Math 490 HW #14

Maxwell Levin

March 16, 2018

Question 1.

Explain what is wrong with each of the following statements

a. The bootstrap distribution is created by resampling with replacement from the population.

The bootstrap distribution is not created by resampling from the population, but rather by resampling with replacement from a sample without replacement of the original population.

b. The bootstrap distribution is created by resampling without replacement from the original sample.

The bootstrap distribution is created by resampling with replacement from the original sample. If we resampled without replacement we would get the same sample every time (order does not matter).

c. When generalizing the resamples, it is best to use a sample size larger than the size of the original sample.

It is actually best to use a sample size that is the same as the sample size of the original sample.

d. The bootstrap distribution will be similar to the sampling distribution in shape, center, and spread

It is true that the bootstrap distribution will be similar to the sampling distribution in shape and spread, but I think that the center will be similar to the center of the original data, not necessarily the center of the sampling distribution.

For Questions 2 and 3, bootstrap the sample mean for each of the datasets given using 1000 resamples, construct a histogram and normal quantile plot to assess the normality of the bootstrap distribution. Do you expect the sampling distribution of the sample means to be close to normal?

Question 2.

**Bootstrap distribution of average IQ score.**

The distribution of the 60 IQ test scores in Table 1.3 is roughly normal and the sample size is large enough that we expect a Normal sampling distribution.

We first read in our data:

```
iqData = read.table("HW14IQ.R", header=TRUE)
attach(iqData)
```

Now that this is done, we bootstrap the sample mean for the IQ data using 1000 resamples by running the following code in R:

```r
bootMean = function(data, rep) {
  len = length(data)
  replicate(rep, mean(sample(data, len, replace=TRUE)))
}

simMeanIQ = bootMean(IQ, 1000)
hist(simMeanIQ, breaks=seq(min(simMeanIQ), max(simMeanIQ) + 1, 1), probability=TRUE)
```

Histogram of simMeanIQ
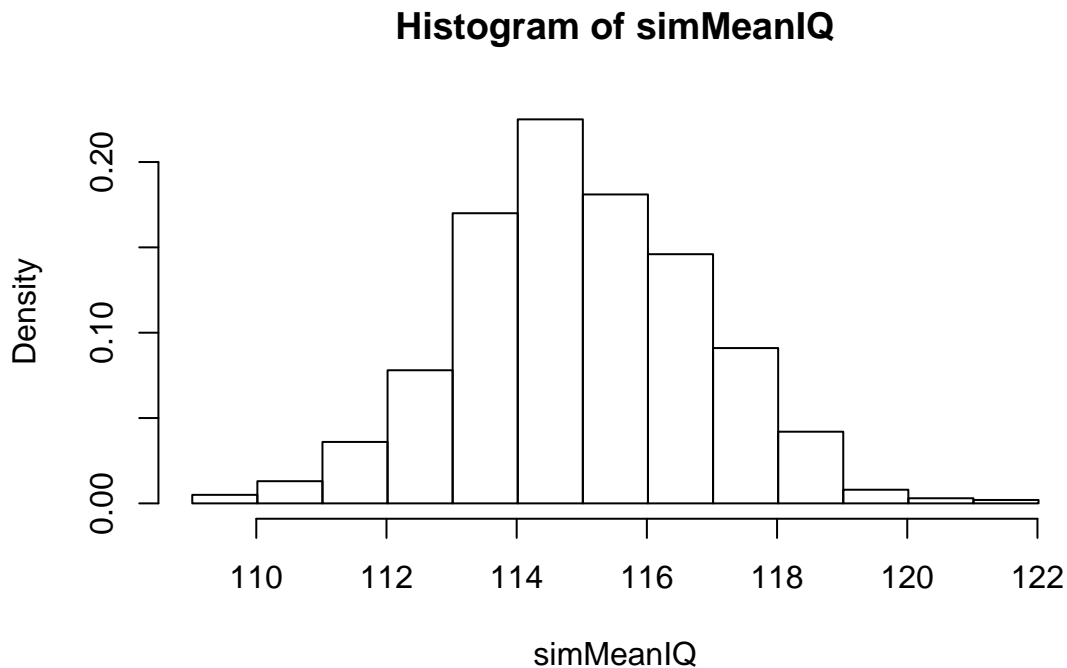
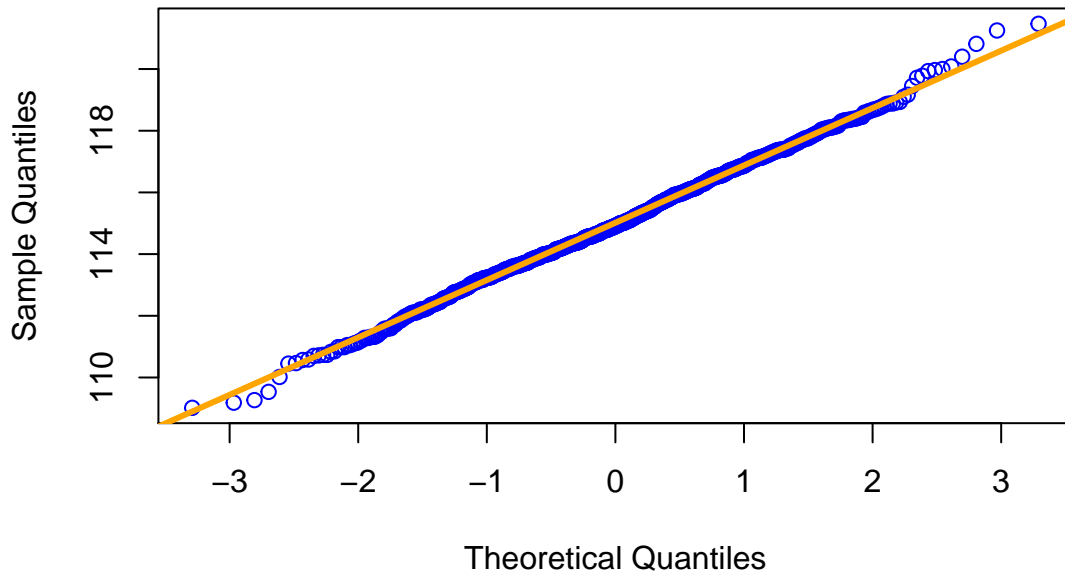Our plot here looks fairly normal. The mean is

```
[1] 114.9962
```

