

Christina Wang

Professor Han

Econ 108

4 October 2022

Problem Set 1

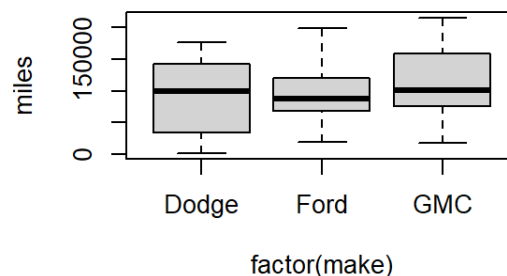
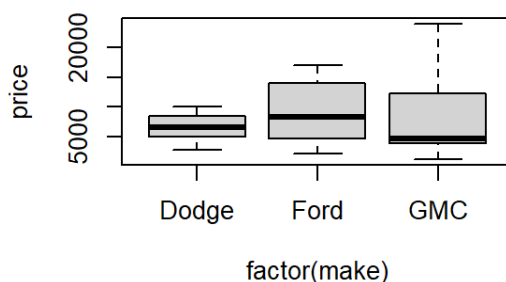
- 1) Frequentist uncertainty is characterized by the thought experiment: “If I were able to see a new sample of data generated by the same processes and scenarios as my current data, how would my estimates change?” This is also known as the thought experiment of repeated experiments.

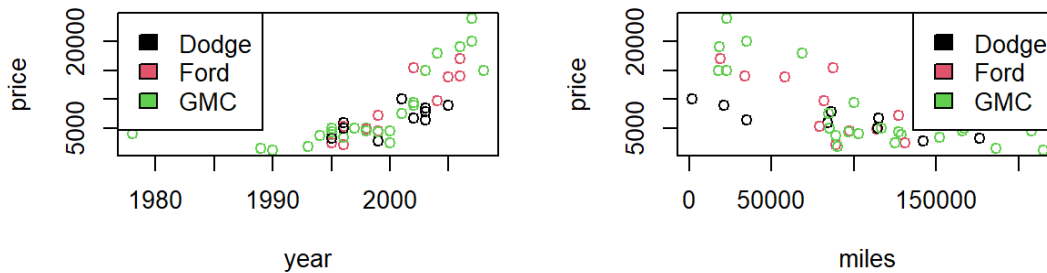
Bayesian uncertainty is characterized by inferences about parameters and the hypotheses are updated as evidence accumulates.

I prefer the school of frequentist uncertainty because it relies on existing data and not inferences.

- 2) The Monte Carlo simulation can be useful for the study of frequentist uncertainty because, given that frequentist uncertainty is characterized by repeated experiments and that in a Monte Carlo simulation, we mimic the data generating process as many times as we can, the Monte Carlo simulation effectively brings the repeated experiments of frequentist uncertainty theory to reality.
- 3) In a Monte Carlo simulation, you mimic the data generating process as many times as you can. In the bootstrap method, you use resampling from your current sample to mimic the sampling distribution. The Monte Carlo simulation is useful for properly evaluating bias. The bootstrap method is also good for reducing bias, but can approximate sampling distributions and construct confidence intervals. Essentially the main difference is that in bootstrapping, you use the original initial sample while the Monte Carlo simulation is based in generating new data.

4)





a. Box plot

```
par(mfrow=c(2,2))
plot(price ~ factor(make), data=trucks)
plot(miles ~ factor(make), data=trucks)
```

Despite the similarities in the syntax of the commands, the `factor()` function categorizes the makes of the trucks, grouping them together and facilitating the creation of the box plot.

Scatterplot

```
par(mfrow=c(2,2))
plot(price ~ year, data=trucks)
plot(price ~ miles, data=trucks)
```

