R Markdown and Basic LATEX

Ethan Tsao

2024-10-01

Introduction

Lab 1

Slide 6

```
x <- c(10, 12, 5, 14, 25)
print(x)
```

```
## [1] 10 12  5 14 25
```

```
mean(x)
```

```
## [1] 13.2
```

```
var(x)
```

```
## [1] 54.7
```

The mean of `x` is 13.2.

Simulations

We will write a for loop to do the following: - Use the sample function to sample from x five times, WITH replacement - Find the mean of this new sample, and store it into the ith spot of a vector. - Do this 10 times, and print the result

Slide 10

```
set.seed(10012024)

reps <- 10
output <- NA # initialize an object to store results
```

```r
for (i in 1:reps){
  output[i] = mean(sample(x, size=5, replace = TRUE, prob = NULL))
}

print(output)
```

```
##  [1] 18.0 18.4 15.0 12.8 11.0 14.0 11.0 15.4 15.0 12.4
```

```r
replicate(n=10, sample(x, size=1))
```

```
##  [1] 14 12 10 25 25 25 14 25 10 14
```

Slide 17

```r
samp_3 <- function(x) {
  sample_elements <- sample(x, size = 3)
  sample_sd <- sd(sample_elements)
  return(sample_sd)
}

set.seed(10012024)

x <- 1:10
result <- samp_3(x)

print(result)
```

```
## [1] 2.516611
```

2) Read in data from an external file

```r
library(tidyverse)
dog_data_cleaned <- read_csv("dog_data_cleaned.csv")
```

```r
library(dplyr)
dog_pre <- dog_data_cleaned %>%
  filter(Stage == "pre")
```

```r
stats_tidy <- dog_pre %>%
  filter(!is.na(Age_Yrs)) %>%
  summarize(n = n(), x_bar=mean(Age_Yrs), sx=sd(Age_Yrs))

library(knitr)
kable(stats_tidy)
```

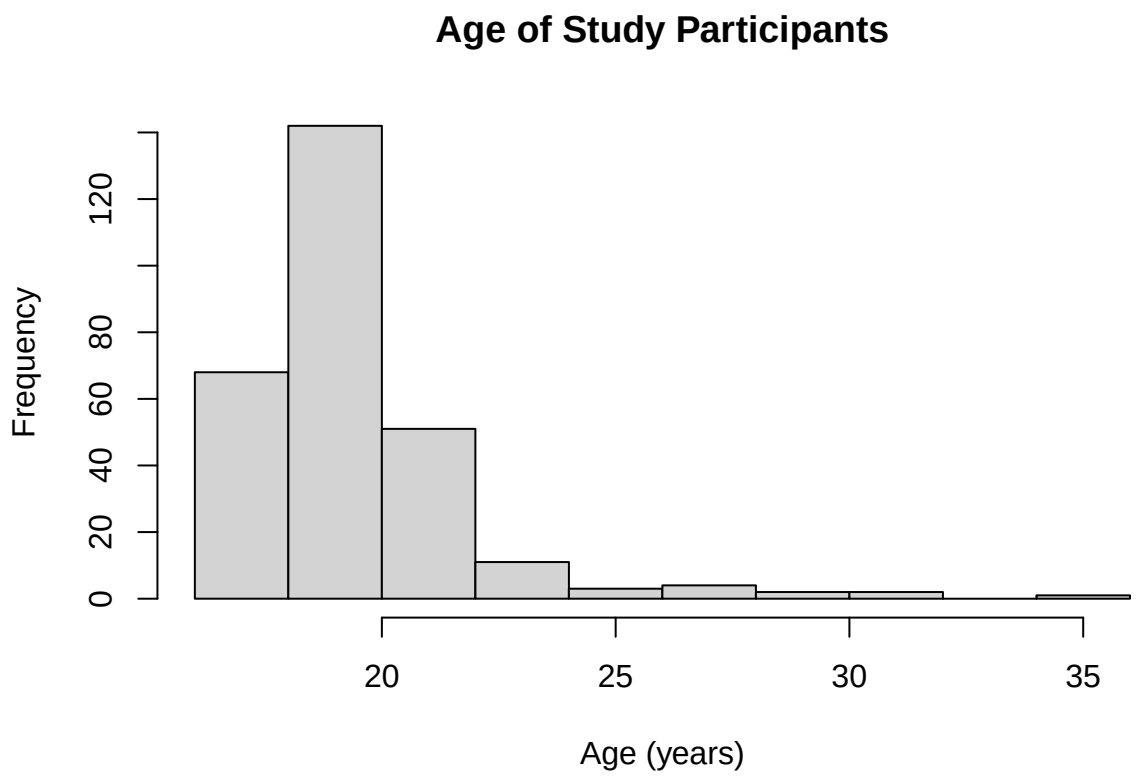
n	x_bar	sx
284	19.88732	2.381016

```
n <- sum(!is.na(dog_pre$Age_Yrs))
x_bar <- mean(dog_pre$Age_Yrs)
sx <- sd(dog_pre$Age_Yrs)

sum_stats <- matrix(c(n, x_bar, sx), nrow=1)
colnames(sum_stats) <- c("n", "mean", "st. dev")

kable(sum_stats)
```

