Problem Set 2

Jonathan A. Pedroza, PhD

2022-02-15

## 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 would be the closest definition for deviance?

A. The distance between the mean and each individual’s score B. The distance between the mean and the median C. The word for inappropriate numbers D. The average distance that scores vary from the mean

**Answer:** A

1. What is the difference between variance and standard deviation? What is similar between variance and standard deviation?

**Answer:** The difference between the two measures are that variance is in squared unit while standard deviation are in more standardized units. Both are measures of dispersion that are the average amount of spread/distance/dispersion/variation of values/scores from the mean.

1. If adding the deviance scores for each participant, what value *should* you get if you did everything correctly?

**Answer:** 0 because there will be values above the mean and below the mean. If adding up both values above and below the mean, you’ll always get zero.

1. What is the range useful for?

**Answer:** The range gives you the value from the lowest value/score subtracted from the highest value/score. While not the most important, the range provides the minimum and maximum values, which you can use to see whether you coded your variable correctly or calculating a variable incorrectly. It is also useful for finding outliers, if there is a value that is extremely high or low compared to the mean and standard deviation.

1. How does the range differ from the interquartile range?

**Answer:** The range provides the highest and lowest value. The interquartile range is the range from the 25 percentile to the 75 percentile.

1. Calculate the sum of deviations. If you’d like to recreate the table that was shown in class you will just fill in the blanks for mean, deviance, and for dev\_squared and just rerun everything every time you fill in more information in the table.

data <- c(9, 7, 8, 3, 5, 2, 4, 7, 4, 9)  
data

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

table\_create <- data.frame(values = c(9, 7, 8, 3, 5, 2, 4, 7, 4, 9),  
 mean = c('', '', '', '', '', '', '', '', '', ''),  
 deviance = c('', '', '', '', '', '', '', '', '', ''),  
 dev\_squared = c('', '', '', '', '', '', '', '', '', ''))   
table\_create

## values mean deviance dev\_squared  
## 1 9   
## 2 7   
## 3 8   
## 4 3   
## 5 5   
## 6 2   
## 7 4   
## 8 7   
## 9 4   
## 10 9

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

## [1] 58

# sum is 58  
  
58/10

## [1] 5.8

# mean is 5.8  
  
9-5.8

## [1] 3.2

7-5.8

## [1] 1.2

8-5.8

## [1] 2.2

3-5.8

## [1] -2.8

5-5.8

## [1] -0.8

2-5.8

## [1] -3.8

4-5.8

## [1] -1.8

7-5.8

## [1] 1.2

4-5.8

## [1] -1.8

9-5.8

## [1] 3.2

3.2+1.2+2.2+-2.8+-.8+-3.8+-1.8+1.2+-1.8+3.2

## [1] 0.000000000000001332268

**Answer:** 0

1. Calculate the sum of deviance squared/sum of squares.

3.2^2+1.2^2+2.2^2+(-2.8)^2+(-.8)^2+(-3.8)^2+(-1.8)^2+1.2^2+(-1.8)^2+3.2^2

## [1] 57.6

# 57.6 is the sum of squares

**Answer:** The sum of squares is 57.6

1. Calculate the variance.

57.6/(10-1)

## [1] 6.4

**Answer:** 6.4 is the variance

1. Calculate the standard deviation.

sqrt(6.4)

## [1] 2.529822

sd(table\_create$values)

## [1] 2.529822

**Answer:** 2.53 is the standard deviation