

Assignment #8 — Due Friday, November 11 by 4:00 P.M.

Turn in homework to your TA's mailbox using this sheet as the cover page.

Fill in your name and also circle the *lecture section in which you are registered* and circle the *discussion section you expect to attend* to pick up this assignment.

Name:

Lecture 1 (Hanlon). **311:** Tu 1:00 - 2:15pm **312:** Th 8:00 - 9:15am **313:** We 1:00 - 2:15pm

Lecture 2 (Larget). **321:** Tu 1:00 - 2:15pm **322:** We 2:30 - 3:45pm **323:** We 1:00 - 2:15pm

Please answer the following questions.

- The following questions ask you to compute confidence intervals for a difference in population means $\mu_1 - \mu_2$. We assume that the data is approximately normal. We use n , \bar{y} , and s to denote the sample size, the sample mean and sample standard deviation, respectively.
 - $n_1 = 10$, $\bar{y}_1 = 20.2$, $s_1 = 5.2$; $n_2 = 34$, $\bar{y}_2 = 30.7$, $s_2 = 4.7$. Compute a 90% confidence interval for $\mu_1 - \mu_2$.
 - $n_1 = 130$, $\bar{y}_1 = 100.11$, $s_1 = 20$; $n_2 = 47$, $\bar{y}_2 = 95.7$, $s_2 = 19.6$. Compute a 95% confidence interval for $\mu_1 - \mu_2$.
 - $n = 17$, $\bar{y} = -5.6$, $s = 1.2$; $n_2 = 25$, $\bar{y}_2 = -4.3$, $s_2 = 1.6$. Compute a 98% confidence interval for $\mu_1 - \mu_2$.
- The following questions ask you to compute the test statistic and p-value for a two-sample independent t-test. We assume that the data is approximately normal. We use n , \bar{y} , and s to denote the sample size, the sample mean and sample standard deviation, respectively.
 - $n_1 = 10$, $\bar{y}_1 = 20.2$, $s_1 = 5.2$; $n_2 = 34$, $\bar{y}_2 = 30.7$, $s_2 = 4.7$. Compute the test statistic and p-value for the test of $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$.
 - $n_1 = 130$, $\bar{y}_1 = 100.11$, $s_1 = 20$; $n_2 = 47$, $\bar{y}_2 = 95.7$, $s_2 = 19.6$. Compute the test statistic and p-value for the test of $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 > \mu_2$.
 - $n = 17$, $\bar{y} = -5.6$, $s = 1.2$; $n_2 = 25$, $\bar{y}_2 = -4.3$, $s_2 = 1.6$. Compute the test statistic and p-value for the test of $H_0 : \mu_1 = \mu_2$ vs. $\mu_1 < \mu_2$.
- Read problem 23 on page 314 for the background, but do the following instead. Data for this problem is in the file `seals.csv`. but do the following instead.
 - Is this a paired study or an independent sample study? Explain.
 - Use statistical software to graph a density plot of the sample differences between oxygen consumption for feeding dives minus nonfeeding dives. Comment briefly features of the plot (shape, skewness, outliers, center, spread).
 - Use statistical software to construct a 99% confidence interval for the mean difference in oxygen consumption (feeding dives minus nonfeeding dives) using t -distribution methods.
 - Use statistical software to construct a 99% confidence interval for the mean difference in oxygen consumption (feeding dives minus nonfeeding dives) using the nonparametric bootstrap.

