Real Data Analysis

Anne STAFF, Bianca BOI

2023-05-01

Real data analysis

Preparation (dependencies & data):

```
library(fitdistrplus)
```

```
## Warning: package 'fitdistrplus' was built under R version 4.2.3
```

```
## Loading required package: MASS
```

```
## Loading required package: survival
```

```
gpigs <- read.table("surv.gpigs.txt", header = T, sep = ";")
gpigs.noncensored <- gpigs[gpigs$censored == 0,]

gpigs.mc <- gpigs.noncensored$lifetime[gpigs.noncensored$regime == "M_C"]
gpigs.mc.with_censored <- gpigs$lifetime[gpigs$regime == "M_C"]
gpigs.m43 <- gpigs.noncensored$lifetime[gpigs.noncensored$regime == "M_4.3"]
```

Description of the data

Provide meaningful descriptive statistics to compare the two treatments M4.3 and M.C.

```
summary(gpigs.mc.with_censored)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    18.0   214.0   621.0   500.9   735.0   735.0
```

```
summary(gpigs.m43)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10.0   108.0   149.5   176.8   224.0   555.0
```

Distribution fit

Using the R-package 'fitdistrplus', fit a distribution to the 'lifetime' variable for both control and treatment taking into account the right censoring. Start with Exponential distribution. Interpret the parameter estimates.
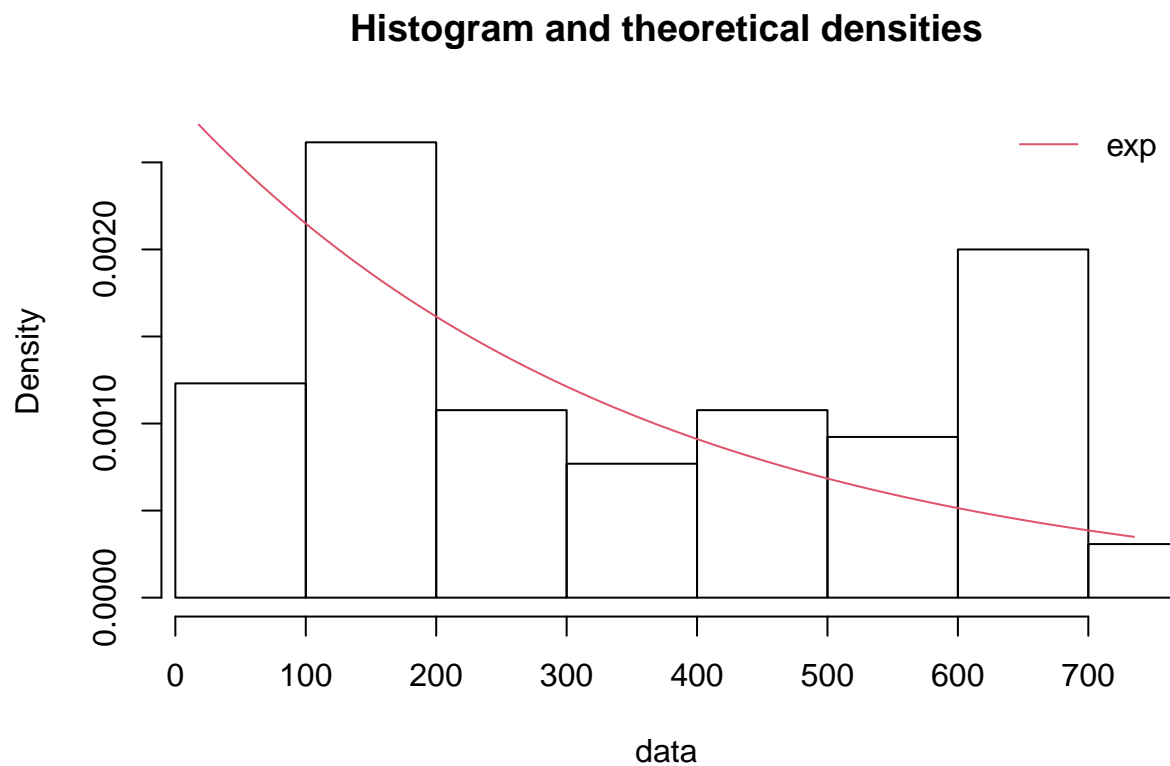
```
fe <- fitdist(gpigs.mc, "exp")
fe$estimate
```