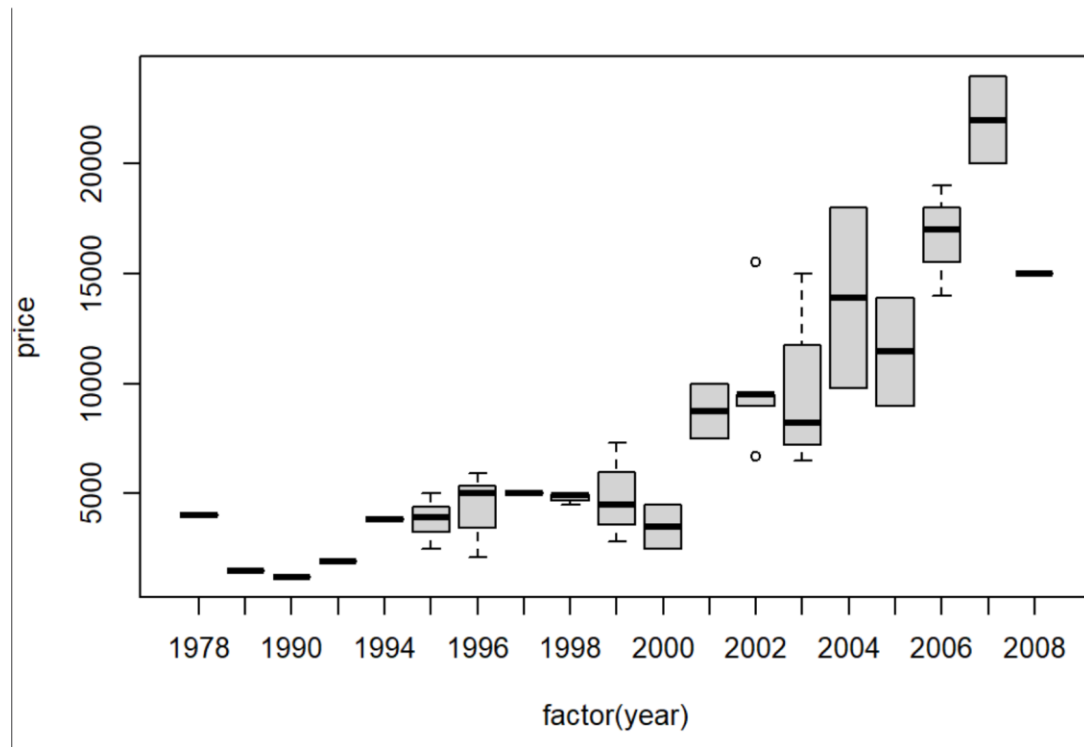
Whereas in the scatterplot, the trucks are not grouped together using the `factor` function; R plots the points individually, creating the scatterplot that we see.

```
trucks$make = factor(trucks$make)
plot(price ~ year, data=trucks, col=make)
plot(price ~ year, data=trucks, col=make)
legend("topleft", fill=1:3, legend=levels(trucks$make))
plot(price ~ miles, data=trucks, col=make)
legend("topright", fill=1:3, legend=levels(trucks$make))
```

The `col=make` colors the dots in based on make, allowing the viewer to see the plots of price against year and miles per make of truck, arguably giving scatterplot the ability to display a greater amount of information at once.

