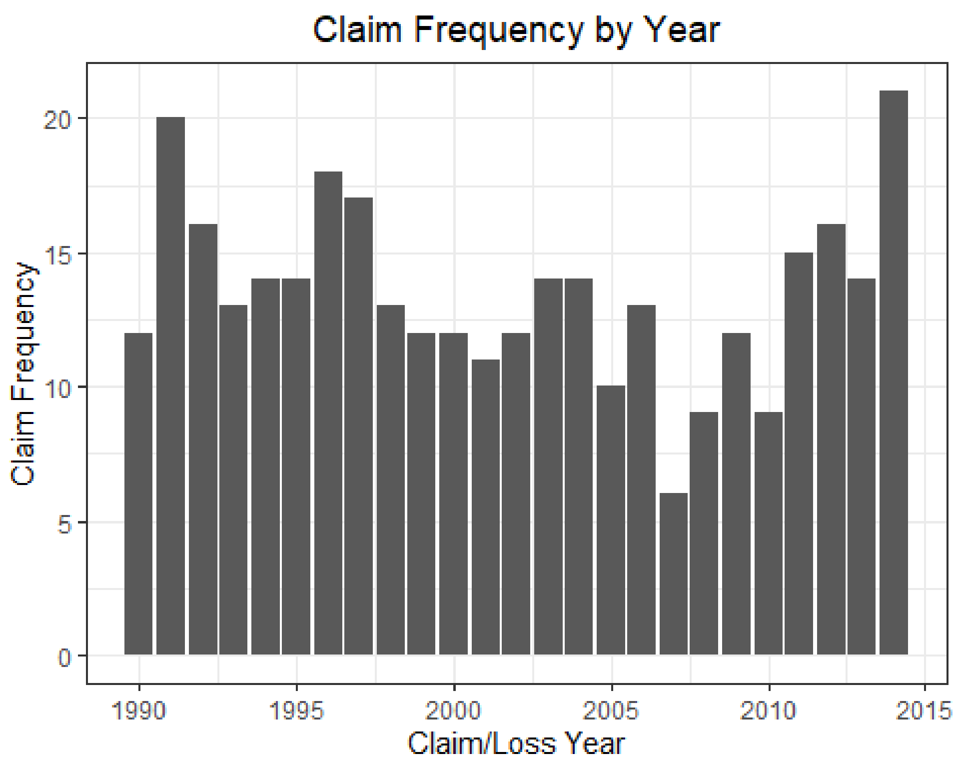
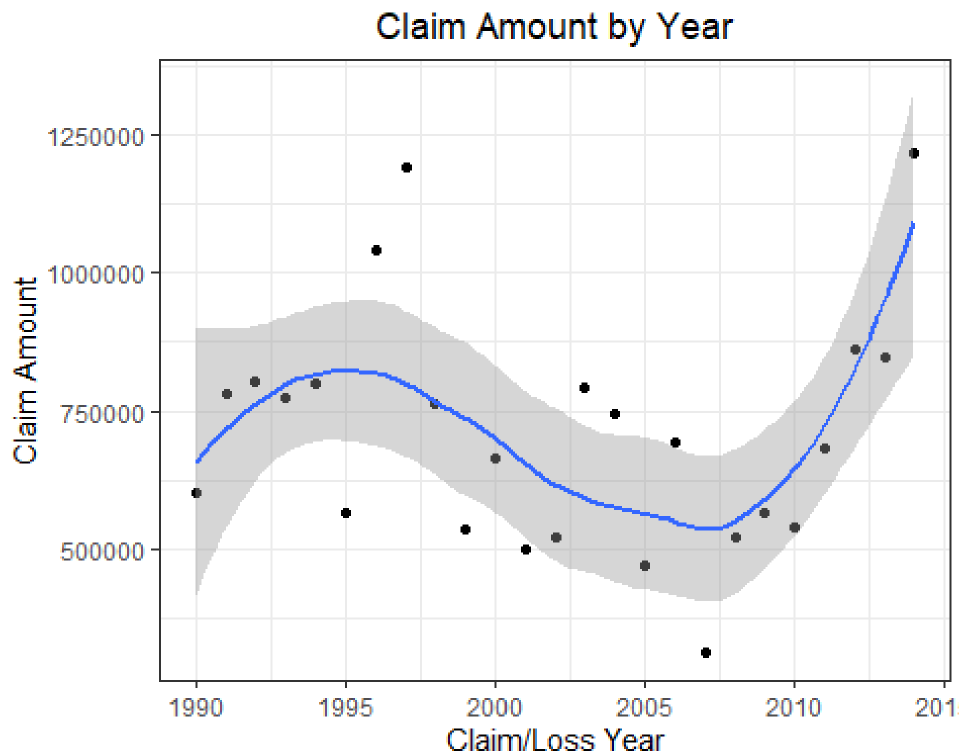
Credibility Model

```r
#EDA: Plotting the data
#1. Plotting the claim frequency for each year
raw_data %>%
  ggplot(aes(x = loss_year)) +
  geom_bar()+
  ylab("Claim Frequency")+
  xlab("Claim/Loss Year")+
  ggtitle("Claim Frequency by Year") +
  theme_bw()+
  theme(plot.title=element_text(hjust=0.5))
```

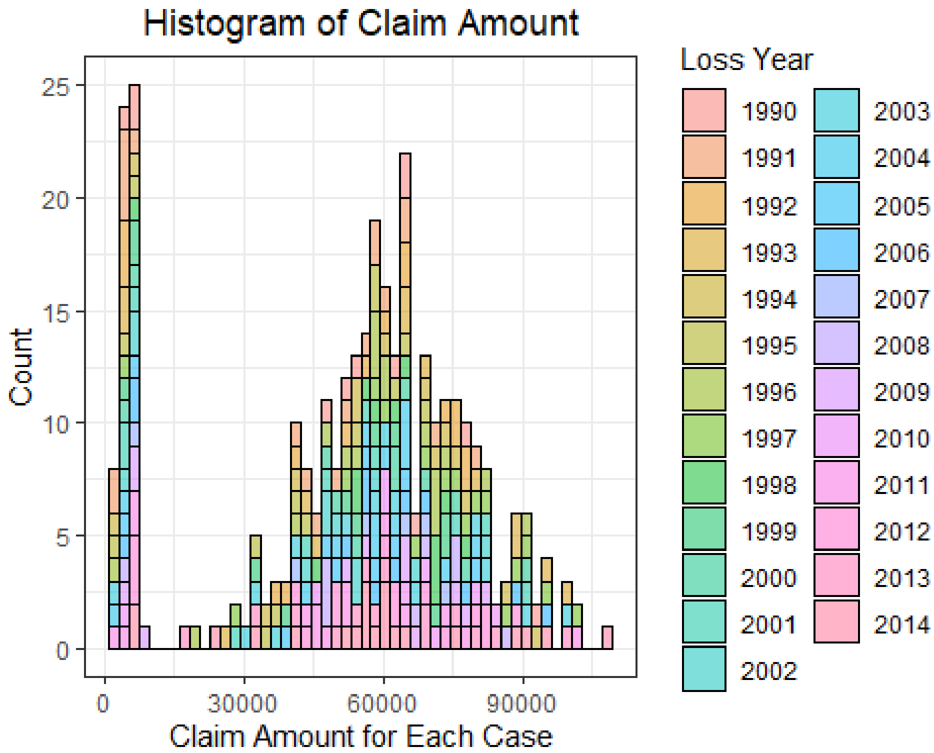



```r
#2. Plotting the claim amount for each year
raw_data%>%
  group_by(loss_year)%>%
  summarise(total_claim_peryear = sum(total_claim_amount))%>%
  ggplot(aes(x = loss_year, y=total_claim_peryear)) +
  geom_point()+
  geom_smooth()+
  ylab("Claim Amount")+
  xlab("Claim/Loss Year")+
  ggtitle("Claim Amount by Year") +
  theme_bw()+
  theme(plot.title=element_text(hjust=0.5))
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
#3. Claim amount per policy ID Histogram
raw_data %>%
  ggplot(aes(x = total_claim_amount, fill = as.factor(loss_year))) +
  geom_histogram(bins = 50, color = "black", alpha = 0.5) +
  labs(fill = "Loss Year")+
  ylab("Count")+
  xlab("Claim Amount for Each Case")+
  ggtitle("Histogram of Claim Amount") +
  theme_bw()+
  theme(plot.title=element_text(hjust=0.5))
```




```
#dari plot histogram, kita bisa lihat bahwa data berbentuk cukup aneh.
#Kita asumsikan data dicatat pada loss year, sehingga perlu ada penyesuaian i
nflasi
```

Dari grafik batang frekuensi klaim per tahun terlihat bahwa frekuensi klaim bervariasi dari tahun ke tahun.Berkenaan dengan hal tersebut, besar klaim aggregat per tahun juga berfluktuasi dari tahun ke tahun dan bentuknya menyurupai grafik batang frekuensi klaim per tahun. Dari histogram untuk besar klaim per polis, terlihat seperti terdapat 2 jenis data yaitu data dengan nilai klaim kecil (dibawah 25000) dan klaim besar (diatas 25000). Dari hasil ini dapat mengindikasikan bahwa terdapat 2 jenis profil risiko pada data.

Inflation Adjusted Data

```
#Inflasi Adjusted data
#data CPI-Urban di US, diambil dari Bureau of Labor Statistics
cpi_data <- c(133.8, 137.9, 141.9, 145.8, 149.7, 153.5, 158.6, 161.3, 163.9,
168.3,
              174.0, 176.7, 180.9, 184.3, 190.3, 196.8, 201.8, 210.036, 210.2
28, 215.949,
              219.179, 225.672, 229.601, 233.049, 234.812)
years <- 1990:2014
cpi_df <- data.frame(year = years, cpi = cpi_data)

#menambahkan kolom CPI ke raw_data
data_adj <- left_join(x=raw_data,
                      y=cpi_df,
                      by =  c("loss_year" = "year"))

#membuat kolom claim yang sudah di adjust dengan inflasi
data_adj <- data_adj%>%
  mutate(total_claim_amount_adj = total_claim_amount*234.812/cpi)

data_adj%>%
  arrange(loss_year)%>%
  head()%>%
  gt()%>%
  fmt_number(columns    =   5, decimals =   3)  %>%
  cols_label("policy_ID"    =   "Policy ID",
             "loss_year"    =   "Loss Year",
             "total_claim_amount"   =   "Claim Amount (not adjusted)",
             "cpi"  =   "CPI",
             "total_claim_amount_adj" = "Claim Amount (inflation adjusted)")
%>%
  tab_header(title  =   md("**Claim Amount per Policy**"),
             subtitle   =   "The first 5 data") %>%
  gt_theme_538()
```

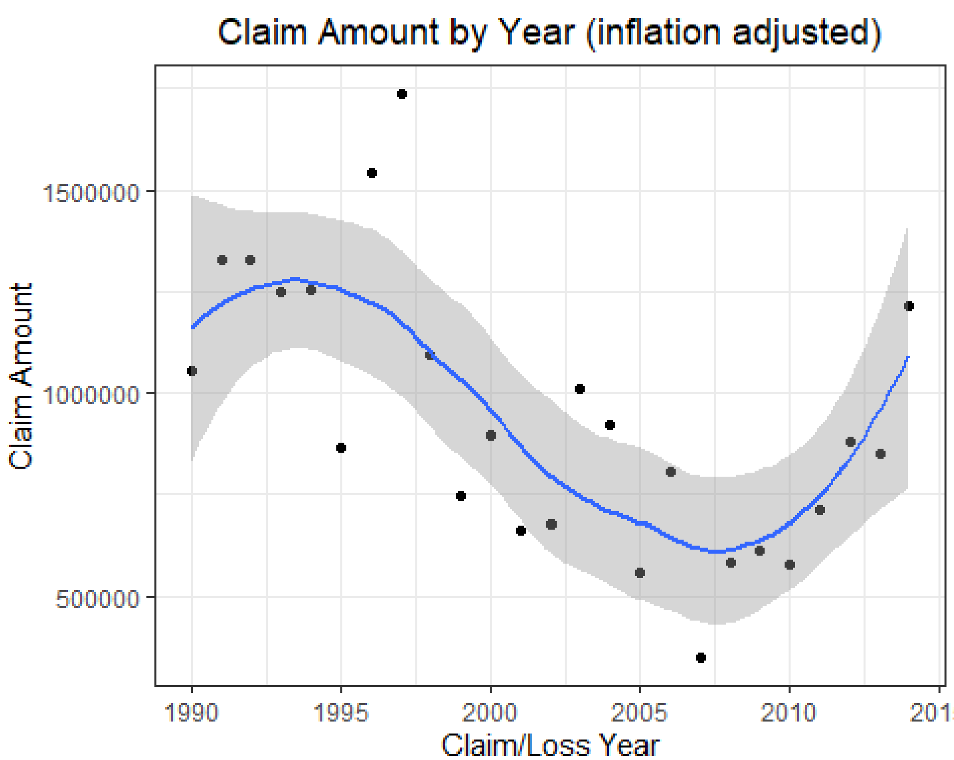
Table 1: **Claim Amount per Policy**

The first 5 data

110084	1990	55000	133.8	96,522.123
113516	1990	67320	133.8	118,143.078
227811	1990	63400	133.8	111,263.683
307447	1990	93720	133.8	164,473.697
429027	1990	51590	133.8	90,537.751
432740	1990	4900	133.8	8,599.244