Exponential distribution

```
##        rate
## 0.002859783
```
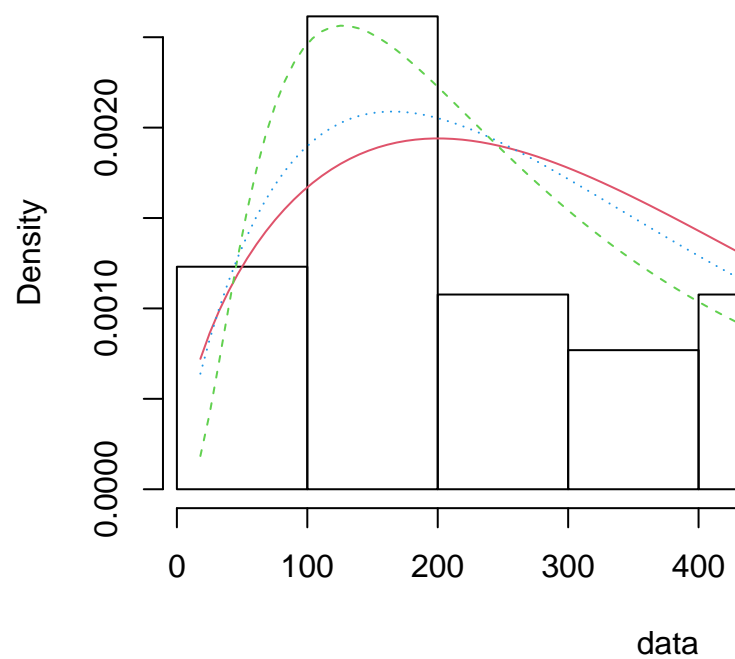
TO DO: interpret the rate estimate

```
denscomp(list(fe))
```

