



ALY 6000
Intro To Analytics
Counting and Probability problems
“Assignment-5”

PROFESSOR – ROY WADA

Submitted By

ITI ROHILLA

On

Tue Oct 24

OVERVIEW

We gained knowledge of probability, permutation, combination, and the concept probability with and without replacement through this assignment. In this assignment we perform various data analysis and calculations based on a dataset containing information about balls. Each ball in the dataset is represented by a color (red, blue, green, or yellow) and a label (A, B, C, D, or E). The key actions we took for this assignment are listed below.

1)Data Loading

2)Frequency Tables

3)Data Visualization

4)Probability Calculations

5)Factorial Function

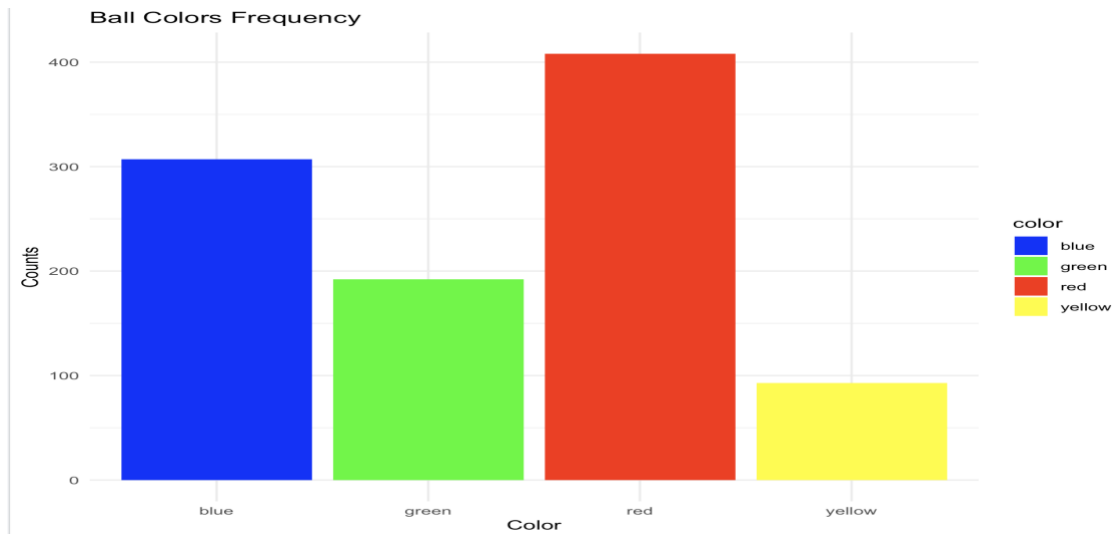
6)Coin Outcomes

7)Probability Distribution.

8)Soccer Game Probability Calculations.

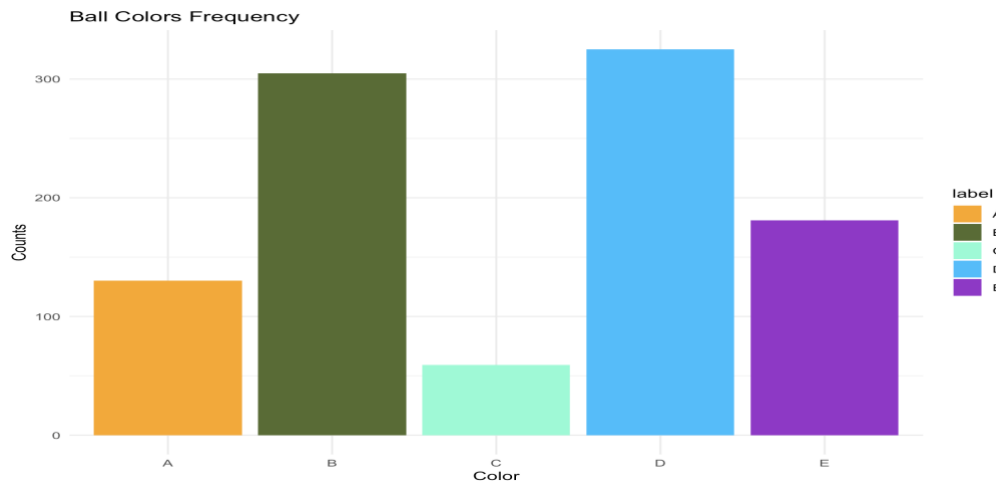
Q4.