The standard deviation is

```
[1] 1.873184
```

We now produce a normal quantile plot to assess the normality of our bootstrapped IQ data by running the following code in R:

```r
qqnorm(simMeanIQ, col="blue")
qqline(simMeanIQ, col="orange", lwd=3)
```

Normal Q−Q Plot

This indicates that our bootstrapped data is pretty normally distributed.

We can double check this result by running a Shapiro-Wilk test:

```
    Shapiro-Wilk normality test

data:  simMeanIQ
W = 0.99833, p-value = 0.4518
```

Since the P-value here is larger than 0.05 by quite a bit, we can safely conclude that our bootstrapped IQ distribution is normal. Because it is normal, we can make the inference that our sample distribution is also normal.

Question 3.

Bootstrap distribution of average CO_2 emissions.

The distributiion of Carbon Dioxide emissions in Table 1.6 is strongly skewed to the right. The United States and several other contries appear to be high outliers.

we first read-in our data:

```
co2Data = read.table("HW14CO2.R", header=TRUE)
attach(co2Data)
```
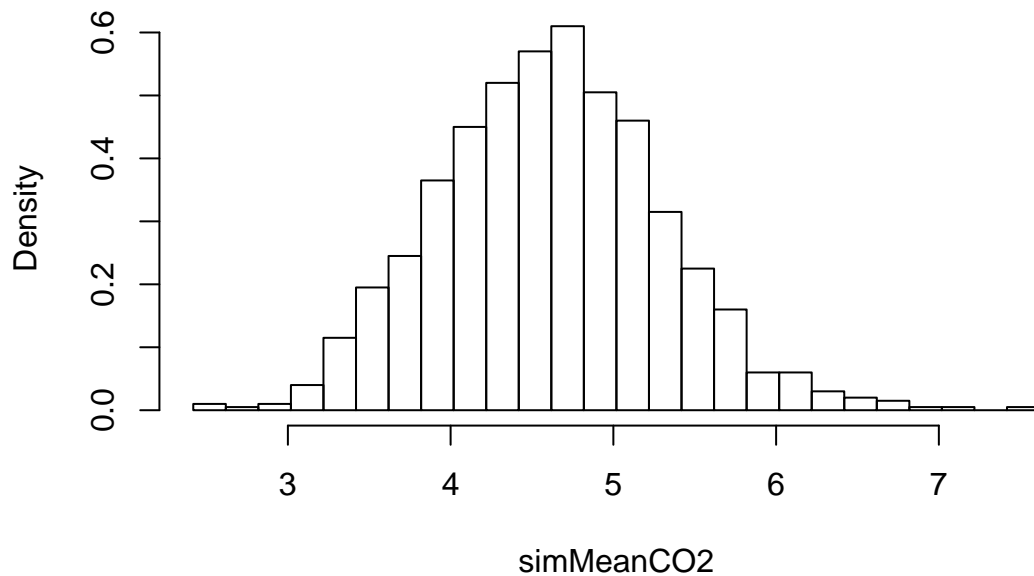
```
The following object is masked from package:datasets:

    CO2
```