Histogram and theoretical densities
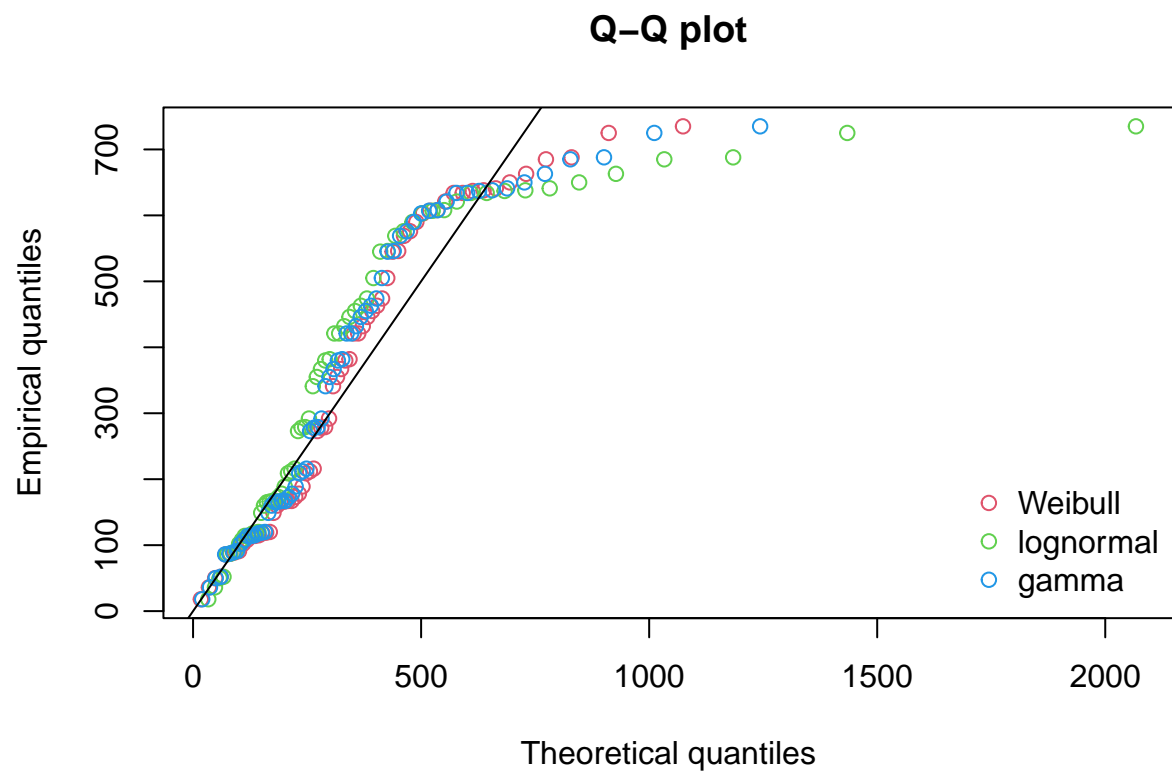

TO DO: describe why it's not adapted

```
fw <- fitdist(gpigs.mc, "weibull")
fg <- fitdist(gpigs.mc, "gamma")
fln <- fitdist(gpigs.mc, "lnorm")
plot.legend <- c("Weibull", "lognormal", "gamma")
denscomp(list(fw, fln, fg), legendtext = plot.legend)
```
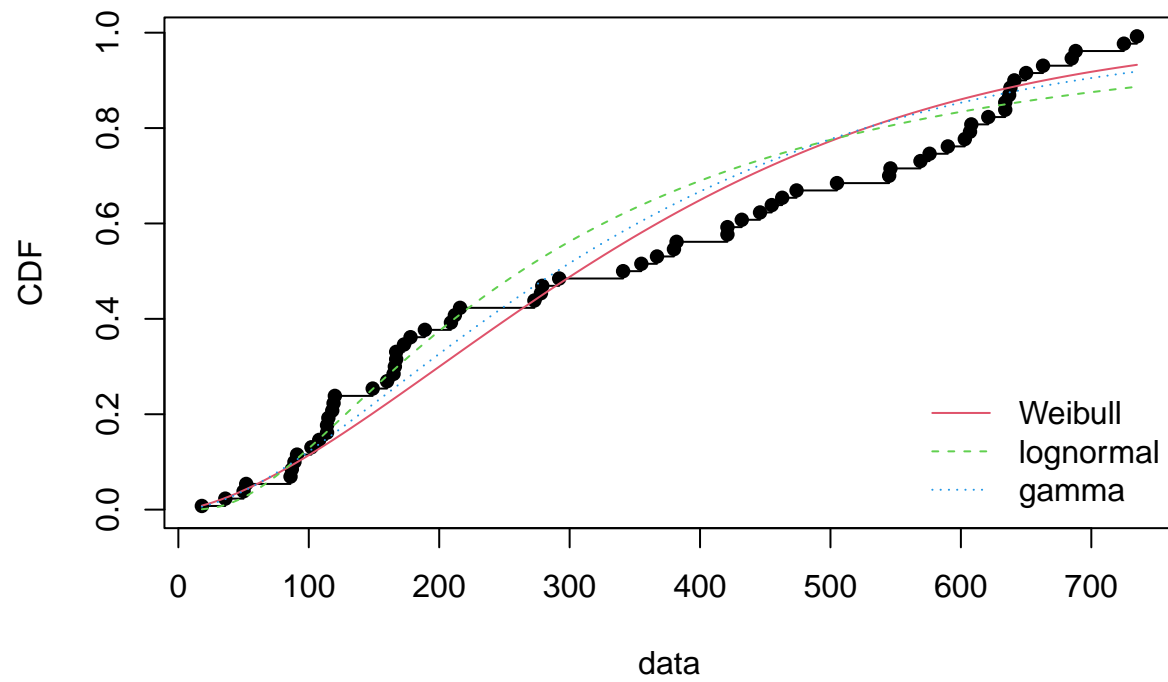
data

