STAT509 Homework 5 Solution

Chong Ma

March 30, 2017

Problem 1

(a) Note that the sample of $n = 84$ bearings is assumed being randomly selected from a population with unknown population mean μ and unknown population variance σ^2. Then, point estimators for μ and σ are

$$\hat{\mu} = \bar{Y} = 6810.3774$$
$$\hat{\sigma} = S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 9153.454$$

(b) Based on the Central Limit Theorem, the sampling distributino of sample mean $\bar{Y} \sim \text{AN}(\mu, \frac{\sigma^2}{n})$. Thus, the standard error of the sample mean $\bar{Y}$ is $\text{se}(\bar{Y}) = \frac{\sigma}{\sqrt{n}}$. Since σ is unknown and the sample standard deviation S is an point estimator of σ, therefore a point estimator of the standard error of the sample mean $\bar{Y}$ is

$$\hat{\text{se}(\bar{Y})} = \frac{S}{\sqrt{n}} = 998.7237$$

Problem 2

(a) The population can be considered as all of children in USA or all of children in SC. The sample is the 633 fourth-grade students selected from schools in Columbia, SC.

(b) The histogram of BMI for the 633 fourth-grade students appears unimodal and slightly right skewed. An straightforward guess for the underlying distribution of BMI is a normal distribution. Nevertheless, Gamma and Weibull distributions are also appealing candidates.

(c) Note that $\bar{y} = 24.26$ and $s = 5.91$,

- Under the normal assumption, that is, $Y_1, Y_2, \ldots, Y_n \overset{i.i.d}{\sim} \text{N}(\mu, \sigma^2)$. Since $E(Y) = \mu$ and $\text{Var}(Y) = \sigma^2$,then a point estimator for μ and σ are

$$\hat{\mu} = \bar{y} = 24.26, \hat{\sigma} = s = 5.91$$

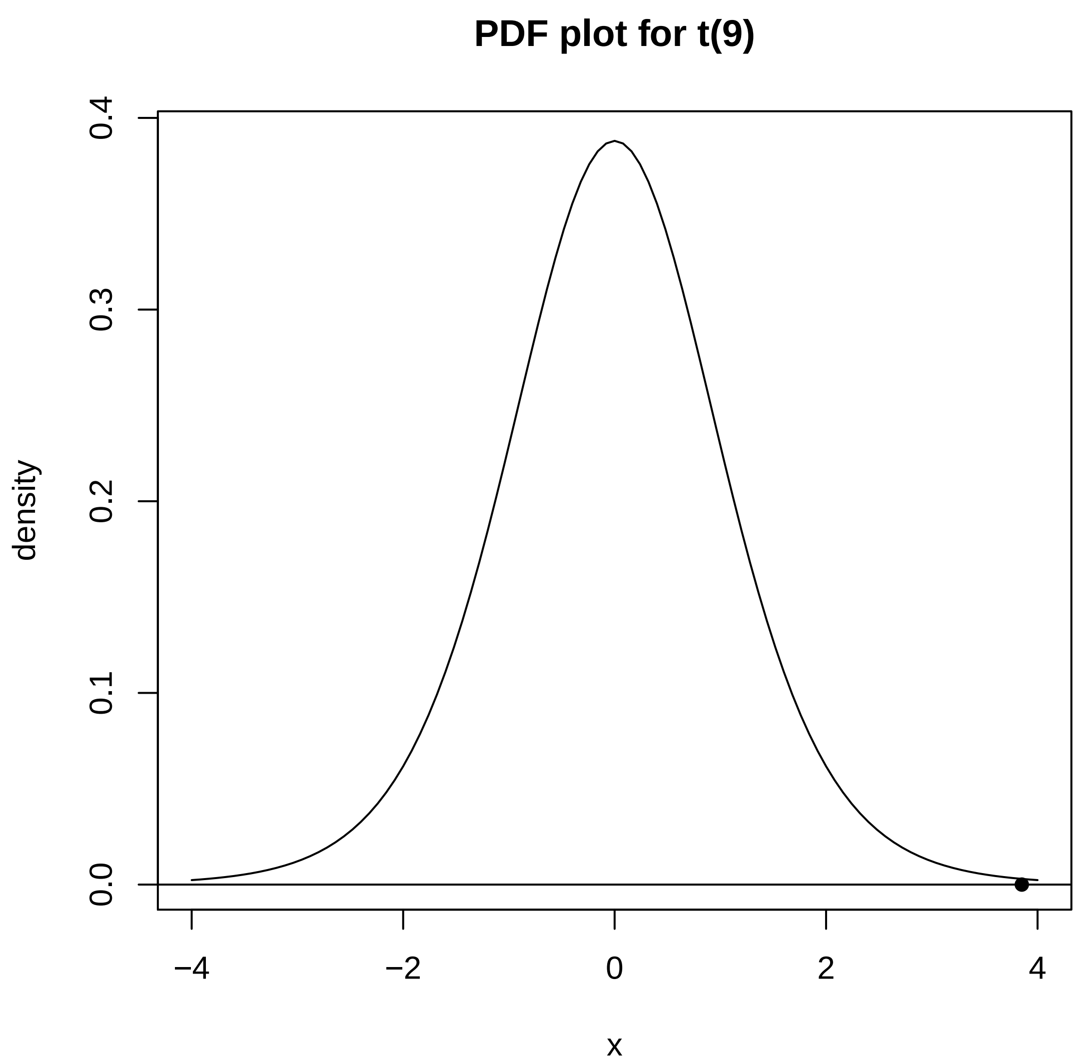
- Under the Gamma assumption, that is, $Y_1, Y_2, \ldots, Y_n \overset{i.i.d}{\sim} \text{Gamma}(\alpha, \lambda)$. Note that $E(Y) = \frac{\alpha}{\lambda}$ and $\text{Var}(Y) = \frac{\alpha}{\lambda^2}$. Thus, point estimators for α and λ are