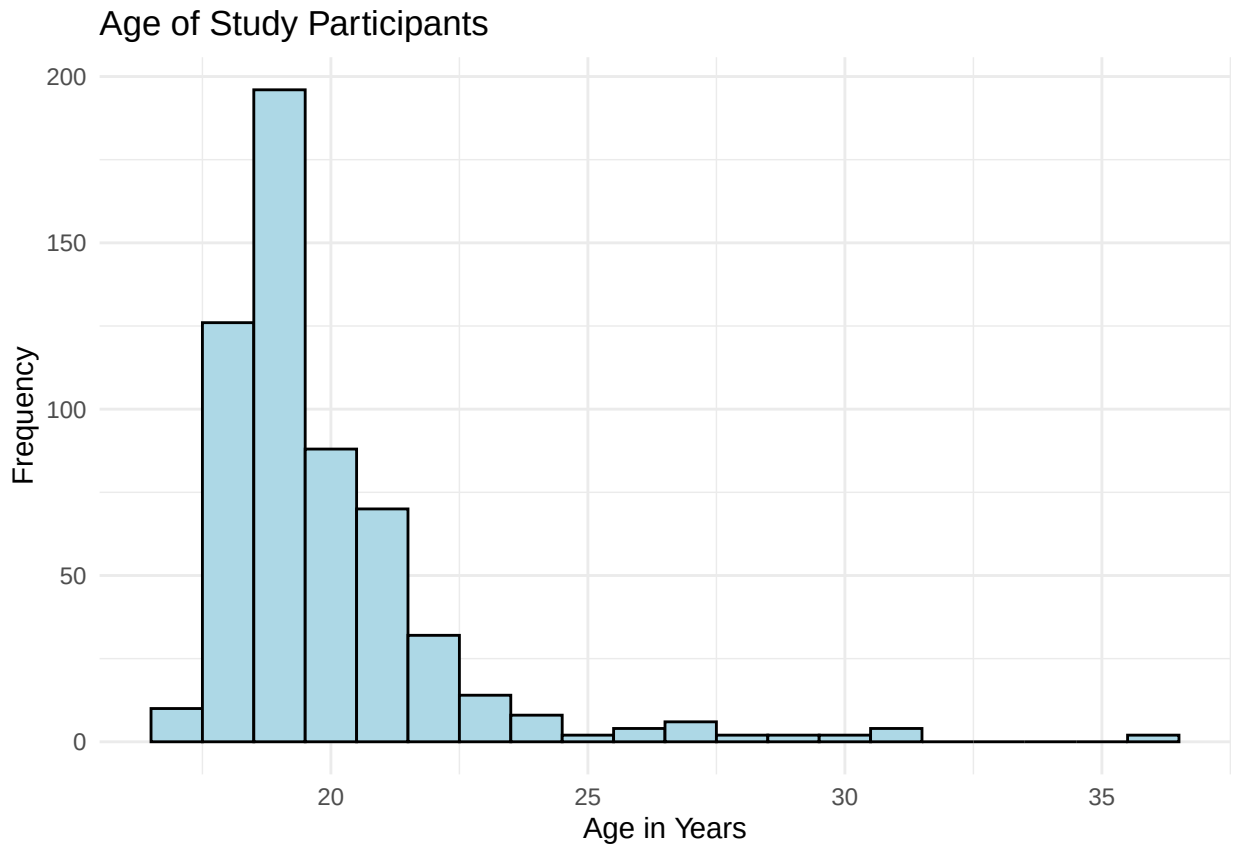
n	mean	st. dev
284	19.88732	2.381016

There are 284 observations, with a mean of 19.8873239 and a standard deviation of 2.3810162.

```
hist(dog_pre$Age_Yrs,
     xlab="Age (years)",
     main="Age of Study Participants")
```



```
ggplot(dog_data_cleaned, aes(x = Age_Yrs)) +
