

STA135 homework5

Zhen Zhang

January 31, 2016

4.25

```
sales <- c(108.28, 152.36, 95.04, 65.45, 62.97, 263.99, 265.19, 285.06, 92.01, 165.68)
profits <- c(17.05, 16.59, 10.91, 14.14, 9.52, 25.33, 18.54, 15.73, 8.10, 11.13)
assets <- c(1484.10, 750.33, 766.42, 1110.46, 1031.29, 195.26, 193.83, 191.11, 1175.16, 211.15)
top <- data.frame(sales, profits, assets)
```

The mean vector is:

```
(top_mean <- apply(top, 2, mean))
```

```
##   sales profits assets
## 155.603  14.704 710.911
```

The covariance matrix:

```
top_d <- t(apply(top, 1, function(x) x-top_mean))
(top_variance <- apply(top_d, 2, function(x) {
  apply(top_d, 2, function(y) {
    x%*%y/(length(top_d[,1])-1)
  })
}))
```

```
##           sales      profits      assets
## sales      7476.4532    303.61862 -35575.960
## profits     303.6186     26.19032 -1053.827
## assets    -35575.9596  -1053.82739 237054.270
```

Now calculate the statistical distance, and order them:

```
(top_unorder_distance <- apply(top, 1, function(x) {
  t(x-top_mean)%*%solve(top_variance)%*%(x-top_mean)
}))
```

```
## [1] 4.8364455 0.3142263 1.2894373 2.0194923 1.4072664 4.9090461 1.6418144
## [8] 4.3520262 3.0411050 3.1891403
```

```
(top_order_variance <- sort(top_unorder_distance))
```

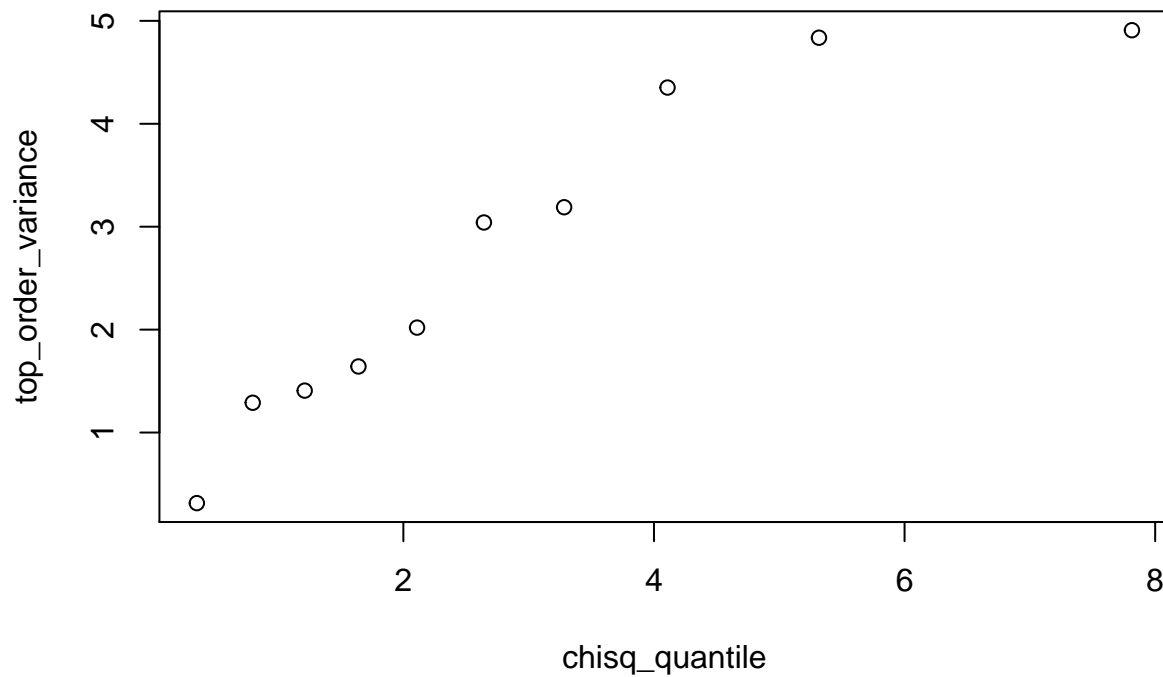
```
## [1] 0.3142263 1.2894373 1.4072664 1.6418144 2.0194923 3.0411050 3.1891403
## [8] 4.3520262 4.8364455 4.9090461
```

The chi-square quantile is:

```
chisq_quantile <- sapply(10:1, function(x) qchisq((10-x+0.5)/10,3))
```

