STAT GU4261/GR5261 - Statistical Methods in Finance - Homework #6 Solutions

3/7/23

Question 1

Problem 1 pp. 127

Cube-root transformation is demonstrated in these solutions.

```r
library(Ecdat)
data(CPSch3)

male.earnings <- CPSch3[CPSch3[, 3] == "male", 2]
sqrt.male.earnings <- sqrt(male.earnings)
log.male.earnings <- log(male.earnings)
cbrt.male.earnings <- male.earnings^(1/3)
```

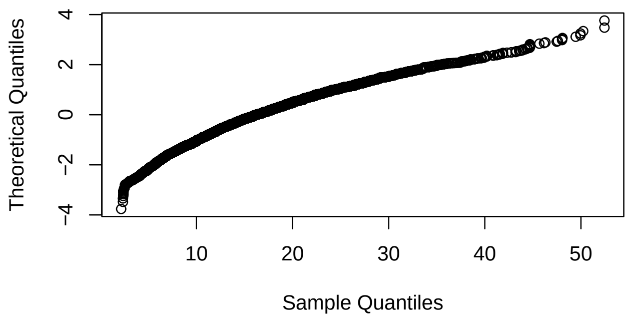
Code for qqplots:

```r
qqnorm(male.earnings, datax = TRUE, main = "")
qqnorm(sqrt.male.earnings, datax = TRUE, main = "")
qqnorm(log.male.earnings, datax = TRUE, main = "")
qqnorm(cbrt.male.earnings, datax = TRUE, main = "")
```

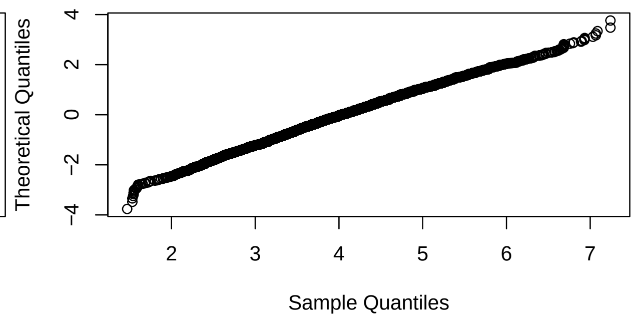
Code for boxplots:

```r
boxplot(male.earnings)
boxplot(sqrt.male.earnings)
boxplot(log.male.earnings)
boxplot(cbrt.male.earnings)
```

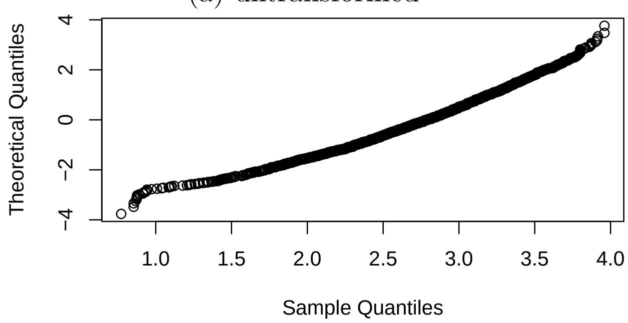
Code for density plots:



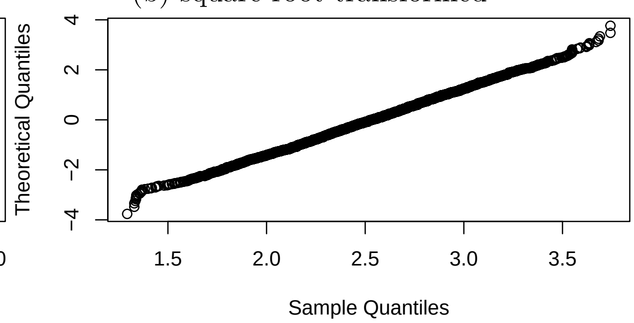
(a) untransformed



(b) square-root transformed

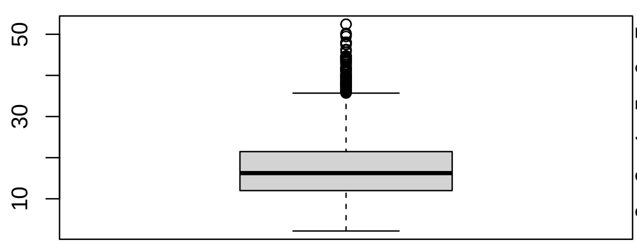


(c) log-transformed

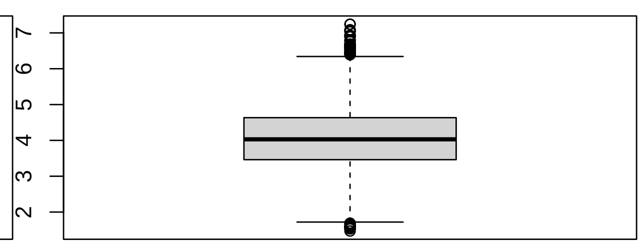


(d) cube-root transformed

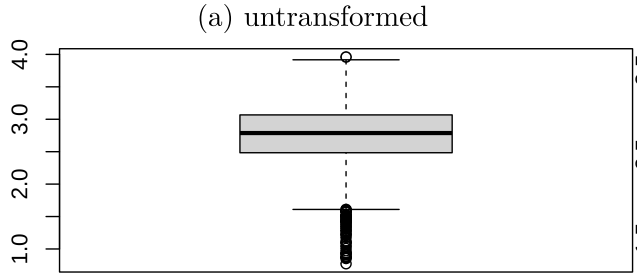
Figure 1: Normal qqplots for each transformation.



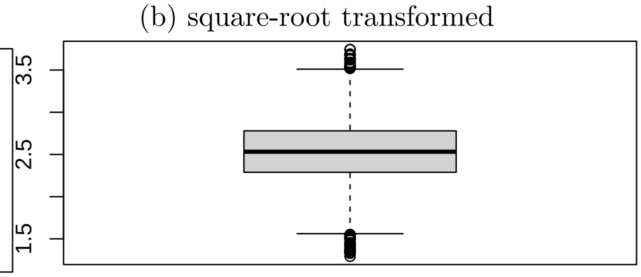
(a) untransformed



(b) square-root transformed



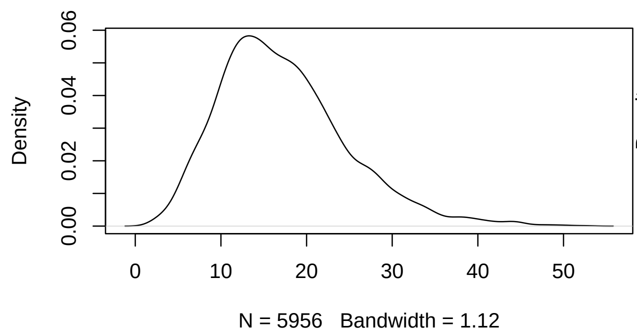
(c) log-transformed



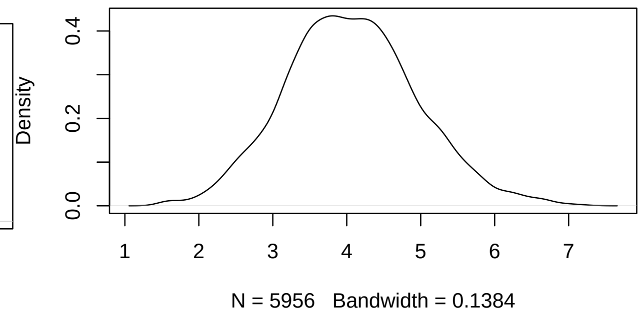
(d) cube-root transformed

Figure 2: Boxplots for each transformation.

