Problem Set 1

Jonathan A. Pedroza, PhD

2022-02-01

## Warning: package 'tidyverse' was built under R version 4.0.5

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.1.1 v dplyr 1.0.5  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.4

## Warning: package 'tibble' was built under R version 4.0.5

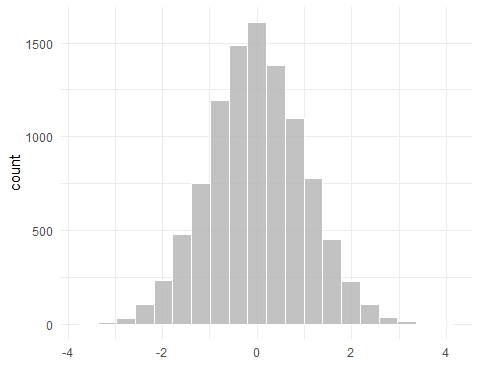
## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'forcats' was built under R version 4.0.5

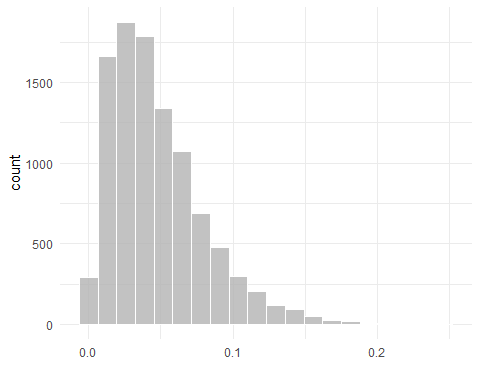
## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

1. What type of distribution is below?



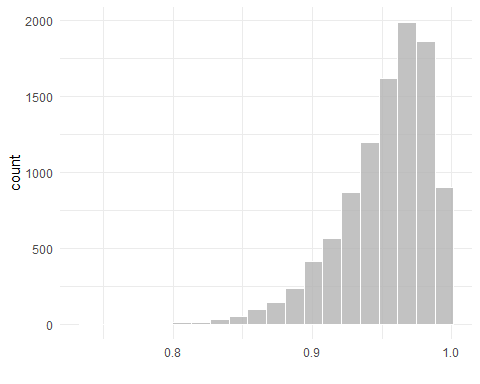
**Answer:** Normal distribution. It looks perfectly symmetrical and has the highest point at the middle with two even tails.

1. What type of distribution is below?



**Answer:** Positively skewed distribution. It has a high frequency of low scores and a low frequency of higher scores.

1. What type of distribution is below?



**Answer:** Negatively skewed distribution. It has a high frequency of high scores and a low frequency of lower scores.

1. Describe what a bimodal distribution is in your own words.

**Answer:** A bimodal distribution is a distribution with two modes, or two most frequent points. It can often show that there are two distinct groups in your distribution.

1. What is the term for the formula below?

**Answer:** This is the formula for relative frequency. It is the frequency/count of a value divided by the number of observations.

1. Calculate the relative frequency for each category. The following data will be for numbers 6-8.**If using R/RStudio, you can copy+paste what is below and run that in R/RStudio to see the values.**

categories <- c('d', 'e', 'a', 'e', 'e', 'e',  
 'b', 'd', 'e', 'c', 'a', 'e',  
 'a', 'd', 'e', 'd', 'e', 'c',  
 'a', 'b', 'e', 'a', 'b', 'd',  
 'd', 'd', 'd', 'e', 'e', 'e')  
categories

## [1] "d" "e" "a" "e" "e" "e" "b" "d" "e" "c" "a" "e" "a" "d" "e" "d" "e" "c" "a"  
## [20] "b" "e" "a" "b" "d" "d" "d" "d" "e" "e" "e"

table(categories)

## categories  
## a b c d e   
## 5 3 2 8 12

**Answer:** 5/30 - A 3/30 - B 2/30 - C 8/30 - D 12/30 - E

1. Get the **proportions** & **percentages** of each category.

**Answer:**

5/30

## [1] 0.1666667

3/30

## [1] 0.1

2/30

## [1] 0.06666667

8/30

## [1] 0.2666667

12/30

## [1] 0.4

.17 - A .10 - B .07 - C .27 - D .40 - E

.17\*100

## [1] 17

.10\*100

## [1] 10

.07\*100

## [1] 7

.27\*100

## [1] 27

.40\*100

## [1] 40

17% A 10% B 07% C 27% D 40% E

1. Calculate the cumulative frequency of categories D and E combined.

**Answer:**

.27 - D .40 - E

.27 + .40

## [1] 0.67

D and E combined = .67 or 67% of the distribution

9.Calculate the median for the following values. The following data will be for numbers 9 & 10. **If using R/RStudio, you can copy+paste what is below and run that in R/RStudio to see the values.**

numbers <- c(3, 5, 7, 6, 4, 9, 1, 8, 4, 6, 6, 4, 5, 2, 3)  
numbers

## [1] 3 5 7 6 4 9 1 8 4 6 6 4 5 2 3

sort(numbers)

## [1] 1 2 3 3 4 4 4 5 5 6 6 6 7 8 9

**Answer:** 5 is the median, as it is the 8th value in order from smallest to largest.

1. Calculate the mean for the following values.

1+2+3+3+4+4+4+5+5+6+6+6+7+8+9

## [1] 73

# sum is 73

73/15

## [1] 4.866667

**Answer:** The mean/average value for the values provided was 4.87.