Plot:

```
plot(chisq_quantile, top_order_variance)
```



4.26

(a)

```
x1 <- c(1,2,3,3,4,5,6,8,9,11)
x2 <- c(18.95, 19.00, 17.95, 15.54, 14.00, 12.95, 8.94, 7.49, 6.00, 3.99)
cars <- data.frame(x1, x2)
# the mean
(cars_mean <- apply(cars, 2, mean))
```

```
##      x1      x2
## 5.200 12.481
```

```
cars_d <- t(apply(cars, 1, function(x) x-cars_mean))
# the variance
(cars_variance <- apply(cars_d, 2, function(x) {
```

```

apply(cars_d, 2, function(y) {
  x%*%y/(length(cars_d[,1])-1)
})
}))

```

```

##           x1           x2
## x1  10.62222 -17.71022
## x2 -17.71022  30.85437

```

```

# the statistical distance
(cars_unorder_distance <- apply(cars, 1, function(x) {
  t(x-cars_mean)%*%solve(cars_variance)%*%(x-cars_mean)
})))

```

```

## [1] 1.8753045 2.0203262 2.9009088 0.7352659 0.3105192 0.0176162 3.7329012
## [8] 0.8165401 1.3753379 4.2152799

```

(b)

```

sum(cars_unorder_distance < qchisq(0.5, 2))

```

```

## [1] 5

```

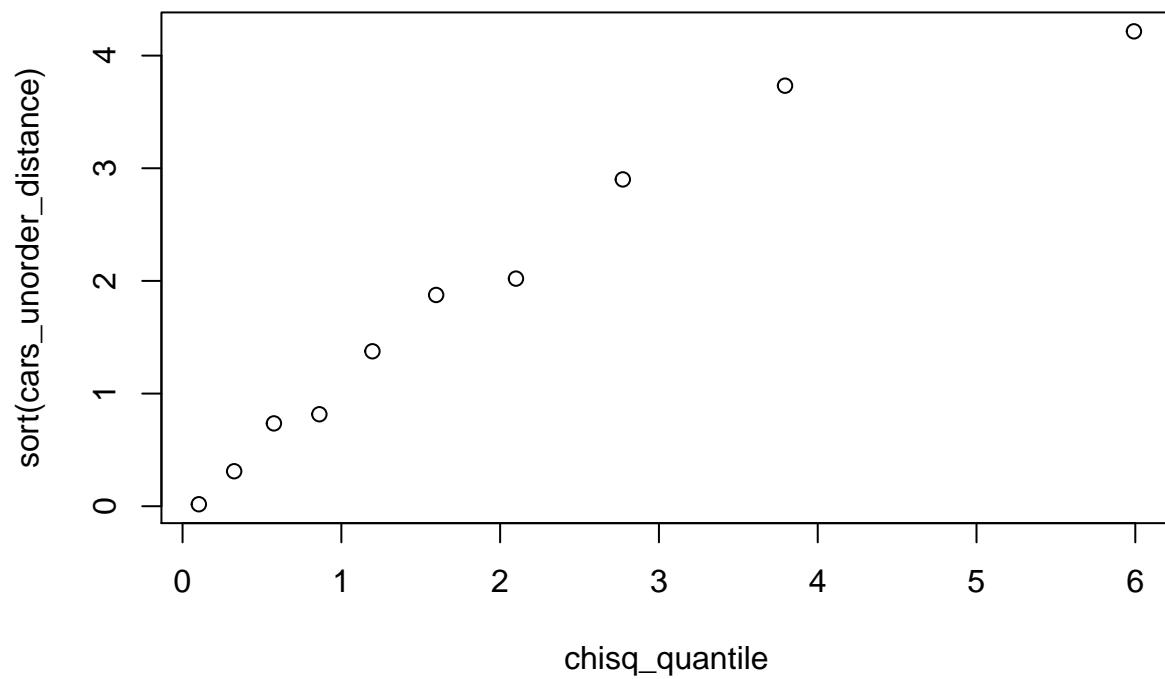
So there are 5 observations within the 50% contour.