Weibull, Gamma, log-Normal fit on the control group

```
qqcomp(list(fw, fln, fg), legendtext = plot.legend)
```
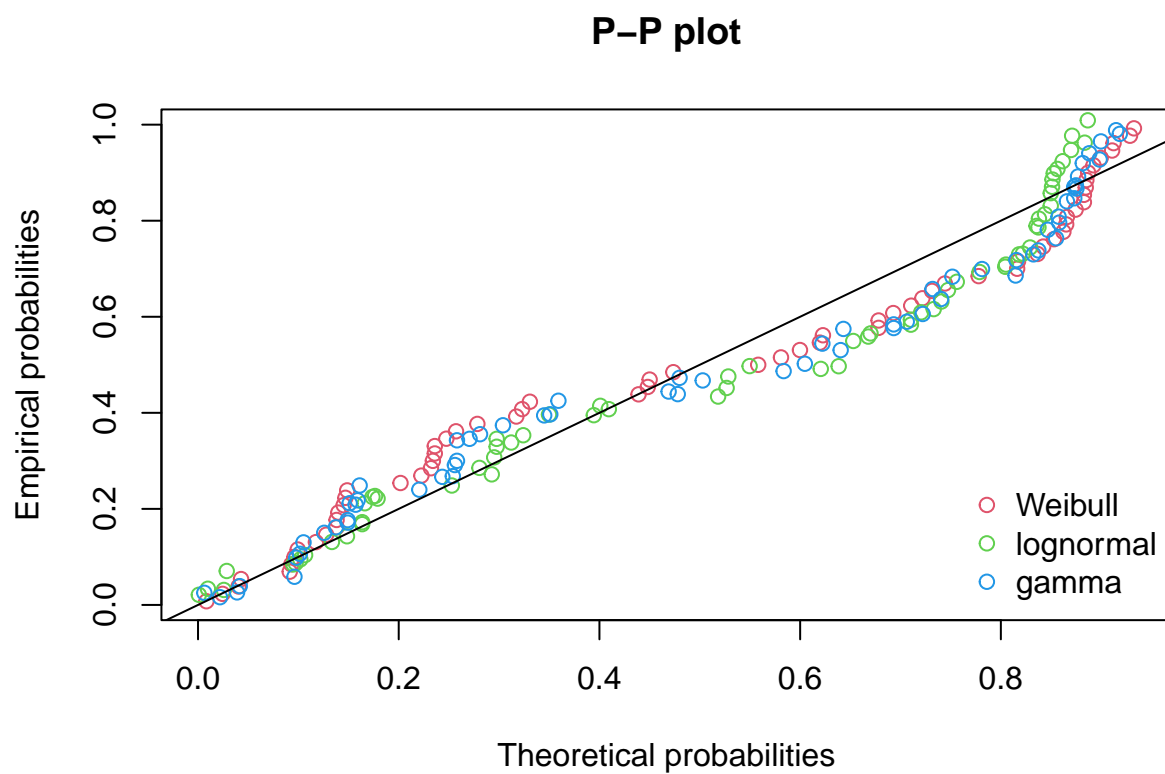
Q–Q plot


```
cdfcomp(list(fw, fln, fg), legendtext = plot.legend)
```

