

ST511 HW #1

Reading: Browse Chapters 1- 3 of Ott & Longnecker.

See Canvas Calendar for due date.

12 points total: Grading will be just a check-off (i.e. full credit for anything that looks reasonable).

Most Students prefer to submit HW through Canvas in pdf format.

Local Students have the additional options of submitting a paper copy at class or to the grader's mailbox in Statistics 102.

Important reminders:

- Use RStudio and R Markdown.
- Download the datasets from the Ott & Longnecker companion site (see Canvas > Modules > Additional Resources). Download the zip file and then **unzip the file**. We will use the CSV files (“**ASCII-comma**”). The file extension is .TXT even though the files are actually CSV!
- The files can be imported into R using the read.csv() function:

```
InData <- read.csv("C:/hess/STAT511_Fall/O&L CSV Data/ASCII-comma/CH03/ex3-30.txt", quote = "'")
```

or

```
InData <- read.csv(file.choose(), quote = "'")
```

The file.choose() function allows files to be chosen interactively (without typing out the file path location). The quote option is used because the column names in the original data are (single) quoted.

1. Use the data described in Problem 3.30 regarding lumber. From the files you downloaded above, you will find the data under CH03, named ex3-30.txt. Use the following commands to import and summarize the data.

```
lumber <- read.csv("ex3-30.txt", quote = "'")  
str(lumber)  
hist(lumber$Number, col = "grey", main = "Lumber Histogram")  
mean(lumber$Number)  
median(lumber$Number)
```

- A. Include the histogram in your assignment.
- B. Give the mean and median of the sample.
- C. Does the data appear to be “normal” (bell-shaped)? Justify your response based on your histogram from above.

2. Use the data described in Problem 3.7 regarding survival times. You will find the data under CH03, named ex3-7.txt. Notice that the two therapies are in two different columns. (There are ways to reformat, but we will work with the data “as is” for now.) Use the following code to get you started:

```
survival <- read.csv("ex3-7.txt", quote = "'")  
str(survival)  
mean(survival$StandardTherapy)  
sd(survival$StandardTherapy)
```

```
boxplot(survival)
```

- A. What is the sample mean and sample standard deviation for each of the therapies?
- B. Construct side-by-side boxplots showing the survival times for each therapy.