(c)

```

chisq_quantile <- sapply(10:1, function(x) qchisq((10-x+0.5)/10, 2))
plot(chisq_quantile, sort(cars_unorder_distance))

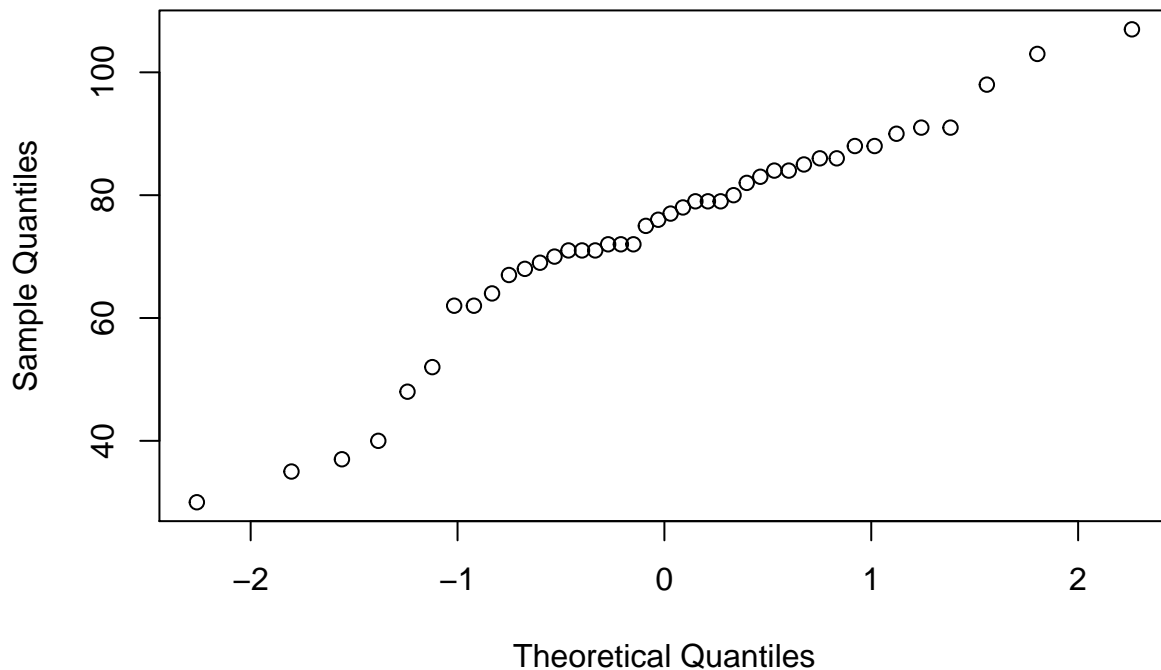
```



4.28

```
pollution <- read.table("T1-5.dat")
pollution_solar_radiation <- pollution[,2]
pollution_solar_radiation_qqnorm <- qqnorm(pollution_solar_radiation)
```

Normal Q-Q Plot



```
cor(pollution_solar_radiation_qqnorm$x, pollution_solar_radiation_qqnorm$y)
```

```
## [1] 0.9693258
```

The correlation is 0.9693258. For $n = 40$, $\alpha = 0.05$, the coefficient should be 0.9726 (from table 4-2), so we would reject the hypothesis of normality at 5% level.

4.29

(a)

```
pollution_cut <- pollution[, 5:6]
pollution_cut_mean <- apply(pollution_cut, 2, mean)
pollution_cut_d <- t(apply(pollution_cut, 1, function(x) x-pollution_cut_mean))
(pollution_cut_variance <- apply(pollution_cut_d, 2, function(x) {
  apply(pollution_cut_d, 1, function(y) {
    x*y/(length(pollution_cut_d[,1])-1)
  })
}))
```

```
##           V5           V6
## V5 11.363531  3.126597
## V6  3.126597 30.978513
```

```
# the statistical distance
(pollution_cut_unorder_distance <- apply(pollution_cut, 1, function(x) {
  t(x-pollution_cut_mean)%*%solve(pollution_cut_variance)%*%(x-pollution_cut_mean)
}))
```

```
## [1] 0.4606524 0.6592206 2.3770610 1.6282902 0.4135364 0.4760726
## [7] 1.1848895 10.6391792 0.1388339 0.8162468 1.3566301 0.6228096
## [13] 5.6494392 0.3159498 0.4135364 0.1224973 0.8987982 4.7646873
## [19] 3.0089122 0.6592206 2.7741416 1.0360061 0.7874152 3.4437748
## [25] 6.1488606 1.0360061 0.1388339 0.8856041 0.1379719 2.2488867
## [31] 0.1901188 0.4606524 1.1471939 7.0857237 1.4584229 0.1224973
## [37] 1.8984708 2.7782596 8.4730649 0.6370218 0.7032485 1.8013611
```

(b)

```
sum(pollution_cut_unorder_distance < qchisq(0.5, 2))
```

```
## [1] 26
```

The number of observations that fall within the 50% probability contour of bivariate normal distribution.

(c)

```
chisq_quantile <- sapply(42:1, function(x) qchisq((42-x+0.5)/42, 2))
plot(chisq_quantile, sort(pollution_cut_unorder_distance))
```

