STAT 636, Fall 2015 - Assignment 3
SOLUTIONS

1. Consider a bivariate normal population with $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_{11} = 2$, $\sigma_{22} = 1$, and $\rho_{12} = 0.5$.

   (a) Write out the bivariate normal density.
   THE DENSITY IS

$$
\begin{aligned}
f(x_1, x_2) &= \frac{1}{2\pi\sqrt{(2)(1)(1-0.5^2)}} \\
&\quad \times \exp\left\{ -\frac{1}{2(1-0.5^2)} \left[ \left(\frac{x_1 - 0}{\sqrt{2}}\right)^2 + \left(\frac{x_2 - 2}{\sqrt{1}}\right)^2 \right.\right. \\
&\qquad\qquad \left.\left. - 2(0.5)\left(\frac{x_1 - 0}{\sqrt{2}}\right)\left(\frac{x_2 - 2}{\sqrt{1}}\right) \right] \right\} \\
&= \frac{1}{\sqrt{6}\pi} \exp\left\{ -\frac{2}{3}\left[ \frac{x_1^2}{2} + (x_2 - 2)^2 - \frac{x_1(x_2 - 2)}{\sqrt{2}} \right] \right\}
\end{aligned}
$$

   (b) Write out the squared generalized distance expression $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ as a function of $x_1$ and $x_2$.
   AS SEEN ABOVE,

$$
(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \frac{x_1^2}{2} + (x_2 - 2)^2 - \frac{x_1(x_2 - 2)}{\sqrt{2}}
$$

   (c) Draw / plot the constant-density contour that contains 50% of the probability.
   SEE FIGURE 1.

2. Let $\mathbf{X}$ be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [-3, 1, 4]$ and

$$
\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}
$$

   Which of the following random variables are independent? Explain.

   (a) $X_1$ and $X_2$.
   WE HAVE $\mathrm{COV}(X_1, X_2) = \sigma_{12} = -2 \neq 0$, SO $X_1$ AND $X_2$ ARE NOT INDEPENDENT.

   (b) $X_2$ and $X_3$.
   WE HAVE $\mathrm{COV}(X_2, X_3) = \sigma_{23} = 0$, SO $X_2$ AND $X_3$ ARE INDEPENDENT.

   (c) $(X_1, X_2)$ and $X_3$.
   WE CAN WRITE $(X_1, X_2)$ AS $\mathbf{AX}$, WHERE

$$
\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}
$$

SIMILARLY, $X_3 = \mathbf{b}'\mathbf{X}$, WHERE $\mathbf{b}' = [0,0,1]$. SO,

$$\text{Cov}((X_1, X_2), X_3) = \text{Cov}(\mathbf{AX}, \mathbf{b}'\mathbf{X}) = \mathbf{A\Sigma b} = \mathbf{0}$$

AND $(X_1, X_2)$ AND $X_3$ ARE INDEPENDENT.

(d) $\frac{X_1+X_2}{2}$ and $X_3$.
WE CAN WRITE $\frac{X_1+X_2}{2}$ AS $\mathbf{a}'\mathbf{X}$, WHERE $\mathbf{a}' = [1/2, 1/2, 0]$, AND $X_3$ AS $\mathbf{b}'\mathbf{X}$, WHERE $\mathbf{b}' = [0,0,1]$. SO,

$$\text{Cov}\left(\frac{X_1 + X_2}{2}, X_3\right) = \text{Cov}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \mathbf{a}'\mathbf{\Sigma b} = 0$$

AND $\frac{X_1+X_2}{2}$ AND $X_3$ ARE INDEPENDENT.

(e) $X_2$ and $X_2 - \frac{5}{2}X_1 - X_3$.
WE CAN WRITE $X_2$ AS $\mathbf{a}'\mathbf{X}$, WHERE $\mathbf{a}' = [0,1,0]$, AND $X_2 - \frac{5}{2}X_1 - X_3$ AS $\mathbf{b}'\mathbf{X}$, WHERE $\mathbf{b}' = [-5/2, 1, -1]$. SO,

$$\text{Cov}\left(X_2, X_2 - \frac{5}{2}X_1 - X_3\right) = \text{Cov}(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{X}) = \mathbf{a}'\mathbf{\Sigma b} = 10 \neq 0$$

AND $X_2$ AND $X_2 - \frac{5}{2}X_1 - X_3$ ARE NOT INDEPENDENT.

3. Let $\mathbf{X}$ be distributed as $N_3(\boldsymbol{\mu}, \mathbf{\Sigma})$, where $\boldsymbol{\mu}' = [1, -1, 2]$ and

$$\mathbf{\Sigma} = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

Specify each of the following.

(a) The conditional distribution of $X_1$, given that $X_3 = x_3$.
THIS WILL BE UNIVARIATE NORMAL WITH

$$\text{MEAN} = 1 + (-1)(2)^{-1}(x_3 - 2) = 2 - \frac{x_3}{2}$$
$$\text{VARIANCE} = 4 - (-1)(2)^{-1}(-1) = \frac{7}{2}$$

(b) The conditional distribution of $X_1$, given that $X_2 = x_2$ and $X_3 = x_3$.
THIS WILL BE UNIVARIATE NORMAL WITH

$$\text{MEAN} = 1 + \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \left( \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right) = 2 - \frac{x_3}{2}$$
$$\text{VARIANCE} = 4 - \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \frac{7}{2}$$

4. Find the maximum likelihood estimates of the $2 \times 1$ mean vector $\boldsymbol{\mu}$ and the $2 \times 2$ covariance matrix $\boldsymbol{\Sigma}$ based on the random sample

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

THE MLES ARE

$$\bar{\mathbf{x}} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} \quad \text{AND} \quad \mathbf{S}_n = \begin{bmatrix} 0.50 & 0.25 \\ 0.25 & 1.50 \end{bmatrix}$$

5. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{20}$ be a random sample of size $n = 20$ from an $N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Specify each of the following.

