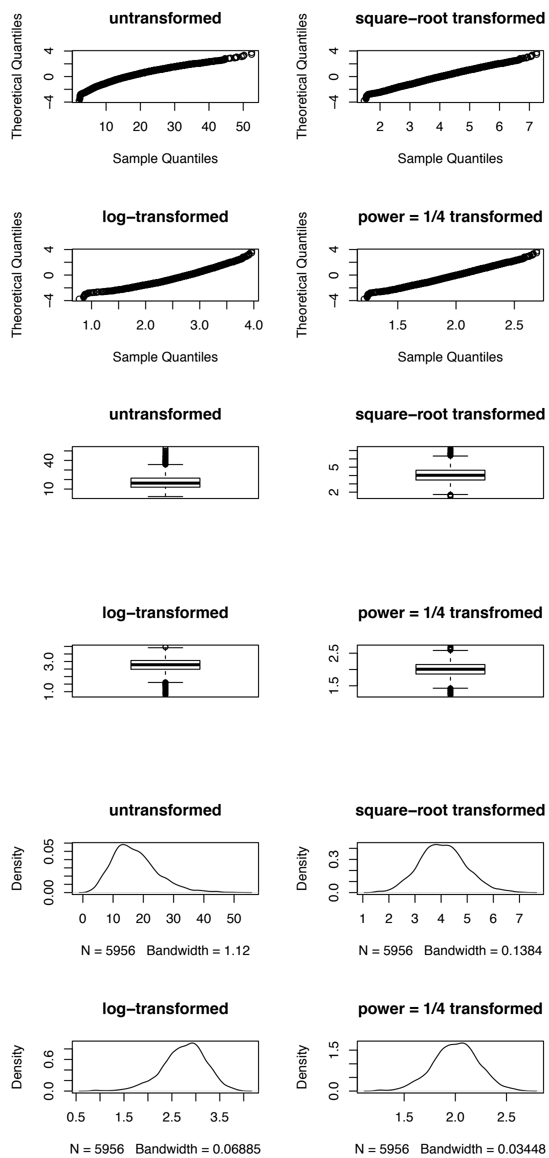
Problem P2: Chapter 5 R-lab

Problem 1
The square-root transformed distribution is the most symmetric. The transformation of power ¼ is good as, if not worse, the square-root transformed distribution.



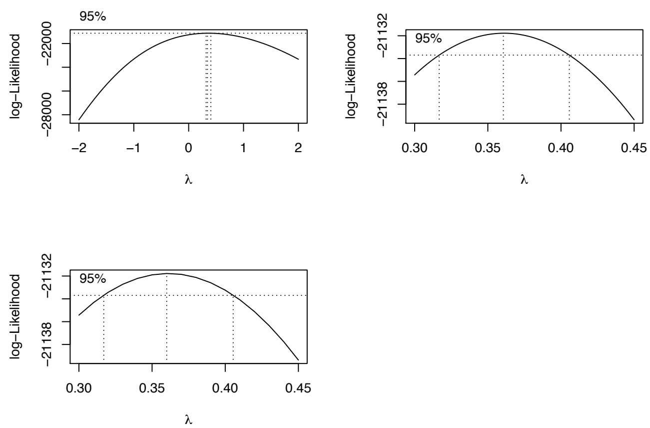
Problem 2
(a) What are ind and ind2 and what purposes do they serve?
“ind” and “ind2“ are indices that meet the described conditions.

(b) What is the effect of interp on the output from boxcox?
“interp” specifies if spline interpolation is used. Default is “true” when lambda of length is less than 100.

(c) What is the MLE of λ?
MLE of λ is 0.36 (bc$x[ind] = 0.36).

(d) What is a 95% confidence interval for λ?
The 95% confidence interval is [0.32, 0.40] (bc$x[ind2] = [0.32, 0.40]).

(e) Modify the code to find a 99% confidence interval for λ.
The 99% confidence interval is [0.31, 0.41]

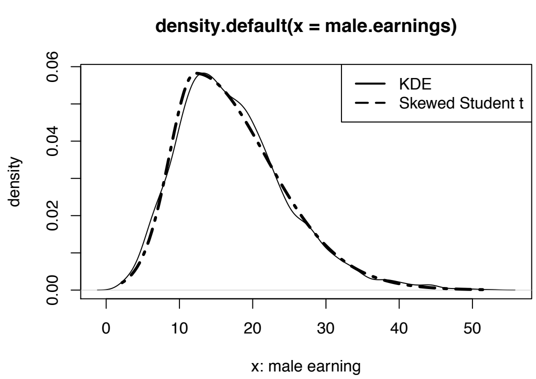


Problem 3
What are the estimates of the degrees-of-freedom parameter and of ξ?