  geom_histogram(binwidth = 1, fill = "lightblue", color = "black") +
  labs(title = "Age of Study Participants",
       x = "Age in Years",
       y = "Frequency") +
  theme_minimal()
```



3) Repeat without the TA on another dataset

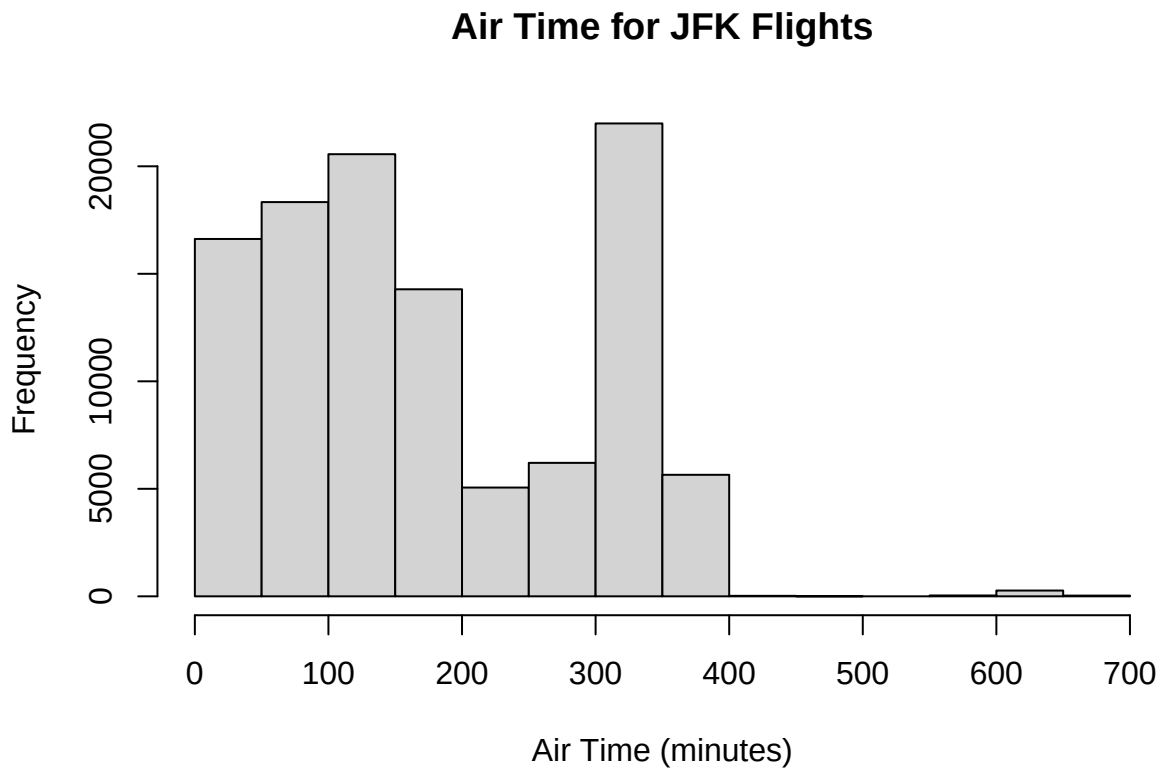
```
library(nycflights13)
```

```
library(dplyr)
jfk_flights <- flights %>%
  filter(origin == "JFK")
```