   (a) The distribution of $(\mathbf{X}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu})$.
   THIS IS A $\chi^2_6$ RANDOM VECTOR.

   (b) The distributions of $\bar{\mathbf{X}}$ and $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu})$.
   WE HAVE $\bar{\mathbf{X}} \sim N_6\left(\boldsymbol{\mu}, \frac{1}{6}\boldsymbol{\Sigma}\right)$ AND $\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}) \sim N_6(\mathbf{0}, \boldsymbol{\Sigma})$.

   (c) The distribution of $(n-1)\mathbf{S}$.
   THIS IS A $W_5(\boldsymbol{\Sigma})$ RANDOM MATRIX.

6. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{60}$ be a random sample of size $n = 60$ from an $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Specify each of the following.

   (a) The distribution of $n(\bar{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$.
   THIS IS A $\chi^2_4$ RANDOM VARIABLE.

   (b) The approximate distribution of $n(\bar{\mathbf{X}} - \boldsymbol{\mu})\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$.
   THIS IS ALSO (APPROXIMATELY, SINCE $n$ IS LARGE RELATIVE TO $p$) A $\chi^2_4$ RANDOM VARIABLE.

7. Given the air pollution data in Table 1.5 of the textbook ("T1-5.DAT"), examine the pairs $X_5 = \mathrm{NO}_2$ and $X_6 = \mathrm{O}_3$ for bivariate normality.

   (a) Calculate statistical distances $(\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$, $j = 1, 2, \ldots, 42$, where $\mathbf{x}_j' = [x_{j5}, x_{j6}]$.
   SEE CODE BELOW.

   (b) Determine the proportion of observations $\mathbf{x}_j' = [x_{j5}, x_{j6}]$, $j = 1, 2, \ldots, 42$, falling within the approximate 50% probability contour of a bivariate normal distribution.
   ABOUT 60% OF THE $d_j^2$ ARE WITHIN THE INNER 50% PROBABILITY ELLIPSE. THIS IS A FAIRLY LARGE DISCREPANCY FROM THE EXPECTED VALUE OF 50%.

   (c) Construct a chi-square plot of the ordered distances in part (a).
   SEE FIGURE 2 BELOW. THERE IS SUBSTANTIAL DEVIATION FROM THE LINE OF EQUALITY. THIS, COUPLED WITH THE RESULT FROM THE PREVIOUS PART, SUGGESTS THAT THE DATA ARE NOT DISTRIBUTED AS BIVARIATE NORMAL.

8. Consider the used-car data from Exercise 4.26 in the textbook.

   (a) Determine the power transformation $\hat{\lambda}_1$ that makes the $x_1$ values approximately normal. Construct a Q-Q plot for the transformed data.

   See code below. The optimal transformation corresponds to $\hat{\lambda}_1 = 0.38$. The resulting Q-Q plot is in Figure 3. The Q-Q plot looks nice, but the histogram doesn't look particularly normal.

   (b) Determine the power transformation $\hat{\lambda}_2$ that makes the $x_2$ values approximately normal. Construct a Q-Q plot for the transformed data.

   See Figure 4. The Q-Q plot suggests that the transformed data are now light-tailed.

   (c) Determine the power transformations $\hat{\boldsymbol{\lambda}}' = [\hat{\lambda}_1, \hat{\lambda}_2]$ that make the $[x_1, x_2]$ values jointly normal. Compare the results with those from above.

