data preprocessing

Yunfeng Yang

2020/10/10

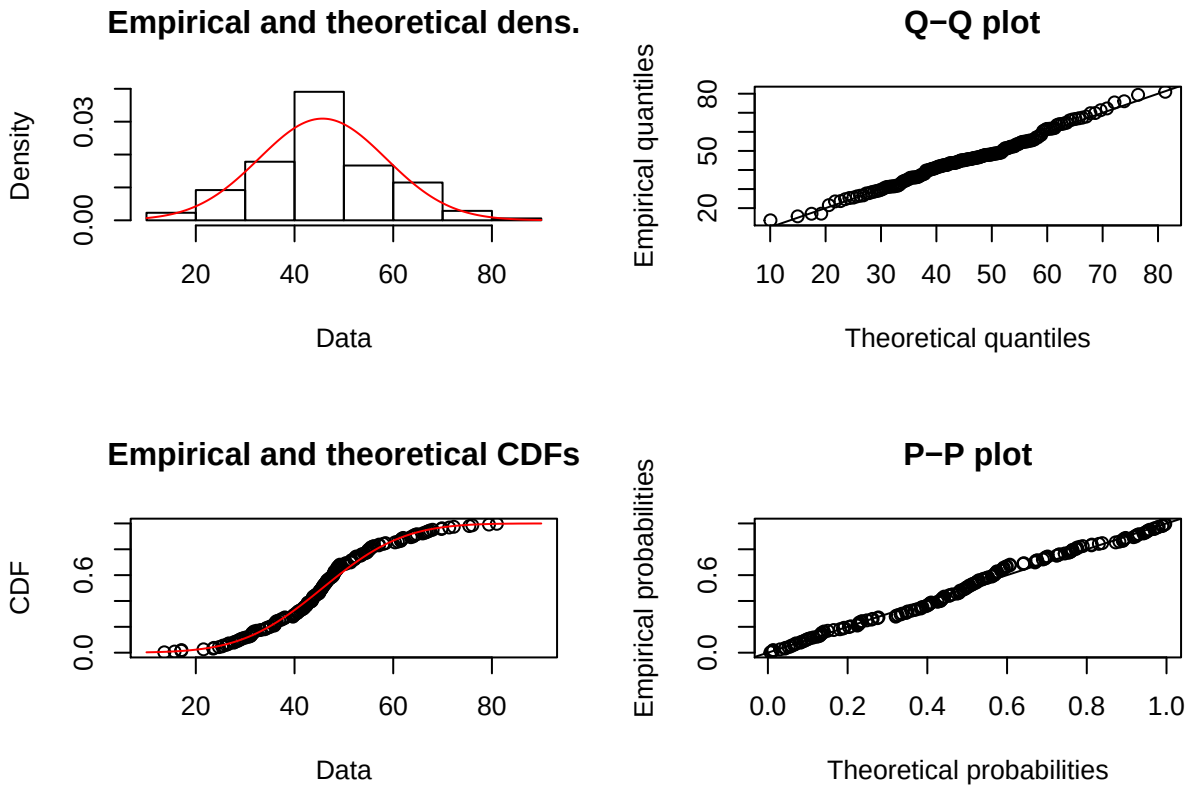
```
setwd("F:/study/Research/Bayesian-lumber-strength/Data")

require(rstan)
require(MASS)

###------------------data preprocessing--------------
library(readxl)
library(dplyr)
summary_all08122013 <- read_excel("summary_all08122013.xlsx")

summary_all08122013 <- select(summary_all08122013,"Group","Broken","MOR","UTS")

## T group
T100_data <-summary_all08122013[which(summary_all08122013$Group == "T100"),]
T60_data <-summary_all08122013[which(summary_all08122013$Group == "T60"),]
T40_data <-summary_all08122013[which(summary_all08122013$Group == "T40"),]
T20_data <-summary_all08122013[which(summary_all08122013$Group == "T20"),]
T100_data$UTS <- as.numeric(T100_data$UTS)
T100_data <- T100_data$UTS
## G group
R100_data <-summary_all08122013[which(summary_all08122013$Group == "R100"),]
R60_data <-summary_all08122013[which(summary_all08122013$Group == "R60"),]
R40_data <-summary_all08122013[which(summary_all08122013$Group == "R40"),]
R20_data <-summary_all08122013[which(summary_all08122013$Group == "R20"),]
R100_data <-as.numeric(R100_data$MOR)


##-----------------T substitute NA to 0 -----------

id <- T60_data$Broken == 1
T60_data$MOR[id] = 0
T60_data$UTS[id] = as.numeric(T60_data$UTS[id])
T60_data$UTS[!id] = 0
T60_data$UTS <- as.numeric(T60_data$UTS)
T60_data$MOR <- as.numeric(T60_data$MOR)


id <- T40_data$Broken == 1
T40_data$MOR[id] = 0
T40_data$UTS[id] = as.numeric(T40_data$UTS[id])
T40_data$UTS[!id] = 0
T40_data$UTS <- as.numeric(T40_data$UTS)
```

```r