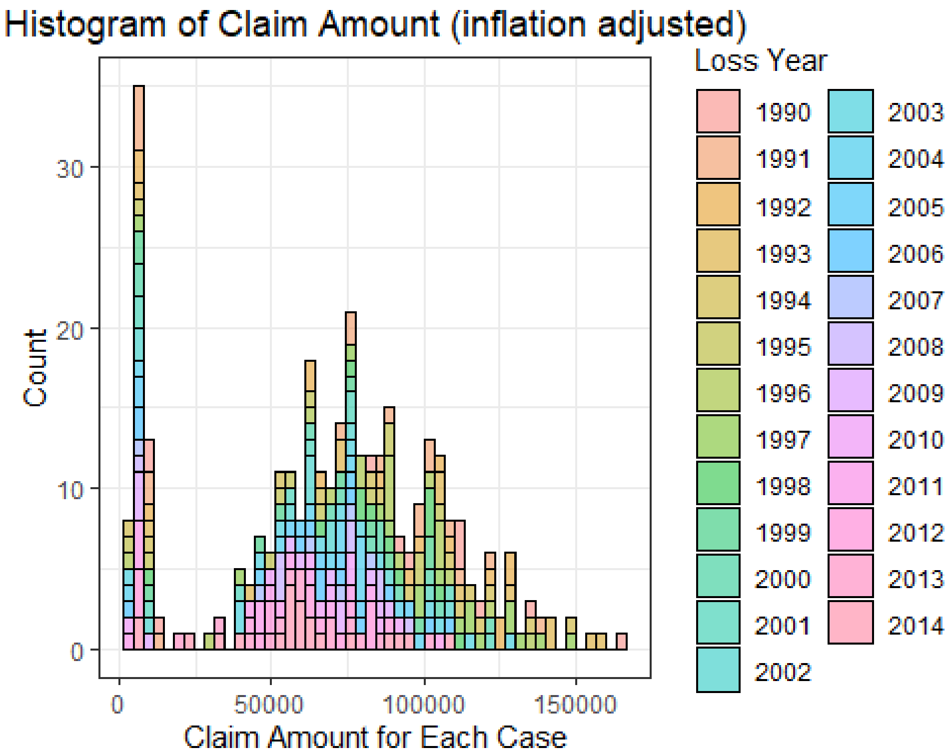
Dengan asumsi data besar klaim (severitas klaim) tertulis dalam US dollar, telah kita hitung nilai total klaim per polis. Perhitungan didasarkan dari CPI (consumer price index) yang diperoleh melalui situs resmi BLS (Bureau of Labor Statistics). Tingkat inflasi dihitung dengen (CPI 2014)/(CPI tahun klaim)

```
#EDA: data setelah diadjust dengan inflasi
#4. Plotting the claim amount for each year
data_adj%>%
  group_by(loss_year)%>%
  summarise(total_claim_peryear = sum(total_claim_amount_adj))%>%
  ggplot(aes(x = loss_year, y=total_claim_peryear)) +
  geom_point()+
  geom_smooth()+
  ylab("Claim Amount")+
  xlab("Claim/Loss Year")+
  ggtitle("Claim Amount by Year (inflation adjusted)") +
  theme_bw()+
  theme(plot.title=element_text(hjust=0.5))

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
#5. Claim amount per policy ID Histogram
data_adj %>%
  ggplot(aes(x = total_claim_amount_adj, fill = as.factor(loss_year))) +
  geom_histogram(bins = 50, color = "black", alpha = 0.5) +
  labs(fill = "Loss Year")+
  ylab("Count")+
```

```
  xlab("Claim Amount for Each Case")+
  ggtitle("Histogram of Claim Amount (inflation adjusted)") +
  theme_bw()+
  theme(plot.title=element_text(hjust=0.5))
```



Dari hasil plot data yang sudah disesuaikan dengan tingkat inflasi, terlihat bahwa fluktuasi besar klaim pertahun tetap terlihat dan cukup mirip dengan data yang tidak disesuaikan. Pada grafik kedua, terlihat pula bentuk bahwa terdapat 2 macam kelas risiko untuk besar klaim per polis. Bentuk grafik untuk klaim polis yang cukup tinggi terlihat lebih memiliki kemiringan (ekor kanan) yang lebih tinggi dibanding data yang tidak disesuaikan dengan inflasi. Hal ini mengindikasikan bahwa data memiliki distribusi yang berbeda antara sebelum dan sesudah disesuaikan dengan inflasi.

Pemisahan Data Low Risk dan High Risk

```
#2. Pemisahan data risiko tinggi dan rendah data non inflation adjusted
low_nonadj <- data_adj %>%
  filter(total_claim_amount <= (max(total_claim_amount) - min(total_claim_amo
unt)) / 50 * 4)%>%
  arrange(loss_year)%>%
  select(policy_ID, loss_year, total_claim_amount)%>%
  group_by(loss_year)%>%
  summarise(total_claim_nonadj = sum(total_claim_amount),
            claim_freq = n())%>%
  bind_rows(tibble(loss_year = 2013, total_claim_nonadj = NaN, claim_freq = 0
))%>%
  arrange(loss_year)


high_nonadj <- data_adj %>%
  filter(total_claim_amount > (max(total_claim_amount) - min(total_claim_amou
nt)) / 50 * 4)%>%
  arrange(loss_year)%>%
  select(policy_ID, loss_year, total_claim_amount)%>%
  group_by(loss_year)%>%
  summarise(total_claim_nonadj = sum(total_claim_amount),
            claim_freq = n())


#3. Pemisahan data risiko tinggi dan rendah data inflation adjusted
low_adj <- data_adj %>%
  filter(total_claim_amount_adj <= (max(total_claim_amount_adj) - min(total_c
laim_amount_adj)) / 50 * 4)%>%
  arrange(total_claim_amount_adj)%>%
  select(policy_ID, loss_year, total_claim_amount_adj)%>%
  group_by(loss_year)%>%
  summarise(total_claim_adj = sum(total_claim_amount_adj),
            claim_freq = n())%>%
  bind_rows(tibble(loss_year = 2013, total_claim_adj = NaN, claim_freq = 0))%
>%
  arrange(loss_year)

high_adj <- data_adj %>%
  filter(total_claim_amount_adj > (max(total_claim_amount_adj) - min(total_cl
aim_amount_adj)) / 50 * 4)%>%
  arrange(total_claim_amount_adj)%>%
  select(policy_ID, loss_year, total_claim_amount_adj)%>%
  group_by(loss_year)%>%
  summarise(total_claim_adj = sum(total_claim_amount_adj),
            claim_freq = n())
```

Pemisahan data dilakukan pada data dengan nilai klaim pada selang batas atas batang histogram ke 4. Hal ini didasari dari grafik histrogram besar klaim per polis yang sudah ditunjukkan sebelumnya yang menunjukkan adanya lompatan besar klaim pada batas tersebut. Setelah data sebelum dan sesudah inflasi dipisahkan, terlihat bahwa pemotongan data terjadi pada satu titik datum yang sama, maka analisis pada bagian selanjutnya dilakukan berdasarkan 1 data yang sama hanya memisahkan nilai sebelum dan sesudah disesuaikan dengan tingkat inflasi pada kolom yang berbeda.

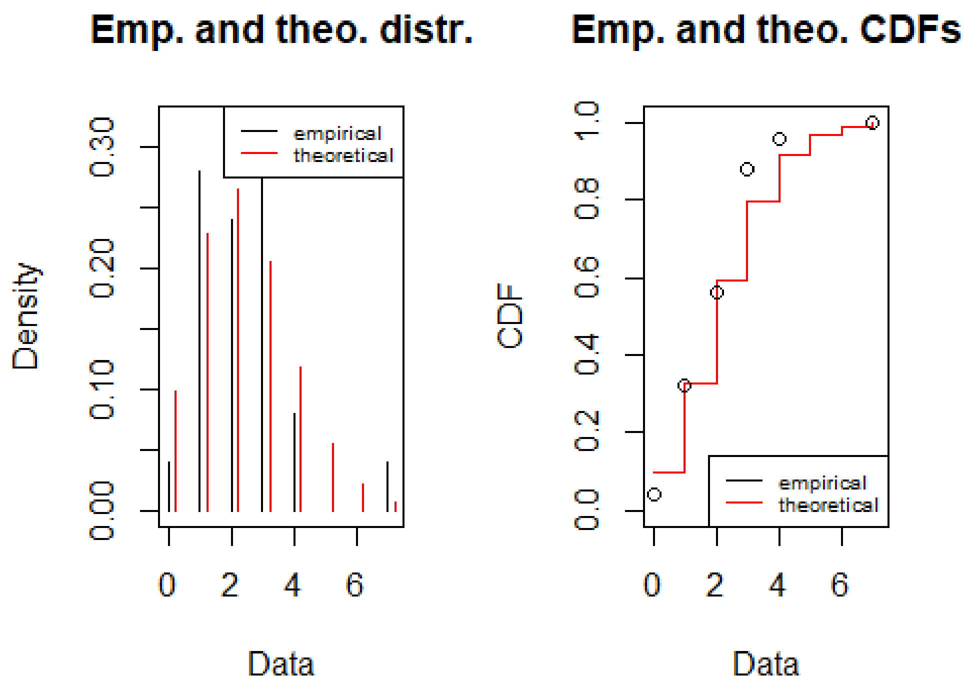
Peluang terpilihnya low dan high risk type

```
#Finding the  probability of low and high risk type
## Nonadjusted data
n_data <- low_nonadj %>%
  left_join(high_nonadj,
            by = "loss_year",
            suffix = c("_low", "_high"))

p_R1 <- sum(n_data$claim_freq_low, na.rm = T)/
  (sum(n_data$claim_freq_high, na.rm=T)+sum(n_data$claim_freq_low, na.rm = T)
)
p_R2 <- sum(n_data$claim_freq_high, na.rm = T)/
  (sum(n_data$claim_freq_high, na.rm=T)+sum(n_data$claim_freq_low, na.rm = T)
)

paste0("Peluang terpilihnya tipe risiko rendah: ", p_R1 )

## [1] "Peluang terpilihnya tipe risiko rendah: 0.172106824925816"

paste0("Peluang terpilihnya tipe risiko tinggi: ", p_R2 )

## [1] "Peluang terpilihnya tipe risiko tinggi: 0.827893175074184"

#p_R1 = 0.1721068
#p_R2 = 0.8278932
```