b.

```
par(mfrow=c(2,2))  
plot(price ~ factor(year), data=trucks)
```



5) a. EY_i : statistical notation for mean

$$Y_i = e^{(z_i)}$$

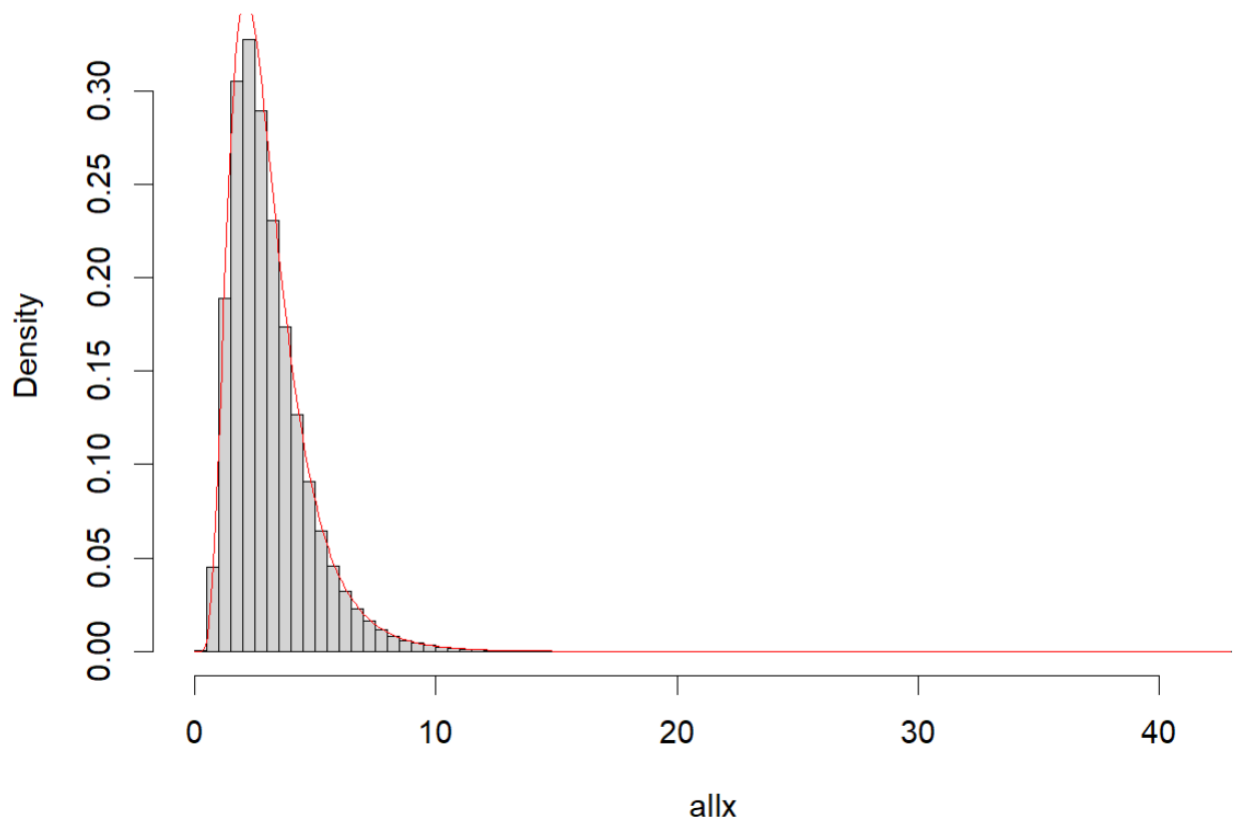
$$\text{Mean } \mu = 1$$

$$\text{SD } \sigma = .5$$

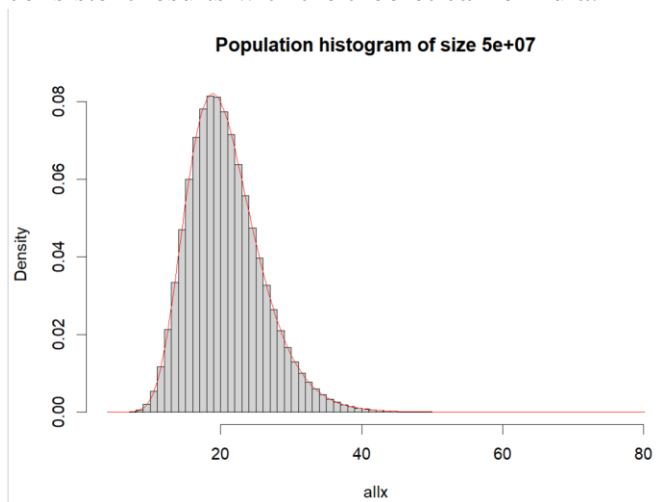
$$EY_i = Ee^{(z_i)} = e^{(\mu + ((\sigma^2)/2))} = e^{1.125} \text{ which is around } 3.08.$$

By looking at the histogram, this information appears consistent.

Population histogram of size 5e+07

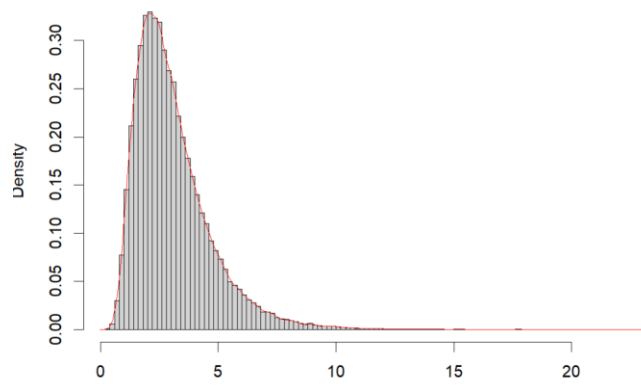


b. I tried Mean $\mu = 3$; SD $\sigma = .25$. $e^{(3 + (1/32))} = 20.72$. The graph below displays consistent results with the theoretical formula.