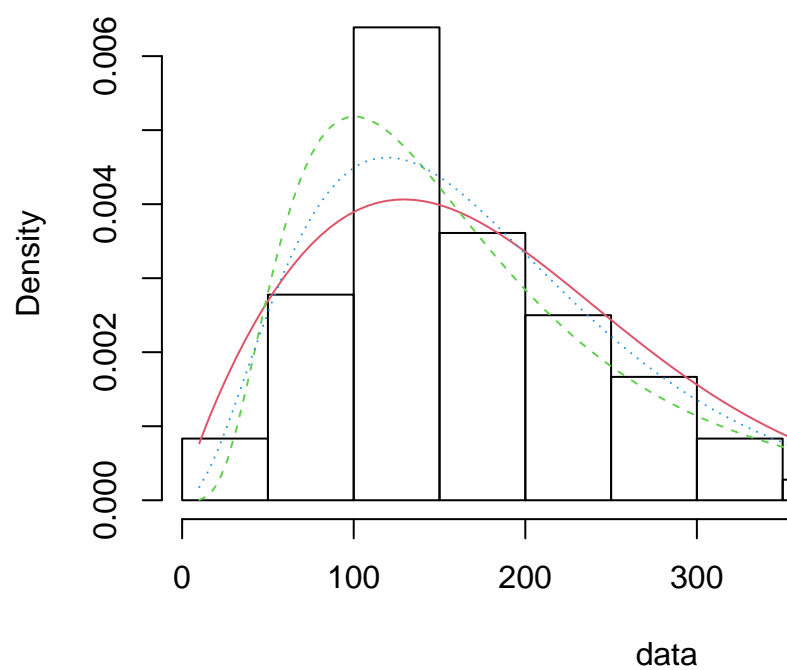**Empirical and theoretical CDFs**


```
ppcomp(list(fw, fln, fg), legendtext = plot.legend)
```
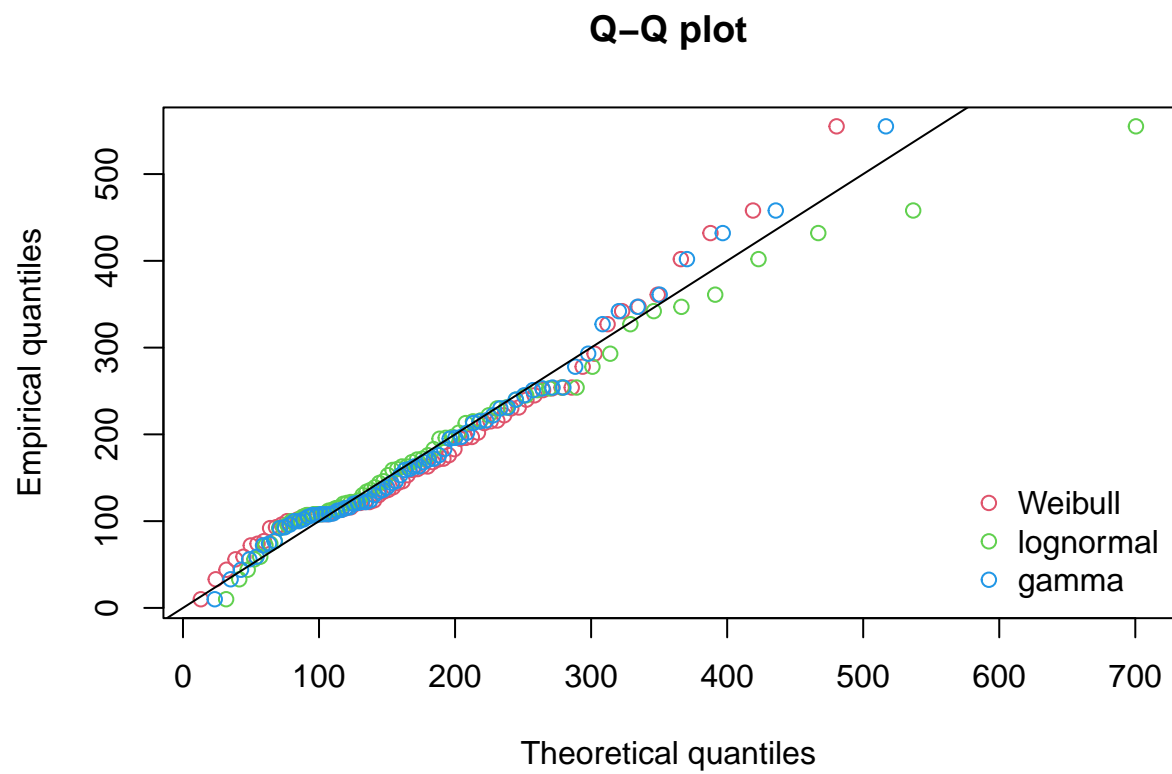
P–P plot


```
fw <- fitdist(gpigs.m43, "weibull")
fg <- fitdist(gpigs.m43, "gamma")
fln <- fitdist(gpigs.m43, "lnorm")
plot.legend <- c("Weibull", "lognormal", "gamma")
denscomp(list(fw, fln, fg), legendtext = plot.legend)
```