We now bootstrap the sample means for the CO2 data by running the following code in R:

```
simMeanCO2 = bootMean(CO2, 1000)
hist(simMeanCO2, breaks=seq(min(simMeanCO2), max(simMeanCO2) + 0.2, 0.2), probability=TRUE)
```
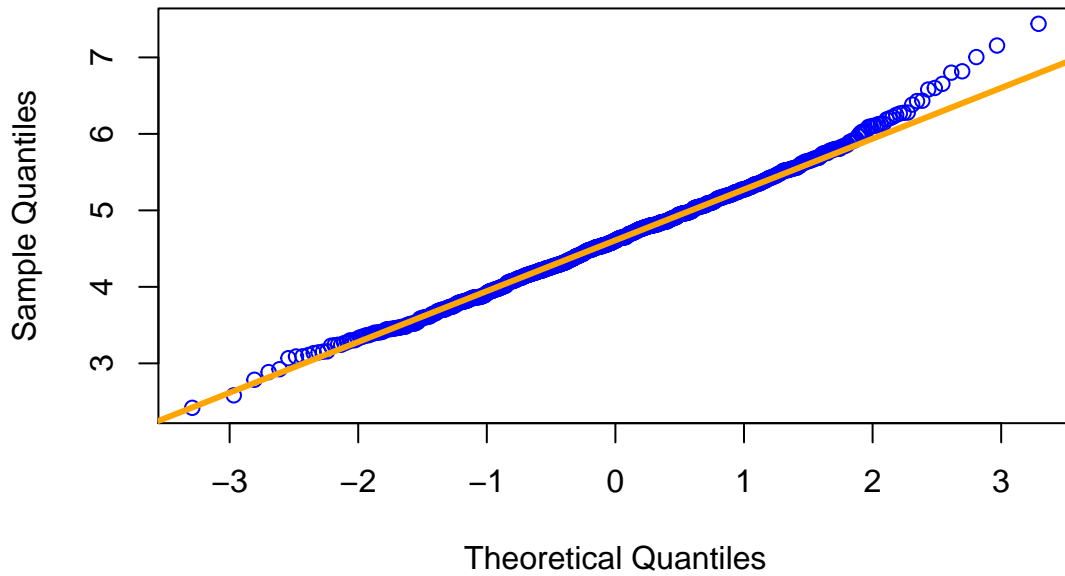
Histogram of simMeanCO2

This plot apprears somewhat skewed to the right, but still fairly normal. To get a better measure of normality, we make a normal quantile plot in R:

```
qqnorm(simMeanCO2, col="blue")
qqline(simMeanCO2, col="orange", lwd=3)
```

Normal Q–Q Plot

This still looks pretty normal. Just to be safe we can double check this by running a Shapiro-Wilk normality test:

```
	Shapiro-Wilk normality test

data:  simMeanCO2
W = 0.99548, p-value = 0.00476
```

With just 1000 repititions this P-value varies wildly with each run of our code. Most of the P-values I have seen while running this code have been much smaller than 0.05. Thus I believe this test is indicates that our bootstrapped mean distribution is not normal.

Based on the appearance of normality from our histogram and quantile plot, I would say that the normality of our boostrapped mean distribution implies that our sample distribution is fairly normal as well. However, with the inconsistency from the Shapiro-Wilk test I cannot be very certain of this.