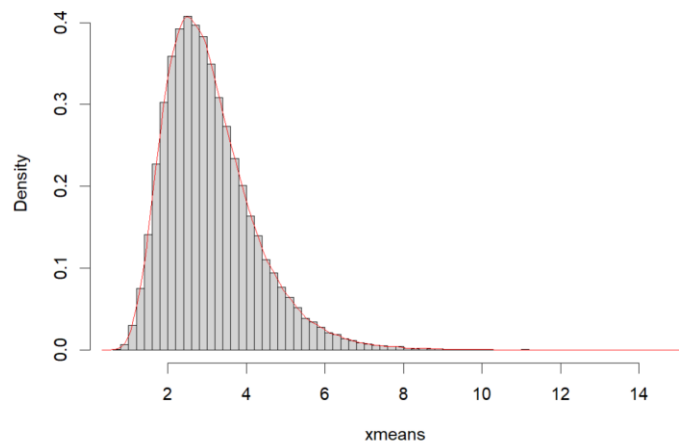


6)

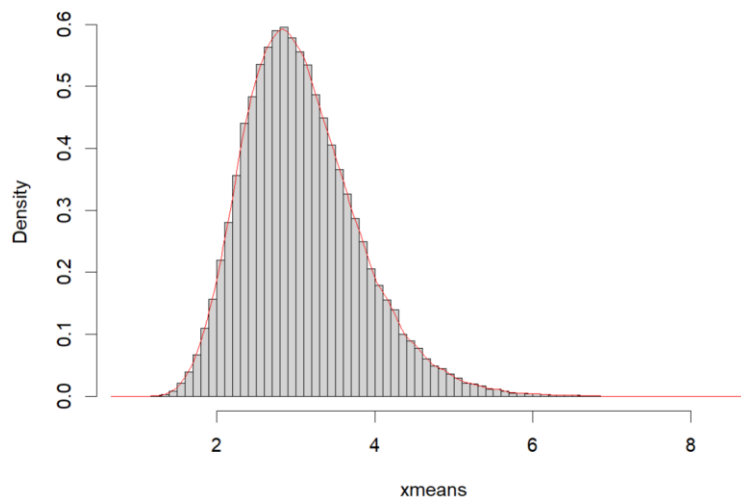
histogram of sample mean of size 1

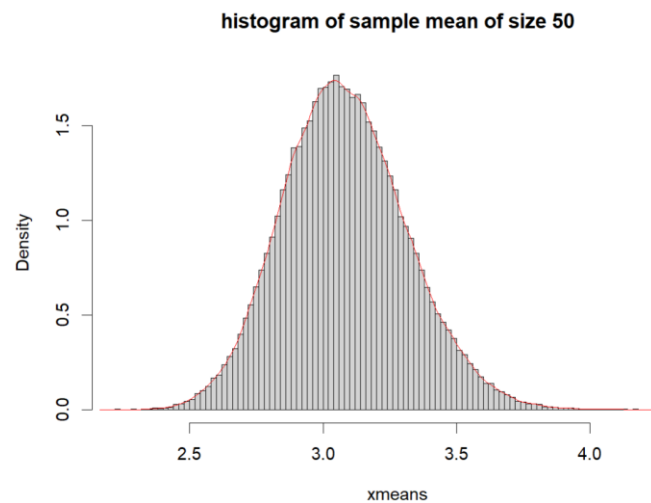
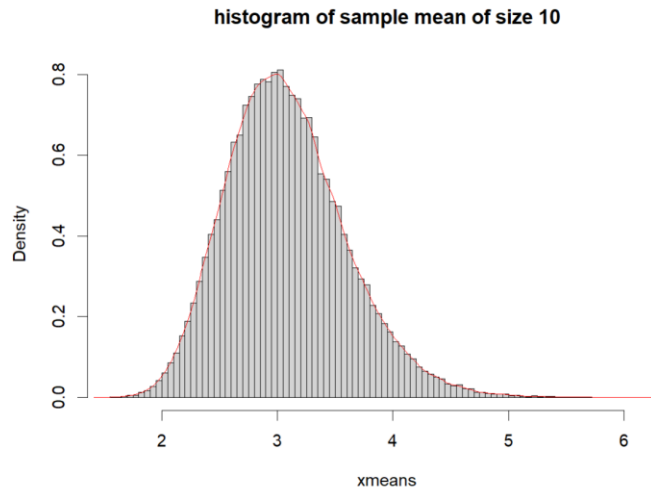


