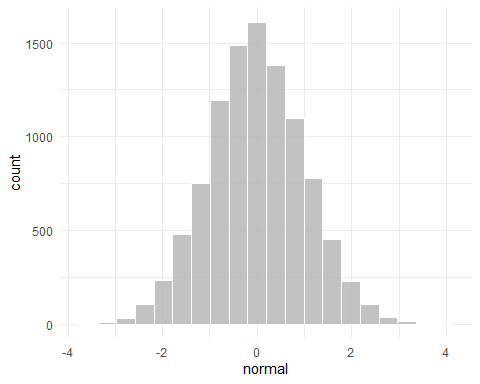
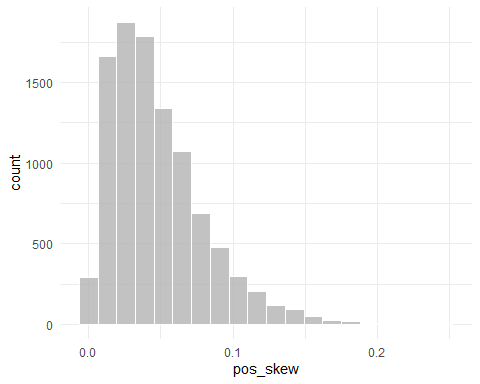
homework\_problem\_set\_1

JP

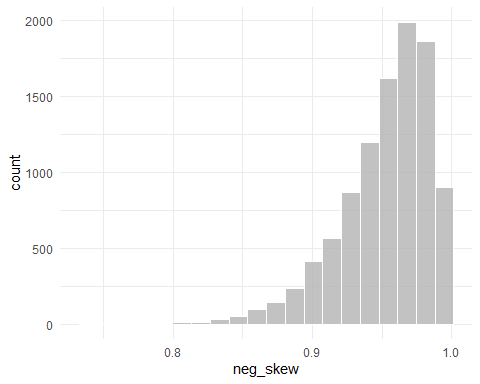
8/26/2021

1. What type of distribution is below? 

**Answer:** This is a normal distribution. It is also known as a normal curve or a bell-shaped curve. The reason why it is a normal distribution is also because the distribution is mainly symmetrical.

1. What type of distribution is below? 

**Answer:** The distribution is positively skewed. From the notes, it is positively skewed because the majority of scores/values are lower. Another way of stating this is that there are higher frequency of lower scores/values.

1. What type of distribution is below? 

**Answer:** The distribution is negatively skewed. From the notes, it is negatively skewed because the majority of scores/values are higher. Another way of stating this is that there are higher frequency of higher scores/values.

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

**Answer:** A bimodal distribution is when there are two peaks, or areas of your distribution that have the highest frequencies. For a perfect bimodal distribution, there will be two values that are the mode for your distribution. This distribution is a sign that you may have two different groups separated by the valley between the two highest points.

1. What is the term for the formula below?

**Answer:** Relative frequency. This tells you the count/frequency of a single category. This also gets you the proportion of how much this category may be of the total amount of participants. An example would be finding the relative frequency or the proportion of how many participants are female out all participants.

1. What is the difference between a histogram and a bar graph?

**Answer:** A histogram is for continuous data that counts the frequency of each value and has all the values together with the bars being adjacent to one another. A bar graph is to show the frequency of each category. Also, a bar graph has the bars separated by spaces.

1. What is this term and what does it mean?

**Answer:** This term is sigma and it is used to refer to the sum. We have learned about sigma from the sum of the scores when calculating the mean.

1. What measure of central tendency is best for skewed data?

**Answer:** The median is best for skewed data. That is because the extreme scores at the tail of the distribution will pull the mean toward those values.

1. What is xbar and what is mu (
2. )

**Answer:** Xbar is for the mean of a sample while mu is the population mean.

# Copy the code below from set.seed to N = 30

### Make sure to run everything together by highlighting everything and then pressing ctrl+R or command+R

# copy and paste this into your R  
  
set.seed(082421)  
categories = rep(letters[1:5], times = c(5, 3, 2, 8, 12))  
test\_scores = rnorm(n = 30, mean = 4.21, sd = 1.16)  
  
example = tibble(categories, test\_scores) %>%   
 rowid\_to\_column()  
   
categories = as.factor(categories)  
  
N = 30  
  
# If doing it by hand, the test scores and categories are below  
test\_scores

## [1] 2.736386 2.275197 3.914406 3.832424 1.438419 4.248828 7.135805 4.573231  
## [9] 3.048463 3.834346 2.795695 5.414713 5.043057 4.601812 5.427571 4.800366  
## [17] 3.643789 5.168277 5.126070 6.255702 2.291076 4.038750 6.849778 4.289894  
## [25] 6.104763 4.601148 4.167243 2.675319 3.541787 3.419616

categories

## [1] a a a a a b b b c c d d d d d d d d e e e e e e e e e e e e  
## Levels: a b c d e

1. Provide the frequencies/proportions of categories A, B, C, D, & E

table(categories)