A4. The 'ggplot2' library is used in this code to create a bar chart. It loads the library before defining 'color_chart' as a plot that makes use of the 'freq_color' data frame. Each bar is filled in accordance with its color using the 'aes' function, which maps the 'color' variable to the x-axis and 'counts' to the y-axis. 'geom_bar(stat = "identity")' adds bars to the plot, where 'stat = "identity"' denotes that the data values correspond to the bar heights. The chart's title and axis labels are set by the 'labs' function. 'theme_minimal()' alters the plot's appearance to be minimal, and 'scale_fill_manual' manually assigns colors to different categories in the 'color' variable to make sure that each color appears in the appropriate hue. The outcome is an eye-catching bar chart showing the frequency of various ball colors.



Q5.

A5. The 'ggplot2' library is used in this code to generate a bar chart. The 'freq_label' data frame is used to initialize the 'label_chart' plot as a plot. The 'aes' function in this plot maps the 'label' variable to the x-axis and the 'counts' variable to the y-axis, with bars filled according to their corresponding labels. Adding bars to the plot with the command "geom_bar(stat = "identity")" indicates that the data values correspond to the bar heights (hence, "identity"). The chart's title and axis labels are set by the 'labs' function. The plot has a neat and minimalistic appearance thanks to the use of "theme_minimal()". The last option, "scale_fill_manual," manually assigns colors to the "A," "B," "C," "D," and "E" label categories. This code creates a bar chart with a specific color for each label to show the frequency of each label category.



Q14. [Challenge]

A

```
> filtered_A <- ball[ball$label == "A",]
```

```
> nrow(filtered_A)
```

```
[1] 130
```

```
> filtered_E <- ball[ball$label == "E",]
```

```
> nrow(filtered_E)
```

```
[1] 181
```

```
> filtered_C <- ball[ball$label == "C",]
```

```
> nrow(filtered_C)
```

```
[1] 59
```

```
> prob14_result <- nrow(filtered_A)*nrow(filtered_E)*nrow(filtered_C)
```

The 'label' column in the 'ball' dataset serves as the basis for the specific filtering and counting operations covered by this code segment. The dataset is first filtered to produce three distinct subsets: one for records with label "A," another for records with label "E," and a third for records with label "C." The code uses `nrow()` to determine the number of rows in each filtered subset after each

filter. It then multiplies the counts of the records in the "A," "E," and "C" subsets to arrive at prob14_result. This code basically counts the records with the labels "A," "E," and "C" separately, and then adds the counts together, probably for a counting or probability scenario, as indicated by the variable name.

Q 16 . [Challenge]

A

```
> green_counts <- freq_color[freq_color$color == "green", "counts"]  
  
> prob16_result <- (green_counts/Total)*(green_counts-1)/(Total-1)*(green_counts-2)/(Total-2)  
  
> prob16_result
```