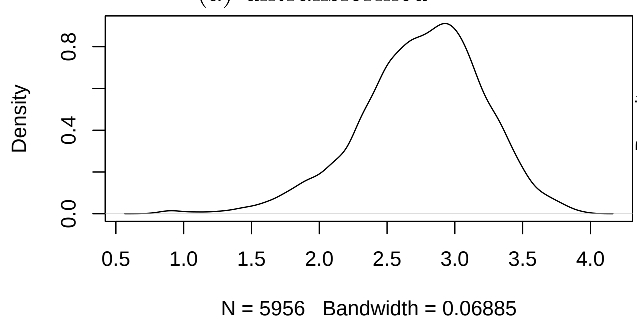
```r
plot(density(male.earnings), main = "")
plot(density(sqrt.male.earnings), main = "")
plot(density(log.male.earnings), main = "")
plot(density(cbrt.male.earnings), main = "")
```



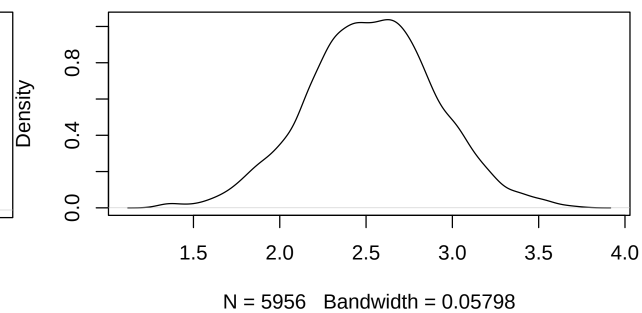
(a) untransformed



(b) square-root transformed



(c) log-transformed



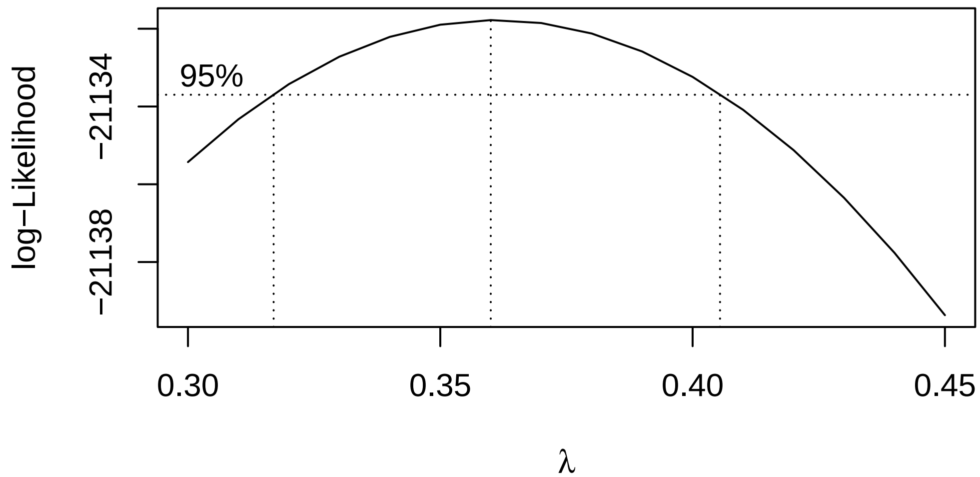
(d) cube-root transformed

Figure 3: Density plots for each transformation.

From the above plots we see that both square-root and cube-root transformations are the most symmetric. However, the square-root transformation seems skewed to the left and in this case I would prefer the cube-root transformation for symmetrisation.

Problem 2 pp. 128

```r
library("MASS")
bc <- boxcox(male.earnings ~ 1,
    lambda = seq(0.3, 0.45, by = 1 / 100),
    interp = FALSE
)
```




```r
ind <- (bc$y == max(bc$y))
ind2 <- (bc$y > max(bc$y) - qchisq(0.95, df = 1) / 2)
ind3 <- (bc$y > max(bc$y) - qchisq(0.99, df = 1) / 2)
```

(a)
`ind` is a logical vector with `TRUE` corresponding to the index for the maximum log-likelihood. `ind2` is a logical vector with `TRUE` corresponding to the indices where the log-likelihood is in a theoretical 95% confidence interval.

(b)
`interp` being `TRUE` corresponds to whether spline interpolation is used to compute a smoothened version of the log-likelihood.

(c)
The MLE is 0.36.

(d)
A 95% confidence interval for λ is $[0.32, 0.4]$.

(e)
A 99% confidence interval for λ is $[0.31, 0.41]$.