$$\hat{\alpha} = \frac{\bar{y}^2}{s^2} = 16.8445$$
$$\hat{\lambda} = \frac{\bar{y}}{s^2} = 0.6943$$

(d) Under the Gamma asumption, $P(Y > 30) = 1 - \text{pgamma}(30, 16.84, 0.69) = 0.1693$, which is quite close to 100/633(0.158).

Problem 3

```r
##inpute the data named after cheese
cheese=c(-0.541,-0.538,-0.532,-0.533,-0.526,
         -0.543,-0.537,-0.528,-0.538,-0.549)
##density plot for t(9)
curve(dt(x,9),-4,4,xlab="x",ylab="density",main="PDF plot for t(9)")
abline(h=0)
points(3.851,0,pch=16)
```



(a) The sample mean $\bar{y} = \text{mean(cheese)} = -0.5365$ and the sample standard deviation $s = \text{sd(cheese)} = 0.007$.

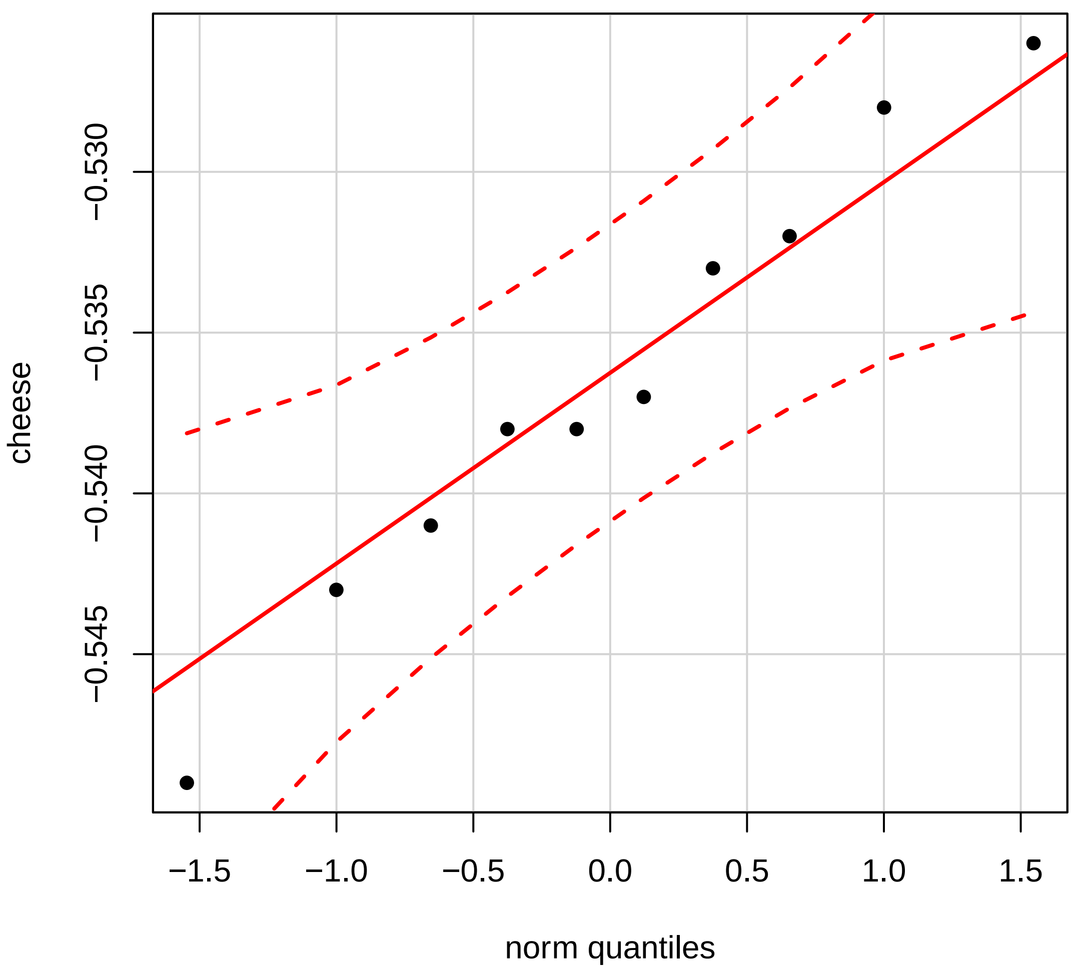
(b) Given $\mu = -0.545$, the t-statistic is

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}} = 3.8508$$

Since the t-statistic is 3.851, which locates on the far right-tail of the distribution t_9, it implies that under the hypothesized mean value $\mu = -0.545$ there is a very small likelihood to observe such a sample. Put in another way, the observed sample make the claimed mean value $\mu = -0.545$ doubtful.

(c) The normal qqplot shows that the data points locate around a straight line and are all within the bound, that is, there is no apparent evidence to support that the normality assmption is violated by the cheese data.

```
qqPlot(cheese,distribution="norm",pch=16)
```