[1] 0.006988632

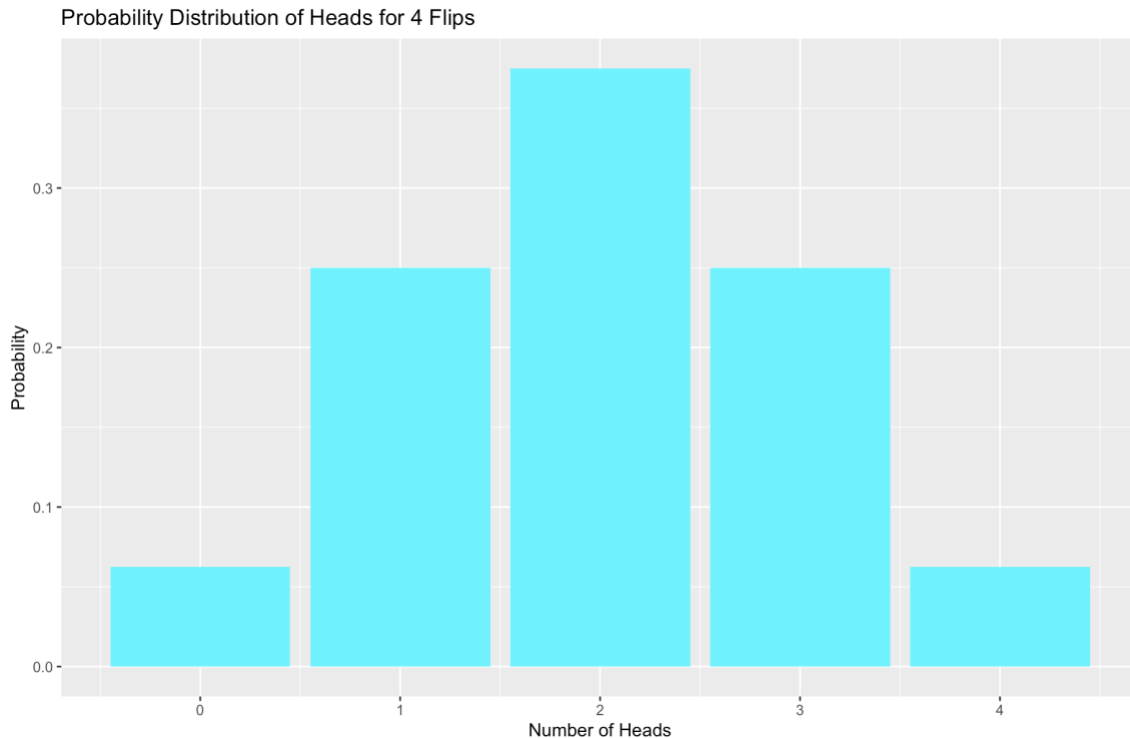
The main goal of this program is to determine the likelihood of a particular event connected to the color "green" in the dataset for "balls." The count of instances of the "green" color is first extracted from the "freq_color" data frame and placed in the variable "green_counts." Then, it multiplies a number of times to determine the "prob16_result." In essence, the expression determines the likelihood that a "green" ball will be drawn in a series of three draws without a replacement. With each draw, the number of "green" balls gets smaller, so it subtracts one from the total to account for this. From a finite pool of balls with various colors, this code seeks to determine the likelihood of drawing three consecutive "green" balls.

Q24. Bar chart- Probability Distribution of the Results of four-coin flips

A24.

- This program creates a bar graph representing the probability distribution of the results of four coin flips. It begins by setting 'n_flips,' which is the number of coin flips, to 4. Using the binomial probability distribution function "dbinom," it then determines the odds of getting any of the possible numbers of heads (0 to 4). The 'probabilities' variable contains the calculated probabilities.
- The next step is to create a data frame called "data" that has two columns: "Heads" for the number of heads (from 0 to 4), and "Probability" for the probabilities associated with each head.
- The code generates a bar chart called "bar_chart" with "Heads" on the x-axis and "Probability" on the y-axis using the "ggplot2" library. In order to specify that the data represents the actual bar heights, the geom_bar function is used to build the bars. It also changes the color of the bar to "turquoise1."

- The y-axis label reads "Probability," the x-axis label "Number of Heads," and the chart's title reads "Probability Distribution of Heads for 4 Flips." The code then uses the expression `print(bar_chart)` to output the resultant bar chart. This code displays the probability distribution of how many times a coin will land on its head after four flips, with each potential result's corresponding probability being represented by a bar on the chart.



CITATIONS

- Permutation and Combination Concept

<https://www.britannica.com/science/permutation>

- Create a Data Frame from All Combinations of Factor Variables

<https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/expand.grid>

- A Guide to dbinom, pbinom, qbinom, and rbinom in R

<https://www.geeksforgeeks.org/a-guide-to-dbinom-pbinom-qbinom-and-rbinom-in-r/>

- Probability: Independent and Mutually Exclusive Events

https://stats.libretexts.org/Courses/Los_Angeles_City_College/Introductory_Statistics/03%3A_A_Probability_Topics/3.03%3A_Independent_and_Mutually_Exclusive_Events#:~:text=When%20sampling%20is%20done%20with,may%20be%20chosen%20only%20once.