Problem 3 pp. 128

```r
library("fGarch")
fit <- sstdFit(male.earnings, hessian = TRUE)
```

The estimated degrees-of-freedom parameter is 21.599837 and for ξ is 1.6516521.

Question 2

```r
capm <- readxl::read_excel("CAPM-DATA-1.xlsx")
capm_returns <- capm[, c("MSOFT", "GE", "GM", "IBM")]

library(ggplot2)
fit_densities <- function(x) {
    #function to fit three densities to `x`. Calculates aic/bic values and
    #returns list of qq plots for each fitted density.
    n <- length(x)
    #fit t
    start <- c(mean(x), sd(x), 5)
    loglik_t <- function(beta) {
        sum(-dt((x - beta[1]) / beta[2], beta[3], log = TRUE) + log(beta[2]))
    }
    fit_t <- optim(
        start, loglik_t, hessian = T, method = "L-BFGS-B",
        lower = c(-0.1, 0.01, 2.1)
    )
    results_t <- list(
        par = fit_t$par,
        sd = sqrt(diag(solve(fit_t$hessian)))
    )
    df_t <- data.frame(
        dist = "t",
        AIC = 2 * fit_t$value + 2 * 3,
        BIC = 2 * fit_t$value + log(n) * 3
    )

    #fit norm
    start <- c(mean(x), sd(x))
    loglik_norm <- function(beta) {
        sum(-dnorm(x, beta[1], beta[2], log = TRUE))
    }
    fit_norm <- optim(
        start, loglik_norm, hessian = T, method = "L-BFGS-B",
        lower = c(-1, 0.001)
    )
    results_norm <- list(
        par = fit_norm$par,
        sd = sqrt(diag(solve(fit_norm$hessian)))
    )
    df_norm <- data.frame(
        dist = "norm",
```

```
        AIC = 2 * fit_norm$value + 2 * 2,
        BIC = 2 * fit_norm$value + log(n) * 2
    )

    #fit ged
    start <- c(mean(x), sd(x), 1)
    loglik_ged <- function(beta) {
        sum(-dged(x, beta[1], beta[2], beta[3], log = TRUE))
    }
    fit_ged <- optim(
        start, loglik_ged, hessian = T, method = "L-BFGS-B",
        lower = c(-0.1, 0.01, 1)
    )
    results_ged <- list(
        par = fit_ged$par,
        sd = sqrt(diag(solve(fit_ged$hessian)))
    )
    df_ged <- data.frame(
        dist = "ged",
        AIC = 2 * fit_ged$value + 2 * 3,
        BIC = 2 * fit_ged$value + log(n) * 3
    )

    myqt <- function(p) {
        qt(p = p, df = results_t$par[3]) * results_t$par[2] + results_t$par[1]
    }

    myqnorm <- function(p) {
        qnorm(p, results_norm$par[1], results_norm$par[2])
    }

    myqged <- function(p) {
        qged(p, results_ged$par[1], results_ged$par[2], results_ged$par[3])
    }

    qq_plot_list <- lapply(
        list("t" = myqt, "norm" = myqnorm, "ged" = myqged), function(distr) {
            ggplot(data.frame(x = x), aes(sample = x)) +
                stat_qq(distribution = distr) +
                stat_qq_line(distribution = distr)
        }
    )

    list(
```

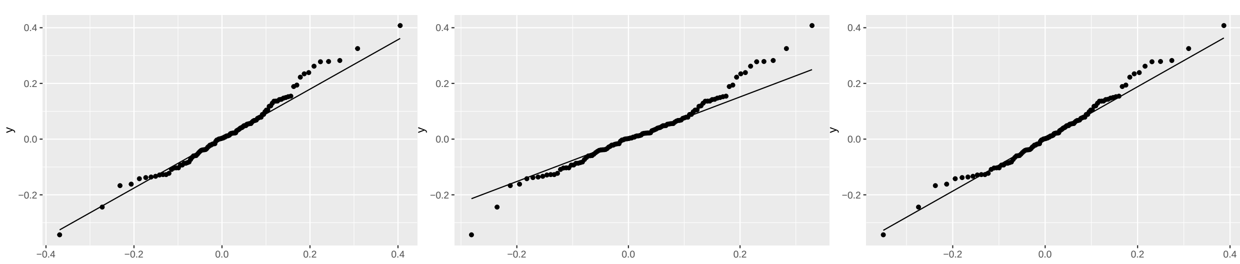
```r
            "results" = list(
                "t" = results_t, "norm" = results_norm, "ged" = results_ged
            ),
            "aicbic" = rbind(df_t, df_norm, df_ged),
            "qq_plot_list" = qq_plot_list
        )
}

fit_list <- lapply(capm_returns, fit_densities)
```