   See Figures 5, 6, and 7. The optimal transformations correspond now to $\hat{\lambda}_1 = 1.27$ and $\hat{\lambda}_2 = 0.02$, substantially different from those obtained by individual Box-Cox applications. The contour plot in Figure 5 suggests that the multivariate objective function is rather flat over a large region which includes both sets of transformations. The histograms and Q-Q plots of the transformed variables in Figures 6 and 7 suggest that the age variable is now right-skewed, and the price variable is now left-skewed. Bottom line: It doesn't appear that Box-Cox performs well in this example. Remember that we have no guaranty that <u>any</u> transformation will be able to achieve normality. That's why it's always a good idea to check Q-Q plots after transformations.

```
####
#### (1)
####

library(mvtnorm)
library(plotrix)
library(MASS)

mu <- c(0, 2)
sg <- c(2, 1)
rho <- 0.5
Sigma <- matrix(c(sg[1], rho * prod(sqrt(sg)), rho * prod(sqrt(sg)), sg[2]), nrow = 2)

## Probability density function.
f <- function(x) {
  (1 / (sqrt(6) * pi)) * exp(-(2 / 3) * (x[1] ^ 2 / 2 + (x[2] - 2) ^ 2 -
    x[1] * (x[2] - 2) / sqrt(2)))
}

f(c(0, 0)); dmvnorm(c(0, 0), mu, Sigma)
f(c(0, 2)); dmvnorm(c(0, 2), mu, Sigma)
f(c(2, 0)); dmvnorm(c(2, 0), mu, Sigma)

## Ellipse containing the inner 50% probability.
ee <- eigen(Sigma)
lambda <- ee$values
ee <- ee$vectors
theta <- -acos(t(ee[, 1]) %*% c(1, 0)) * (360 / (2 * pi))

pdf("figures/(1)_ellipse.pdf")
plot(c(-2, 2), c(0, 4), xlab = expression(x[1]), ylab = expression(x[2]), asp = 1,
  type = "n")
draw.ellipse(mu[1], mu[2], sqrt(qchisq(0.5, 2) * lambda[1]),
  sqrt(qchisq(0.5, 2) * lambda[2]), angle = theta, deg = TRUE, lwd = 2)
dev.off()

####
#### (4)
####

X <- matrix(c(3, 4, 5, 4, 6, 4, 7, 7), ncol = 2)

colMeans(X)
var(X) * 3 / 4
```

```
####
#### (7)
####

##
## Input data.
##

X <- read.delim("T1-5.DAT", header = FALSE, sep = "")
colnames(X) <- c("Wind", "Solar", "CO", "NO", "NO_2", "O_3", "HC")
X <- X[, c("NO_2", "O_3")]
n <- nrow(X)
p <- ncol(X)

## Statistical distances, d_j^2.
x_bar <- colMeans(X)
S <- ((n - 1) / n) * var(X)
D <- X - matrix(rep(x_bar, each = n), ncol = 2)
d2_j <- apply(D, 1, function(x) { t(as.numeric(x)) %*% solve(S) %*% as.numeric(x) })

mean(d2_j <= qchisq(0.5, 2))

## Chi-square plot.
d2_j_o <- sort(d2_j)
q_j <- qchisq((1:n - 0.5) / n, 2)

pdf("figures/(7)_chisquare.pdf")
plot(q_j, d2_j_o, xlab = expression(paste(chi[2]^2, " percentiles")),
  ylab = expression(paste("Ordered ", d[j]^2)))
abline(0, 1, lty = 2)
dev.off()

####
#### (8)
####

## Load data.
X <- data.frame("Age" = c(1, 2, 3, 3, 4, 5, 6, 8, 9, 11), "Price" = c(18.95, 19.00,
  17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99))
n <- nrow(X)
attach(X)

## Function for evaluating univariate Box-Cox objective function.
csld <- colSums(log(X))
```

```
obj_f_univ <- function(lambda, j) {
  x_l <- X[, j]
  if(lambda != 0) {
    x_l <- (x_l ^ lambda - 1) / lambda
  } else {
    x_l <- log(x_l)
  }
  s <- ((n - 1) / n) * var(x_l)

  return(-(n / 2) * log(s) + (lambda - 1) * csld[j])
}

##
## Transforming the age variable.
##

## Somewhat right-skewed as-is.
par(mfrow = c(1, 2))
hist(Age)
qqnorm(Age); qqline(Age)

## Optimal power transformation with lambda about 0.38.
par(mfrow = c(1, 1))
bc_Age <- boxcox(Age ~ 1); title(main = "Age")
lambda_Age <- bc_Age$x[which.max(bc_Age$y)]

Age_tr <- (Age ^ lambda_Age - 1) / lambda_Age

## Verify the above manually.
lambda_seq <- seq(from = -2, to = 2, length = 100)
obj <- rep(NA, 100)
for(i in 1:100)
  obj[i] <- obj_f_univ(lambda_seq[i], 1)
lambda_seq[which.max(obj)]

plot(lambda_seq, obj, type = "l")

## After which, QQ plot looks better. Histogram doesn't look particularly Normal, though.
pdf("figures/(8)_age.pdf")
par(mfrow = c(1, 2))
hist(Age_tr, xlab = "Transformed Age", main = "")
qqnorm(Age_tr); qqline(Age_tr)
dev.off()

##
```

```
## Transforming the price variable.
##

## Left-skewed as-is.
par(mfrow = c(1, 2))
hist(Price)
qqnorm(Price); qqline(Price)

## Optimal power transformation with lambda about 0.95.
par(mfrow = c(1, 1))
bc_Price <- boxcox(Price ~ 1); title(main = "Price")
lambda_Price <- bc_Price$x[which.max(bc_Price$y)]

Price_tr <- (Price ^ lambda_Price - 1) / lambda_Price

## Verify the above manually.
obj <- rep(NA, 100)
for(i in 1:100)
  obj[i] <- obj_f_univ(lambda_seq[i], 2)
lambda_seq[which.max(obj)]

plot(lambda_seq, obj, type = "l")

## After which, the data almost look light-tailed.
pdf("figures/(8)_price.pdf")
par(mfrow = c(1, 2))
hist(Price_tr, xlab = "Transformed Price", main = "")
qqnorm(Price_tr); qqline(Price_tr)
dev.off()

##
## Transforming both variables jointly. While the optimal individual transformations
## correspond to lambda_1 = 0.38 and lambda_2 = 0.95, the optimal joint transformation
## corresponds to lambda' = [1.27, 0.02]. I guess the two approaches don't always match
## after all.
##

## Function to evaluate bivariate Box-Cox objective function.
obj_f_biv <- function(lambda) {
  X_l <- X

  for(k in 1:2) {
    if(lambda[k] != 0) {
      X_l[, k] <- (X_l[, k] ^ lambda[k] - 1) / lambda[k]
    } else {
```

```
      X_l[, k] <- log(X_l[, k])
    }
  }
  S <- var(X_l)

  return(-(n / 2) * log(det(S)) + (lambda - 1) %*% csld)
}

lambda_seq <- seq(from = -2, to = 2, length = 100)
obj <- matrix(NA, nrow = 100, ncol = 100)

for(i in 1:100) {
  for(j in 1:100) {
    obj[i, j] <- obj_f_biv(lambda_seq[c(i, j)])
  }
}

## Optimal lambda vector.
lambda_biv <- c(matrix(rep(lambda_seq, times = 100), nrow = 100)[which.max(obj)],
  matrix(rep(lambda_seq, each = 100), nrow = 100)[which.max(obj)])

pdf("figures/(8)_biv_contour.pdf")
par(mfrow = c(1, 1))
contour(lambda_seq, lambda_seq, obj, xlab = expression(lambda[1]),
  ylab = expression(lambda[2]))
points(lambda_Age, lambda_Price, pch = 20, col = "red", cex = 2)
points(lambda_biv[1], lambda_biv[2], pch = 20, col = "blue", cex = 2)
dev.off()

## Apply transformations.
Age_tr_biv <- (Age ^ lambda_biv[1] - 1) / lambda_biv[1]
Price_tr_biv <- (Price ^ lambda_biv[2] - 1) / lambda_biv[2]

pdf("figures/(8)_biv_age.pdf")
par(mfrow = c(1, 2))
hist(Age_tr_biv, xlab = "Transformed Age", main = "")
qqnorm(Age_tr_biv); qqline(Age_tr_biv)
dev.off()

pdf("figures/(8)_biv_price.pdf")
par(mfrow = c(1, 2))
hist(Price_tr_biv, xlab = "Transformed Price", main = "")
qqnorm(Price_tr_biv); qqline(Price_tr_biv)
dev.off()
```

