1.Bamba Cisse, Jason Seda, Ava Droznik, Marwan Kenawy

2.Lab 2 Testing R:

how\_many\_rolls <- 1000

sim\_rolls <- sample(1:6, how\_many\_rolls, replace = TRUE)

# for first time

lots\_of\_sim\_rolls <- sample(1:6,how\_many\_rolls, replace = TRUE)

# do 49 more simulations

for (indx in 1:99) {

sim\_rolls <- sample(1:6,how\_many\_rolls, replace = TRUE)

lots\_of\_sim\_rolls <- data.frame(lots\_of\_sim\_rolls,sim\_rolls)

}

six\_counts <- colSums(lots\_of\_sim\_rolls == 6)

# Plot histogram of 6's per simulation

hist(six\_counts,

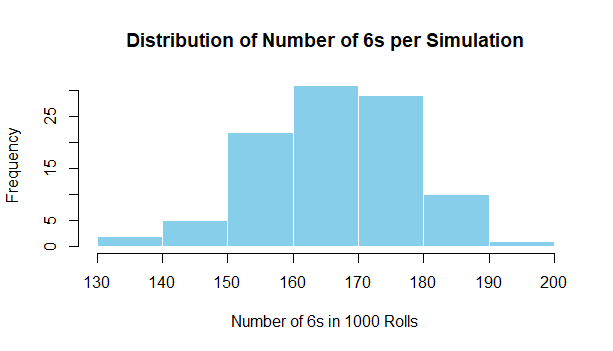
main = "Distribution of Number of 6s per Simulation",

xlab = "Number of 6s in 1000 Rolls",

ylab = "Frequency",

col = "skyblue",

border = "white")



\*No Chance to discuss Final Outcome\*- ran into syntax errors to create testing result-

3. Hypothesis testing:

H0: We will Have an uneven distribution of songs played due to limitation on number of play throughs, ie. # of songs listened to < or >1

H1: We will have an even distribution of songs played across each cycle, ie. #of songs=1

Data seen:

Standard Deviation from sample size= 1.348

Mean=1

N=100

Control in place: average hypothesis testing: # of songs if played through via 1000 times(generated).

H0: N times average= 10

H1: N times average is not =10.

\*\*DATA SET SEE EXCEL\*\*

Overview and Conclusion:

Null Hypothesis (H₀) and Expected Value  
The core of this investigation is a statistical Null Hypothesis, which represents the default position of no effect or no bias.

* Formal Definition: H1 states that the probability of the subject song being played is equal to that of any other song in the set. If the total number of songs is N, the probability (p) for the subject song on any single play is:  
  p = 1/N
* Predicted Outcome (μ): The expected average number of times the song is played over a series of trials is derived from this probability:  
  μ = (Total Trials) × (p) = (Total Trials) × (1/N)
* Control Validation: A control data set confirmed that the empirical average aligns with this prediction, yielding an average of 10 plays per session, thus solidifying μ = 10 as the benchmark for H1.

Experimental Results and Interpretation  
The observed data from the experimental trial was compared to the expected value of 10.

* Observed Outcome: The experimental result was not equal to exactly 1 (i.e., it was either less than or greater than 1).
* Statistical Agreement with H1: Crucially, this deviation is not evidence against the Null Hypothesis. Probability theory dictates that in any finite sample, it is highly unlikely to hit the expected value precisely due to inherent random variation.
* Conclusion: The observed deviation was small enough to be attributable to normal chance variation within the sample. Therefore, we fail to reject the Null Hypothesis. The data does not provide statistically significant evidence to suggest the song selection is biased.(N variation)

The Complication of Sample Size  
The most significant factor affecting the confidence and precision of this conclusion is the sample size.

* The Principle: Law of Large Numbers  
  This fundamental law of probability states that as the number of experimental trials (sample size) increases, the sample average will converge toward the true theoretical average (the expected value, μ).
* Application to This Experiment:
  + Current Sample (100 trials): The observed average is a reasonable estimate of the true probability but is subject to noticeable random fluctuation. A result of 5 or 7 is plausible and not considered significant for this sample size.
  + Larger Sample (1000 trials): Repeating the experiment with 1,000 trials would produce a sample average much closer to the expected value of 10. The larger sample dilutes the effect of random outliers, providing a more precise and reliable measure of the true probability.
* Why Sample Size Matters:
  + Precision: Larger samples yield more precise estimates, narrowing the confidence interval around the average.
  + Power: A larger sample size increases the statistical power of the test. This means it becomes easier to detect a true deviation from H1, if one actually exists. A small effect (e.g., a true average of 10.2) might be invisible in a small sample but detectable in a large one.

Final Conclusion  
The experiment's results are consistent with a random song selection process. The initial finding of a deviation from the expected value of 10 is a typical product of random sampling in a limited data set and does not invalidate the Null Hypothesis. To enhance the robustness and precision of this finding, future research should employ a significantly larger sample size, which would allow the Law of Large Numbers to ensure the result aligns more closely with the underlying probability.

\*\*More Testing is needed with an enlarged sample size for active test\*\*