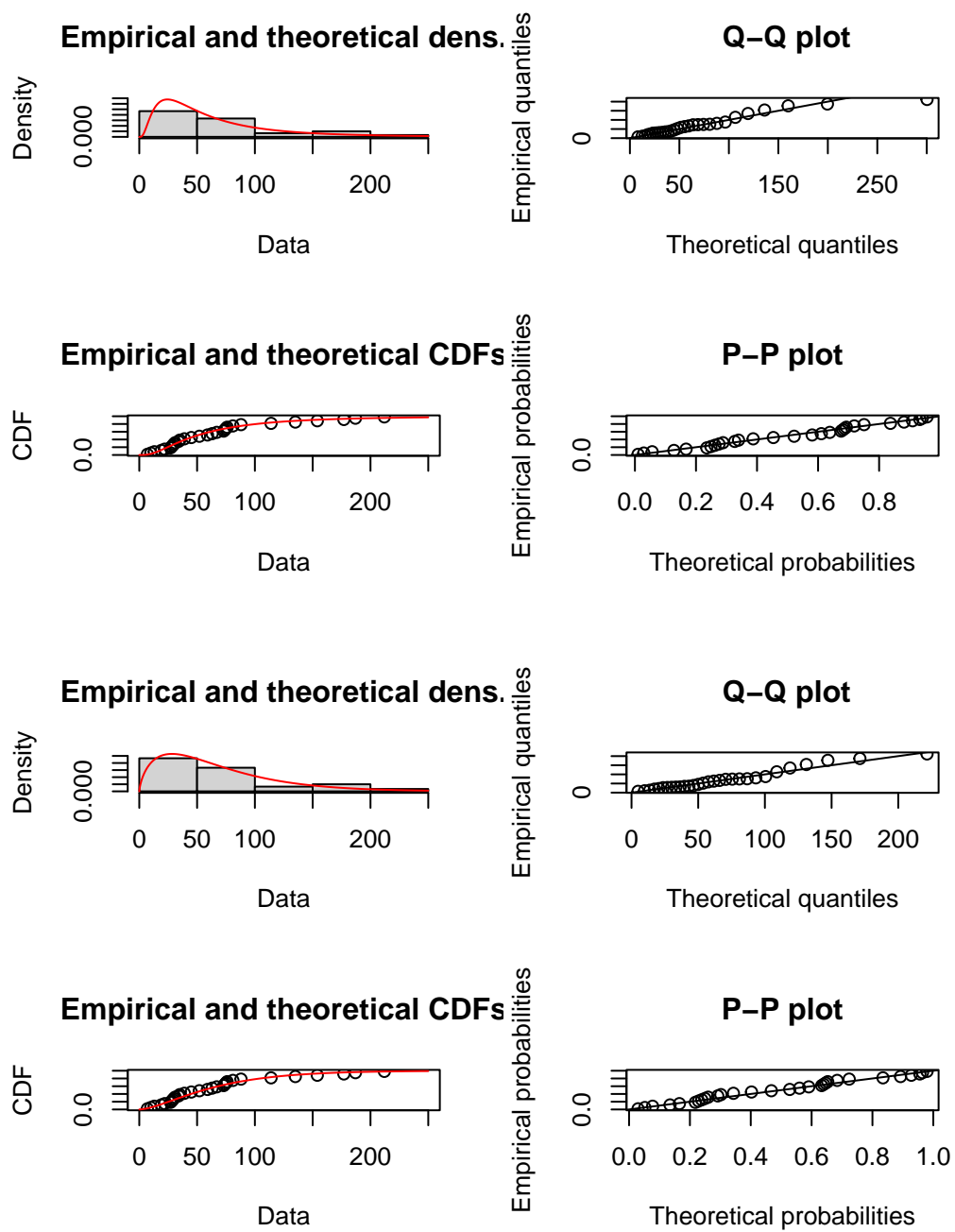
median: 55.5

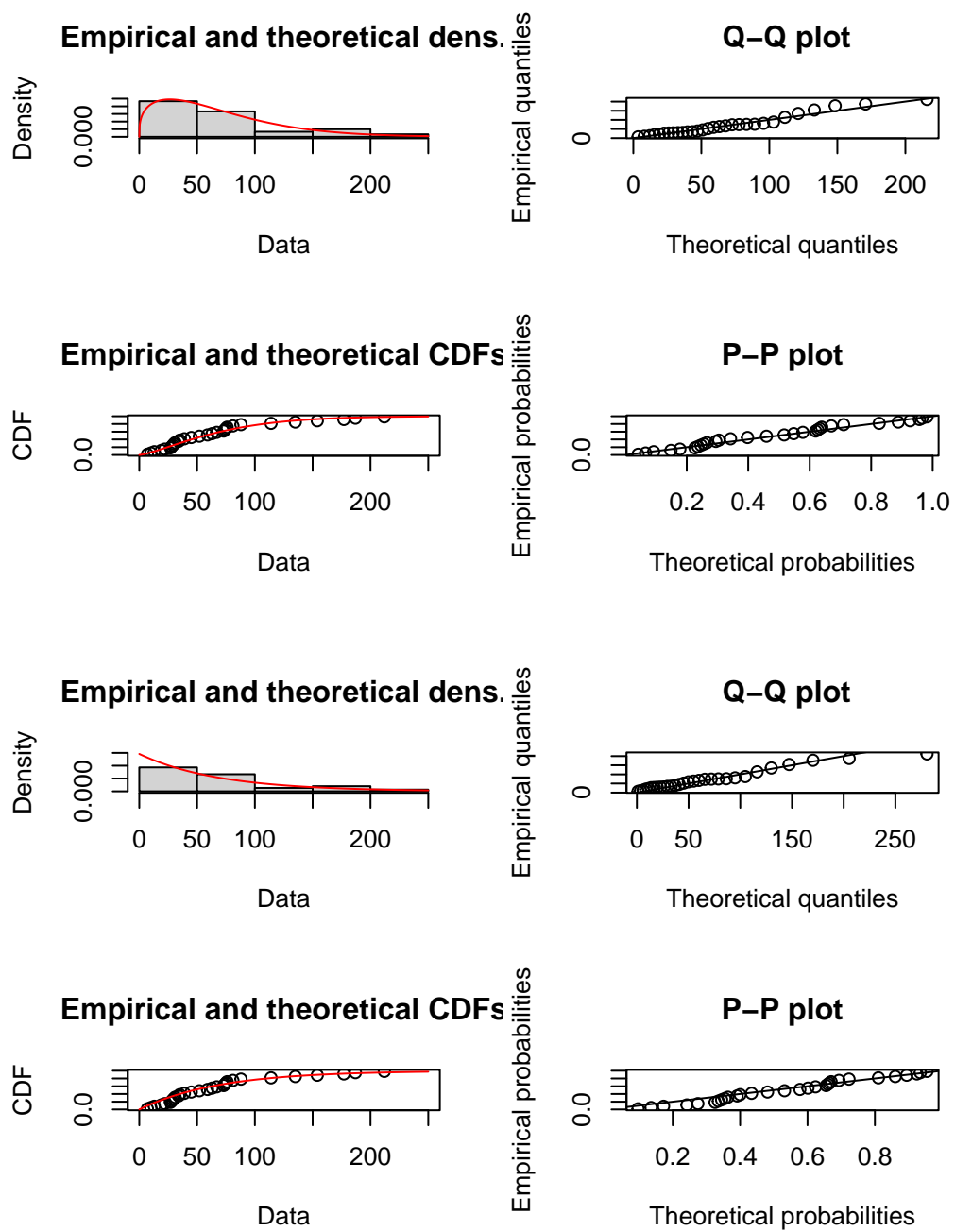
mean: 68.53333

estimated sd: 55.00265

estimated skewness: 1.255227

estimated kurtosis: 3.828367





Maximum likelihood estimates for the Gamma model

shape	rate
1.70124	0.02482

Beim Vergleich fällt schnell auf, dass Gamma- und Weibullverteilungen die genauesten Modelle sind. Gewählt wird hier die Gammaverteilung, da diese sowohl der Empfehlung von UnivariateML entspricht als auch die unteren Quantile im Vergleich besser abdeckt.

