Riley Payung

Professor Romanelli

CDS 290

December 12, 2020

Homework 11

Problem 1 (10.35 – Body Dissatisfaction)

Sample Mean for Body dissatisfaction = 13.2 (s = 8.0) for 16 lean sport athletes

Sample Mean for body dissatisfaction = 7.3 (s = 6.0) for 68 nonlean sport athletes.

Assume equal population standard deviations.

1. Find the standard error:

SE for 16 lean sport athletes: 8 / sqrt(16) = 8 / 4 = 2.0

SE for 68 nonlean sport athletes: 6.0 / sqrt(68) = 6.0 / 8.2462 = 0.72761

2. Construct a 95% confidence interval for the difference:

Mean 1 = 13.2. Mean 2 = 7.3. CI = 95%.

Difference between means: 13.2 – 7.3 = 5.9

Standard Error of Difference: sqrt(4 + 0.72761) = 2.1743

Confidence Interval: 5.9 (+-) (1.96 * 2.1743) = 5.9 (+-) 4.262 = [1.638, 10.162]

3. Interpret:

We can be 95% confident that the difference between the sample mean of female lean sport athletes and female non-lean sport athletes lies between 1.638 and 10.162.

Problem 2 (10.51 Social Activities for Students)

N = 10.

1. Independent or Dependent?

Independent, because attending a movie means they did not attend a sporting event. I suppose this would be weird if the movies were of the sports genre or a documentary about a specific sporting event in history.

2. Does the dot plot show any irregularities that would make statistical inference unreliable?

Yes, there are outliers, which would skew the data heavily, specifically the one that is less than -20 and the one that's greater than +30. One person obviously attended movies A LOT more than sports, which is person number 7. Person number 7 also seems like he/she may be a part of Greek life.

3. Use the MINITAB output to show how the confidence interval was obtained.

Movies mean = sum of all movies over n = 130 / 10 = 13.00

Movies std = 13.17

Standard error: 13.17 / sqrt(13) = 13.17 / 3.60555

The book here is wrong in the confidence interval. I actually used the compare two means to get the confidence interval and this is what it output:

Descriptive Statistics:

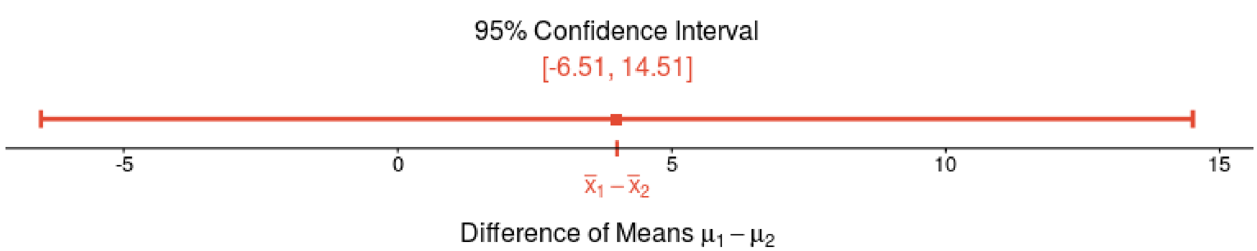
Group	Sample Size	Mean	Std. Dev.
Movies	10	13.00	13.17
Sports	10	9.00	8.38

Estimate of Difference of Means:

Point Estimate	Standard Error	t-score (df=15.3)	Margin of Error
4.00	4.94	2.13	± 10.51

Confidence Interval:

Population Parameter	Lower Bound	Upper Bound	Confidence Level
Difference $\mu_1 - \mu_2$	-6.51	14.51	95%



4. Interpret

Based on the resulting confidence interval, and not on the MINITAB one, we are 95% confident that the true difference of the means lies between the confidence interval of -6.51 to 14.51. Based on this MINITAB one, we are 95% confident that the true difference of the means lies between the confidence interval of -7.56 and 15.56.

Problem 3 (10.67 Basketball Paradox)

2-Pointers:

- Barry (58.8%) (237/403)
- O'Neal (58.0) (712/1228)

3-Pointers:

- Barry (42.4%) (164/387)
- O'Neal (0%) (0/1)

Overall:

- Barry (50.8%) (401/790)
- O'Neal (57.9%) (712/1229)

1. Simpsons Paradox:

This is an example of Simpson's paradox because Shaq does better in the two pointers than three pointers because it seems like he only made two pointers, whereas Barry made both two and three pointers frequently; however, if you add up each player's shots over their attempted shots, it seems like Shaq does better, even though he only attempted one three-pointer. This illustrates Simpsons because the better player is reversed when you add up everything together for each player.

2. Explain O'Neal's higher overall

Again, like in the first one, he has more shots overall, even though he does not make any three pointers whereas Barry does, but if you add up all their shots, the better percentage is going to go toward the person with more made shots. A way to prevent this (because its bias) is to add weights to the type of shot, where a three-pointer shot is worth a little bit more than a two pointer.

Problem 4 (R3.4 Random Variability in Baseball)

A professional baseball team wins 92 of its 162 games this year. Can we conclude that this team is a better than average team, having greater than a 0.50 chance of winning a game? Answer by putting this in the context of a one-sided significance test, stating all assumptions.

Win rate: 92/162 = 0.5679

N = 162

We assume that the wins of the games are normally distributed.

H0: mean = 0.50

Ha: mean > 0.50

Alpha: 0.10

One sided test:

Z = 0.56 – 0.50 / sqrt(0.5*0.5 / 162) = 0.06 / 0.03928 = 1.527494

P(Z > 1.527) = 1 – 0.9345 = 0.0655

P = 0.0655

P < alpha in this case, so we can reject the null hypothesis and say that there is nothing proving that they are not better than average.