

Lab 8

Adam Orr

October 20, 2017

Section: 91973 Friday 9am

Question 1

Script

```
# a. Estimate the mean turning angle and calculate its exact 95% confidence
# interval.
setwd("~/code/biometry-lab/lab8")
turning <- read.csv('turning.csv')
turning <- turning$angle #turn list into vector

#estimate the mean
(xbar <- mean(turning))
s <- sd(turning)
n <- length(turning)
se <- s / sqrt(n)
df <- n - 1

#calculate exact confidence interval
critvals <- qt(c(0.025,.975), df = df) #get vector of critical values
(ci <- xbar + se * critvals) #since one critical value is positive and
#the other negative, this is equivalent to xbar +/- se * critvals

# b. Test the hypothesis that people tend to turn in one direction more on
# average than the other direction. That is, test whether the mean turning angle
# differs from zero. Carry out the test "by hand" in R, calculating each step
# yourself. Explicitly carry out the five steps of hypothesis testing presented
# in lecture. Clearly state your conclusion, along with the essential
# statistical information in parentheses, including the name of the test, value
# of the t statistic, the degrees of freedom, and the P value).

#step 1 - hypotheses
#h0: mean = 0
#ha: mean != 0

#step 2 - alpha
#alpha = 0.05

#step 3 - define test statistic
# t = (xbar - null) / se

#step 4 - calculate test statistic
null <- 0
(t <- (xbar - null)/se)
```

```

#step 5 - calculate the p-value and make a conclusion
(p <- 2 * (1 - pt(abs(t),df)))

#We fail to reject the null hypothesis that the mean equals zero with an alpha
#of 0.05 (t-test: t = -0.288, df = 13, p = 0.778)

# c. Repeat the test using R's t.test function. Make sure that your answers from
# part b match those given by t.test.
t.test(turning)

```

Output

```

# a. Estimate the mean turning angle and calculate its exact 95% confidence
# interval.
#estimate the mean

## [1] -0.135

#calculate exact confidence interval

## [1] -1.1463515  0.8763515

# b. Test the hypothesis that people tend to turn in one direction more on
# average than the other direction. That is, test whether the mean turning angle
# differs from zero. Carry out the test "by hand" in R, calculating each step
# yourself. Explicitly carry out the five steps of hypothesis testing presented
# in lecture. Clearly state your conclusion, along with the essential
# statistical information in parentheses, including the name of the test, value
# of the t statistic, the degrees of freedom, and the P value).

#step 1 - hypotheses
#h0: mean = 0
#ha: mean != 0

#step 2 - alpha
#alpha = 0.05

#step 3 - define test statistic
# t = (xbar - null) / se

#step 4 - calculate test statistic

## [1] -0.2883763

#step 5 - calculate the p-value and make a conclusion

## [1] 0.7776054

#We fail to reject the null hypothesis that the mean equals zero with an alpha
#of 0.05 (t-test: t = -0.288, df = 13, p = 0.778)

# c. Repeat the test using R's t.test function. Make sure that your answers from
# part b match those given by t.test.

##

```

```
## One Sample t-test
##
## data: turning
## t = -0.28838, df = 13, p-value = 0.7776
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -1.1463515 0.8763515
## sample estimates:
## mean of x
## -0.135
```

Answers

Question 2

Script

```
# a. Import the data into R and estimate the mean difference in number of
# species between areas upstream and downstream of a tributary. Calculate a 95%
# confidence interval for the estimate.
tributaries <- read.csv('tributaries.csv')

d <- tributaries$upstream - tributaries$downstream

#estimate the mean
(xbar <- mean(d))
s <- sd(d)
n <- length(d)
se <- s / sqrt(n)
df <- n - 1

#calculate exact confidence interval
critvals <- qt(c(0.025,.975), df = df) #get vector of critical values
(ci <- xbar + se * critvals) #since one critical value is positive and
#the other negative, this is equivalent to xbar +/- se * critvals

# b. Carry out (by hand) a test of the hypothesis that the number of species
# differs between upstream and downstream locations. Be sure to clearly state
# null and alternative hypotheses and to give a full statement of the
# conclusions of your test.

#step 1 - hypotheses
#h0: d = 0
#ha: d != 0

#step 2 - alpha
#alpha = 0.05

#step 3 - define test statistic
# t = xbar_d / se_d

#step 4 - calculate test statistic
```

```

(t <- (xbar/se))

#step 5 - calculate the p-value and make a conclusion
(p <- 2 * (1 - pt(abs(t),df)))

#We reject the null hypothesis that the mean difference between the number of
#species is 0 with an alpha of 0.05 (paired t-test: t = -2.2669, df = 14, p =
#0.0398)

# c. Perform the same test using R's t.test function.
t.test(tributaries$upstream, tributaries$downstream, paired = T)

```

Output

```

# a. Import the data into R and estimate the mean difference in number of
# species between areas upstream and downstream of a tributary. Calculate a 95%
# confidence interval for the estimate.
#estimate the mean

```

```
## [1] -1.8
```

```
#calculate exact confidence interval
```

```
## [1] -3.50301491 -0.09698509
```

```

# b. Carry out (by hand) a test of the hypothesis that the number of species
# differs between upstream and downstream locations. Be sure to clearly state
# null and alternative hypotheses and to give a full statement of the
# conclusions of your test.