## Simulation

### Generieren

Das Generieren der Daten wurde in R umgesetzt. Dabei wurden die gewählten Verteilungen gewählt, welche im Kapitel Verteilungen gewählt wurden. Es wurde sich dafür entschieden die Anzahl Steinschläge auf  $10^5$  Iteration festzulegen.

```
set.seed(8128) # The 4. perfect number
zone_1_generated = data.frame(
  mass = runif(amount_iteration, min=0, max=1),
  velocity = runif(amount_iteration, min=0, max=1),
  time_diff = runif(amount_iteration, min=0, max=1)
)

zone_2_generated = data.frame(
  mass = runif(amount_iteration, min=0, max=1),
  velocity = runif(amount_iteration, min=0, max=1),
  time_diff = runif(amount_iteration, min=0, max=1)
)
```

Die Kinetische Energie wird anhand der Geschwindigkeit und Masse eines Steinschlags, mittels der Physikalischen Formel dafür berechnet und in Kilo Newton Meter umgewandelt.

$$E_{\text{kin}} = \frac{1}{2} \times m \times v^2$$

```
zone_1_generated$kin_energy =
  zone_1_generated$mass * zone_1_generated$velocity^2 * 0.5 / 1000

zone_2_generated$kin_energy =
  zone_2_generated$mass * zone_2_generated$velocity^2 * 0.5 / 1000
```

// TODO: Show Plots with densetiy: ExtraGrid

## Wahrscheinlichkeitsberechnung

Dass bei einem Steinschlag die Sicherheitsnetze reissen, müsste die kinetische Energie höher sein als 1200 kNm oder 600 kNm und die Masse im Sicherheitsnetz grösser oder gleich 2000 kg sein. Das Sicherheitsnetz wird alle 24 Stunden geleert.

Um die Masse im Netz bei einem Steinschlag zu berechnen, wird rekursiv durch alle vergangenen Steinschläge, welche in den letzten 24 Stunden stattgefunden haben, gegangen und die Masse summiert. Dadurch wird es ermöglicht, alle Steinschläge herauszufiltern, welche eine kinetische Energie von mindestens 1200 kNm haben oder mindestens 600 kNm und im Sicherheitsnetz eine Gesamtmasse von 2000 kg liegen.

Mit diesen Angaben ist es möglich, die Wahrscheinlichkeit, dass das Sicherheitsnetz reisst, auszurechnen. Dafür verwenden wir folgende Formel.

$$P_{\text{Netz reisst}} = \frac{\text{Anzahl Netz Versagen}}{\text{Anzahl Steinschläge}}$$

Das ein Steinschlag in einem Unfall endet muss zusätzlich ein Auto vor Ort sein. Um die Wahrscheinlichkeit zu berechnen wird die Zeit Benötigt, welches ein Auto braucht um für einen Möglichen Steinschlag zu bremsen. Dies lässt sich aus dem Bremsweg, was sich bei einer [Gefahrenbesmung bei 18m](#) liegt, bei einer Geschwindigkeit von 60 km/h und der Geschwindigkeit, wie folgt berechnen.

$$t = \frac{2s}{v} \Rightarrow \frac{2 \times 18m}{16.6m/s} \approx 2.16s$$

Dazu wird noch die [Vorbremszeit von 1.2 Sekunde](#) addiert, welche ein Mensch braucht um auf die Bremse zu drücken. Damit kommt man auf *3.36 Sekunden*, welche ein Auto in der Gefahrenzone ist. Durch diesen Wert können wir die Wahrscheinlichkeit berechnen, dass ein Auto zu einem beliebigen Zeitpunkt in der Gefahrenzone ist.

$$P_{\text{Auto anwesend}} = \frac{3.36 \times \text{Anzahl Autos Pro Tag}}{\text{Anzahl Sekunden Pro Tag}} \Rightarrow \frac{3.36 \times 600}{86400} = \frac{7}{300}$$

Mit der Wahrscheinlichkeit, dass das Sicherheitsnetz reisst, ein Auto anwesend ist und der durchschnittlichen Steinschläge pro Jahr, lässt sich die Wahrscheinlichkeit eines tödlichen Unfalls in einer Zone für ein Jahr berechnen.

$$P_{\text{Tödlicher Unfall pro Jahr}} = P_{\text{Auto anwesend}} \times P_{\text{Netz reisst}} \times \text{Durchschnittliche Steinschläge pro Jahr}$$

## **Aussagekraft der Simulation**

// TODO: mit Varianz und so. Könnte noch triky werden.

## Ergebniss



## Empfehlung