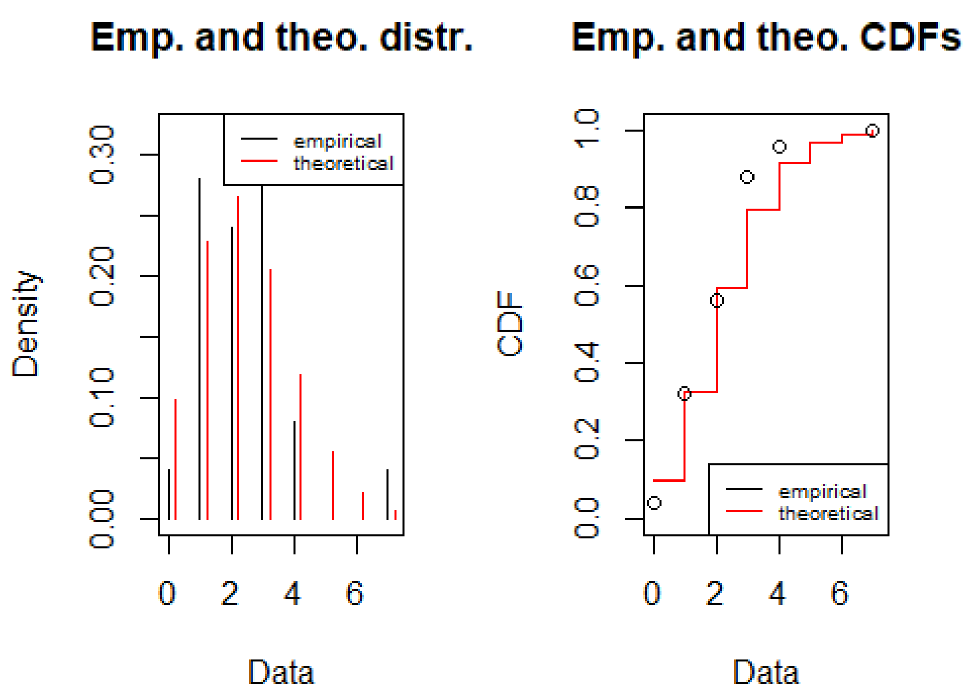
#1. Mencari Distribusi dari Besar Claim ## Distribusi dari N (Data inflation adjusted dan noninflation adjusted tidak mempengaruhi banyak N) ### N low risk

```
#mencari distribusi dari N

## N untuk low risk nonadj
# Fit Poisson Distribution
fit_pois <- fitdist(n_data$claim_freq_low, "pois")

# Fit Negative Binomial Distribution
fit_nb <- fitdist(n_data$claim_freq_low, "nbinom")
```

```r
#model Summary
summary(fit_pois)

## Fitting of the distribution ' pois ' by maximum likelihood
## Parameters :
##        estimate Std. Error
## lambda     2.32  0.3046309
## Loglikelihood:  -42.56333   AIC:  87.12666   BIC:  88.34554

summary(fit_nb)

## Fitting of the distribution ' nbinom ' by maximum likelihood
## Parameters :
##          estimate Std. Error
## size 3.460543e+06  46.758593
## mu   2.319931e+00   0.304622
## Loglikelihood:  -42.56333   AIC:  89.12666   BIC:  91.56442
## Correlation matrix:
##               size            mu
## size  1.000000e+00 -9.883549e-09
## mu   -9.883549e-09  1.000000e+00

mean(n_data$claim_freq_low)

## [1] 2.32

var(n_data$claim_freq_low)

## [1] 2.06

# Plot the fits
plot(fit_pois)
```

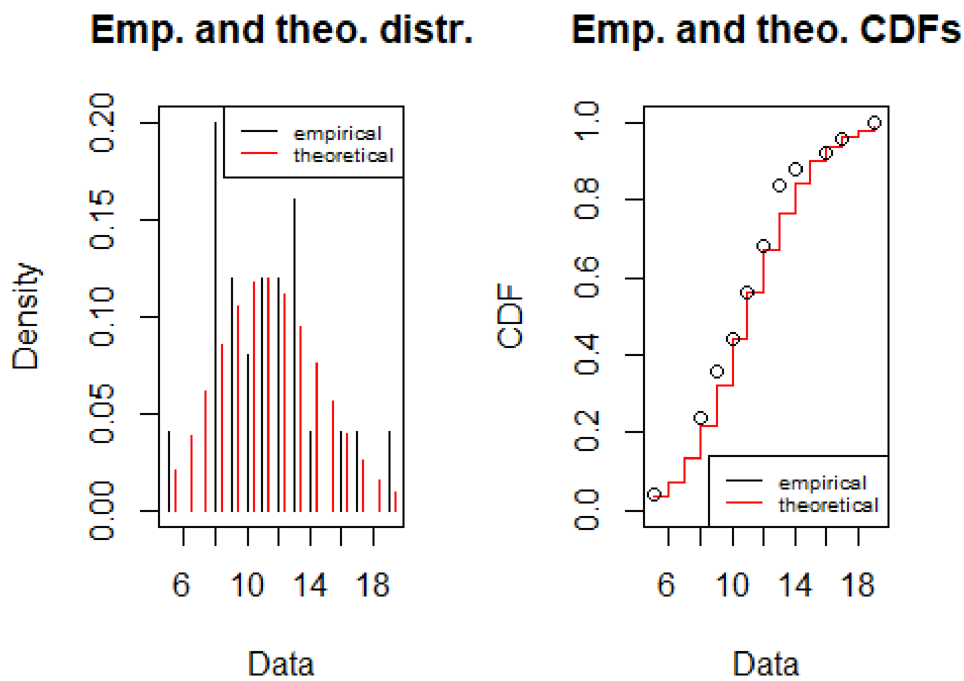

Emp. and theo. distr.
Density
Data
empirical
theoretical
0.00
0.10
0.20
0.30
0
2
4
6
Emp. and theo. CDFs
CDF
Data
0.0
0.2
0.4
0.6
0.8
1.0

```
plot(fit_nb)
```

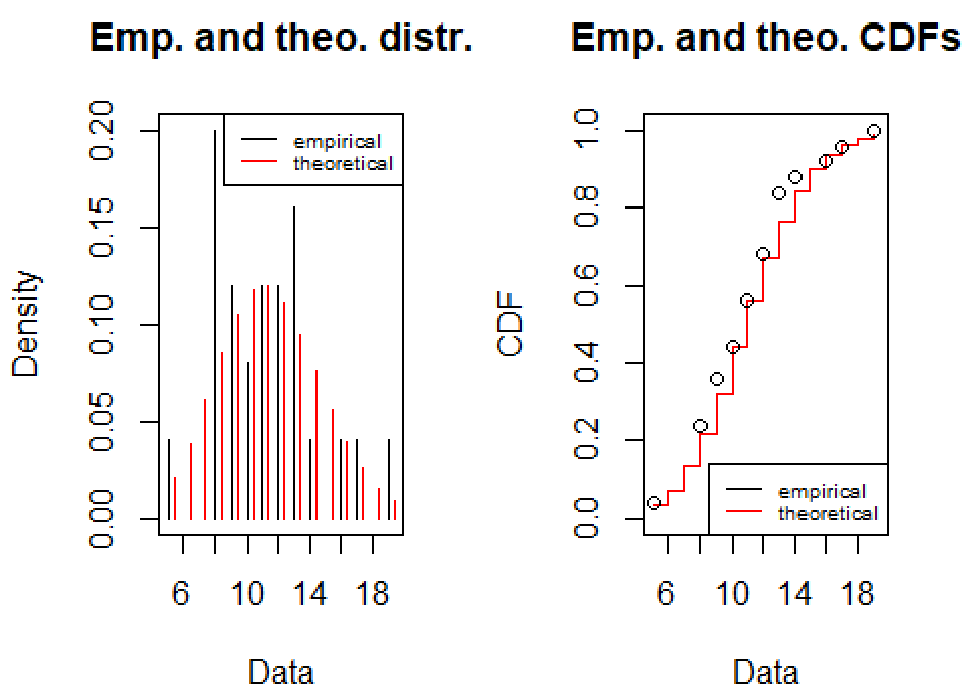

Emp. and theo. distr.
Density
Data
empirical
theoretical
0.00
0.10
0.20
0.30
0
2
4
6
Emp. and theo. CDFs
CDF
Data
0.0
0.2
0.4
0.6
0.8
1.0

N high risk

```r
#mencari distribusi dari N

## N untuk high risk nonadj
# Fit Poisson Distribution
fit_pois <- fitdist(n_data$claim_freq_high, "pois")

# Fit Negative Binomial Distribution
fit_nb <- fitdist(n_data$claim_freq_high, "nbinom")

## Warning in cov2cor(varcovar): diag(V) had non-positive or NA entries; the
## non-finite result may be dubious

## Warning in sqrt(diag(varcovar)): NaNs produced

#model summary
summary(fit_pois)

## Fitting of the distribution ' pois ' by maximum likelihood
## Parameters :
##        estimate Std. Error
## lambda    11.16  0.6681317
## Loglikelihood:  -63.76834   AIC:  129.5367   BIC:  130.7556

summary(fit_nb)

## Fitting of the distribution ' nbinom ' by maximum likelihood
## Parameters :
##          estimate Std. Error
## size 4.555023e+06        NaN
## mu   1.115918e+01  0.6680835
## Loglikelihood:  -63.76835   AIC:  131.5367   BIC:  133.9744
## Correlation matrix:
##      size  mu
## size    1 NaN
## mu    NaN   1

mean(n_data$claim_freq_high)

## [1] 11.16

var(n_data$claim_freq_high)

## [1] 10.30667

# Plot the fits
plot(fit_pois)
```


Emp. and theo. distr.
Emp. and theo. CDFs
Density
CDF
Data
empirical
theoretical
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
6
10
14
18

```r
plot(fit_nb)
```


Emp. and theo. distr.
Emp. and theo. CDFs
Density
CDF
Data
empirical
theoretical
0.00
0.05
0.10
0.15
0.20
0.0
0.2
0.4
0.6
0.8
1.0
6
10
14
18

Distribusi dari X

Memisahkan data low risk dan high risk

```
# Fitting the distributions to claim severity data
#membuat data low risk dan high risk untuk data non adjusted
x_low_nonadj <- data_adj%>%
  filter(total_claim_amount <= (max(total_claim_amount) - min(total_claim_amo
unt)) / 50 * 4)%>%
  arrange(loss_year)

x_high_nonadj <- data_adj%>%
  filter(total_claim_amount > (max(total_claim_amount) - min(total_claim_amou
nt)) / 50 * 4)%>%
  arrange(loss_year)

x_low_nonadj

## # A tibble: 58 × 5
##    policy_ID loss_year total_claim_amount   cpi total_claim_amount_adj
##        <dbl>     <dbl>              <dbl> <dbl>                  <dbl>
##  1    432740      1990               4900  134.                  8599.
##  2    476839      1990               7200  134.                 12636.
##  3    477177      1990               6100  134.                 10705.
##  4    159243      1991               4900  138.                  8344.
##  5    253085      1991               3190  138.                  5432.
##  6    552788      1991               4290  138.                  7305.
##  7    553436      1991               5060  138.                  8616.
##  8    582480      1991               3840  138.                  6539.
##  9    808544      1991               7370  138.                 12549.
## 10    865839      1991               3190  138.                  5432.
## # ℹ 48 more rows

x_high_nonadj

## # A tibble: 279 × 5
##    policy_ID loss_year total_claim_amount   cpi total_claim_amount_adj
##        <dbl>     <dbl>              <dbl> <dbl>                  <dbl>
##  1    110084      1990              55000  134.                 96522.
##  2    113516      1990              67320  134.                118143.
##  3    227811      1990              63400  134.                111264.
##  4    307447      1990              93720  134.                164474.
##  5    429027      1990              51590  134.                 90538.
##  6    506333      1990              76890  134.                134938.
##  7    512813      1990              64350  134.                112931.
##  8    687755      1990              64200  134.                112668.
##  9    793948      1990              47430  134.                 83237.
## 10    131478      1991              58200  138.                 99101.
## # ℹ 269 more rows
```

```
#membuat data low risk dan high risk untuk data adjusted
x_low_adj <- data_adj%>%
  filter(total_claim_amount_adj <= (max(total_claim_amount_adj) - min(total_c
laim_amount_adj)) / 50 * 4)%>%
  arrange(loss_year)

x_high_adj <- data_adj%>%
  filter(total_claim_amount_adj > (max(total_claim_amount_adj) - min(total_cl
aim_amount_adj)) / 50 * 4)%>%
  arrange(loss_year)

#Perhatikan bahwa data nonadjusted dan adjusted terpotong pada titik nilai cl
aim yg sama, sehingga cukup digunakan satu data saja
x_data_low <- x_low_nonadj
x_data_high <- x_high_nonadj
```

Distribusi untuk noniflation adjusted data

```
# Distribusi untuk low risk
fit_lnorm <- fitdist(x_data_low$total_claim_amount, "lnorm")
fit_exp <- fitdist(x_data_low$total_claim_amount, "exp")
fit_gamma <- fitdist(x_data_low$total_claim_amount, "gamma")
fit_weibull <- fitdist(x_data_low$total_claim_amount, "weibull")
fit_pareto <- fitdist(x_data_low$total_claim_amount, "pareto", start=list(sha
pe = 1, scale = 1))

# Summary of the fitted models
summary(fit_lnorm)

## Fitting of the distribution ' lnorm ' by maximum likelihood
## Parameters :
##          estimate Std. Error
## meanlog 8.5123877 0.03846772
## sdlog   0.2929614 0.02719936
## Loglikelihood:  -504.8095   AIC:  1013.619   BIC:  1017.74
## Correlation matrix:
##         meanlog sdlog
## meanlog       1     0
## sdlog         0     1

summary(fit_exp)

## Fitting of the distribution ' exp ' by maximum likelihood
## Parameters :
##          estimate   Std. Error
## rate 0.0001930566 4.315837e-08
## Loglikelihood:  -554.0466   AIC:  1110.093   BIC:  1112.154

summary(fit_gamma)
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##           estimate   Std. Error
## shape 12.622196788 0.7760857814
## rate   0.002436965 0.0001185655
## Loglikelihood:  -503.2571   AIC:  1010.514   BIC:  1014.635
## Correlation matrix:
##           shape      rate
## shape 1.0000000 0.8079604
## rate  0.8079604 1.0000000
```

