Predictions

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
# Loading in the tidyverse library and the Among Us Game Data
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.6     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.4     ✓ stringr 1.4.0
## ✓ readr   2.1.1     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
amongus <- readr::read_csv("User1.csv")
```

```
## Rows: 100 Columns: 13
```

```
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (13): Game Completed Date, Team, Outcome, Task Completed, All Tasks Comp...
```

```
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
amongus %>% glimpse()
```

```
## Rows: 100
## Columns: 13
## $ `Game Completed Date`        <chr> "12/13/2020 at 1:26:56 am EST", "12/13/20…
## $ Team                         <chr> "Crewmate", "Crewmate", "Crewmate", "Impo…
## $ Outcome                      <chr> "Win", "Loss", "Win", "Win", "Loss", "Los…
## $ `Task Completed`             <chr> "3", "7", "3", "-", "4", "7", "8", "8", "…
## $ `All Tasks Completed`        <chr> "No", "Yes", "No", "-", "No", "Yes", "Yes…
## $ Murdered                     <chr> "Yes", "No", "No", "-", "No", "Yes", "Yes…
## $ `Imposter Kills`             <chr> "-", "-", "-", "2", "-", "-", "-", "-", "…
## $ `Game Length`                <chr> "07m 04s", "16m 21s", "11m 33s", "08m 05s…
## $ Ejected                      <chr> "No", "No", "No", "No", "No", "No", "No",…
## $ `Sabotages Fixed`            <chr> "2", "1", "0", "N/A", "0", "0", "1", "0",…
## $ `Time to complete all tasks` <chr> "-", "09m 48s", "-", "-", "-", "12m 16s",…
## $ `Rank Change`                <chr> "++", "--", "++", "+++", "---", "--", "--…
## $ `Region/Game Code`           <chr> "NA / WYMSBP", "NA / WYMSBP", "NA / WYMSB…
```

```
# Retrieving important values from the data that will help us analyze and make predictions
# Determining the total number of games and the number of wins and losses

total <- amongus %>% count()
total[[1]]
```

```
## [1] 100
```

```
Outcome <- amongus %>% group_by(Outcome)
mogus_tibble <- tibble(Outcome)
mogus_tibble %>% group_by(Outcome) %>% summarize(count=n())
```

```
## # A tibble: 2 × 2
##   Outcome count
##   <chr>   <int>
## 1 Loss       42
## 2 Win        58
```

```
# Further visualizing the predictions
# Using the sample function to select random samples from the data and find similar probabilities for winning a g
ame of Among Us based on the data

mogus_tibble %>%
  sample_n(size = total[[1]], replace = TRUE) %>%
  group_by(Outcome) %>%
  filter(Outcome == "Win") %>%
  summarize(n(), p_win = n()/total[[1]])
```

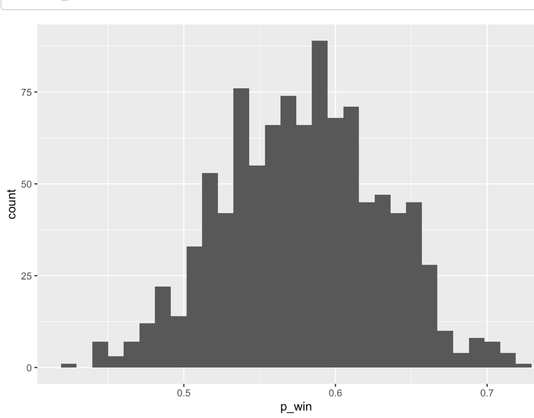
```
## # A tibble: 1 × 3
##   Outcome `n()` p_win
##   <chr>   <int> <dbl>
## 1 Win        62  0.62
```

```
# Estimating the sampling distribution of probability to win the next game of Among Us based on the data (wins/lo
sses of previous games)
# The sampling distribution is the distribution of all values that the probability of winning a game could be for
random samples from the data
# To estimate the sampling distribution, I took 1000 values from the probabilty of winning and calculated 1000 ra
ndoms samples from the population (main data)

sample_pwin <- rep(NA, 1000)

for(i in 1:1000){
  sample_100 <- mogus_tibble %>% sample_n(size=total[[1]], replace = TRUE)
  sample_pwin[i] <-
    sample_100 %>%
    filter(Outcome == "Win") %>%
    summarise(n()/total[[1]]) %>%
    as.numeric()
}

sample_pwin <- tibble(p_win = (sample_pwin)) %>% glimpse()
```

```
## Rows: 1,000
## Columns: 1
## $ p_win <dbl> 0.56, 0.57, 0.62, 0.63, 0.56, 0.55, 0.53, 0.62, 0.58, 0.63, 0.58…
```

```
# This is a histogram displaying the sampling distribution of the probability of winning a game of Among Us
# The graph is normal distribution
# Also calculated the confidence interval of the sampling distribution, where we can estimate with a 95% confiden
ce what the probability of winning the next game of Among Us

ggplot(sample_pwin, aes(x = p_win)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
quantile_tibble <- quantile(sample_pwin$p_win, c(0.025, 0.975))
quantile_tibble
```

```
##  2.5% 97.5%
##  0.48  0.67
```