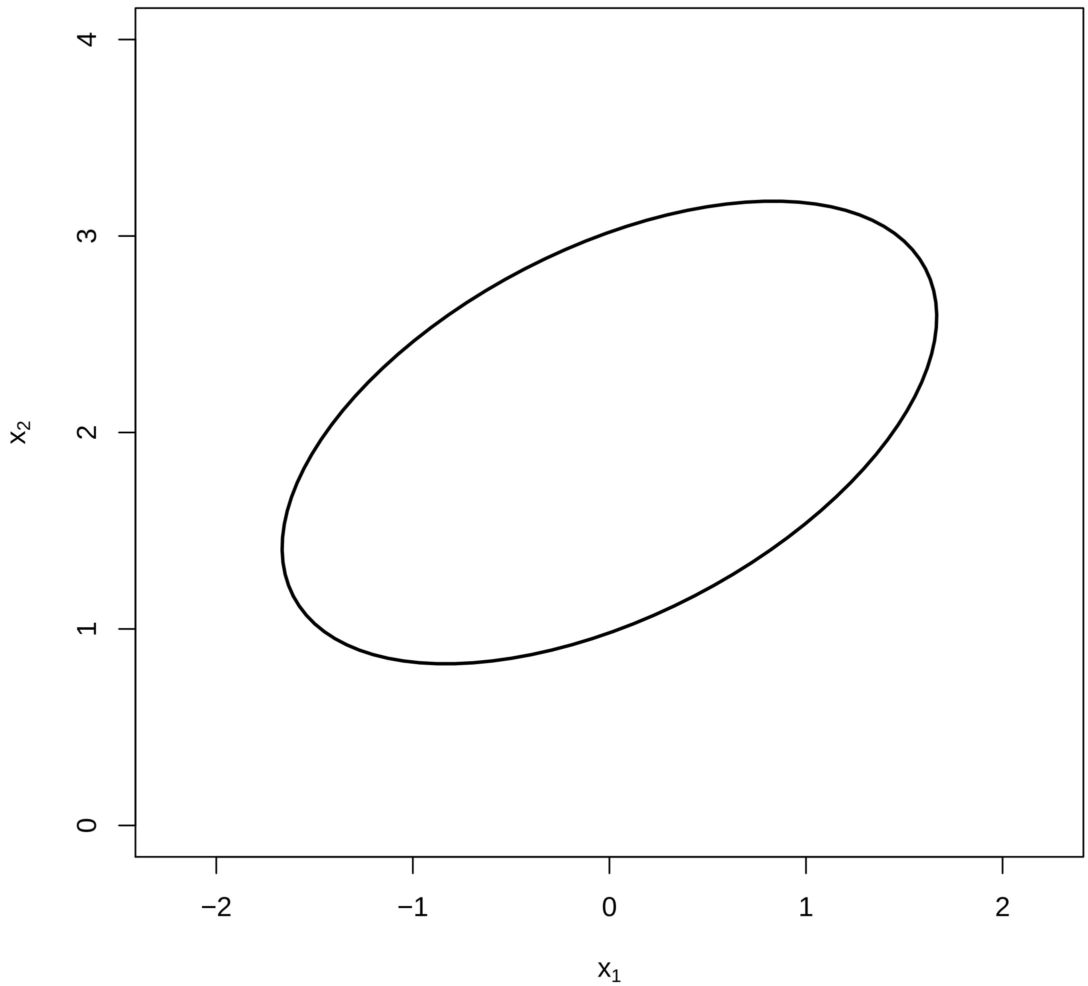


Figure 1: Ellipse containing inner 50% probability for bivariate normal distribution in (1).

