

STAT22000 Summer 2020 Homework 11

All page, section, and exercise numbers below refer to the course text (*OpenIntro Statistics*, 3rd edition, by Diez, Barr, and Cetinkaya-Rundel.).

Reading: Section 7.1

Problems for Self-Study : (Do Not Turn In)

- Exercise 5.1, 5.3, 5.7, 5.11 on p.257-259. The answers can be found at the end of the book.

Problems to Turn In: due **midnight on Monday, July 20, on Canvas.**

1. If the 95% confidence interval for a parameter μ is (3.1, 5.1) and the 99% confidence interval for μ is (2.8, 5.4). One wants to test the hypotheses $H_0: \mu = 3$ v.s. $H_a: \mu \neq 3$.
 - (a) Will $H_0: \mu = 3$ be rejected at significance level $\alpha = 0.05$? Why or why not?
 - (b) Will $H_0: \mu = 3$ be rejected at significance level $\alpha = 0.01$? Why or why not?
2. The P -value for a two-sided test of the null hypothesis $H_0: \mu = 10$ is 0.03.
 - (a) Does the 95% confidence interval for μ include 10? Why or why not?
 - (b) Does the 99% confidence interval for μ include 10? Why or why not?
3. It has been hypothesized that allergies result from a lack of early childhood exposure to antigens. If this hypothesis were true, then we would expect allergies to be more common in very hygienic households with low levels of bacteria and other infectious agents. To test this theory, researchers at the University of Colorado sampled the houses of 61 children 9-24 months old and recorded two variables: (1) whether the child tested positive for allergies and (2) the concentration of bacterial endotoxin in the house dust (endotoxin units per ml, EU/ml)¹. The following are the endotoxin levels at the homes of the 51 children tested negative for allergies.

```
708.23  911.60  976.81 1316.63  262.74 9772.08  370.76  229.16 2570.51
891.19 3163.20 1777.65 1288.57  436.23 2631.63 1173.52  911.67 7942.42
740.32  356.92 1175.48 1480.55 2754.61  575.62  573.89  468.26 1000.71
364.22 1025.26 1022.04  645.41  363.57  977.47 1022.75 1860.63  371.13
174.73  399.68 1479.77 2882.96  601.99 1697.32 2291.00  646.49 1176.27
1995.43  955.54 1480.05  456.71 1174.70 5494.22
```

- (a) Make a boxplot for the endotoxin levels at the homes of the 51 children without allergy. Comment on whether it is appropriate to construct a t -confidence interval for the mean endotoxin level at the homes of children without allergy.

Endotoxin =

```
c(708.23, 911.60, 976.81, 1316.63, 262.74, 9772.08, 370.76, 229.16, 2570.51,
891.19, 3163.20, 1777.65, 1288.57, 436.23, 2631.63, 1173.52, 911.67, 7942.42,
740.32, 356.92, 1175.48, 1480.55, 2754.61, 575.62, 573.89, 468.26, 1000.71,
364.22, 1025.26, 1022.04, 645.41, 363.57, 977.47, 1022.75, 1860.63, 371.13,
174.73, 399.68, 1479.77, 2882.96, 601.99, 1697.32, 2291.00, 646.49, 1176.27,
1995.43, 955.54, 1480.05, 456.71, 1174.70, 5494.22)
library(mosaic)
bwplot(Endotoxin, horizontal=T)
```

¹Gereda JE, Leung DYM, Thatayatikom A, Streib JE, Price MR, Klinnert MD, and Liu AH. (2000). Relation between house-dust endotoxin exposure, type 1 T-cell development, and allergen sensitisation in infants at high risk of asthma. *The Lancet*, **355**: 1680-1683.

- (b) Make a boxplot of the log of the endotoxin levels at the homes of the 51 children without allergy.

```
bwplot(log(Endotoxin), horizontal=T)
```

Comment on whether it is more appropriate to construct a t confidence interval for the mean of the log endotoxin level at the homes of children without allergy.

- (c) Construct a 95% t -confidence interval for the mean of the log endotoxin level at the homes of children without allergy.

Important Note: *You may find the sample mean and the sample SD of the log endotoxin levels using R, but you are NOT suppose to find the confidence interval using the R function `t.test`. Please write down the formula for the confidence interval and show your calculation.*

4. Please complete the “On Your Own” part of Lab 7:

<http://www.stat.uchicago.edu/~yibi/s220/labs/lab07.html>

- (a) Now use the 260 women in the data set as the population. Make a histogram for the height of the 260 women. Do you find any outlier or clear skewness? Find the population mean and SD of the height of the 260 women.
- (b) Take 100 samples of size 5 from the population of the 260 women. For each of the sample, construct the 3 kinds of confidence intervals ($\bar{x} \pm z^* \sigma / \sqrt{n}$, $\bar{x} \pm z^* s / \sqrt{n}$, and $\bar{x} \pm t^* s / \sqrt{n}$) for the mean height of the 260 women at 90% level. Use the `plot_ci` function in Lab 7 to plot the 3 kinds of confidence intervals. For each of the 3 kinds of intervals, calculate the proportion of intervals that include the true population mean. Are those percentages close to the nominal level 90%?