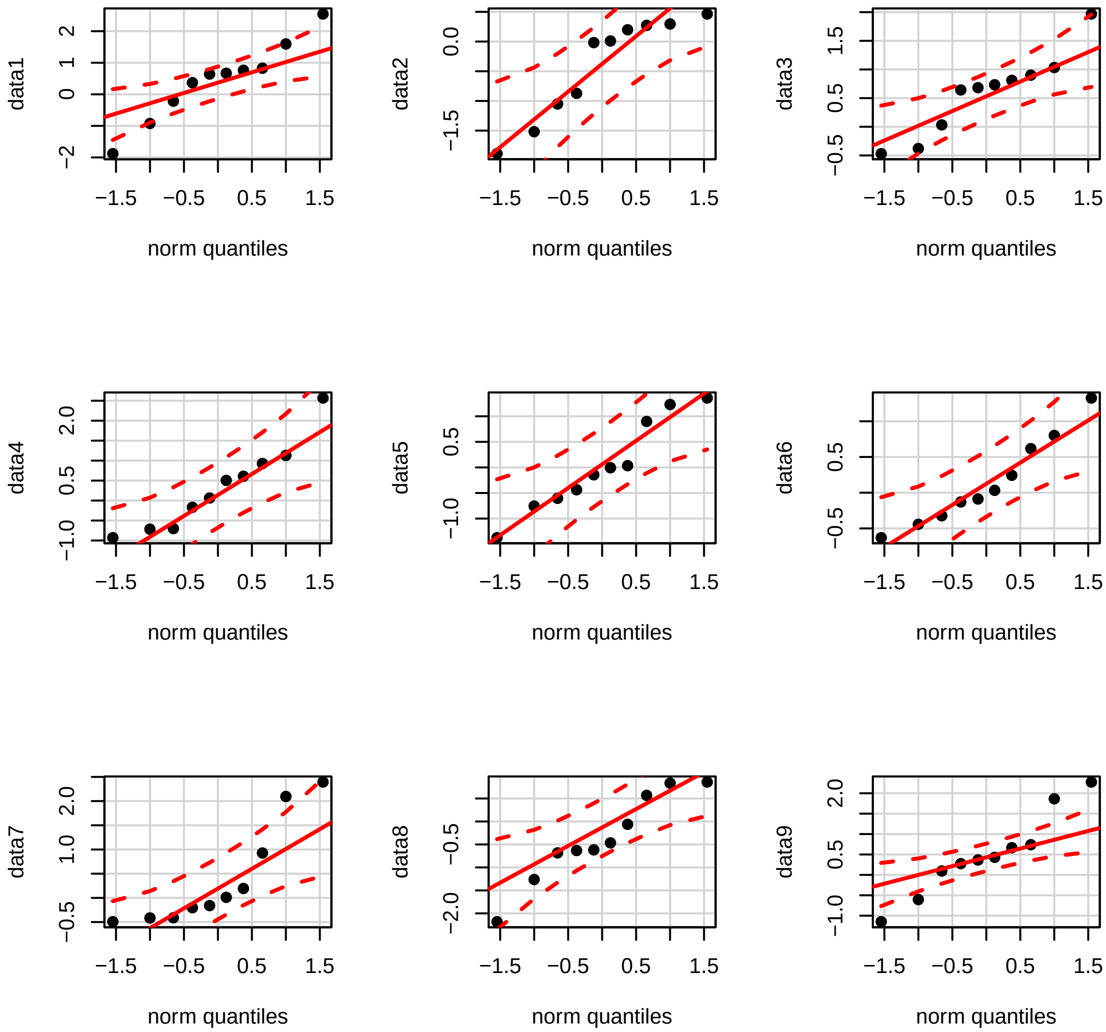


Problem 4

Overall, this simulation tells us that the sampled data can never be perfectly matched with the theoretical distribution because of the natural sampling variability. The departure for the sampled data from its underlying theoretical distribution depends on how large the sampled data. The larger the sampled data, the closer it is to the underlying distribuion.