T40_data$MOR <- as.numeric(T40_data$MOR)


id <- T20_data$Broken == 1
T20_data$MOR[id] = 0
T20_data$UTS[id] = as.numeric(T20_data$UTS[id])
T20_data$UTS[!id] = 0
T20_data$UTS <- as.numeric(T20_data$UTS)
T20_data$MOR <- as.numeric(T20_data$MOR)



##-----------------R substitute NA to 0 ------

id <- R60_data$Broken == 0
R60_data$MOR[id] = 0
R60_data$UTS[id] = as.numeric(R60_data$UTS[id])
R60_data$UTS[!id] = 0
R60_data$UTS <- as.numeric(R60_data$UTS)
R60_data$MOR <- as.numeric(R60_data$MOR)



id <- R40_data$Broken == 0
R40_data$MOR[id] = 0
R40_data$UTS[id] = as.numeric(R40_data$UTS[id])
R40_data$UTS[!id] = 0
R40_data$UTS <- as.numeric(R40_data$UTS)
R40_data$MOR <- as.numeric(R40_data$MOR)

id <- R20_data$Broken == 0
R20_data$MOR[id] = 0
R20_data$UTS[id] = as.numeric(R20_data$UTS[id])
R20_data$UTS[!id] = 0
R20_data$UTS <- as.numeric(R20_data$UTS)
R20_data$MOR <- as.numeric(R20_data$MOR)



##------proof loading--------------
R_pf = c(4.956690733, 6.110714122, 7.092435407)
T_pf = c(2.962390379, 3.986497991, 4.916102264)




##---------------Convert psi to Mpa------------

# 1 thousand psi = 6.895 MPa

c = 6.895
```

```
R20_data$MOR <- c*R20_data$MOR
R20_data$UTS <- c*R20_data$UTS


R40_data$MOR <- c*R40_data$MOR
R40_data$UTS <- c*R40_data$UTS

R60_data$MOR <- c*R60_data$MOR
R60_data$UTS <- c*R60_data$UTS

R100_data <- c*R100_data

T20_data$MOR <- c*T20_data$MOR
T20_data$UTS <- c*T20_data$UTS

T40_data$MOR <- c*T40_data$MOR
T40_data$UTS <- c*T40_data$UTS

T40_data$MOR <- c*T40_data$MOR
T40_data$UTS <- c*T40_data$UTS


T100_data <- c*T100_data
```

