zscore-hw

Alex Steiner

2024-05-03

### Descriptive Stats Wednesday Module 5 Demonstration

### NR 2990

### Fall 2023

#Make sure we are in the right place  
#setwd(dir = "~/Desktop/NR2990")  
getwd()

## [1] "C:/Users/alste/Documents/Documents/IntroGradStats/Module-5-zdistribution"

#and make sure our libraries are loaded:  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3

## Warning: package 'ggplot2' was built under R version 4.3.3

## Warning: package 'readr' was built under R version 4.3.2

## Warning: package 'dplyr' was built under R version 4.3.2

## Warning: package 'stringr' was built under R version 4.3.2

## Warning: package 'forcats' was built under R version 4.3.2

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

#install.packages("lessR")  
library(lessR)

## Warning: package 'lessR' was built under R version 4.3.3

##   
## lessR 4.3.2 feedback: gerbing@pdx.edu   
## --------------------------------------------------------------  
## > d <- Read("") Read text, Excel, SPSS, SAS, or R data file  
## d is default data frame, data= in analysis routines optional  
##   
## Learn about reading, writing, and manipulating data, graphics,  
## testing means and proportions, regression, factor analysis,  
## customization, and descriptive statistics from pivot tables  
## Enter: browseVignettes("lessR")  
##   
## View changes in this and recent versions of lessR  
## Enter: news(package="lessR")  
##   
## Interactive data analysis  
## Enter: interact()  
##   
##   
## Attaching package: 'lessR'  
##   
## The following objects are masked from 'package:dplyr':  
##   
## recode, rename

library(readxl)  
library(dplyr)

Question 1.

#Import our data  
depression\_data <- read\_xlsx("Depression\_and\_the\_Internet.xlsx")  
str(depression\_data)

## tibble [169 × 8] (S3: tbl\_df/tbl/data.frame)  
## $ Internet use (mean hours per week): chr [1:169] "1.070487" "1.910293" "0" "1.923E-3" ...  
## $ DepressionBefore : chr [1:169] "1.7331540000000001" "0.13333100000000001" "0.59997599999999995" "1.866455" ...  
## $ DepressionAfter : chr [1:169] "0.92846700000000004" "0.14285300000000001" "0.35711700000000002" "0.85705600000000004" ...  
## $ Gender : chr [1:169] "male" "male" "female" "female" ...  
## $ Race (white = 1, minority = 0) : num [1:169] 1 1 1 1 1 1 1 1 0 1 ...  
## $ Age : chr [1:169] "Adult" "Adult" "Adult" "Adult" ...  
## $ $Household income ($000) : chr [1:169] "85" "85" "85" "85" ...  
## $ Household size : chr [1:169] "5" "5" "5" "5" ...

head(depression\_data)

## # A tibble: 6 × 8  
## `Internet use (mean hours per week)` DepressionBefore DepressionAfter Gender  
## <chr> <chr> <chr> <chr>   
## 1 1.070487 1.7331540000000001 0.928467000000… male   
## 2 1.910293 0.133331000000000… 0.142853000000… male   
## 3 0 0.599975999999999… 0.357117000000… female  
## 4 1.923E-3 1.866455 0.857056000000… female  
## 5 0.98567700000000003 0.399963000000000… 0.285706000000… male   
## 6 2.8928989999999999 0.733276000000000… 0.642822 male   
## # ℹ 4 more variables: `Race (white = 1, minority = 0)` <dbl>, Age <chr>,  
## # `$Household income ($000)` <chr>, `Household size` <chr>

Question 2.

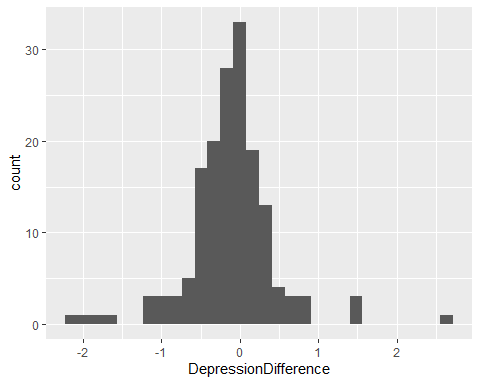
depression\_data\_num <- depression\_data |>   
 mutate\_at(c('DepressionAfter', 'DepressionBefore'), as.numeric)   
  
depression\_data\_diff <- depression\_data\_num |>   
 mutate(DepressionDifference = DepressionAfter - DepressionBefore)  
head(depression\_data\_diff)

## # A tibble: 6 × 9  
## `Internet use (mean hours per week)` DepressionBefore DepressionAfter Gender  
## <chr> <dbl> <dbl> <chr>   
## 1 1.070487 1.73 0.928 male   
## 2 1.910293 0.133 0.143 male   
## 3 0 0.600 0.357 female  
## 4 1.923E-3 1.87 0.857 female  
## 5 0.98567700000000003 0.400 0.286 male   
## 6 2.8928989999999999 0.733 0.643 male   
## # ℹ 5 more variables: `Race (white = 1, minority = 0)` <dbl>, Age <chr>,  
## # `$Household income ($000)` <chr>, `Household size` <chr>,  
## # DepressionDifference <dbl>

Question 3. Question 3.1

ggplot(data = depression\_data\_diff, aes(x =DepressionDifference ))+  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 The data seems a little skinny and tall

Question 3.2

pivot(depression\_data\_diff, c(IQR, skew, kurtosis, mean, sd, var), DepressionDifference)

## n na IQR skew kurt mean sd var  
## 162 7 0.476 0.208 5.981 -0.118 0.553 0.305

summary(depression\_data\_diff$DepressionDifference)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -2.17609 -0.35230 -0.09045 -0.11807 0.12378 2.60004 7

#Question 3.3

#normality test  
shapiro.test(depression\_data\_diff$DepressionDifference)

##   
## Shapiro-Wilk normality test  
##   
## data: depression\_data\_diff$DepressionDifference  
## W = 0.88718, p-value = 0.0000000009229

P < 0.05, data is not normal

Question 4. What is the probability of seeing a person in the general population with an increase of more than 1 depression unit?

pnorm(q = 1, mean = -0.11807, sd = 0.553, lower.tail = F)

## [1] 0.02159714

The code I saw in the example did not make sense to me. The sd is 0.553, and the variance is 0.305. Additionally, logically, the probability of having an increase of 1 point would be low, because the mean difference (After-before) is negative.

Based on this data, what is the probability of seeing someone decrease in their depression rating

pnorm(q = 0, mean = -0.11807, sd = 0.553, lower.tail = T)

## [1] 0.5845347

If we wanted to flag people whose depression increased significantly (using a 1-tailed 0.05 alpha threshold) what is the depression value we would use as the threshold to flag people for additional follow-up?