Histogram and theoretica

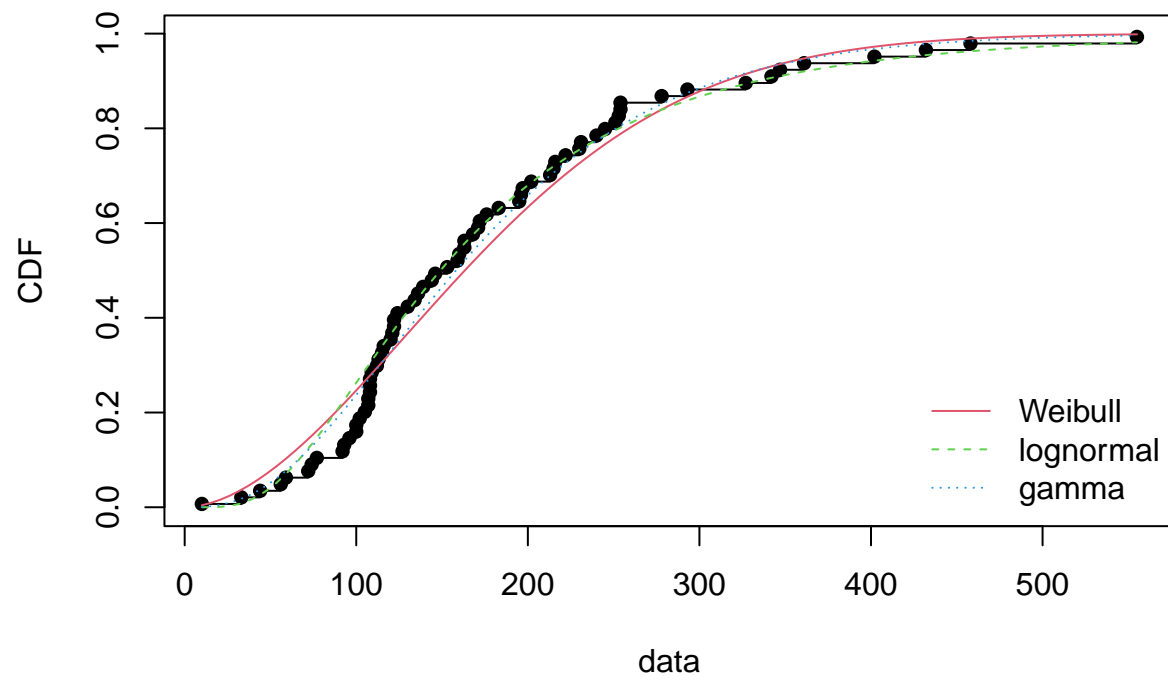Weibull, Gamma, log-Normal fit on the 4.3 group

```
qqcomp(list(fw, fln, fg), legendtext = plot.legend)
```

