Problem Set 4

JP

2022-02-24

You are interested in knowing how a sample of college students compare to all college students in regard to their GPAs. You decide to test to see whether your group is significantly different from the rest of the university. You have data for 15 students that you are using for your sample. From information you gathered on the college’s website, you know that the average college GPA is 3.5 for all 1000 students. You also break into the registrar’s office and find information about how much students deviate from the GPA for all college students. You find out the standard deviation for all college students is 1.2. Answer the following questions.

problem\_set <- data.frame(gpa = c(3.7, 3.2, 4, 2.7, 3.1,  
 2.5, 2.9, 3.3, 1.9, 3.1,  
 3.6, 3, 2.5, 2.3, 3.5),  
 mean\_value = c('', '', '', '', '',  
 '', '', '', '', '',  
 '', '', '', '', ''),  
 deviance = c('', '', '', '', '',  
 '', '', '', '', '',  
 '', '', '', '', ''),  
 dev\_squared = c('', '', '', '', '',  
 '', '', '', '', '',  
 '', '', '', '', ''))  
problem\_set

## gpa mean\_value deviance dev\_squared  
## 1 3.7   
## 2 3.2   
## 3 4.0   
## 4 2.7   
## 5 3.1   
## 6 2.5   
## 7 2.9   
## 8 3.3   
## 9 1.9   
## 10 3.1   
## 11 3.6   
## 12 3.0   
## 13 2.5   
## 14 2.3   
## 15 3.5

1. What is the number of participants?

problem\_set$n <- 15  
unique(problem\_set$n)

## [1] 15

1. What is the mean gpa score?

problem\_set$mean\_value <- mean(problem\_set$gpa)  
unique(problem\_set$mean\_value)

## [1] 3.02

1. What is the deviance between participants’ scores and the mean gpa score?

3.7-3.02

## [1] 0.68

3.2-3.02

## [1] 0.18

4.0-3.02

## [1] 0.98

2.7-3.02

## [1] -0.32

3.1-3.02

## [1] 0.08

2.5-3.02

## [1] -0.52

2.9-3.02

## [1] -0.12

3.3-3.02

## [1] 0.28

1.9-3.02

## [1] -1.12

3.1-3.02

## [1] 0.08

3.6-3.02

## [1] 0.58

3.0-3.02

## [1] -0.02

2.5-3.02

## [1] -0.52

2.3-3.02

## [1] -0.72

3.5-3.02

## [1] 0.48

problem\_set$deviance <- problem\_set$gpa - problem\_set$mean\_value  
unique(problem\_set$deviance)

## [1] 0.68 0.18 0.98 -0.32 0.08 -0.52 -0.12 0.28 -1.12 0.58 -0.02 -0.72  
## [13] 0.48

1. Calculate the sum of errors/total deviation.

(0.68) + (0.18) +(0.98) +(-0.32) +(0.08)+(-0.52)+(-0.12)+  
 (0.28)+(-1.12) +(0.08)+(0.58)+(-0.02)+(-0.52)+(-0.72) +(0.48)

## [1] 0

sum(problem\_set$deviance)

## [1] 0

1. Calculate the deviance squared for each participant.

(0.68)^2

## [1] 0.4624

(0.18)^2

## [1] 0.0324

(0.98)^2

## [1] 0.9604

(-0.32)^2

## [1] 0.1024

(0.08)^2

## [1] 0.0064

(-0.52)^2

## [1] 0.2704

(-0.12)^2

## [1] 0.0144

(0.28)^2

## [1] 0.0784

(-1.12)^2

## [1] 1.2544

(0.08)^2

## [1] 0.0064

(0.58)^2

## [1] 0.3364

(-0.02)^2

## [1] 4e-04

(-0.52)^2

## [1] 0.2704

(-0.72)^2

## [1] 0.5184

(0.48)^2

## [1] 0.2304

problem\_set$dev\_squared <- problem\_set$deviance^2  
  
unique(problem\_set$dev\_squared)

## [1] 0.4624 0.0324 0.9604 0.1024 0.0064 0.2704 0.0144 0.0784 1.2544 0.3364  
## [11] 0.0004 0.5184 0.2304

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

(0.68)^2 + (0.18)^2 +(0.98)^2 +(-0.32)^2 +(0.08)^2+(-0.52)^2+(-0.12)^2+  
 (0.28)^2+(-1.12)^2 +(0.08)^2+(0.58)^2+(-0.02)^2+(-0.52)^2+(-0.72)^2 +(0.48)^2

## [1] 4.544

problem\_set$ss <- sum(problem\_set$dev\_squared)  
  
unique(problem\_set$ss)

## [1] 4.544

1. Calculate the degrees of freedom.

15 - 1

## [1] 14

problem\_set$df <- problem\_set$n - 1  
unique(problem\_set$df)

## [1] 14

1. Calculate the variance/mean squared error for your model.

4.54/14

## [1] 0.3242857

problem\_set$var <- problem\_set$ss/problem\_set$df  
unique(problem\_set$var)

## [1] 0.3245714

1. Calculate the standard deviation for your sample.

sqrt(.32)

## [1] 0.5656854

problem\_set$sd\_value <- sqrt(problem\_set$var)  
unique(problem\_set$sd\_value)

## [1] 0.5697117

1. Conduct a z-test using your sample and the population mean and population standard deviation. (*HINT* You’ll need to calculate the population standard error.)

problem\_set$pop\_sd <- 1.2  
problem\_set$mu <- 3.5  
problem\_set$pop\_n <- 1000  
  
1.2/sqrt(1000)

## [1] 0.03794733

# se is .04  
  
problem\_set$sigma\_se <- problem\_set$pop\_sd/sqrt(problem\_set$pop\_n)  
unique(problem\_set$sigma\_se)

## [1] 0.03794733

(3.02 - 3.5)/.04

## [1] -12

problem\_set$z\_obt <- (problem\_set$mean\_value - problem\_set$mu)/problem\_set$sigma\_se  
unique(problem\_set$z\_obt)

## [1] -12.64911

1. What is the z-critical value for a two-tailed z-test?

Answer:

Now, for some reason you lost the standard deviation for the entire college. You decide that the z-test probably wasn’t the best way to compare your sample to the whole college. You decide to estimate the population standard error by using your sample data. (The population mean will remain the same).

1. Calculate the standard error from your calculations using the mean. (Using S for the standard error to not get confused with Sigma. They are the same.)

.57/sqrt(15)

## [1] 0.1471734

# se is .15  
  
problem\_set$se\_value <- problem\_set$sd\_value/sqrt(problem\_set$n)  
unique(problem\_set$se\_value)

## [1] 0.1470989

1. Calculate a one-sample t-test.

(3.02 - 3.5)/.15

## [1] -3.2

problem\_set$t\_obt <- (problem\_set$mean\_value - problem\_set$mu)/problem\_set$se\_value  
unique(problem\_set$t\_obt)

## [1] -3.26311

1. Tell me whether the finding is statistically significant or not. (Look at the t-table)