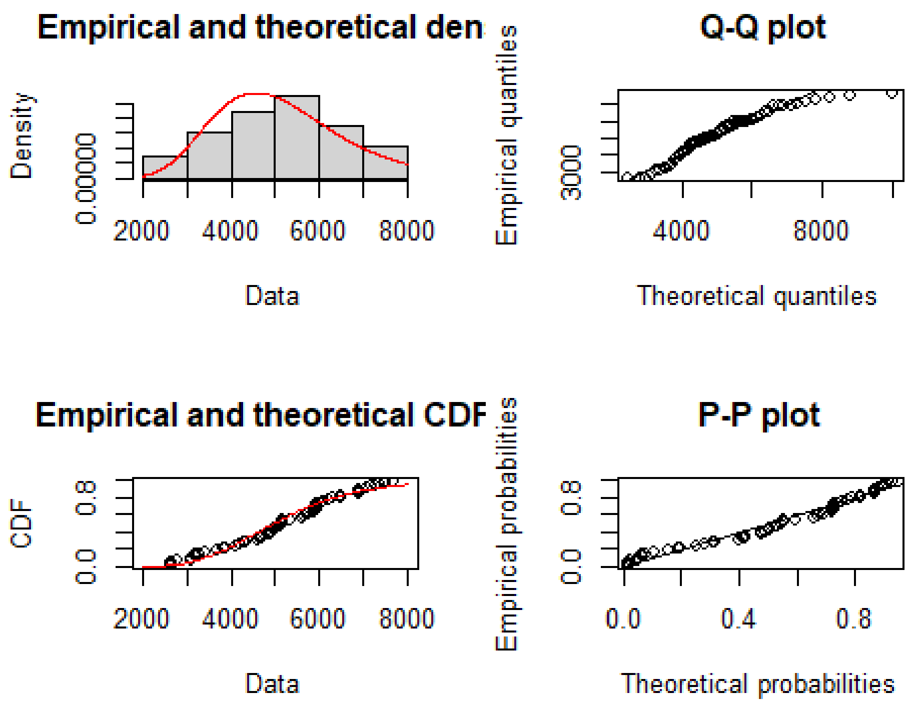
```r
summary(fit_weibull)
```

```
## Fitting of the distribution ' weibull ' by maximum likelihood
## Parameters :
##          estimate  Std. Error
## shape    4.322965   0.4554651
## scale 5702.441158 182.4638059
## Loglikelihood:  -500.9818   AIC:  1005.964   BIC:  1010.084
## Correlation matrix:
##           shape     scale
## shape 1.0000000 0.3132377
## scale 0.3132377 1.0000000
```

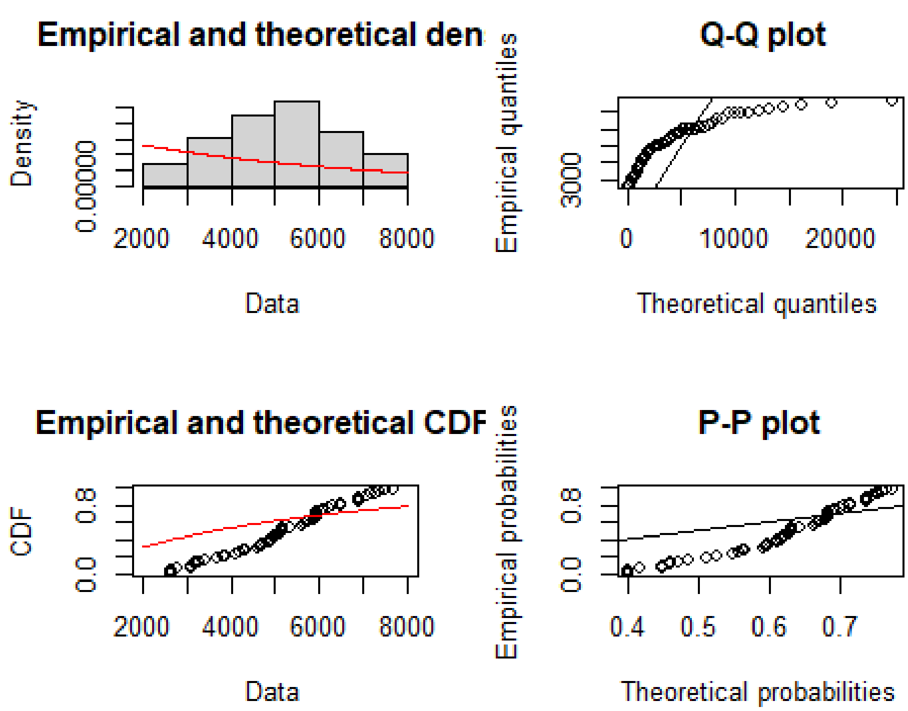
```r
summary(fit_pareto)
```

```
## Fitting of the distribution ' pareto ' by maximum likelihood
## Parameters :
##          estimate
## shape     3306714
## scale 17155212255
## Loglikelihood:  -554.0466   AIC:  1112.093   BIC:  1116.214
```

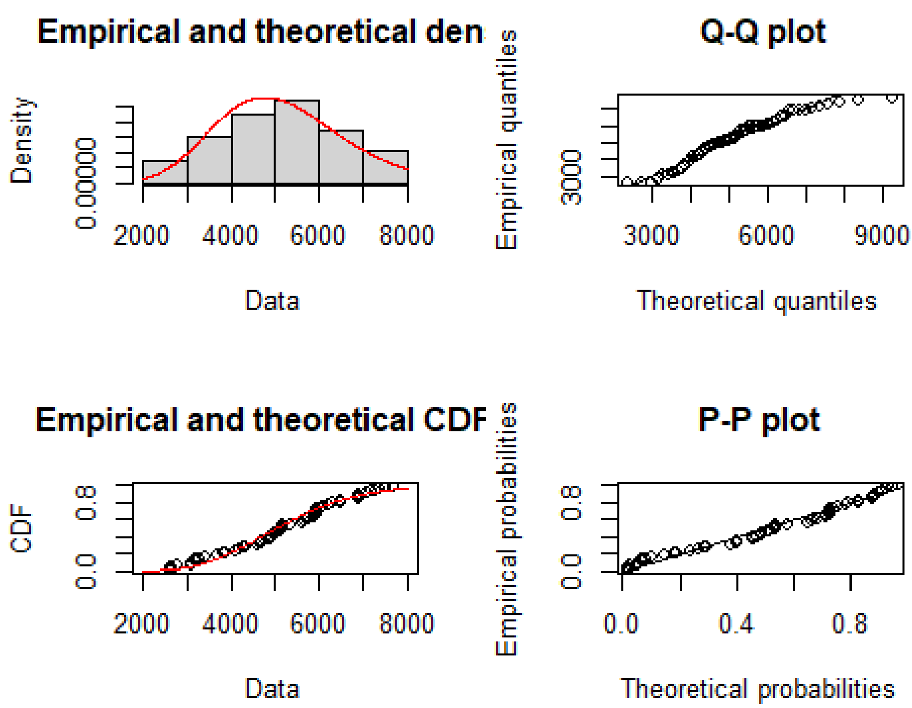
```r
plot(fit_lnorm)
```



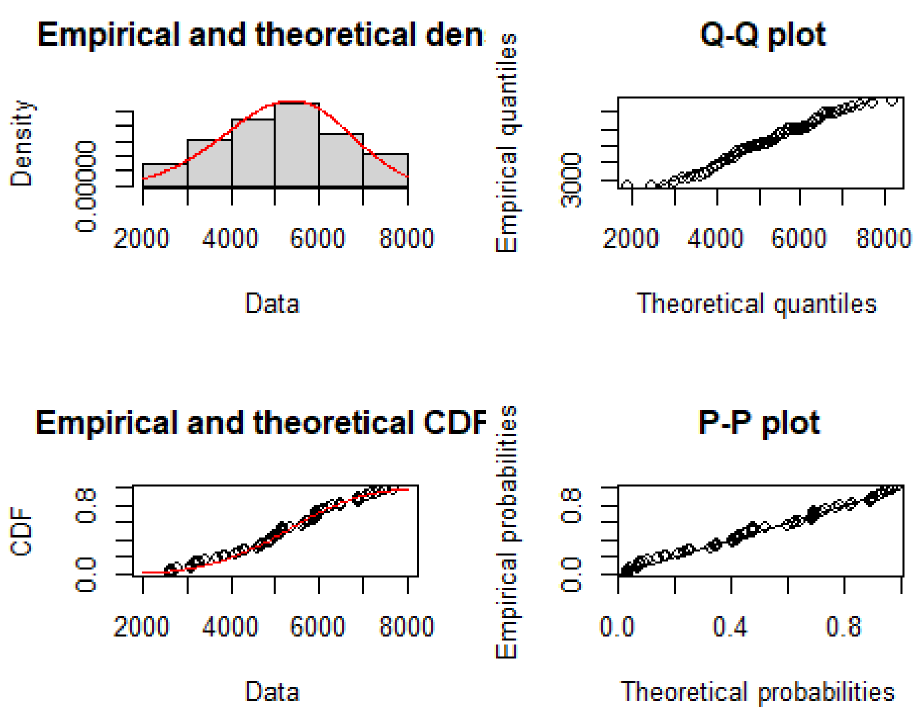

```r
plot(fit_exp)
```



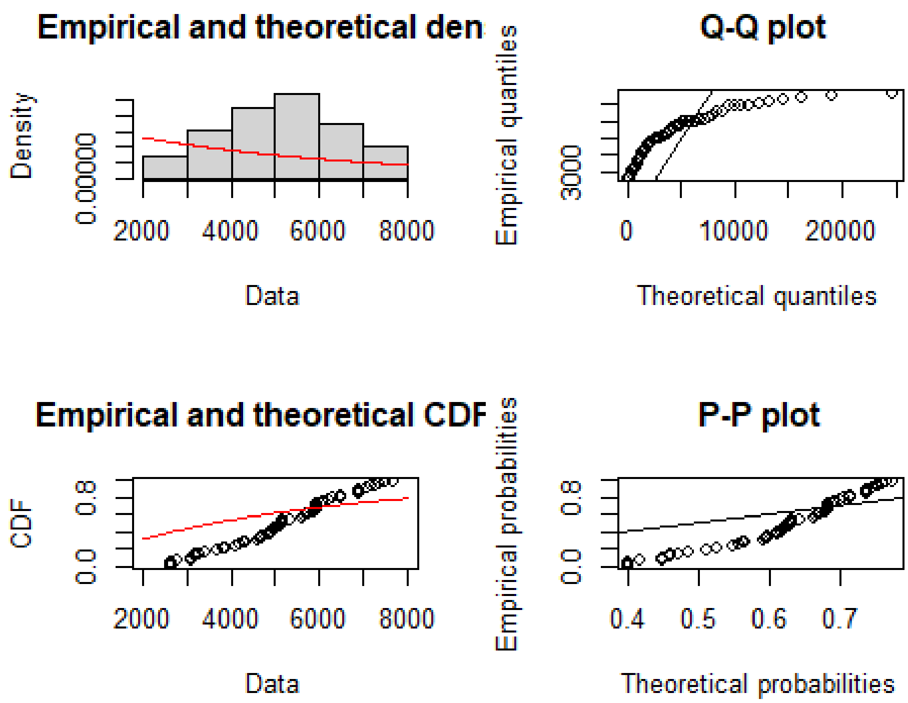

```r
plot(fit_gamma)
```




```
plot(fit_weibull)
```




```
plot(fit_pareto)
```




```
#yg terbaik adalah weibull dengan
#shape = 4.322965
#scale = 5702.441158
```