Q–Q plot

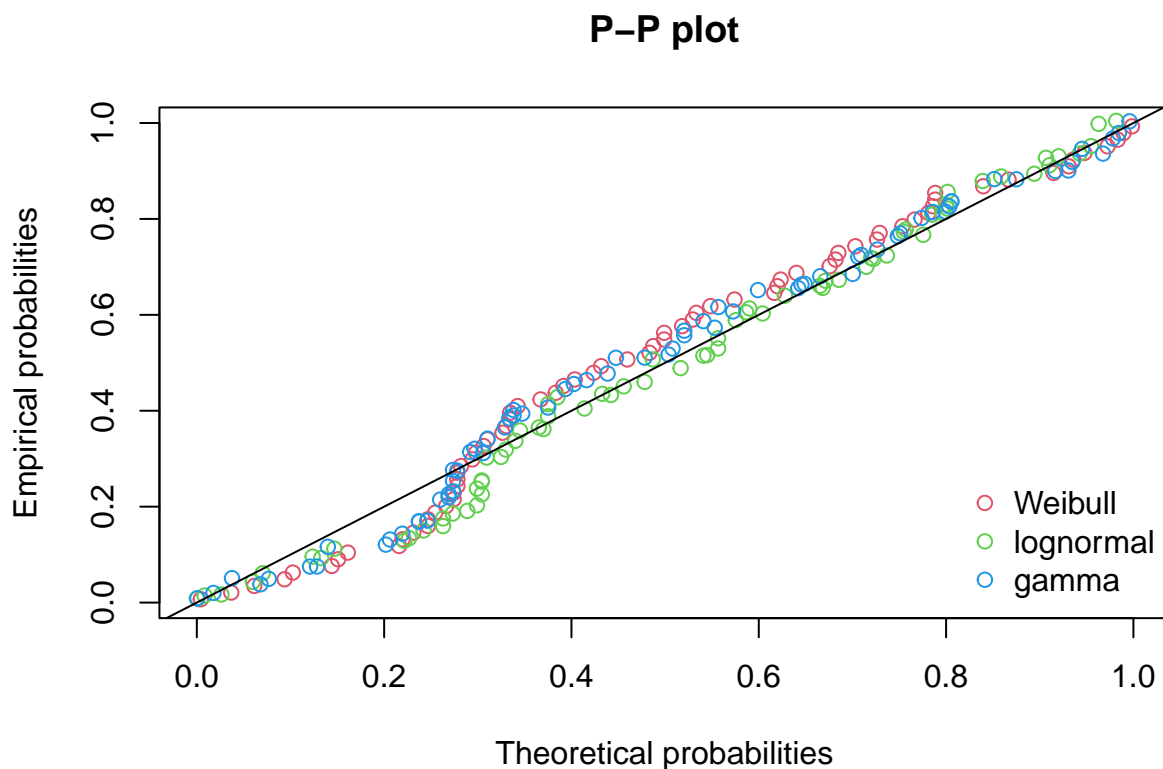```
cdfcomp(list(fw, fln, fg), legendtext = plot.legend)
```

# Empirical and theoretical CDFs



```
ppcomp(list(fw, fln, fg), legendtext = plot.legend)
```

P–P plot


```
fw$estimate
```

```
##      shape      scale
##   1.825211 199.586910
```

TO DO : interpret estimates

```
fg$estimate
```

```
##      shape       rate
## 3.08241826 0.01743161
```

TO DO : interpret estimates

```
fln$estimate
```

```
##   meanlog      sdlog
## 5.0042920 0.6290239
```

TO DO : interpret estimates

log-Logistic   Define the distribution function & compagnie

```r
dloglogis <- function(x, alpha, beta) {
  res <- (beta/alpha) * (x/alpha)^(beta-1)*(1+(x/alpha)^beta)^-2
  return(res)
}

ploglogis <- function(q, alpha, beta) {
  res <- 1 / (1 + (q/alpha)^beta)
```

```
  return(res)
}

qloglogis <- function(p, alpha, beta) {
  res <- alpha * ((1/p) - 1)^(1/beta)
  return(res)
}
```