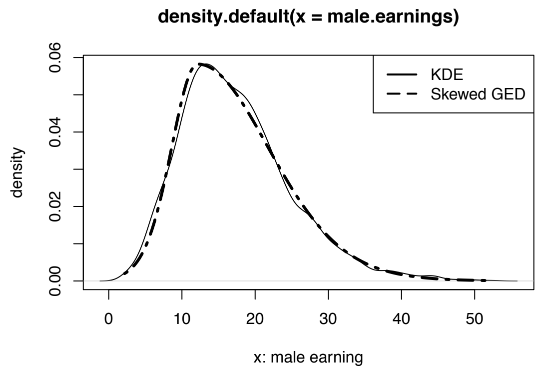
```
$minimum
[1] 20121.41

$estimate
     mean        sd        nu        xi
17.322933  7.492440 21.600108  1.651652
```

From above output, the estimate of ξ is 1.651652

Problem 4
The kernel density estimate resembles the skewed t density, and the skewed t-model provides an adequate fit to male.earnings.



Problem 5
The skewed t-distribution and the skewed GED almost overlap. The skewed t-distribution fits better, if not as good as the skewed GED does.



Problem 6
What are the MLEs of the mean, standard deviation, and the degrees-of-freedom parameter? What is the value of AIC?
"MLE =" "0.00078" "0.01058" "4.03515"
From the above output, the MLEs of the mean, standard deviation, and the degrees-of-freedom are 0.00078, 0.01058, 4.03515, and AIC_std = -11960.47.

Problem 7
Modify the code so that the MLEs for the skewed t-distribution are found:

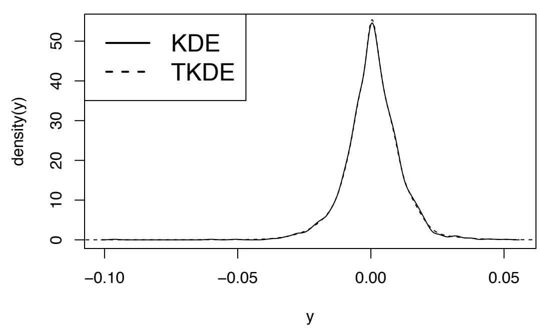
```
loglik_std = function(x) {
  f = -sum(log(dsstd(Y, mean=x[1], sd=x[2], nu=x[3], xi=x[4])))
  f}
start=c(mean(Y),sd(Y),4,1.5) # optimization starting point
fit_sstd = optim(start,loglik_std,method="L-BFGS-B",
              lower=c(-.1,.001,2.1,0.5),
              upper=c(.1,1,20,3))
print(c("MLE =",round(fit_sstd$par,digits=5)))
m_logL_sstd = fit_sstd$value  # minus the log-likelihood
AIC_std = 2*m_logL_sstd+2*length(fit_sstd$par)
```

What are the MLEs? Which distribution is selected by AIC, the t or the skewed t-distribution?
Output:
"MLE =" "0.00075" "0.00979" "7.40648" "1.00014"
From the above output, the MLEs of the mean, standard deviation, and the degrees-of-freedom are 0.00075, 0.00979, 7.40648, 1.00014, and AIC_std = -11937.08

Problem 8

```
x1=pstd(Y,mean=fit_std$par[1],sd=fit_std$par[2],nu=fit_std$par[3])
x=qnorm(x1)
par(mfrow=c(1,1))
```

```
d1=density(Y)
plot(d1$x,d1$y,type="l",lty=1, xlab="y",ylab="density(y)")

d2=density(x)
ginvx=qstd(pnorm(d2$x),mean=fit_std$par[1],sd=fit_std$par[2],nu=fit_std$par[3])
gprime_num=dstd(ginvx,mean=fit_std$par[1],sd=fit_std$par[2],nu=fit_std$par[3])
gprime_den=dnorm(qnorm(pstd(ginvx,mean=fit_std$par[1],sd=fit_std$par[2],nu=f
it_std$par[3])))
gprime=gprime_num/gprime_den
lines(ginvx,d2$y*gprime,type="l",lty=2)
legend("topleft",c("KDE","TKDE"),lty=c(1,2),lwd=2,cex=1.5)
```



Problem 9

The parametric estimates and the TKDE are similar.