fitting for R100_data

```
##-------check normal fitting---------

library(fitdistrplus)

# check R100_data
FIT <- fitdist(R100_data, "norm")    ## note: it is "norm" not "normal"
plot(FIT)    ## use method `plot.fitdist`
```




```
FIT$estimate # mean = 45.679 sd = 12.900
```

```
##     mean       sd
## 45.67857 12.89992
```

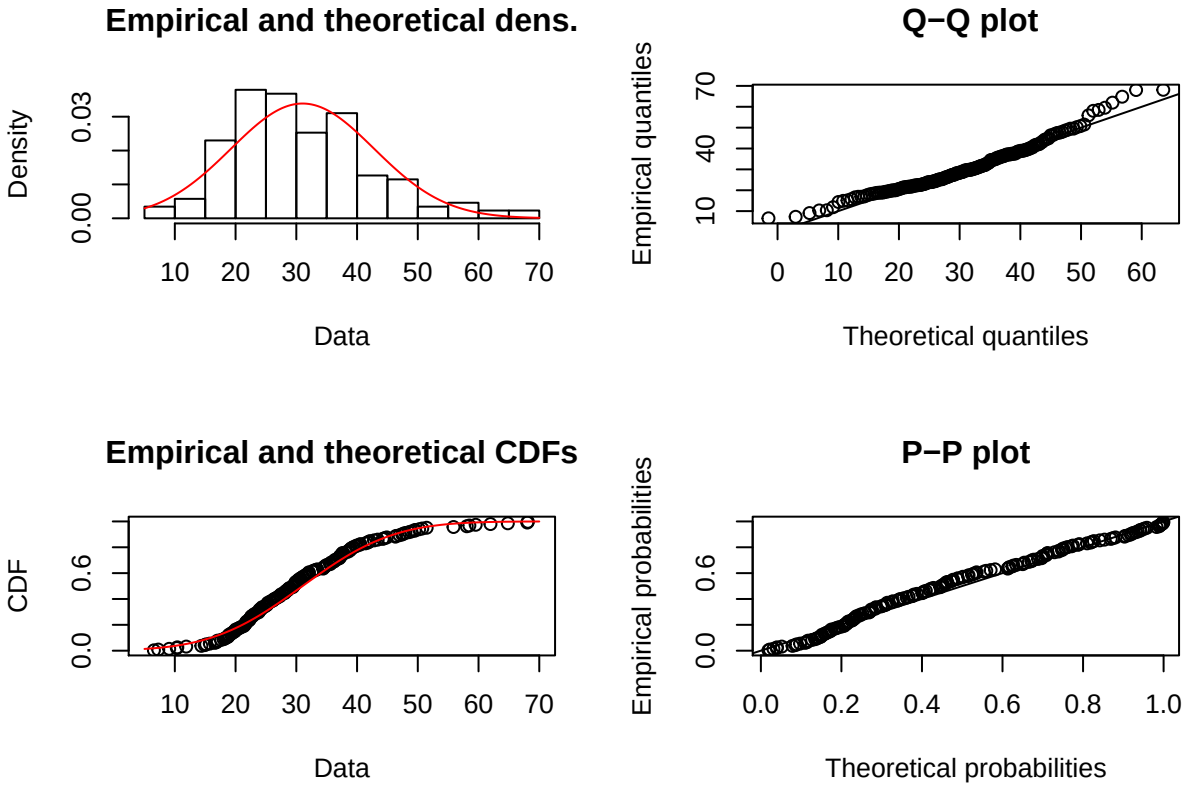
```
FIT$bic # bic = 1394.022
```

```
## [1] 1394.022
```

```
# good normal fitting
```

check T100_data

```
# check T100_data

FIT <- fitdist(T100_data, "norm")    ## note: it is "norm" not "normal"
plot(FIT)    ## use method `plot.fitdist`
```




```
FIT$estimate # mean = 31.029 sd = 11.775
```

```
##     mean       sd
## 31.02680 11.77522
```

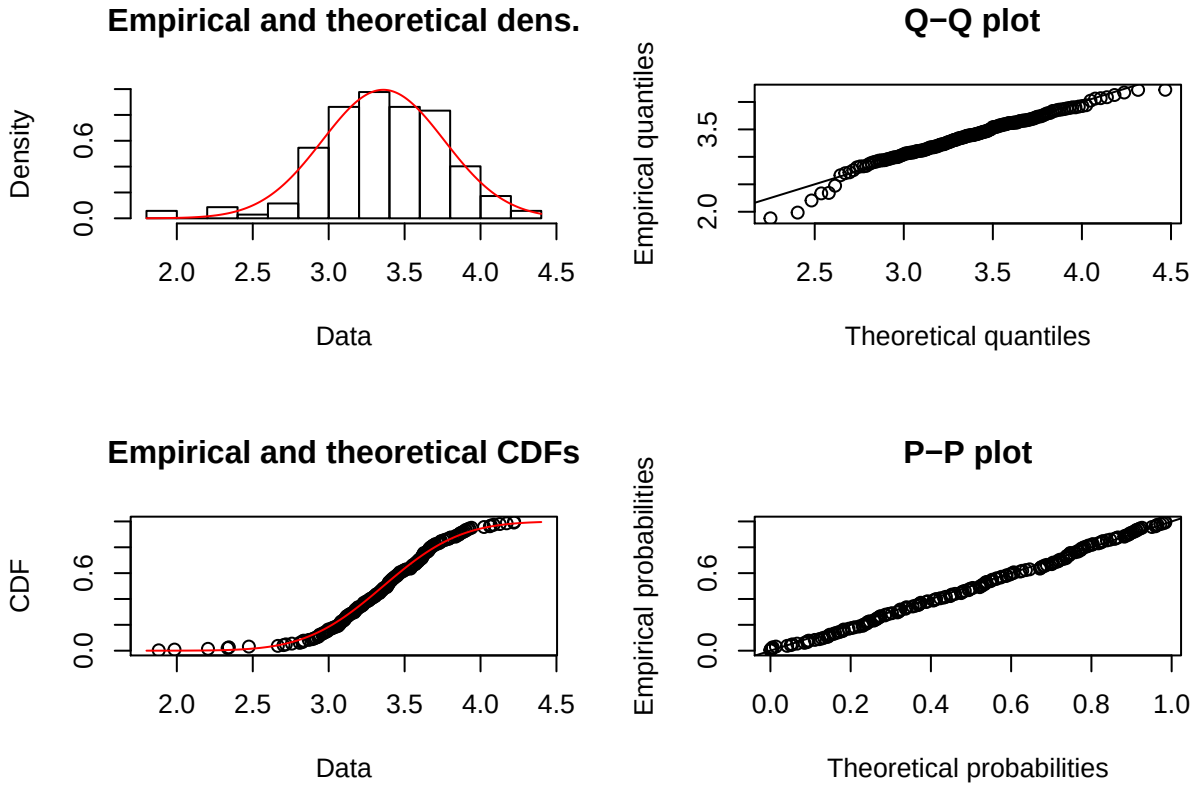
```
FIT$bic # bic = 1362.276
```

```
## [1] 1362.276
```

```
# left-skewed
```