```
# MC regime (sans données censorées pour le dist)
fll <- fitdist(gpigs.mc, "loglogis", method = "mle", start=list(alpha=10, beta=5))
fll$estimate
```
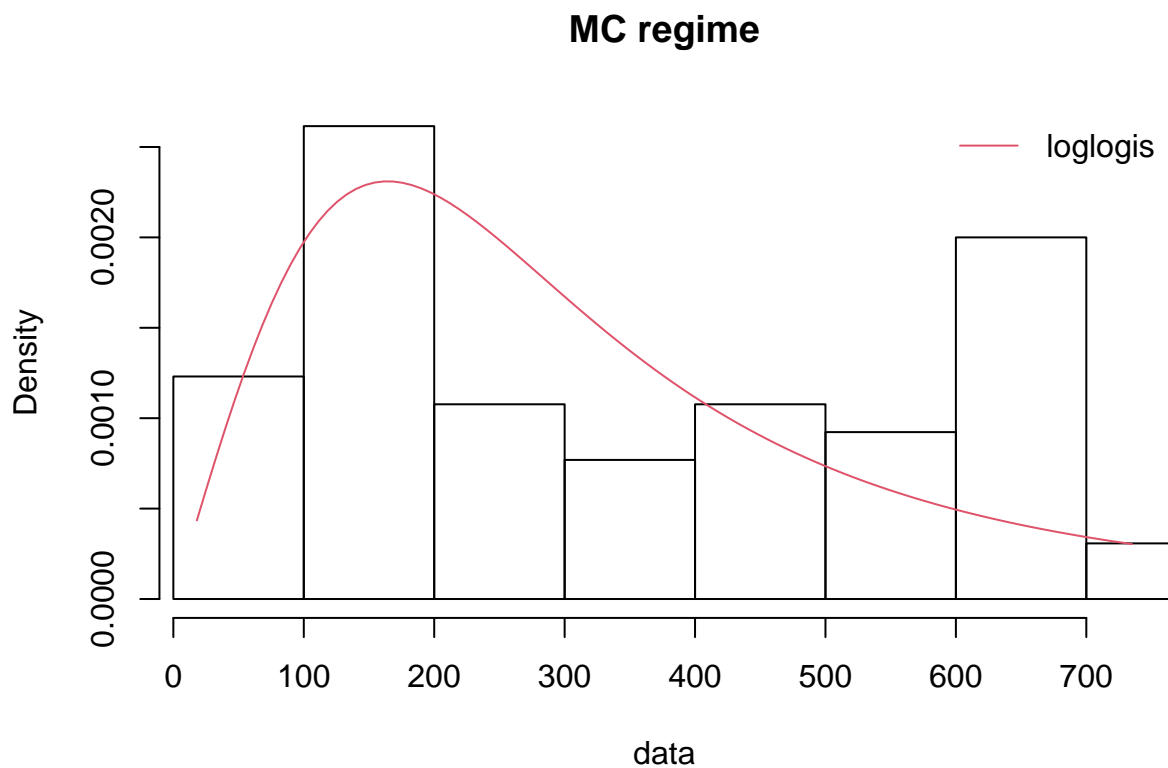
Control group

```
##      alpha        beta
## 282.086118    2.014417
```

```
denscomp(list(fll), main = "MC regime")
```



```
# M4.3 regime
fll <- fitdist(gpigs.noncensored$lifetime[gpigs.noncensored$regime == "M_4.3"], "loglogis", method = "m
fll$estimate
```

M4.3-Regime group

```
##       alpha        beta
## 152.389696    3.012415
```

TO DO: interpret estimates

```
denscomp(list(fll), main = "M4.3 regime")
```

M4.3 regime