STAT 515 hw 7

CIs for mean with σ unknown, sample size calculations

Attach a sheet with the R plots and R code printed on it. You may write out your other answers by hand if you want. Just try to make it easy for me grade!!

1. Open R and enter `data(Loblolly)` into the console. This imports the Loblolly data set into the workspace. Type `?Loblolly` into the console to read a description of the data set.

   (a) On how many trees was data collected?

   14

   (b) How many times was the height of each tree recorded?

   6

   (c) At what ages was the height of each tree recorded?

   At ages $3, 4, 10, 15, 20, 25$.

   (d) Compute the mean $\bar{X}_n$ and the sample standard deviation s for the heights of Loblolly pines which are 3 years old. Hint: Enter the command

   ```
   x <- Loblolly$height[Loblolly$age==3]
   ```

   Then to compute the mean $\bar{X}$, you can simply type `mean(x)` and for the standard deviation, you can type `sd(x)`.

   We get $\bar{X}_n = 4.237857$, $S_n = 0.4036026$.

   (e) Generate a Normal QQ plot of the heights of the Loblolly pines at age 3. Turn in this plot. Use `qqnorm(x)`.

   (f) Based on the QQ plot, do you think that the heights follow a Normal distribution?

   Looks fairly Normal.

   (g) Compute a 95% confidence interval for the mean height of three-year-old Loblolly pines.

   We get

   $$4.237857 \pm t_{13,\alpha/2} 0.4036026/\sqrt{14} = 4.237857 \pm (2.16) 0.4036026/\sqrt{14} = (4.004864, 4.470851)$$

   (h) Interpret this interval.

We are 95% confident that the mean height of three-year-old Loblolly pines is in this interval.

(i) Give a 95% percent confidence interval for the mean height of twenty-year-old Loblolly pines.

We get the interval $(50.19172, 52.74542)$.

(j) If you had constructed 99% confidence intervals for the Loblolly heights, would they have been wider or narrow than the 95% confidence intervals?

The 99% confidence interval would be wider.

(k) You plan to estimate the mean height of 3-year-old Loblolly pines in a different region of North America, and you need to know how many trees to measure. Give a recommended sample size if you want

i. a 95% confidence interval no wider than 0.25 feet.

Using $S_n = 0.404$ as our best guess of σ, we see that to have a margin of error $M^* \leq 0.25/2 = 0.125$ with $\alpha = 0.05$, we would need a sample size of at least

$$\left(\frac{z_{0.05/2} \cdot 0.404}{0.125}\right)^2 = \left(\frac{1.96 \cdot 0.404}{0.125}\right)^2 = 40.12868.$$

So we would need $n = 41$.

ii. a 99% confidence interval with margin of error no greater than 0.10.

With $M^* = 0.10$ and $\alpha = 0.01$, we would need a sample size of at least

$$\left(\frac{z_{0.01/2} \cdot 0.404}{0.10}\right)^2 = \left(\frac{2.576 \cdot 0.404}{0.10}\right)^2 = 108.3065.$$

So we would need $n = 109$.

2. Make a 95% confidence interval for the variance σ^2 of the heights of Loblolly trees which are three years old in the following steps:

(a) Compute S_n^2.

We get $S_n^2 = 0.1628951$

(b) Find the degrees of freedom of the relevant Chi-square distribution.

Use $\nu = 13$.

(c) Find $\chi^2_{\nu,1-\alpha/2}$ and $\chi^2_{\nu,\alpha/2}$, where ν is your answer to part (b).

> We have $\chi^2_{13,.975} = 5.00875$, and $\chi^2_{13,.025} = 24.7356$.

(d) Compute the confidence interval.

> We get the interval $(0.08561085, 0.4227873)$.

3. You wish to estimate the proportion of bees in a beehive that are drones within 0.02 with confidence level 95%. A sample of 307 bees from a previous hive contained 44 drones.

(a) How many bees should you sample?

> With $M^* = 0.02/2 = 0.01$ and $\alpha = 0.05$, and using $44/307 = 0.143$ for p, we would need a sample size of at least
>
> $$\left(1.96\frac{\sqrt{44/307(1-44/307)}}{0.01}\right)^2 = 4716.76,$$
>
> so we would need to sample 4716.76 bees.

(b) If you ignore the data from the previous hive, how many bees would you recommend sampling?

> With $M^* = 0.02/2 = 0.01$ and $\alpha = 0.05$, and using 0.5 for p, we would need a sample size of at least
>
> $$\left(1.96\frac{\sqrt{1/2(1-1/2)}}{0.01}\right)^2 = 9604.$$

Optional (do not turn in) problems for additional study from McClave, J.T. and Sincich T. (2017) *Statistics*, 13th Edition: 7.38, 7.40, 7.50