homework_ch5_JimmyNg

Jimmy Ng

October 15, 2018

Question 5.6

90% confidence interval indicates a critical value of 1.64 n = 25 margin of error: 1.64 +/- (sd / sqrt(n)) # in this case, it is just (upper CI - lower Ci) / 2 = 6 mean: x <- 77 - ((77 - 65) / 2)

working backward to get the sd: # 77 = x + 1.64 * (sd / sqrt(n)) # 77 - 71 = 1.64 * (sd / sqrt(25)) # 6 = 1.64 * (sd / 5) sd = (6 / 1.64) * 5

In summary, the mean (x bar) is equal to 71; the margin of error is equal to 6; and the sd is approximately 18.29268.

Question 5.14

JN: 5.14(a) margin of error is equal to critical value * sd / sqrt(n). In this case, a 90% CI would give a critical value of 1.64. A margin of error equal to 25 or less, plus 90% CI and 250 sd would give us this below,

n = ((250)*(1.64) / 25)^2 n = 268.96

Raina would need to collect 269 samples.

JN: 5.14(b) Luke has to use a bigger sample size to achieve 99% confidence interval (CI). The reasoning is that everything else is constant, in order to get the same margin of error but increasing the CI (which will increase the critical value for 99% jumping from 90%), the standard error (SE) must decrease. Since the SE is equal to SD / sqrt(n) and the SD is also constant, the only way to minimize the SE is to increase the sample size.

JN: 5.15(c) 99% CI would have the critical value of 2.58.

n = ((250)*(2.58) / 25)^2 n = 665.64

Luke would need to collect 666 samples.

Question 5.20

JN: 5.20(a) Not really clear visible difference in terms of average; however, reading seems to have larger variation within data. JN: 5.20(b) No. We would expect there's some sort of correlation, such as students score high in reading would also do well in writing. JN: 5.20(c) The null hypothesis is that there is no difference between reading and writing scores, whereas the alternative hypothesis is that there is statistical difference between the two. JN: 5.20(d) The condition will not be met because reading and writing are not independent from each other. Someone scored high in one would also intuitively score high in another. In addition, reading seems to have larger variance than writing, which makes it hard to use parametric test to compare the mean between the two. JN: 5.20(e) No, there's no convincing evidence in supporting there's difference between reading and writing scores. The SE = 8.887/sqrt(200) = 0.6284058. The mean of difference is -0.545. 95% CI would generate the margin of error = 1.96 * SE = 1.231675. Therefore, 95% CI would be equal to (-1.776675, 0.686675). Since the range of CI crosses 0, there's no significant evidence in supporting the alternative hypothesis. JN: 5.20(f) We could have made a type 2 error, meaning we fail to reject the null hypothesis when there's actually a difference between the two. In this case, that means there could possibly be a difference between reading and writing, but we just fail to capture the difference statistically. JN: 5.20(g) As showed in answer 5.20(e), the range of CI would cover 0.

Question 5.32

JN: the mean difference between manual and automatic is 19.85 - 16.12 = 3.73, whereas the SE is equal to sqrt(((3.58)^2)/26 + ((4.51)^2)/26) = 1.12927. The margin of error for 95% CI is equal to 1.96 * SE = 2.213369. As a result, the 95% confidence interval the mean difference is (1.516631, 5.943369). The result suggested that manual needs more fuel, or automatic is statistically more fuel efficient.

Question 5.48

JN: 5.48(a) the null hypothesis is that the level of education would have no impact on hours of work per week, whereas the alternative hypothesis is that there's an impact on hours of work per week varied by level of education.

JN: 5.48(b) independence, sample size are checked; normality is ok with large sample size (minor skews/outliners are tolerated); however, variance does not seem to check out as these groups (as shown by the boxplots) have inconsistent variance within group.

JN: 5.48(c) see below calculation. dfG = k - 1 = 5 - 1 = 4 // dfE = n - k = 1172 - 5 = 1167 // total Df = 4 + 1167 = 1171 // Sum of Squares between Groups (SSG) = Mean Square between Groups (MSG) * dfG = 501.54 * 4 = 2006.16 // Sum of Squares Total (SST) = SSG + SSE = 2006.16 + 267382 = 269388.2 // Mean Square Error (MSE) = (1 / dfE) * Sum of Squared Errors (SSE) = (1 / 1167) * 267382 = 229.1191 // F value: MSG / MSE = 501.54 / 229.1191 = 2.188993

JN. 5.48(d) since the p-value is above .05, there is not statistical evidence in supporting that there is any significant difference among groups. In other words, level of education does not appear to have an impact on hours of work per week (at least, there's no evidence displayed in this study).