check log(T100_data)

```
# check log(T100_data)

FIT <- fitdist(log(T100_data), "norm")    ## note: it is "norm" not "normal"
plot(FIT)    ## use method `plot.fitdist`
```




```
FIT$estimate # mean = 3.360 sd = 0.401
```

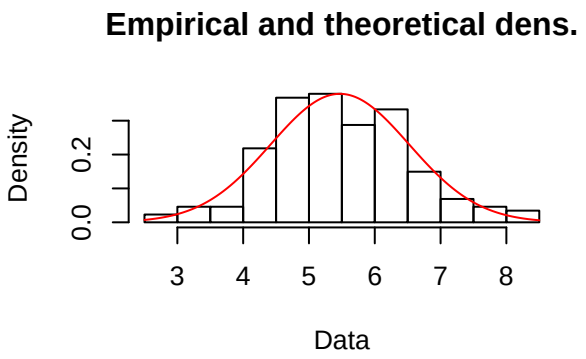
```
##      mean        sd
## 3.3597595 0.4012101
```

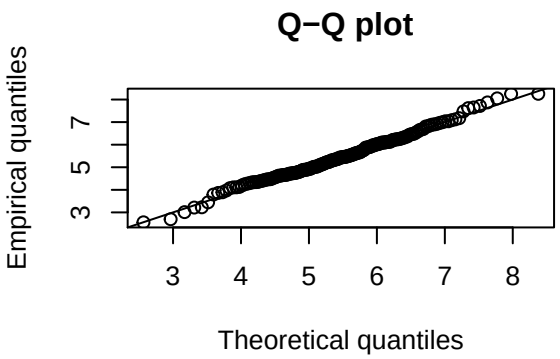
```
FIT$bic # bic = 186.2908
```

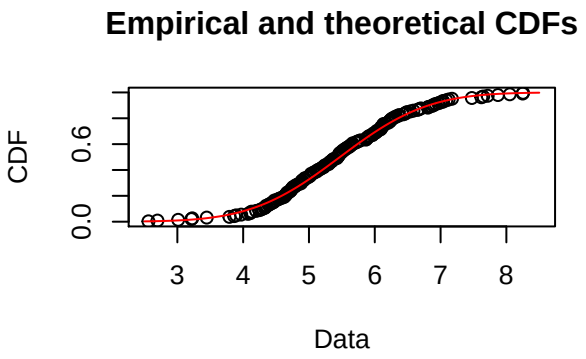
```
## [1] 186.2908
```

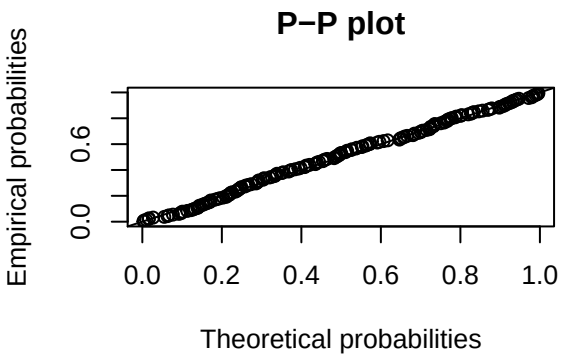
check sqrt(T100_data)

```
# check sqrt(T100_data)
FIT <- fitdist(sqrt(T100_data), "norm")    ## note: it is "norm" not "normal"
plot(FIT)    ## use method `plot.fitdist`
```










```
FIT$estimate # mean = 5.470 sd = 1.052
```

```
##     mean       sd
## 5.469924 1.052010
```

```
FIT$bic # bic = 521.7532
```

```
## [1] 521.7532
```