The values for AIC/BIC for `MSOFT` are as follows:

```r
fit_list$MSOFT$aicbic
##   dist       AIC       BIC
## 1    t -176.6292 -168.2667
## 2 norm -173.1735 -167.5985
## 3  ged -178.6375 -170.2750
```

We would choose the `ged` fit having parameters 0.0182665, 0.1162802, 1.1919819 with standard errors 0.0099705, 0.0103273, 0.2208107. The qqplots for all three fitted distributions are in Figure 4.

```r
fit_list$MSOFT$qq_plot_list
```



(a) t (b) norm (c) ged

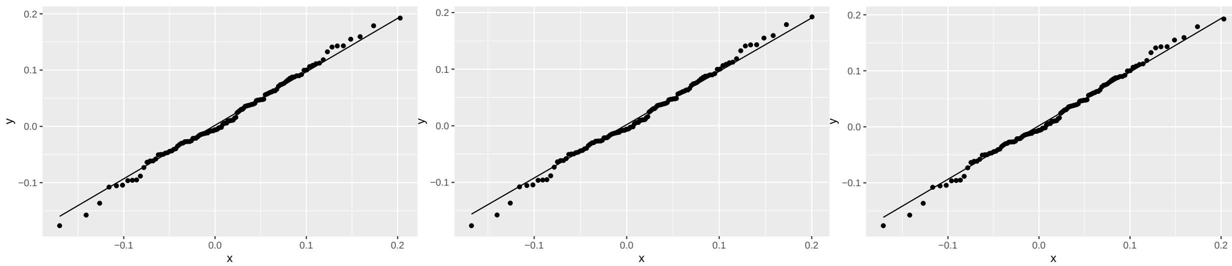
Figure 4: qq plots for MSOFT returns.

The values for AIC/BIC for `GE` are as follows:

```r
fit_list$GE$aicbic
##   dist       AIC       BIC
## 1    t -292.5166 -284.1541
## 2 norm -294.4946 -288.9196
## 3  ged -292.5560 -284.1935
```

We would choose the `norm` fit having parameters 0.0162877, 0.0697739 with standard errors 0.0063695, 0.0045009. The qqplots for all three fitted distributions are in Figure 5.

```
fit_list$GE$qq_plot_list
```



(a) t (b) norm (c) ged

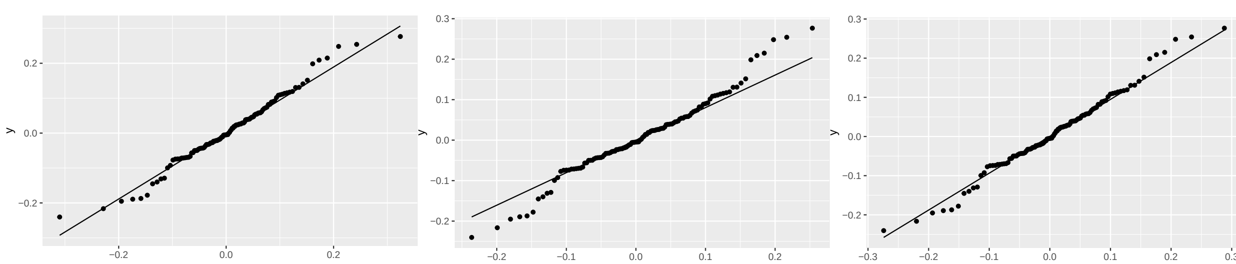
Figure 5: qq plots for GE returns.