Berdasarkan grafik-grafik kecocokan model, terlihat bahwa model distribusi terbaik untuk memodelkan besar klaim per polis untuk kelas risiko rendah sebelum disesuaikan dengan tingkat inflasi adalah distribusi Weibull(4.322965, 5702.441158). Indikasi ini juga dibuktikan dengan nilai AIC (1005.964) dan BIC (1010.084) terendah dibandingkan dengan distribusi-distribusi lain (lognormal, eksponensial, gamma, dan pareto).

```
# Distribusi untuk high risk
fit_lnorm <- fitdist(x_data_high$total_claim_amount, "lnorm")
fit_exp <- fitdist(x_data_high$total_claim_amount, "exp")
fit_gamma <- fitdist(x_data_high$total_claim_amount, "gamma")
fit_weibull <- fitdist(x_data_high$total_claim_amount, "weibull")
fit_pareto <- fitdist(x_data_high$total_claim_amount, "pareto", start=list(sh
ape = 1, scale = 1))

# Summary of the fitted models
summary(fit_lnorm)

## Fitting of the distribution ' lnorm ' by maximum likelihood
## Parameters :
##            estimate Std. Error
## meanlog 11.0074859 0.01729023
```

```
## sdlog    0.2888037 0.01222538
## Loglikelihood:  -3120.452   AIC:  6244.904   BIC:  6252.167
## Correlation matrix:
##                meanlog         sdlog
## meanlog  1.000000e+00 -1.309506e-11
## sdlog   -1.309506e-11  1.000000e+00
```

```r
summary(fit_exp)
```

```
## Fitting of the distribution ' exp ' by maximum likelihood
## Parameters :
##          estimate   Std. Error
## rate 1.595081e-05 4.315807e-08
## Loglikelihood:  -3360.834   AIC:  6723.668   BIC:  6727.3
```

```r
summary(fit_gamma)
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##           estimate   Std. Error
## shape 1.314415e+01 2.129398e-01
## rate  2.096431e-04 4.315825e-08
## Loglikelihood:  -3111.134   AIC:  6226.268   BIC:  6233.531
## Correlation matrix:
##             shape        rate
## shape 1.000000000 0.002246858
## rate  0.002246858 1.000000000
```

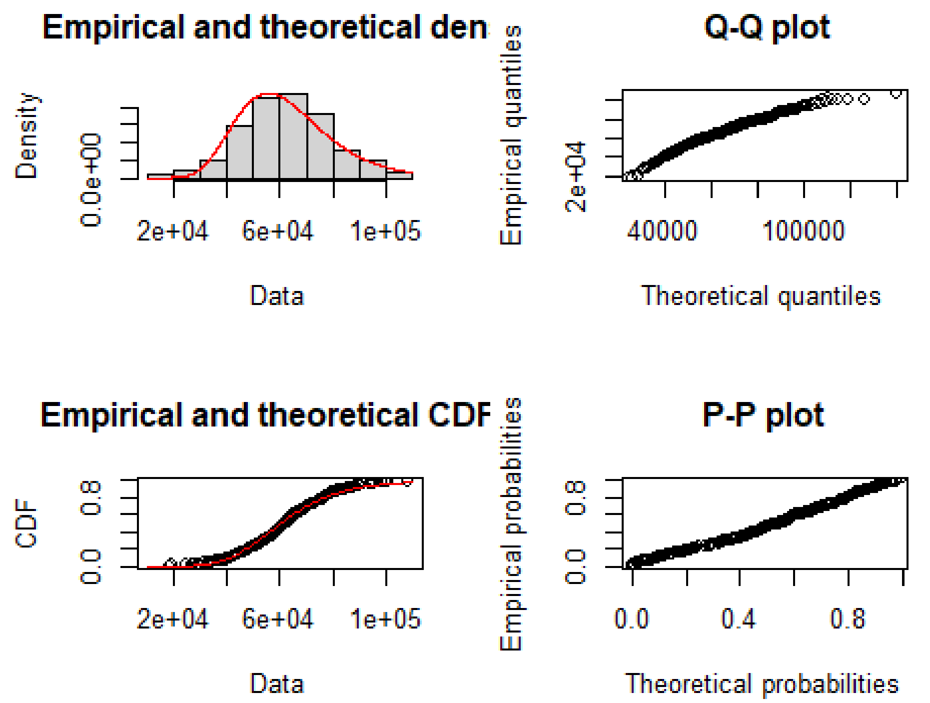
```r
summary(fit_weibull)
```

```
## Fitting of the distribution ' weibull ' by maximum likelihood
## Parameters :
##           estimate   Std. Error
## shape     4.189244    0.1920439
## scale 68925.335137 1062.4719401
## Loglikelihood:  -3106.256   AIC:  6216.511   BIC:  6223.773
## Correlation matrix:
##           shape     scale
## shape 1.0000000 0.3260087
## scale 0.3260087 1.0000000
```

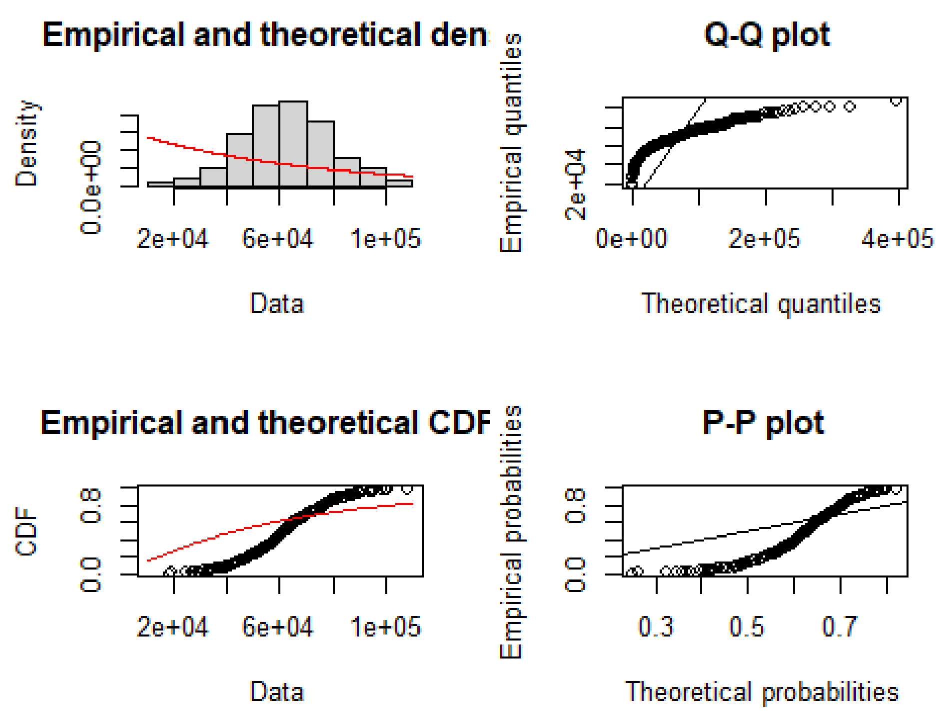
```r
summary(fit_pareto)
```

```
## Fitting of the distribution ' pareto ' by maximum likelihood
## Parameters :
##           estimate
## shape      3124262
## scale 195874114475
## Loglikelihood:  -3360.834   AIC:  6725.668   BIC:  6732.931
```

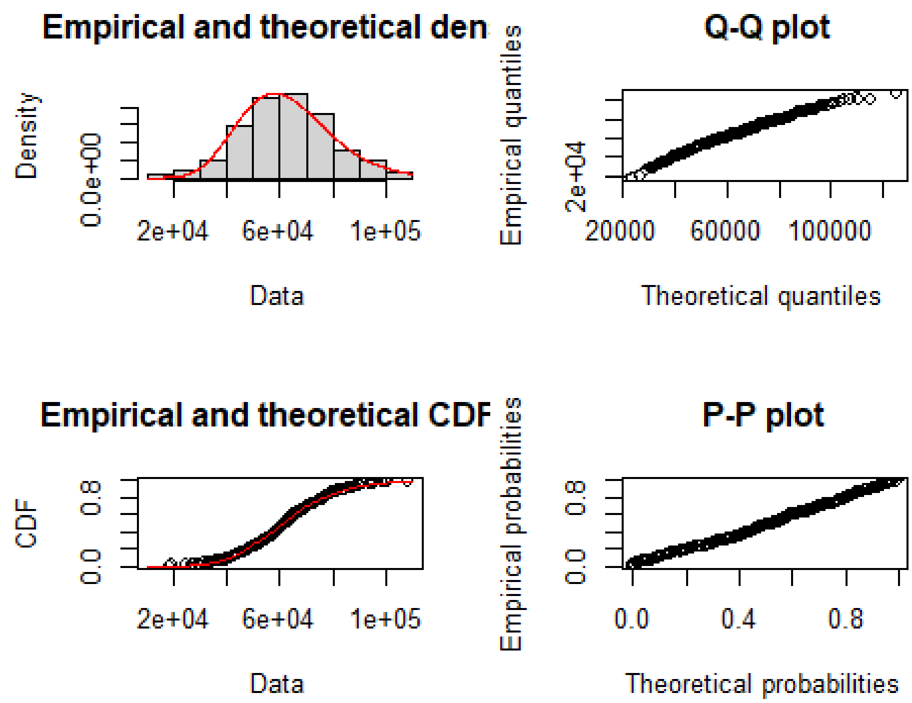
```r
plot(fit_lnorm)
```



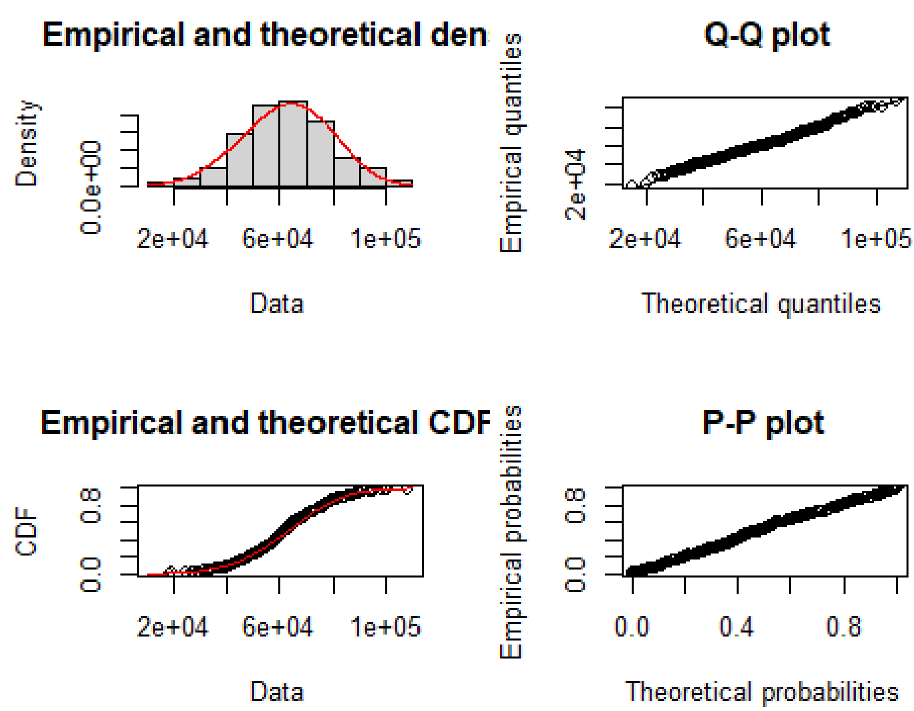

```
plot(fit_exp)
```



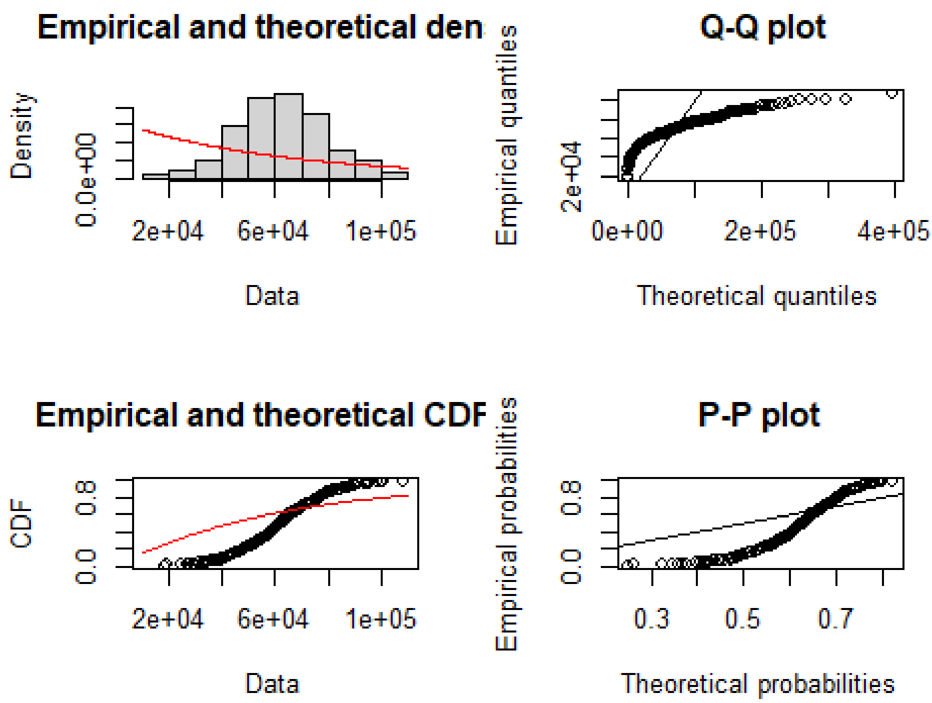

```
plot(fit_gamma)
```




```r
plot(fit_weibull)
```




```r
plot(fit_pareto)
```




```
#fit terbaik adalah weibull dengan
#shape = 4.189244
#scale = 68925.335137
```

Serupa dengan analisis untuk kelas risiko rendah, besar klaim untuk kelas risiko tinggi untuk nilai sebelum disesuaikan dengan inflasi juga terlihat paling cocok dimodelkan dengan distribusi Weibull. Hanya saja, parameter distribusi untuk kelas risiko tinggi berbeda, yaitu Weibull(4.189244, 68925.335137). Hal ini juga didukung dengan nilai AIC(6216.511) dan BIC(6223.773) terendah dibandingkan dengan distribusi-distribusi lain.