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

a = 5/30 #.167  
b = 3/30 #.1  
c = 2/30 #.067  
d = 8/30 #.267  
e = 12/30 #.4  
  
# cbind(a, b, c, d, e)  
  
a + b + c + d + e

## [1] 1

.167 + .1 + .067 + .267 + .4

## [1] 1.001

1. Calculate the cumulative frequency of categories D & E

d + e

## [1] 0.6666667

.267 + .4

## [1] 0.667

1. Provide the mode for the categories.

table(categories)

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

# E is the most common value

1. Calculate the Median for the test scores.

sort(test\_scores)

## [1] 1.438419 2.275197 2.291076 2.675319 2.736386 2.795695 3.048463 3.419616  
## [9] 3.541787 3.643789 3.832424 3.834346 3.914406 4.038750 4.167243 4.248828  
## [17] 4.289894 4.573231 4.601148 4.601812 4.800366 5.043057 5.126070 5.168277  
## [25] 5.414713 5.427571 6.104763 6.255702 6.849778 7.135805

sort(test\_scores)[15:16]

## [1] 4.167243 4.248828

(4.17 + 4.25)/2

## [1] 4.21

# median value is 4.21  
median(test\_scores)

## [1] 4.208036

1. Calculate the Mean for the test scores.

test\_scores

## [1] 2.736386 2.275197 3.914406 3.832424 1.438419 4.248828 7.135805 4.573231  
## [9] 3.048463 3.834346 2.795695 5.414713 5.043057 4.601812 5.427571 4.800366  
## [17] 3.643789 5.168277 5.126070 6.255702 2.291076 4.038750 6.849778 4.289894  
## [25] 6.104763 4.601148 4.167243 2.675319 3.541787 3.419616

2.736386+2.275197+3.914406+3.832424+1.438419+4.248828+7.135805+4.573231 + 3.048463+3.834346+2.795695+5.414713+5.043057+4.601812+5.427571+4.800366+3.643789+5.168277+5.126070+6.255702+2.291076+4.038750+6.849778+4.28989 +6.104763+4.601148+4.167243+2.675319+3.541787+3.419616

## [1] 127.2939

127.29/30

## [1] 4.243

# The mean is 4.24  
  
mean(test\_scores)

## [1] 4.243131

1. Calculate the deviation of scores from the mean score/value.

test\_scores

## [1] 2.736386 2.275197 3.914406 3.832424 1.438419 4.248828 7.135805 4.573231  
## [9] 3.048463 3.834346 2.795695 5.414713 5.043057 4.601812 5.427571 4.800366  
## [17] 3.643789 5.168277 5.126070 6.255702 2.291076 4.038750 6.849778 4.289894  
## [25] 6.104763 4.601148 4.167243 2.675319 3.541787 3.419616

2.736386 - 4.243131

## [1] -1.506745

2.275197 - 4.243131

## [1] -1.967934

3.914406 - 4.243131

## [1] -0.328725

3.832424 - 4.243131

## [1] -0.410707

1.438419 - 4.243131

## [1] -2.804712

4.248828 - 4.243131

## [1] 0.005697

7.135805 - 4.243131

## [1] 2.892674

4.573231 - 4.243131

## [1] 0.3301

3.048463 - 4.243131

## [1] -1.194668

3.834346 - 4.243131

## [1] -0.408785

2.795695 - 4.243131

## [1] -1.447436

5.414713 - 4.243131

## [1] 1.171582

5.043057 - 4.243131

## [1] 0.799926

4.601812 - 4.243131

## [1] 0.358681

5.427571 - 4.243131

## [1] 1.18444

4.800366 - 4.243131

## [1] 0.557235

3.643789 - 4.243131

## [1] -0.599342

5.168277 - 4.243131

## [1] 0.925146

5.126070 - 4.243131

## [1] 0.882939

6.255702 - 4.243131

## [1] 2.012571

2.291076 - 4.243131

## [1] -1.952055

4.038750 - 4.243131

## [1] -0.204381

6.849778 - 4.243131

## [1] 2.606647

4.28989 - 4.243131

## [1] 0.046759

6.104763 - 4.243131

## [1] 1.861632

4.601148 - 4.243131

## [1] 0.358017

4.167243 - 4.243131

## [1] -0.075888

2.675319 - 4.243131

## [1] -1.567812

3.541787 - 4.243131

## [1] -0.701344

3.419616 - 4.243131

## [1] -0.823515

deviation = test\_scores - 4.243131  
  
deviation

## [1] -1.506745199 -1.967934487 -0.328724976 -0.410706596 -2.804711850  
## [6] 0.005697443 2.892673649 0.330100174 -1.194667586 -0.408784544  
## [11] -1.447436468 1.171581858 0.799925613 0.358681378 1.184440300  
## [16] 0.557234524 -0.599341679 0.925145724 0.882939057 2.012570838  
## [21] -1.952055314 -0.204381129 2.606646525 0.046763115 1.861632065  
## [26] 0.358017350 -0.075887783 -1.567812020 -0.701344446 -0.823515344

### Remember to double check your work with the following functions

mean(test\_scores)

median(test\_scores)