Math 300 Lesson 22 Notes

CLT, Normal Distribution

YOUR NAME HERE

June, 2022

Contents

Objectives

1. Explain the central limit theorem.
2. For a normal distribution, use `R` to find probabilities and percentiles.

Reading

Chapter 7.4

Appendix A.2

Lesson

Work through the learning checks LCA.1 - LCA.2.

- The central limit theorem is difficult. The key idea is that if we create a point estimate that is using summation, then the **distribution of this point estimate** tends to be normal. The larger he sample, the closer the distribution is to being normal. If we are finding a mean, then the larger the sample, the smaller the standard error.
- Use `R` to find probabilities. This implies the use of `pnorm()` and `qnorm()`.

Let's do a simulation.

Suppose we record the time of arrival of 4 friends.

```
library(tidyverse)
```

```
set.seed(316)
(time<-round(runif(4,5,6),2))
```

```
## [1] 5.33 5.74 5.29 5.37
```

We are generating 4 random numbers in the interval 5 to 6. We record these and find the average.

```
mean(time)
```

```
## [1] 5.4325
```

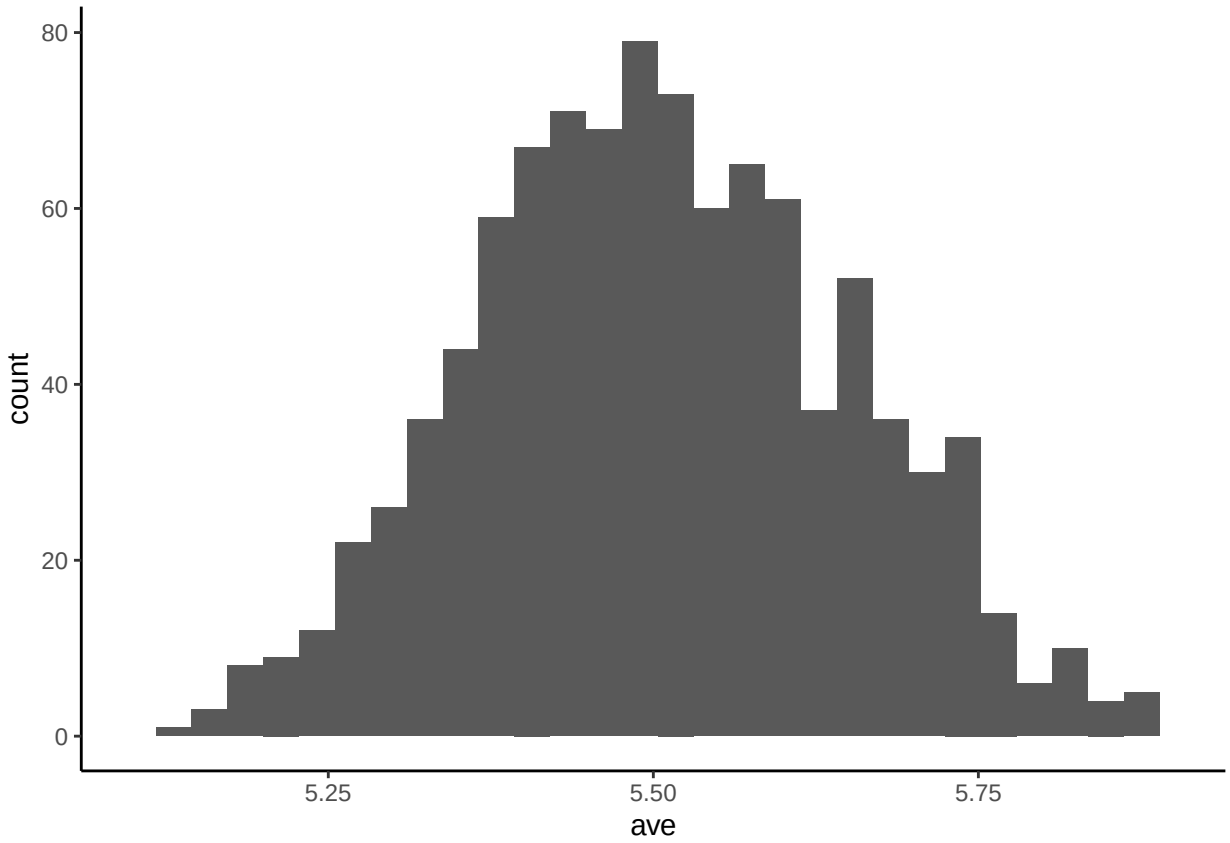
This is just one value. What is the distribution of the mean if we were to repeat this? Since this is a simulation, we can do this many, many times and plot the distribution of the mean. Don't worry about the code for doing this.

```
# Repeat this 1000 times
rep_data<-tibble(ave=replicate(1000,mean(round(runif(4,5,6),2))))
```

```
summary(rep_data)
```

```
##       ave
##  Min.   :5.095
##  1st Qu.:5.402
##  Median :5.500
##  Mean   :5.508
##  3rd Qu.:5.603
##  Max.   :5.928
```