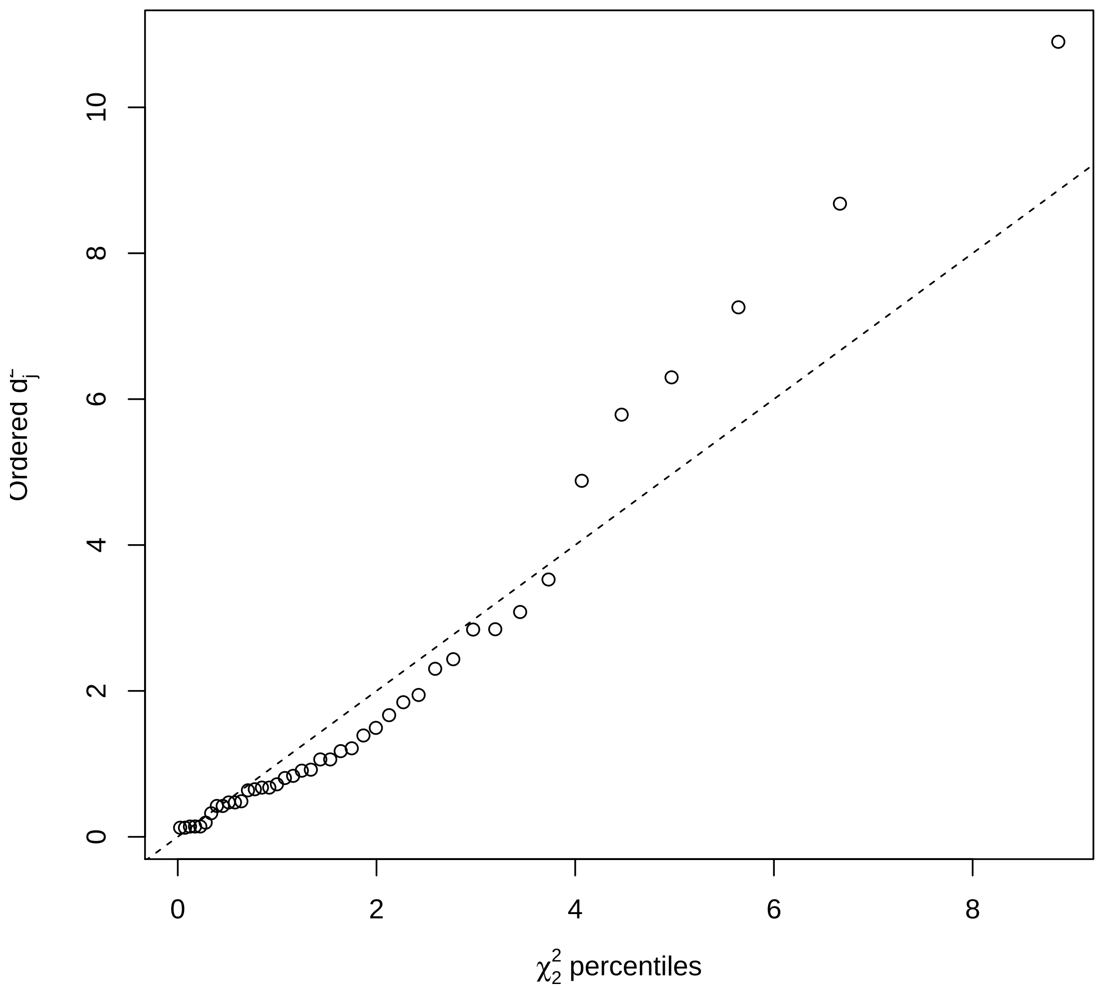


Figure 2: Chi-square plot for the pollution data in (7). Substantial deviation from the line of equality, suggesting non-normality.