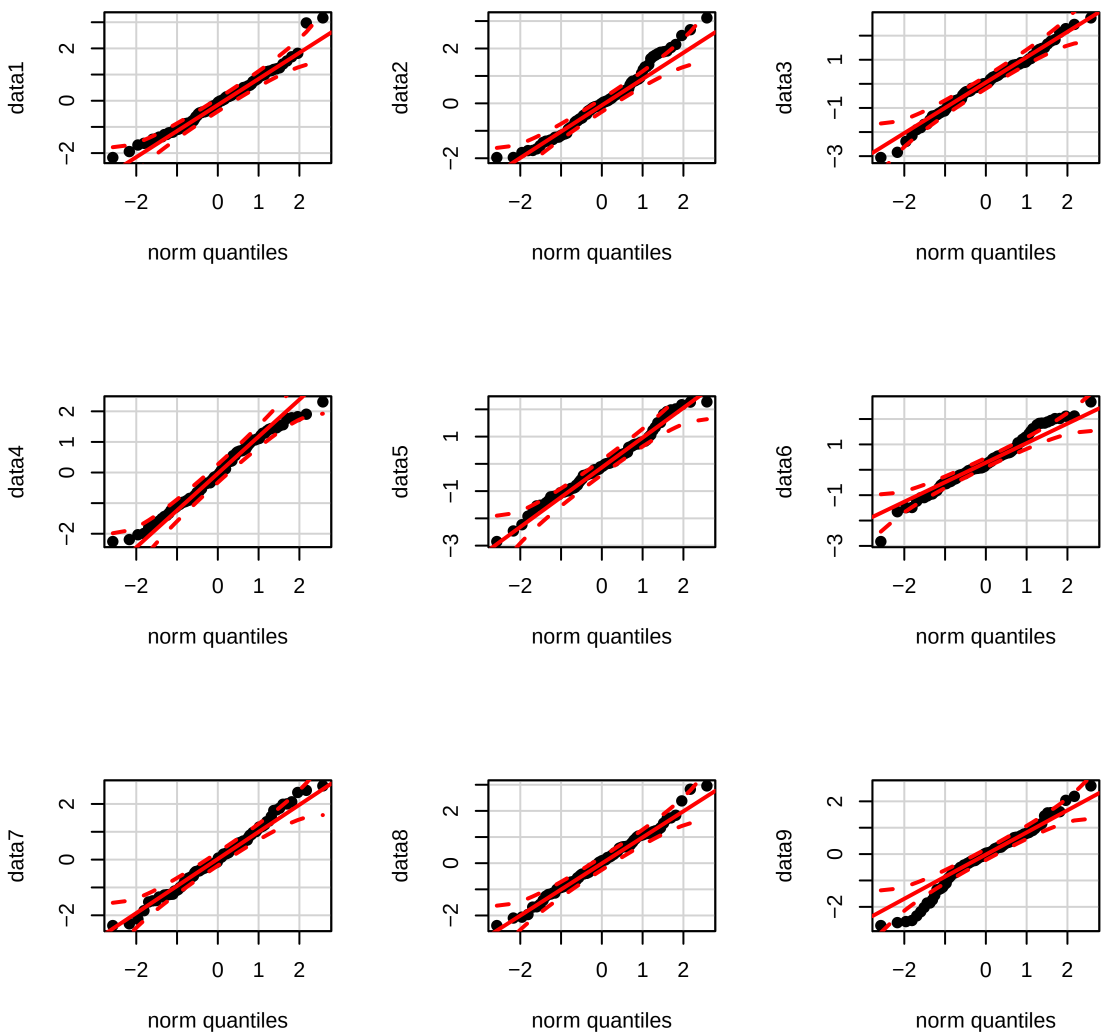
(a)

```
# create 3 by 3 figure
op<-par(mfrow=c(3,3))
B = 9 # number of simulated data sets
n = 10 # sample size
# create matrix to hold all the data
# I removed the round function
# you don't have to print off the data
data = matrix(rnorm(n*B,0,1), nrow = B, ncol = n)
# for loop
# this creates a qq plot for each sample of data
for (i in 1:B){
  qqPlot(data[i,],distribution = "norm",pch=16,cex=1.2,ylab=paste0("data",i))
}
```


data1
2
1
0
−2
data2
0.0
−1.5
data3
1.5
0.5
−0.5
data4
2.0
0.5
−1.0
data5
0.5
−1.0
data6
0.5
−0.5
data7
2.0
1.0
−0.5
data8
−0.5
−2.0
data9
2.0
0.5
−1.0
−1.5
−0.5
0.5
1.5
norm quantiles

(b)

```
# create 3 by 3 figure
op<-par(mfrow=c(3,3))
B = 9 # number of simulated data sets
n = 100 # sample size
# create matrix to hold all the data
# I removed the round function
# you don't have to print off the data
data = matrix(rnorm(n*B,0,1), nrow = B, ncol = n)
# for loop
# this creates a qq plot for each sample of data
for (i in 1:B){
  qqPlot(data[i,],distribution = "norm",pch=16,cex=1.2,ylab=paste0("data",i))
}
```



```r
# create 3 by 3 figure
op<-par(mfrow=c(3,3))
B = 9 # number of simulated data sets
n = 1000 # sample size
# create matrix to hold all the data
# I removed the round function
# you don't have to print off the data
data = matrix(rnorm(n*B,0,1), nrow = B, ncol = n)
# for loop
# this creates a qq plot for each sample of data
for (i in 1:B){
  qqPlot(data[i,],distribution = "norm",pch=16,cex=1.2,ylab=paste0("data",i))
}
```