Distribusi untuk inflation adjusted data

```
# Distribusi untuk low risk
fit_lnorm <- fitdist(x_data_low$total_claim_amount_adj, "lnorm")
fit_exp <- fitdist(x_data_low$total_claim_amount_adj, "exp")
fit_gamma <- fitdist(x_data_low$total_claim_amount_adj, "gamma")
fit_weibull <- fitdist(x_data_low$total_claim_amount_adj, "weibull")
fit_pareto <- fitdist(x_data_low$total_claim_amount_adj, "pareto", start=list
(shape = 1, scale = 1))

# Summary of the fitted models
summary(fit_lnorm)

## Fitting of the distribution ' lnorm ' by maximum likelihood
## Parameters :
```

```
##          estimate Std. Error
## meanlog 8.8009515 0.04045424
## sdlog   0.3080903 0.02860411
## Loglikelihood:  -524.4666   AIC:  1052.933   BIC:  1057.054
## Correlation matrix:
##         meanlog sdlog
## meanlog       1     0
## sdlog         0     1
```

```r
summary(fit_exp)
```

```
## Fitting of the distribution ' exp ' by maximum likelihood
## Parameters :
##          estimate   Std. Error
## rate 0.0001438182 4.315837e-08
## Loglikelihood:  -571.1237   AIC:  1144.247   BIC:  1146.308
```

```r
summary(fit_gamma)
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##           estimate   Std. Error
## shape 11.034058568 4.263307e-01
## rate   0.001586811 4.315839e-08
## Loglikelihood:  -524.0076   AIC:  1052.015   BIC:  1056.136
## Correlation matrix:
##              shape         rate
## shape 1.0000000000 0.0007915727
## rate  0.0007915727 1.0000000000
```

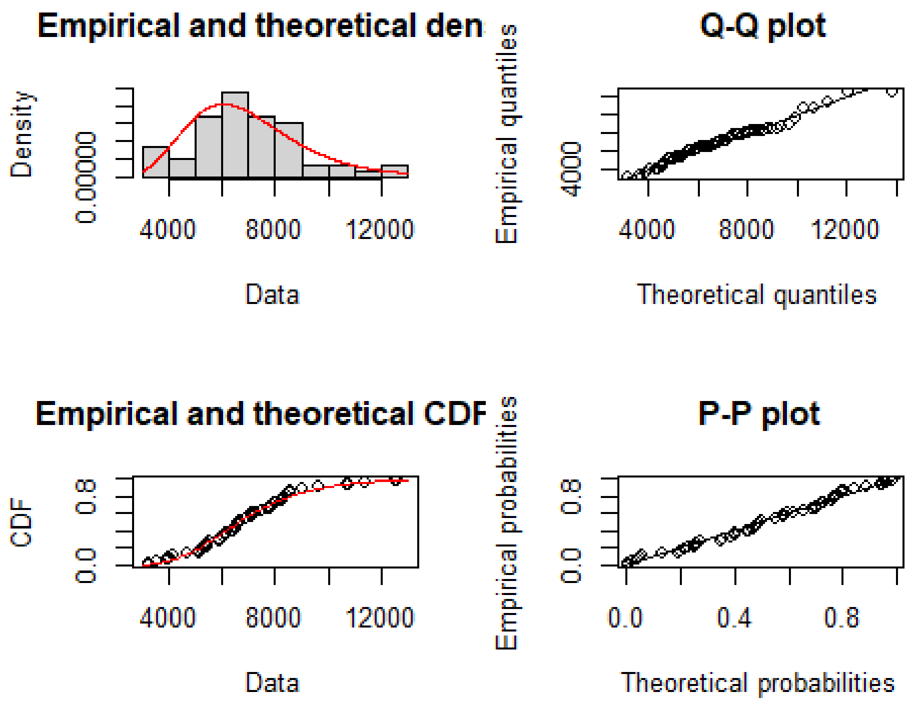
```r
summary(fit_weibull)
```

```
## Fitting of the distribution ' weibull ' by maximum likelihood
## Parameters :
##          estimate  Std. Error
## shape    3.497058   0.3367787
## scale 7715.555044 307.3810307
## Loglikelihood:  -526.2638   AIC:  1056.528   BIC:  1060.649
## Correlation matrix:
##           shape     scale
## shape 1.0000000 0.3296258
## scale 0.3296258 1.0000000
```

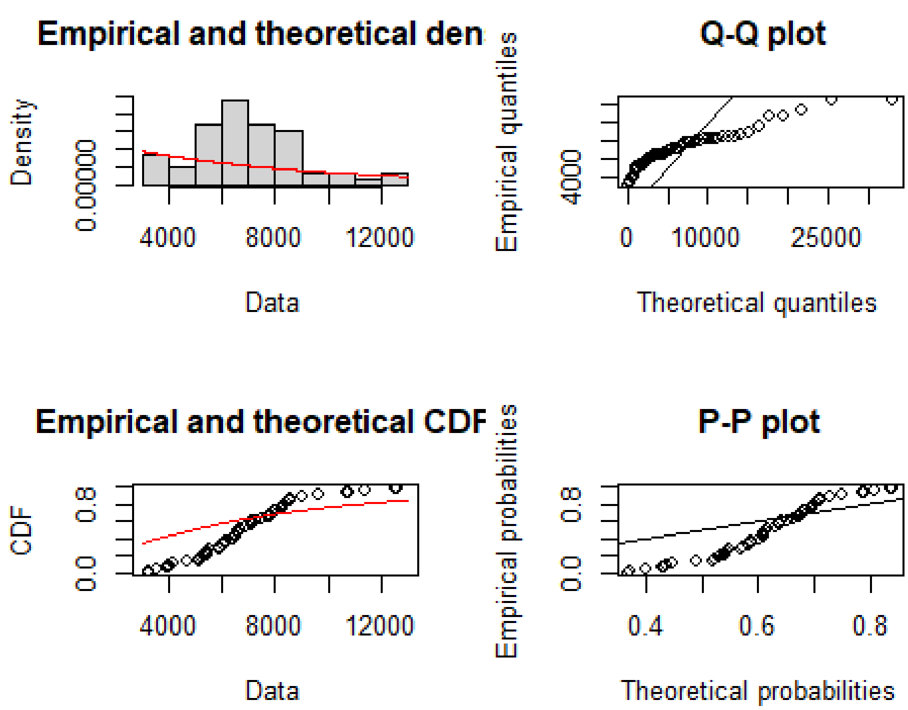
```r
summary(fit_pareto)
```

```
## Fitting of the distribution ' pareto ' by maximum likelihood
## Parameters :
##          estimate
## shape     2233228
## scale 15526829414
## Loglikelihood:  -571.1237   AIC:  1146.247   BIC:  1150.368
```

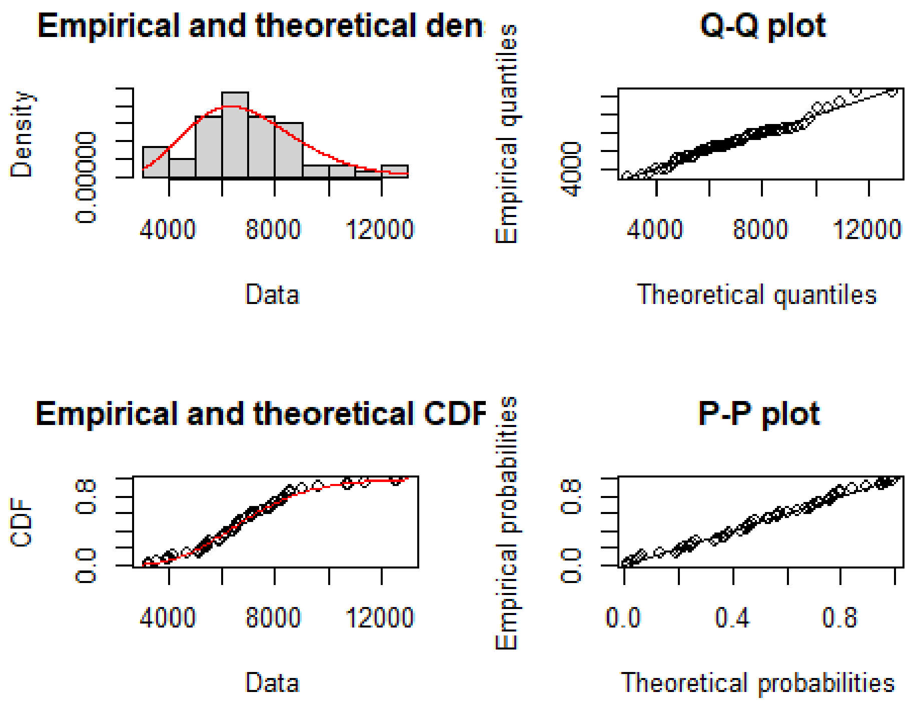
```r
plot(fit_lnorm)
```



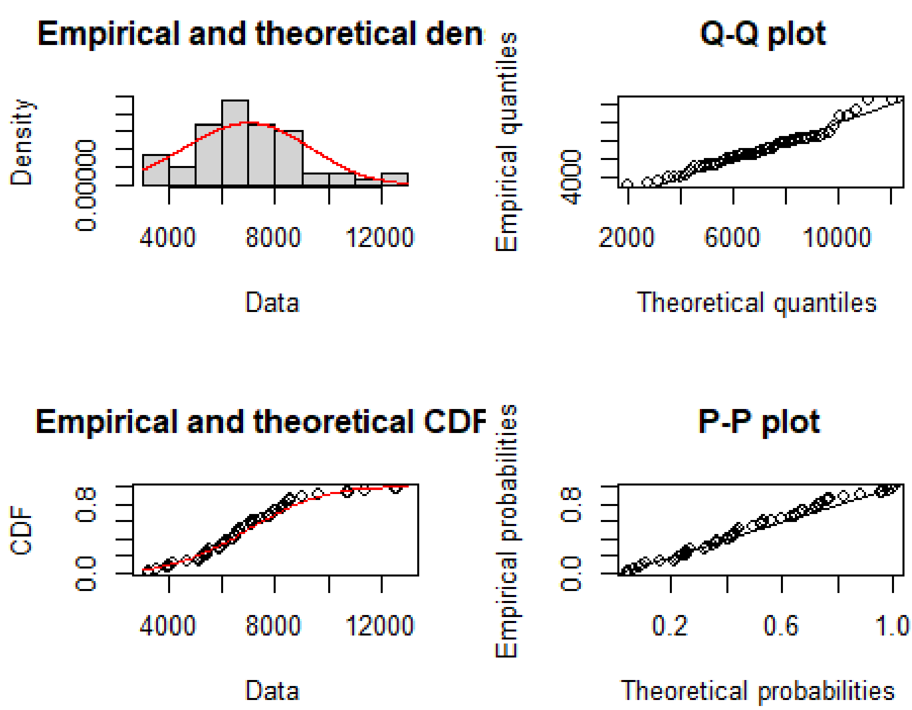

```r
plot(fit_exp)
```



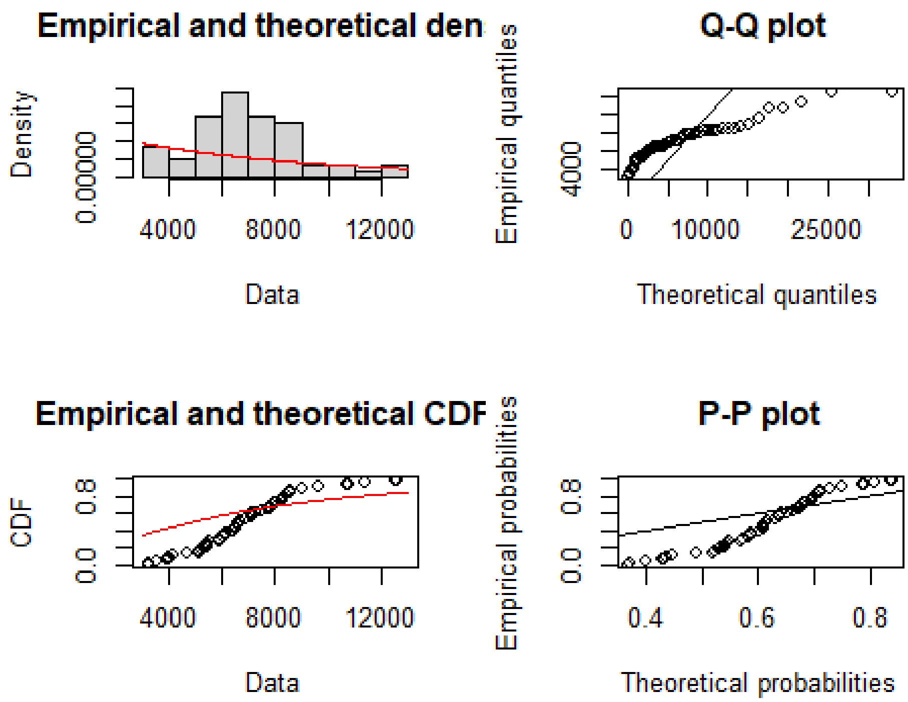
```r
plot(fit_gamma)
```




```r
plot(fit_weibull)
```



```r
plot(fit_pareto)
```




```r
#the best fit adalah gamma dengan
#shape = 11.034058568
#rate = 0.001586811
```