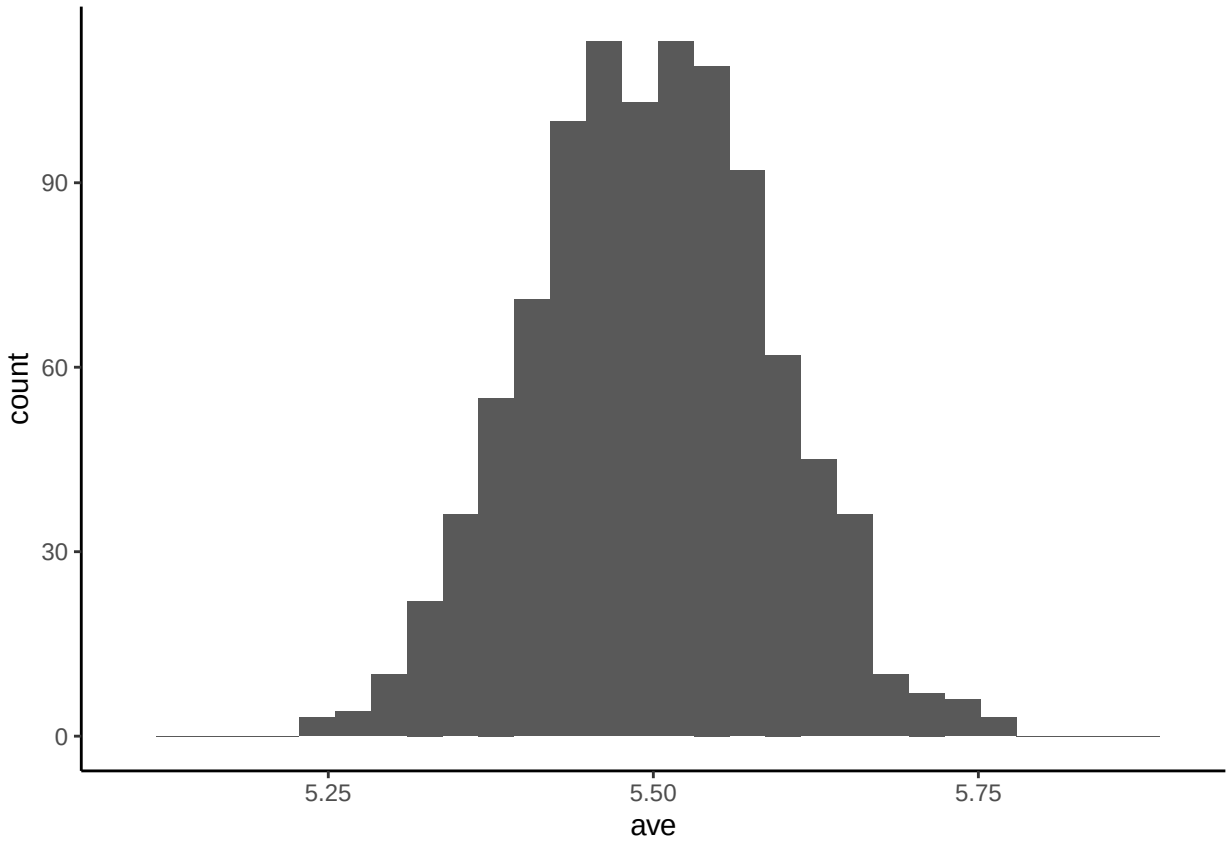
```
ggplot(rep_data,aes(x=ave)) +
  geom_histogram() +
  xlim(5.1,5.9) +
  theme_classic()
```



Notice how this distribution of is approximately normal. That is the central limit theorem. If we used 10 points per sample instead of 4, the standard error would be smaller.

```
# Repeat this 1000 times
rep_data2<-tibble(ave=replicate(1000,mean(round(runif(10,5,6),2))))
```

```
ggplot(rep_data2,aes(x=ave)) +
  geom_histogram() +
  xlim(5.1,5.9) +
  theme_classic()
```



- Run the code again, but change the sample size.

LC A.1 (Objective 2)

(LCA.1) What proportion of the area under the normal curve is less than 3? Greater than 12? Between 0 and 12?

Solution:

LC A.2 (Objective 2)

(LCA.1) What is the 2.5th percentile of the area under the normal curve? The 97.5th percentile? The 100th percentile?

Solution:

Documenting software

- File creation date: 2022-06-10
- R version 4.1.3 (2022-03-10)
- `tidyverse` package version: 1.3.1