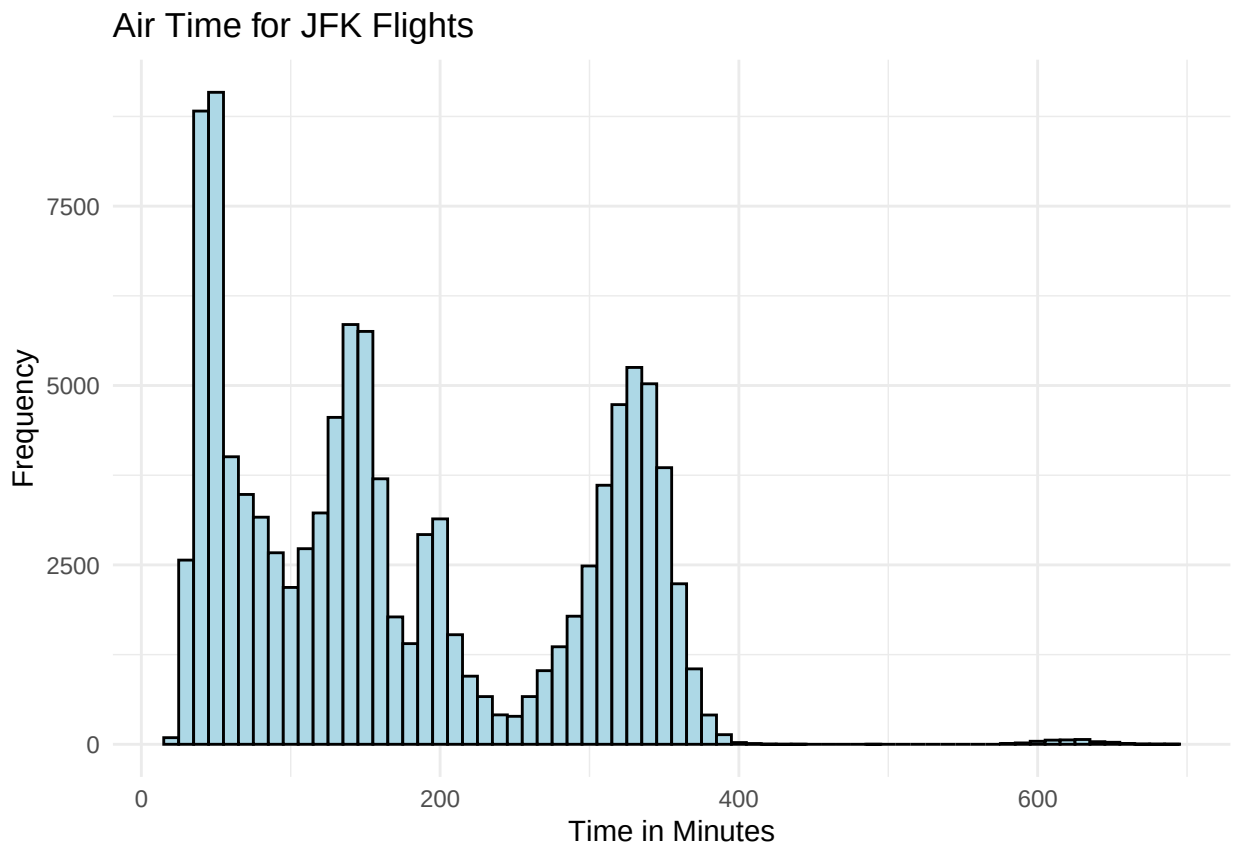
```
stats_tidy <- jfk_flights %>%
  filter(!is.na(air_time)) %>%
  summarize(n = n(), x_bar=mean(air_time), sx=sd(air_time))

library(knitr)
kable(stats_tidy)
```

n	x_bar	sx
109079	178.349	113.7943

```
hist(jfk_flights$air_time,
     xlab="Air Time (minutes)",
     main="Air Time for JFK Flights")
```




```
ggplot(jfk_flights %>% filter(!is.na(air_time)), aes(x = air_time)) +
  geom_histogram(binwidth = 10, fill = "lightblue", color = "black") +
  labs(title = "Air Time for JFK Flights",
       x = "Time in Minutes",
       y = "Frequency") +
  theme_minimal()
```



Basic LATEX coding

A simple linear regression model is $y = \beta_0 + \beta_1 x$

The formula for the sample standard deviation is:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

The null and alternative hypotheses are:

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_A : &\quad \mu \neq \mu_0 \end{aligned}$$