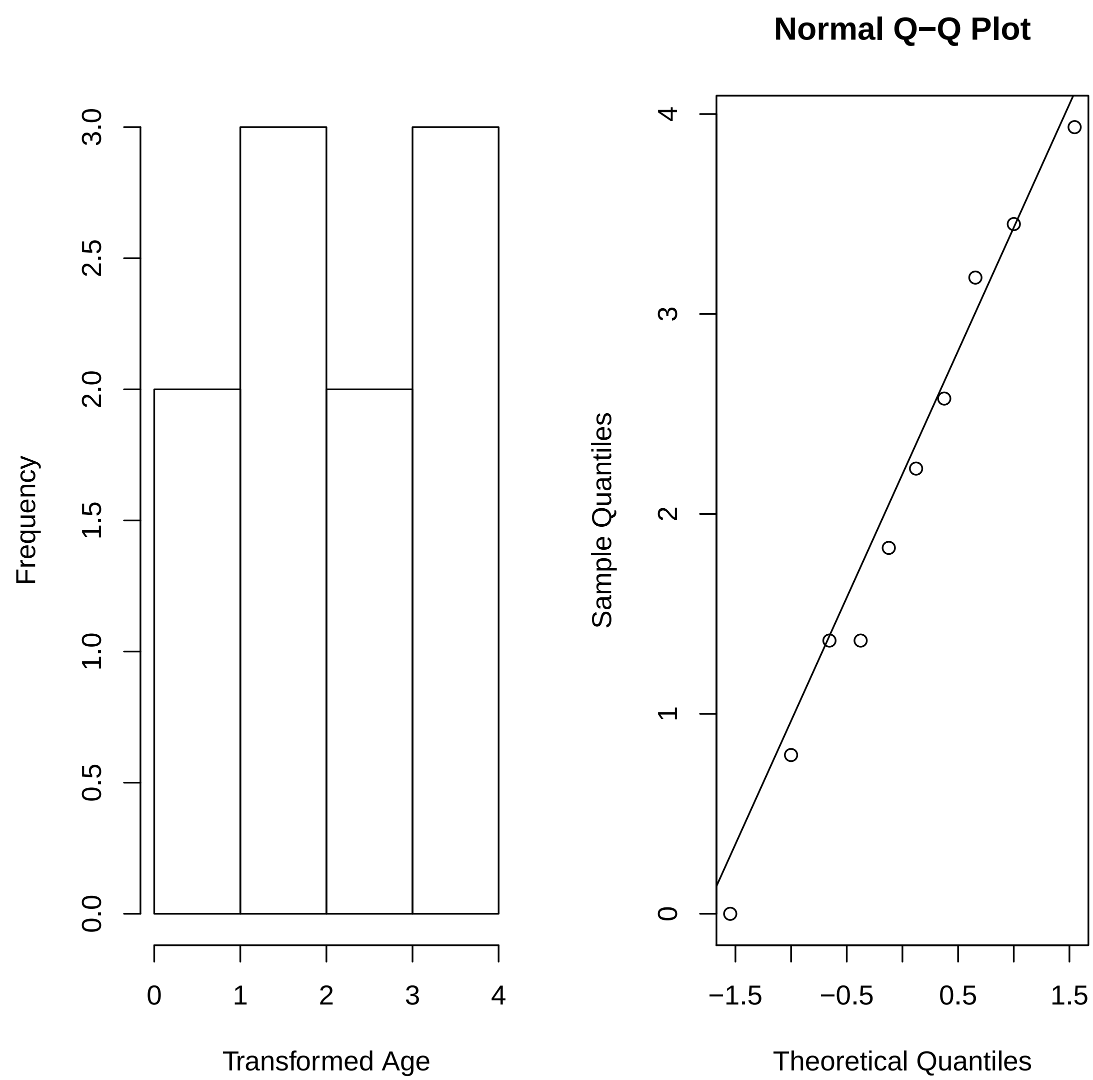


Figure 3: Histogram and normal Q-Q plot for transformed age variable, based on individual application of Box-Cox. The Q-Q plot looks nice, althought the histogram doesn't look particularly normal.

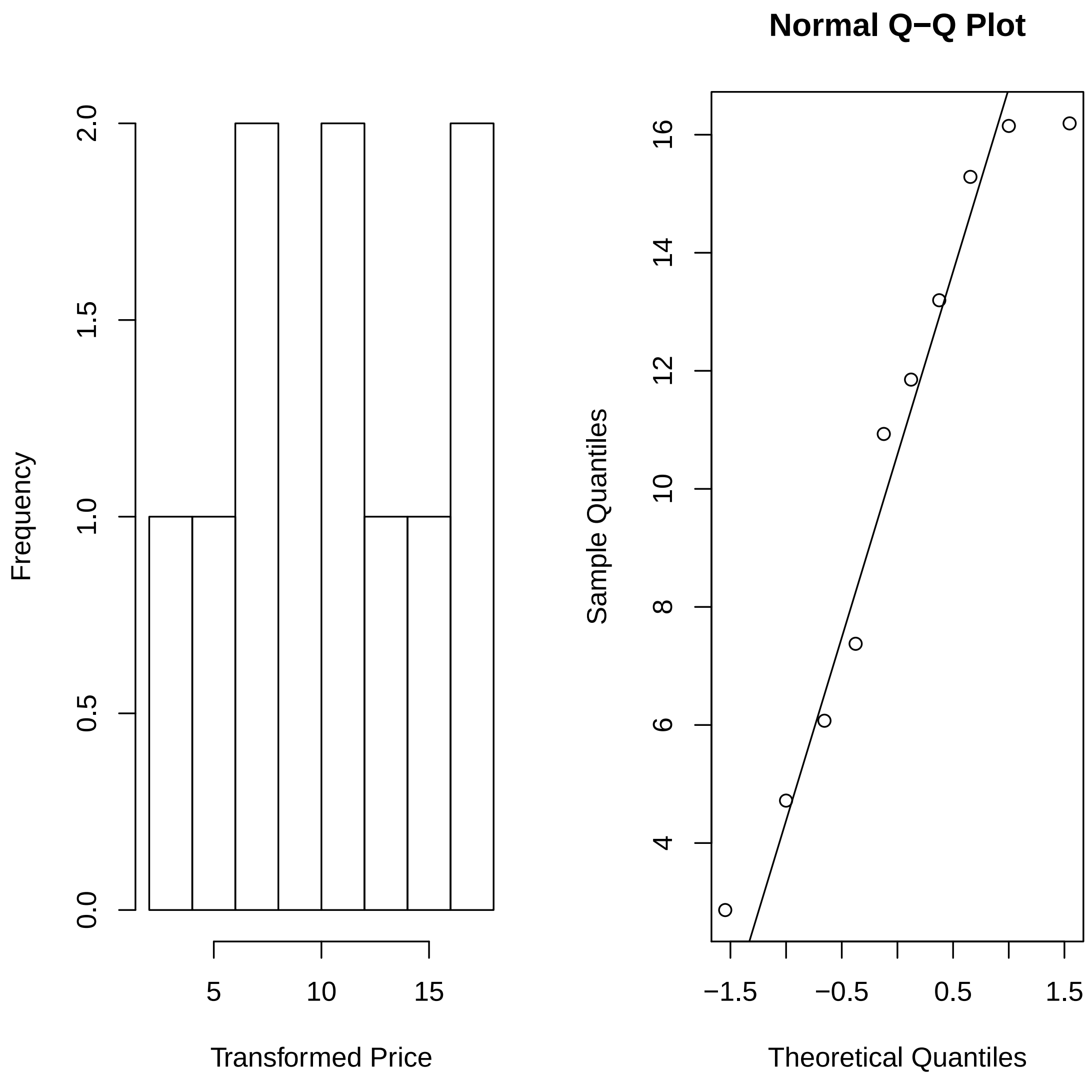


Figure 4: Histogram and normal Q-Q plot for transformed price variable, based on individual application of Box-Cox. Based on the Q-Q plot, the transformed data now appear light-tailed.