- (e) Briefly comment on how the assumptions for the methods to construct the previous two confidence intervals differ.
 - (f) Choose one of the two intervals and interpret the results in the context of the problem.
 - (g) Use statistical software to test the null hypothesis of no difference in oxygen consumption between feeding dives and nonfeeding dives versus the alternative hypothesis that feeding dives require more oxygen. Report the test statistic and p-value from this test.
 - (h) Interpret the result of this hypothesis test in the context of the problem.
4. Read problem 10 on page 309 for the background on a genetics experiment for lengths of tobacco leaves. The data file `tobacco.csv` contains a data set of this size that is similar, except that leaf lengths have not been rounded as much. Do not answer the questions in the textbook, but do the following instead.
- (a) Use statistical software to make a plot that effectively compares the distribution of flower lengths between the F1 and F2 generations.
 - (b) Use statistical software to construct a 95% confidence interval for the difference in population mean flower lengths between the F2 and F1 generations.
 - (c) Interpret this interval in the context of the problem.
 - (d) Use statistical software to test the hypothesis that mean flower length is equal for the F1 and F2 generations versus the alternative that is different.
 - (e) Interpret this test in the context of the problem.
5. Read problem 16 on page 311 for the background on a data set on the eye spans for male flies bred on two different diets. The data file `flies.csv` contains the data. Do not answer the textbook questions, but do this instead.
- (a) Use statistical software to make a plot that effectively compares the eye spans of flies raised on the two diets.
 - (b) Use statistical software to construct a 95% confidence interval for the difference in mean eye span length (corn diet minus cotton diet) using the t -distribution.
 - (c) Interpret the interval in the context of the problem.
 - (d) Use statistical software to test the hypothesis that the population mean eye spans are equal between the two diets versus the alternative that the mean is higher for the corn diet with a t -test.
 - (e) Use statistical software to repeat the previous test, but with a randomization test. Graph the sampling distribution of the test statistic and report a p-value.
 - (f) Interpret one of these tests in the context of the problem.

Note that you may use whichever statistical software package you choose to. The following may be helpful should you elect to use R.

1. Pages 26–29 from the document R for Statistics 571 on the course web page has much useful information relative to the tasks required for this assignment.
 2. The file `boot.R` contains R code for both the bootstrap and permutation tests needed for this assignment.
 3. You can enter this code into R by either (1) using the command `source("boot.R")` from the command line (assuming the working directory contains the file) or (2) using the Source option from the file menu.
 4. The function `boot.ci()` in `boot.R` returns a list with a confidence interval and the entire bootstrap sample of the test statistic. See the sockeye salmon mass example below for usage.
-

5. The function `perm.test()` in `boot.R` returns a list with a p-value and the entire permutation distribution of the test statistic. See the pseudoscorpion example below for usage.
6. The base function `t.test()` can be used for both paired and independent sample *t*-tests and confidence intervals. See slide notes on usage for arguments such as: `paired=T`, `alternative="greater"`, `conf=0.99`, and `var.equal=T`.
7. To read in data, it is necessary to enter a command from the command line; Source from the menu will *not* work as this command is to read in R code, not data.
8. Files saved in comma-separated-variable format (.csv) can be read using `read.csv()`. A header line with variable names is assumed.
9. Files with white space between values can be read using `read.table()`. A header line is *not* the default, so the argument `header=T` needs to be added when a header line is present.
10. The function `str()` is useful to examine the structure of data you have just read in to ensure that R understood it properly.
11. Variables in a data frame in R can be accessed using the `$` operator.
12. The function `with()` allows you to leave off the `data.frame$` prefix to each variable on the line; see examples below.
13. The square bracket operator can be used to grab part of a variable; see examples below.

Examples:

Randomization test with pseudoscorpion data

```
> pseudo = read.csv("pseudoscorpions.csv")
> str(pseudo)
> source("boot.R")
> out = with(pseudo, perm.test(broods[mating == "Different"], broods[mating ==
+   "Same"], alternative = "greater", B = 10000))
> out$p.value
> sum(out$test.sample >= out$test.stat)
> library(lattice)
> plot(densityplot(~out$test.sample, plot.points = F, n = 201))
> with(pseudo, t.test(broods[mating == "Different"], broods[mating ==
+   "Same"], alternative = "greater"))
```

Bootstrap with sockeye salmon data

```
> sockeye = read.table("sockeye.txt", header = T)
> str(sockeye)
> source("boot.R")
> out = boot.ci(sockeye$mass, mean, conf = 0.99)
> out$ci
> library(lattice)
> plot(densityplot(~out$boot.sample, plot.points = F, n = 201))
> t.test(sockeye$mass, conf = 0.99)
```