The values for AIC/BIC for `GM` are as follows:

```
fit_list$GM$aicbic
##   dist       AIC       BIC
## 1    t -228.9665 -220.6041
## 2 norm -226.0596 -220.4846
## 3  ged -229.4180 -221.0555
```

We would choose the `ged` fit having parameters 0.0068829, 0.0926148, 1.3391906 with standard errors 0.0084861, 0.0075181, 0.2263592. The qqplots for all three fitted distributions are in Figure 6.

```
fit_list$GM$qq_plot_list
```



(a) t (b) norm (c) ged

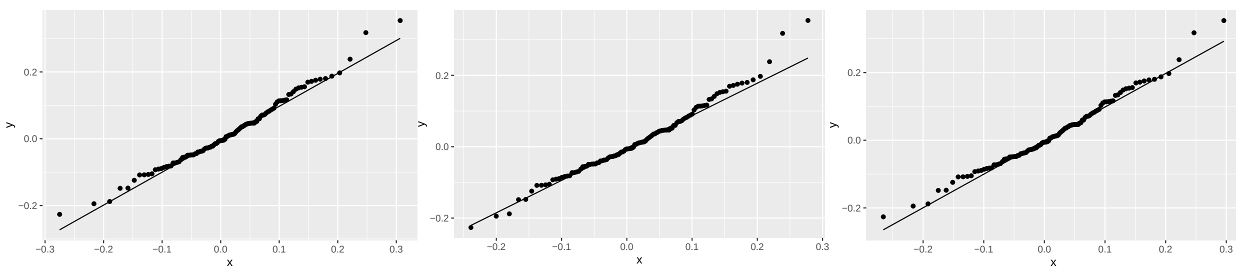
Figure 6: qq plots for GM returns.

The values for AIC/BIC for `IBM` are as follows:

```
fit_list$IBM$aicbic
##   dist       AIC       BIC
## 1    t -213.4973 -205.1348
## 2 norm -213.1882 -207.6132
## 3  ged -212.9874 -204.6250
```

We would choose the `norm` fit having parameters 0.023866, 0.1156602 with standard errors 0.0105583, 0.007464. The qqplots for all three fitted distributions are in Figure 7.

```
fit_list$IBM$qq_plot_list
```



(a) t (b) norm (c) ged

Figure 7: qq plots for IBM returns.

Question 3

Problem 1 pp. 131

```
cov(CRSPday[, 4:6])
##                 ge          ibm        mobil
## ge    1.882164e-04 8.007660e-05 5.270394e-05
## ibm   8.007660e-05 3.061309e-04 3.588748e-05
## mobil 5.270394e-05 3.588748e-05 1.670265e-04
cor(CRSPday[, 4:6])
##              ge       ibm     mobil
## ge    1.0000000 0.3335979 0.2972499
## ibm   0.3335979 1.0000000 0.1587072
## mobil 0.2972499 0.1587072 1.0000000
apply(CRSPday[, 4:6], 2, mean)
##           ge          ibm        mobil
## 0.0010713801 0.0007000767 0.0007788801
```

(a)

The mean of the Mobil returns is 7.79×10^{-4}.

(b)

The variance of the GE returns is 1.88×10^{-4}

(c)

The covariance between the GE and Mobil returns is 5.27×10^{-5}.

(d)

The correlation between the GE and Mobil returns is 0.297.

Question 4

Problem 3 pp. 132

By symmetry around zero,

$$\int_{-\infty}^{\infty} f^*(y|xi)dy = \int_{-\infty}^{0} f(y\xi)dy + \int_{0}^{\infty} f(y/\xi)dy = \int_{0}^{\infty} f(y\xi)dy + \int_{0}^{\infty} f(y/\xi)dy.$$

Consider the second term. By a change of variables $u = y/\xi$,

$$\int_{0}^{\infty} f(y/\xi)dy = \xi \int_{0}^{\infty} f(u)du = \xi/2.$$

Last equality follows from f being symmetric around zero: $\int_{-\infty}^{0} f(u)du = \int_{0}^{\infty} f(u)du = 1/2$. Similarly, for the first term $\int_{0}^{\infty} f(y\xi)dy = \xi^{-1}/2$. Thus, $\int f^*(y)dy = (\xi^{-1} + \xi)/2$.