histogram of sample mean of size 2



histogram of sample mean of size 5





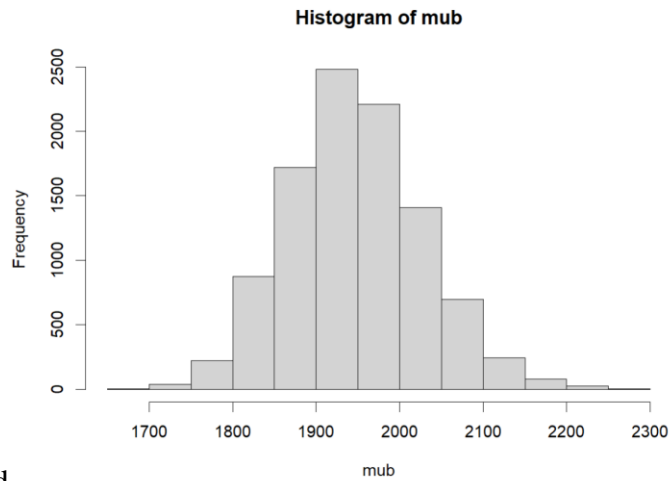
Central Limit Theorem states if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population, the distribution of the sample means will be approximately normally distributed. With smaller sizes, the histogram becomes more long tail, skewing right. With larger sizes, the histogram conforms to Central Limit Theorem and becomes more normally distributed.

7) a.

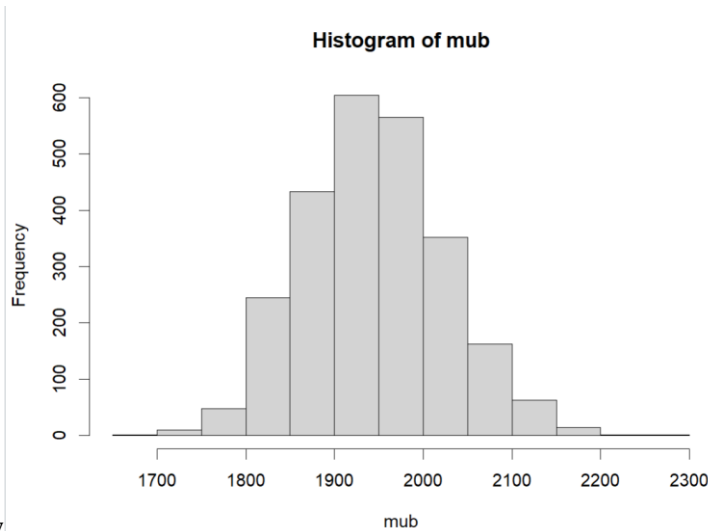
```
B <- 2500
mub <- c()
for (b in 1:B) {
  samp_b = sample.int(nrow(browser), replace=TRUE)
  mub <- c(mub, mean(browser$spend[samp_b]))
}
sd(mub)
```

b. The new bootstrap standard deviation estimate is not terribly different from the original bootstrap standard deviation estimate. The old estimate was 80.79163. The new estimate is 80.06002.

c.



old



new

The histogram shapes and skews are similar. The difference only lies in the scale of frequency, with the highest peaks of the new frequency reaching 600 while the old peaks reach 2500.

d. Central Limit Theorem states if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population, the distribution of the sample means will be approximately normally distributed. Because the sample sizes are still so large, even if one is a fraction of the other, both histograms are normally distributed. The new histogram's frequency is limited due to the smaller

population size. The standard deviation does not change because it measures how variable the population itself is and that has not changed.