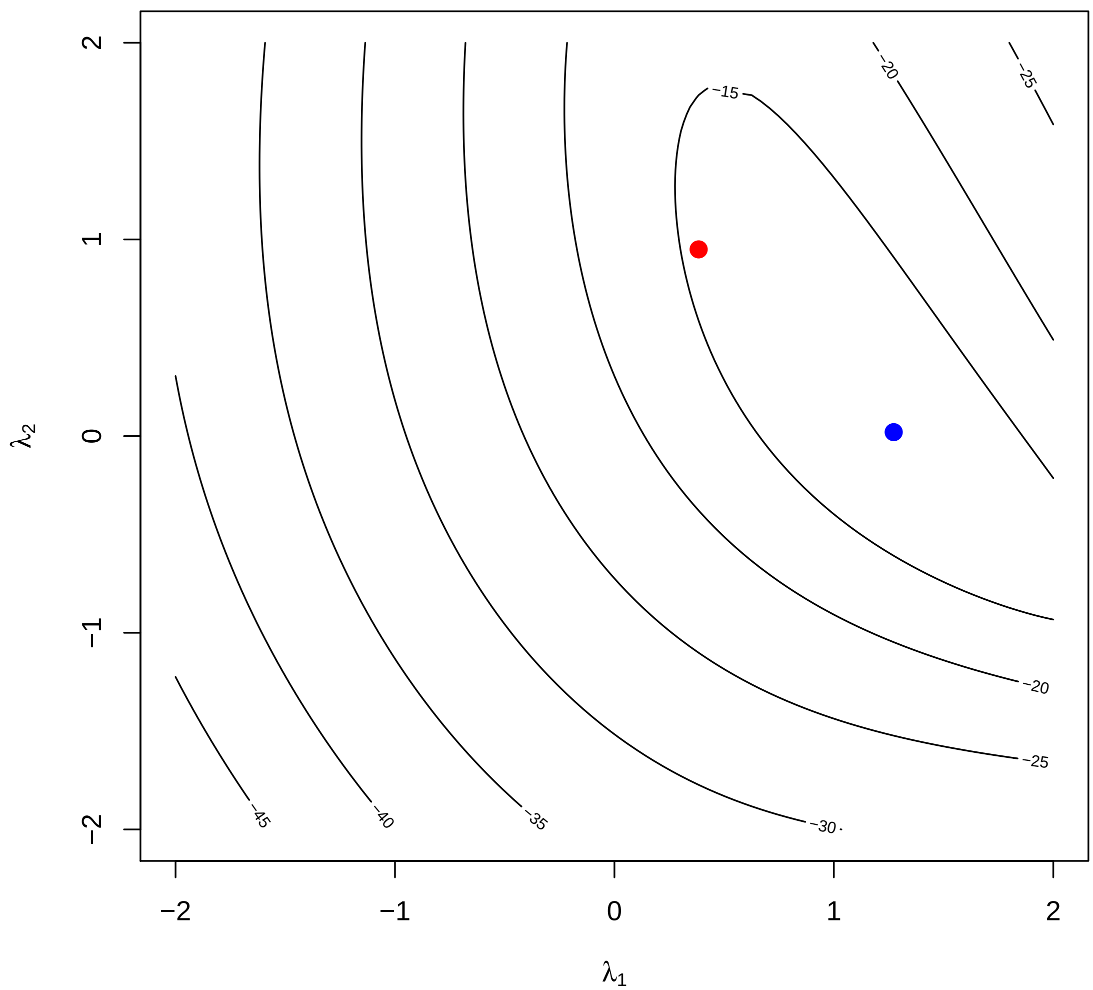


Figure 5: Box-Cox multivariate objective function contour plot. The red dot shows the location of the $\hat{\lambda}$ values from the previous two parts, after individual Box-Cox applications. The blue dot shows the location of the $\hat{\lambda}$ values after multivariate Box-Cox application.