Berbeda dengan data tidak disesuaikan dengan inflasi, data besar klaim per polis yang telah disesuaikan dengan inflasi juga menunjukkan adanya kesesuaian model dengan distribusi Gamma. Data kelas risiko rendah terlihat cocok dimodelkan dengan distribusi Gamma(11.034058568, 1/0.001586811). Hal ini didasari dari kecocokan grafik distribusi Gamma dibandingkan dengan distribusi lainnya serta nilai AIC (1052.015) dan BIC (1056.136) terkecil dibandingkan distribusi lainnya.

```r
# Distribusi untuk high risk
x_data_high

## # A tibble: 279 × 5
##    policy_ID loss_year total_claim_amount   cpi total_claim_amount_adj
##        <dbl>     <dbl>              <dbl> <dbl>                  <dbl>
##  1    110084      1990              55000  134.                 96522.
##  2    113516      1990              67320  134.                118143.
##  3    227811      1990              63400  134.                111264.
##  4    307447      1990              93720  134.                164474.
##  5    429027      1990              51590  134.                 90538.
##  6    506333      1990              76890  134.                134938.
```

```
## 7    512813      1990              64350  134.                 112931.
## 8    687755      1990              64200  134.                 112668.
## 9    793948      1990              47430  134.                  83237.
## 10   131478      1991              58200  138.                  99101.
## # ℹ 269 more rows
```

```r
fit_lnorm <- fitdist(x_data_high$total_claim_amount_adj, "lnorm")
fit_exp <- fitdist(x_data_high$total_claim_amount_adj, "exp")
fit_gamma <- fitdist(x_data_high$total_claim_amount_adj, "gamma")
fit_weibull <- fitdist(x_data_high$total_claim_amount_adj, "weibull")
fit_pareto <- fitdist(x_data_high$total_claim_amount_adj, "pareto", start=list(shape = 1, scale = 1))

# Summary of the fitted models
summary(fit_lnorm)
```

```
## Fitting of the distribution ' lnorm ' by maximum likelihood
## Parameters :
##           estimate Std. Error
## meanlog 11.2711250  0.0202621
## sdlog    0.3384438  0.0143269
## Loglikelihood:  -3238.26   AIC:  6480.52   BIC:  6487.782
## Correlation matrix:
##         meanlog sdlog
## meanlog       1     0
## sdlog         0     1
```

```r
summary(fit_exp)
```

```
## Fitting of the distribution ' exp ' by maximum likelihood
## Parameters :
##          estimate   Std. Error
## rate 1.206884e-05 4.315796e-08
## Loglikelihood:  -3438.643   AIC:  6879.285   BIC:  6882.916
```

```r
summary(fit_gamma)
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##           estimate   Std. Error
## shape 9.4610247795 1.793092e-01
## rate  0.0001141539 4.315815e-08
## Loglikelihood:  -3231.906   AIC:  6467.812   BIC:  6475.074
## Correlation matrix:
##            shape       rate
## shape 1.00000000 0.00245925
## rate  0.00245925 1.00000000
```

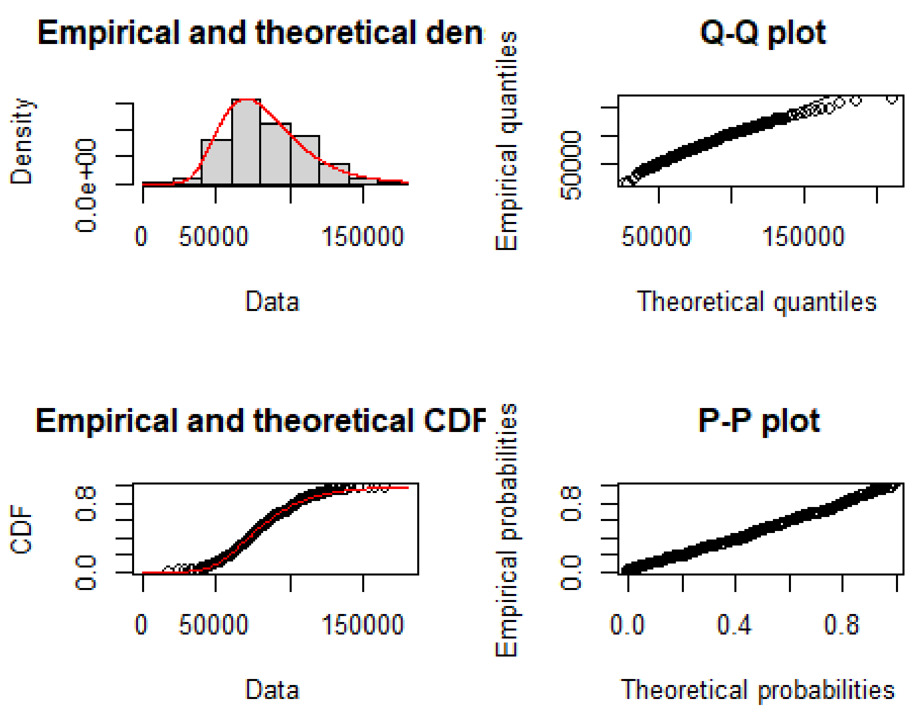
```r
summary(fit_weibull)
```

```
## Fitting of the distribution ' weibull ' by maximum likelihood
## Parameters :
##            estimate   Std. Error
## shape     3.384101    0.1504775
## scale 92212.076079 1480.9816481
## Loglikelihood:  -3235.296   AIC:  6474.591   BIC:  6481.854
## Correlation matrix:
##            shape     scale
## shape 1.0000000 0.2829236
## scale 0.2829236 1.0000000
```

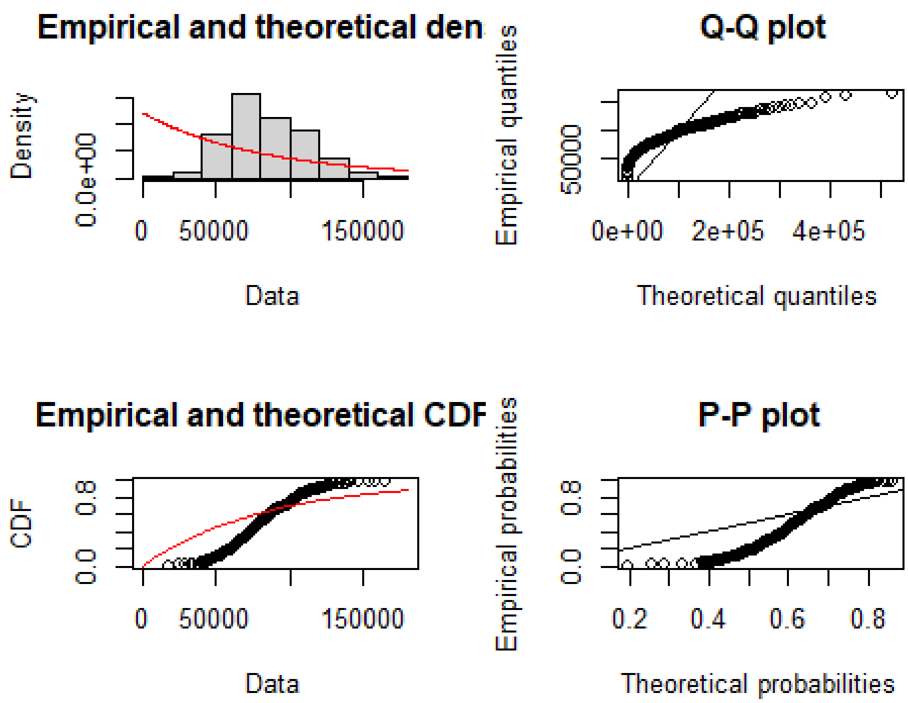
```r
summary(fit_pareto)
```

```
## Fitting of the distribution ' pareto ' by maximum likelihood
## Parameters :
##            estimate
## shape 2.302612e+05
## scale 1.908325e+10
## Loglikelihood:  -3438.643   AIC:  6881.286   BIC:  6888.549
```