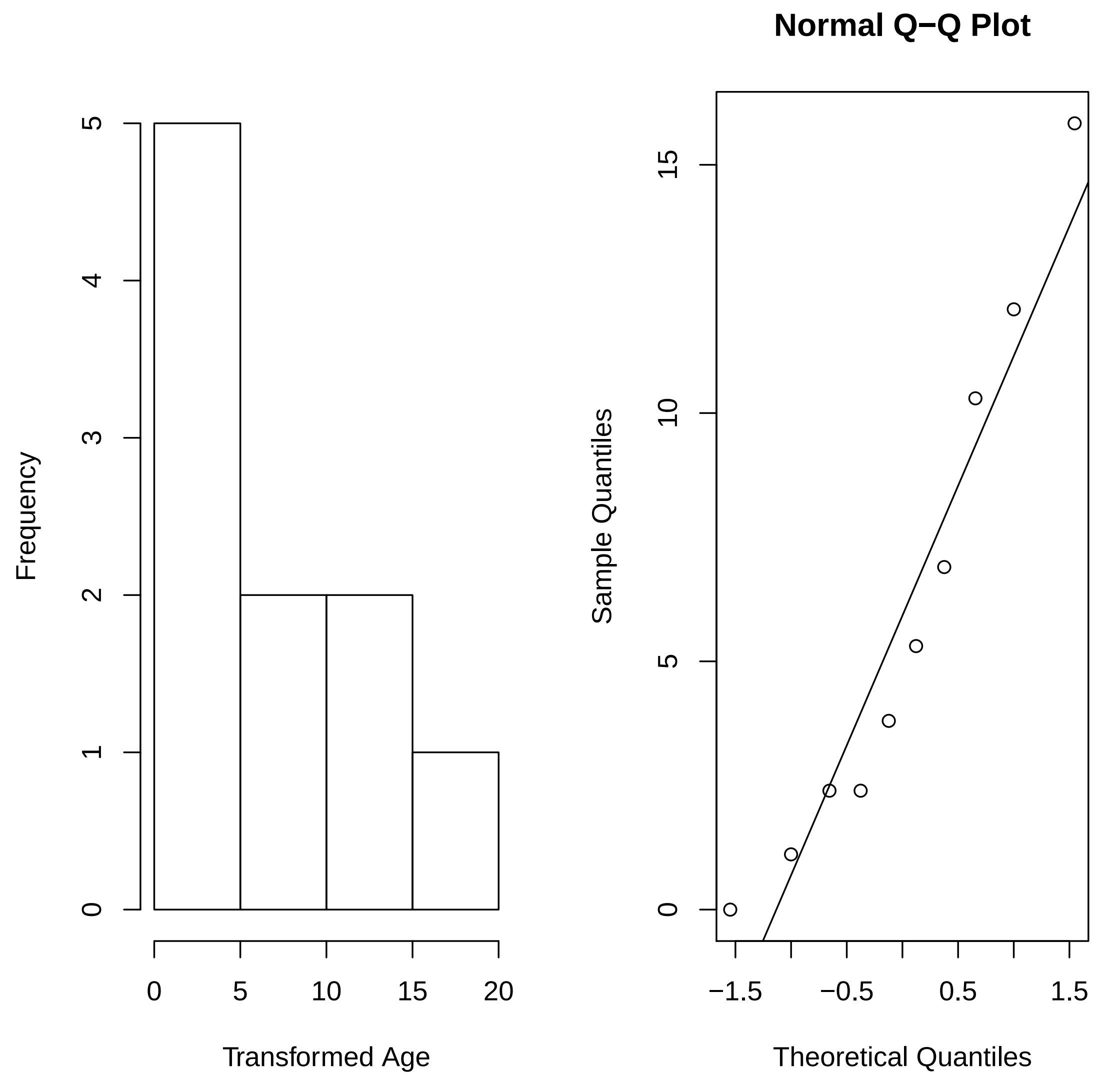


Figure 6: Histogram and normal Q-Q plot for transformed age variable, based on multivariate Box-Cox. Now looks a bit right-skewed.

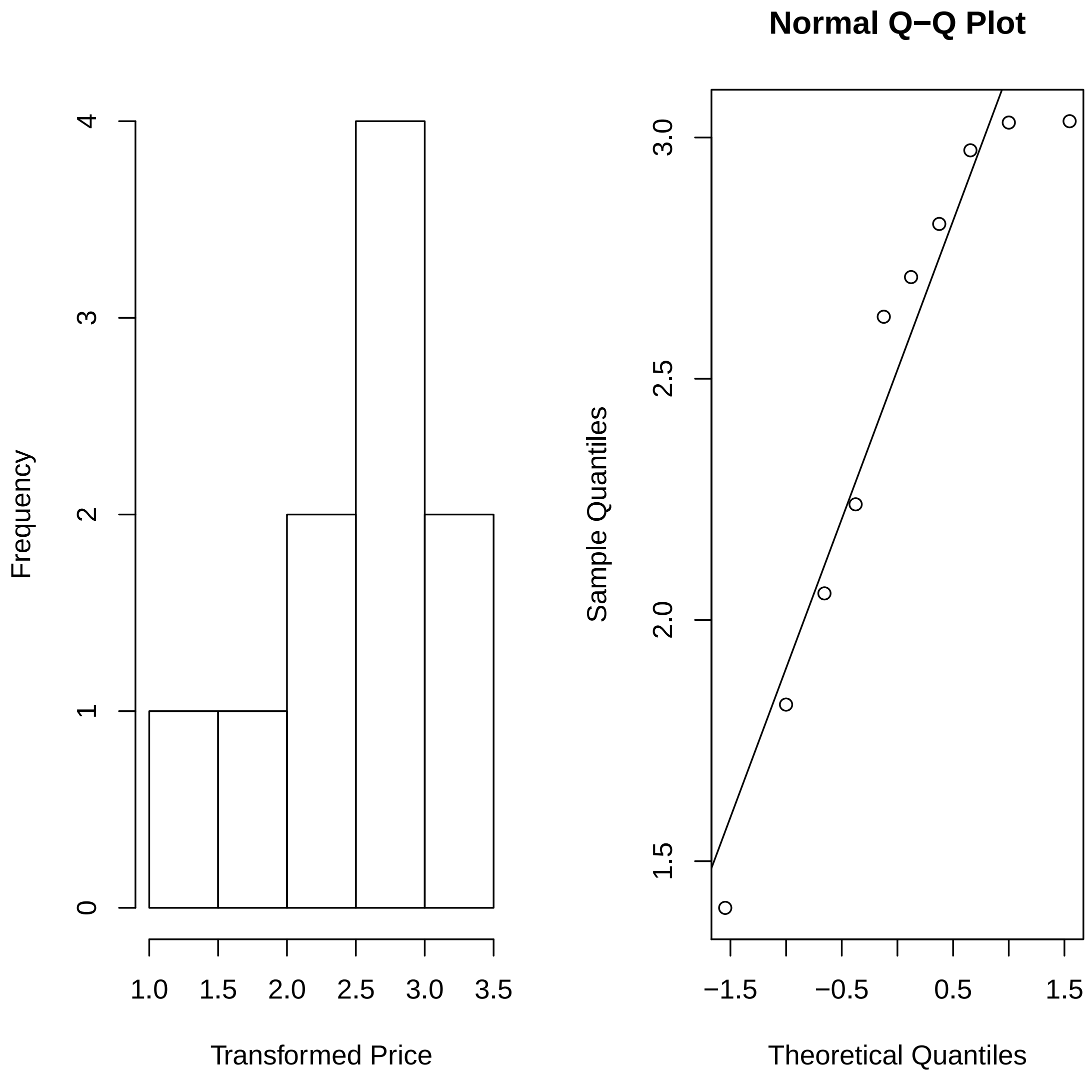


Figure 7: Histogram and normal Q-Q plot for transformed age variable, based on multivariate Box-Cox. Now looks a bit left-skewed.