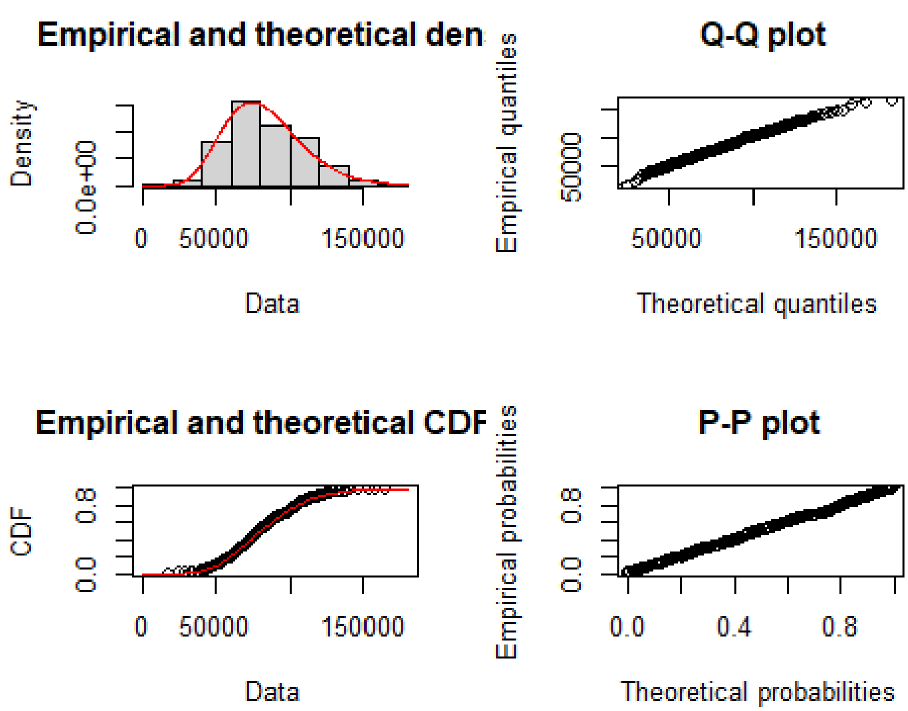
```r
plot(fit_lnorm)
```



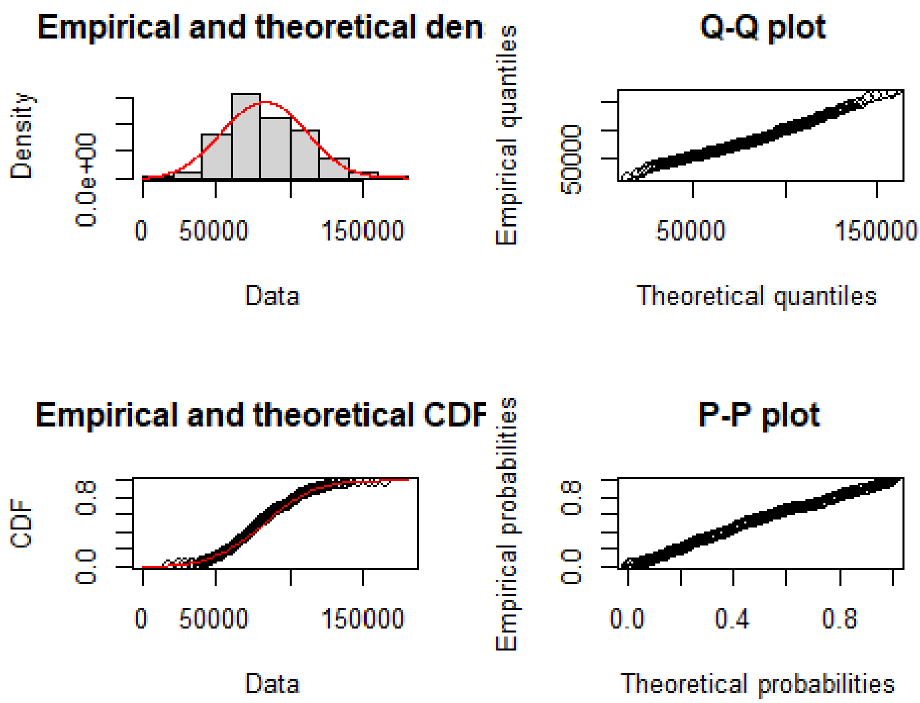

```r
plot(fit_exp)
```



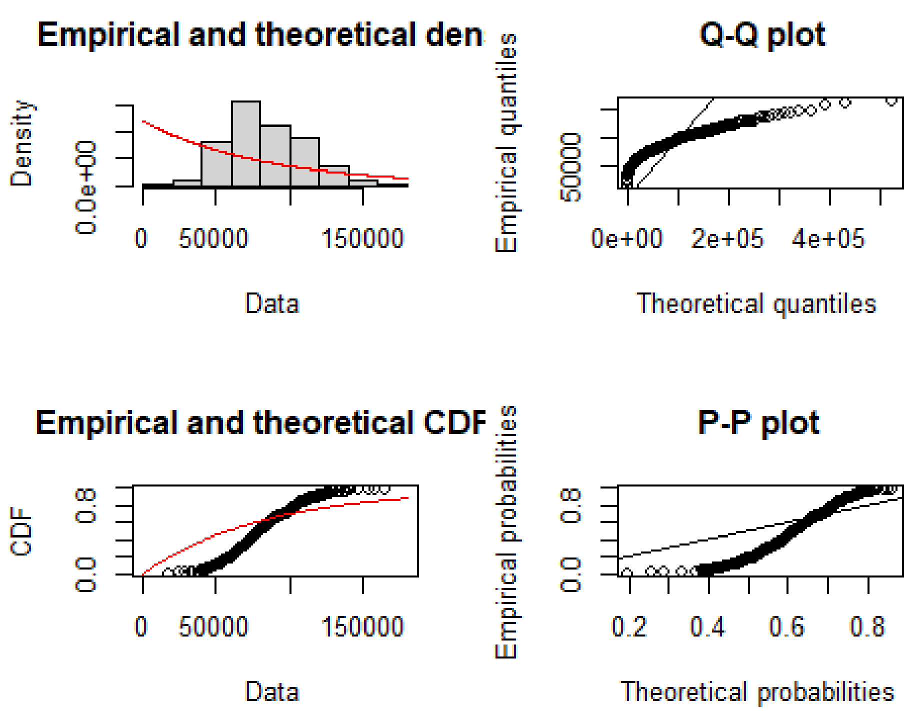

```
plot(fit_gamma)
```




```
plot(fit_weibull)
```




```
plot(fit_pareto)
```



```
#fit terbaik adalah gamma dengan
#shape = 9.4610247795
#rate = 0.0001141539
```

Data besar klaim per polis kelas risiko tinggi yang sudah disesuaikan dengan tingkat inflasi cocok dimodelkan dengan distribusi Gamma(9.4610247795, 1/0.0001141539). Maka baik untuk data yang disesuaikan dengan inflasi maupun tidak, menunjukkan bahwa perbedaan tingkat risiko dari polis dapat dimodelkan dengan distribusi yang sama hanya saja dengan parameter yang berbeda, yaitu distribui Weibull untuk data tidak disesuaikan dan Gamma untuk data yang sudah disesuaikan.

#2. Second Central Moment Kita notasikan R: Peubah acak yang menunjukkan kelas risiko polis R bernilai 1 dan 2 untuk kelas risiko rendah dan risiko tinggi

X|R: Peubah acak yang menunjukkan besar klaim per polis pada kelas risiko r

Noninflation Adjusted

P(R=1) = 0.1721068 P(R=2) = 0.8278932

X|R=1 ~ Weibull(4.322965, 5702.441158) X|R=2 ~ Weibull(4.189244, 68925.335137)

```
#The parameters
p_R1 <- 0.1721068
p_R2 <- 0.8278932
shape_R1_weibull <- 4.322965
scale_R1_weibull <- 5702.441158
shape_R2_weibull <- 4.189244
scale_R2_weibull <- 68925.335137

#Conditional Expectation of X
ex_given_R1_weibull <- scale_R1_weibull * gamma(1 + 1 / shape_R1_weibull)
ex_given_R2_weibull <- scale_R2_weibull * gamma(1 + 1 / shape_R2_weibull)

#Conditional  Variance of X
varx_given_R1_weibull <- scale_R1_weibull^2 * (gamma(1 + 2/shape_R1_weibull))
- ex_given_R1_weibull^2
varx_given_R2_weibull <- scale_R2_weibull^2 * (gamma(1 + 2/shape_R2_weibull))
- ex_given_R2_weibull^2

#Expectation of X
ex_weibull <- ex_given_R1_weibull*p_R1 + ex_given_R2_weibull*p_R2

#EPV of X
```

```
epvx_weibull <- varx_given_R1_weibull*p_R1 + varx_given_R2_weibull*p_R2

#VHM of X
vhmx_weibull <- (ex_given_R1_weibull^2)*p_R1 + (ex_given_R2_weibull^2)*p_R2 -
ex_weibull^2

#Variance of X
varx_weibull <- epvx_weibull + vhmx_weibull

print("Jadi, untuk data yang tidak disesuaikan diperoleh:")

## [1] "Jadi, untuk data yang tidak disesuaikan diperoleh:"

paste0("Besar momen pusat kedua dari X bersyarat R=1: ", varx_given_R1_weibul
l)

## [1] "Besar momen pusat kedua dari X bersyarat R=1: 1843094.94522275"

paste0("Besar momen pusat kedua dari X bersyarat R=2: ", varx_given_R2_weibul
l)

## [1] "Besar momen pusat kedua dari X bersyarat R=2: 284023922.455695"

paste0("Besar momen pusat kedua dari X: ", varx_weibull)

## [1] "Besar momen pusat kedua dari X: 705697899.445023"
```

Inflation Adjusted

P(R=1) = 0.1721068 P(R=2) = 0.8278932

X|R=1 ~ Gamma(11.034058568, 1/0.0015868118) X|R=2 ~ Gamma(9.4610247795, 1/0.0001141539)

```
#The Parameters
shape_R1_gamma <- 11.034058568
rate_R1_gamma <- 1/0.0015868118
shape_R2_gamma <- 9.4610247795
rate_R2_gamma <- 1/0.0001141539

#Conditional Expectation of X
ex_given_R1_gamma <- shape_R1_gamma * rate_R1_gamma
ex_given_R2_gamma <- shape_R2_gamma * rate_R2_gamma

#Conditional  Variance of X
varx_given_R1_gamma <- shape_R1_gamma * rate_R1_gamma^2
varx_given_R2_gamma <- shape_R2_gamma * rate_R2_gamma^2

#Expectation of X
```

```
ex_gamma <- ex_given_R1_gamma*p_R1 + ex_given_R2_gamma*p_R2

#EPV of X
epvx_gamma <- varx_given_R1_gamma*p_R1 + varx_given_R2_gamma*p_R2

#VHM of X
vhmx_gamma <- (ex_given_R1_gamma^2)*p_R1 + (ex_given_R2_gamma^2)*p_R2 - ex_ga
mma^2

#Variance of X
varx_gamma <- epvx_gamma + vhmx_gamma

print("Jadi, untuk data yang telah disesuaikan diperoleh:")

## [1] "Jadi, untuk data yang telah disesuaikan diperoleh:"

paste0("Besar momen pusat kedua dari X bersyarat R=1: ", varx_given_R1_gamma)

## [1] "Besar momen pusat kedua dari X bersyarat R=1: 4382121.7759802"

paste0("Besar momen pusat kedua dari X bersyarat R=2: ", varx_given_R2_gamma)

## [1] "Besar momen pusat kedua dari X bersyarat R=2: 726033521.683174"

paste0("Besar momen pusat kedua dari X: ", varx_gamma)

## [1] "Besar momen pusat kedua dari X: 1423228963.89507"
```

#3. Credibility at 80% confidence level ## For non inflation adjusted data

```
#maximum k (deviation from true mean)
## for noninflation adjusted data
k_nonadj <- qnorm((0.8+1)/2)*(varx_weibull/337)^(1/2)/ex_weibull

## for inflation adjusted data
k_adj <- qnorm((0.8+1)/2)*(varx_gamma/337)^(1/2)/ex_gamma

paste0("Deviasi maksimum untuk data tidak disesuaikan: ", k_nonadj)

## [1] "Deviasi maksimum untuk data tidak disesuaikan: 0.0351550794126899"

paste0("Deviasi maksimum untuk data telah disesuaikan: ", k_adj)

## [1] "Deviasi maksimum untuk data telah disesuaikan: 0.0377248300678088"
```

Dari hasil ini terlihat bahwa kedua deviasi maksimum yang dapat digunakan agar kredibilitas klasik memperoleh kredibilitas penuh adalah 0.0352 dan 0.0377 untuk data tidak disesuaikan dan data telah disesuaikan. Dari hasil ini, nilai k yang masih dibawh 4% masih cukup masuk akal dan dapat diterima. Maka dari itu, pada tingkat kepercayaan 80%, data yang digunakan memiliki nilai kredibilitas yang cukup tinggi, yaitu kredibilitas penuh dengan deviasi dari mean X sebenarnya lebih kecil dari 4%.

4. Buhlman Credibility

```
#The Buhlmann Credibility
z_nonadj = 337/(337+(epvx_weibull/vhmx_weibull))
z_adj = 337/(337+(epvx_gamma/vhmx_gamma))

#Estimate of future claim
x_nonadj = ex_weibull*(1-z_nonadj) + z_nonadj*mean(data_adj$total_claim_amoun
t)
x_adj = ex_gamma*(1-z_adj) + z_adj*mean(data_adj$total_claim_amount_adj)

paste0("Kredibilitas Buhlmann data tidak sesuaikan: ", z_nonadj)

## [1] "Kredibilitas Buhlmann data tidak sesuaikan: 0.998516385186431"

paste0("Kredibilitas Buhlmann data telah sesuaikan: ", z_adj)

## [1] "Kredibilitas Buhlmann data telah sesuaikan: 0.997830550280848"

paste0("Estimasi nilai klaim selanjutnya data tidak sesuaikan: ", x_nonadj)

## [1] "Estimasi nilai klaim selanjutnya data tidak sesuaikan: 52794.29980580
4"

paste0("Estimasi nilai klaim selanjutnya data telah sesuaikan: ", x_adj)

## [1] "Estimasi nilai klaim selanjutnya data telah sesuaikan: 69794.32663287
"
```

Baik data yang sudah disesuaikan dengan tingkat inflasi maupun belum, keduanya menunjukkan keberadaan perbedaan kelas risiko. Karena data peluang terpilihnya masing-masing kelas risiko dapat diperoleh (diestimasi) dengan proporsi frekuensi klaim dengan nilai rendah terhadap frekuensi klaim dengan nilai tinggi, yaitu diperoleh $P(R=1) = 0.1721068$ dan $P(R=2) = 0.8278932$, maka kredibilitas Buhlmann dapat dihitung. Untuk data yang tidak sesuaikan, diperoleh kredibilitas Buhlmann sebesar 0.9985 dan 0.9978 untuk data yang telah disesuaikan. Kedua nilai tersebut mendorong kita untuk percaya terhadap data yang sudah diperoleh. Artinya, nilai kredibilitas data cukup tinggi berdasarkan kredibilitas Buhlmann. Dari nilai ini, dapat pula diperoleh estimasi ekspektasi nilai klaim selanjutnya yaitu 52794.2998 (data tidak disesuaikan) dan 69794.3266 (data telah disesuaikan).