```

```
#step 1 - hypotheses
```

```
#h0: d = 0
```

```
#ha: d != 0
```

```
#step 2 - alpha
```

```
#alpha = 0.05
```

```
#step 3 - define test statistic
```

```
# t = xbar_d / se_d
```

```
#step 4 - calculate test statistic
```

```
## [1] -2.26693
```

```
#step 5 - calculate the p-value and make a conclusion
```

```
## [1] 0.03976278
```

```

#We reject the null hypothesis that the mean difference between the number of
#species is 0 with an alpha of 0.05 (paired t-test: t = -2.2669, df = 14, p =
#0.0398)

```

```
# c. Perform the same test using R's t.test function.
```

```
##
```

```
## Paired t-test
```

```
##
## data: tributaries$upstream and tributaries$downstream
## t = -2.2669, df = 14, p-value = 0.03976
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.50301491 -0.09698509
## sample estimates:
## mean of the differences
## -1.8
```

Answers

Question 3

Script

```
# a. Read this data into R and use it to estimate mean drinking time for each
# treatment (straight and curved).
beer <- read.csv('beer.csv')

#straight glass mean
(xbar_straight <- mean(beer$Straight))

#curved glass mean
(xbar_curved <- mean(beer$Curved))

# b. Calculate the pooled sample variance. Use it to calculate the 95%
# confidence intervals of the mean drinking time for each treatment.
nstraight <- length(beer$Straight)
ncurved <- length(beer$Curved)

s2_pooled <- (sum((beer$Straight - xbar_straight)^2) + sum((beer$Curved - xbar_curved)^2))/(nstraight +
se_diff <- sqrt(s2_pooled/nstraight + s2_pooled/ncurved)
df <- nstraight + ncurved - 2

critvals <- qt(c(0.025,.975), df = df) #get vector of critical values
(ci_straight <- (xbar_straight + se_diff * critvals))
(ci_curved <- (xbar_curved + se_diff * critvals))

# c. Why is it better to use the pooled variance rather than the separate sample
# variances for each treatment?

#Since a t-test assumes that both population standard deviations are equal, by
#using more samples to estimate the the standard deviation increases the
#precision of the estimate.

# d. Make an error bar plot showing the two estimates and their confidence
# intervals. Based on your inspection of these intervals, does it seem likely
# that the mean drinking times are different for the two glass shapes?
library('gplots')
plotCI(c(xbar_straight,xbar_curved),
      uiw = (critvals[2]*se_diff),
```

```

xaxt="n",
yaxt="n",
xlab = "Glass Type",
xlim = c(0.5, 2.5),
ylim = c(2,20),
ylab = "Time to consume beer (minutes)")
axis(1, at = c(1,2), labels = c("Straight", "Curved"))
axis(2, at = seq(2,20,2))

#Based on a visual inspection, it seems that the mean drinking times are
#different, though there is a slight overlap of the intervals.

# e. Now estimate the difference in mean drinking time between the two
# treatments and calculate a 95% confidence interval for this difference.
(xbar <- xbar_straight - xbar_curved)
(ci_diff <- xbar + critvals * se_diff)

# f. Carry out a t-test of the hypothesis that the mean drinking time differs
# between the two groups. Be sure to clearly state null and alternative
# hypotheses and to give a full statement of the conclusions of your test.

#step 1 - hypotheses
#h0: xbar_straight = xbar_curved
#ha: xbar_straight != xbar_curved

#step 2 - alpha
#alpha = 0.05

#step 3 - define test statistic
# t = xbar_diff / se_p

#step 4 - calculate test statistic
(t <- (xbar/se_diff))

#step 5 - calculate the p-value and make a conclusion
(p <- 2 * (1 - pt(abs(t),df)))

#We reject the null hypothesis that the mean minutes to finish a beer in a
#straight glass is equal to the mean minutes to finish a beer in a curved glass
#with an alpha of 0.05 (two-sample t-test: t = 3.577, df = 18, p = 0.0022)

# g. Perform the same test using R's t.test function.
t.test(beer$Straight, beer$Curved, alternative = 't', var.equal = TRUE)

```

Output

```

# a. Read this data into R and use it to estimate mean drinking time for each
# treatment (straight and curved).
#straight glass mean

```

```
## [1] 14.913
```

```
#curved glass mean
```

```
## [1] 7.624
```

```
# b. Calculate the pooled sample variance. Use it to calculate the 95%  
# confidence intervals of the mean drinking time for each treatment.
```

```
## [1] 10.6313 19.1947
```

```
## [1] 3.342299 11.905701
```

```
# c. Why is it better to use the pooled variance rather than the separate sample  
# variances for each treatment?
```

```
#Since a t-test assumes that both population standard deviations are equal, by  
#using more samples to estimate the the standard deviation increases the  
#precision of the estimate.
```

```
# d. Make an error bar plot showing the two estimates and their confidence  
# intervals. Based on your inspection of these intervals, does it seem likely  
# that the mean drinking times are different for the two glass shapes?
```

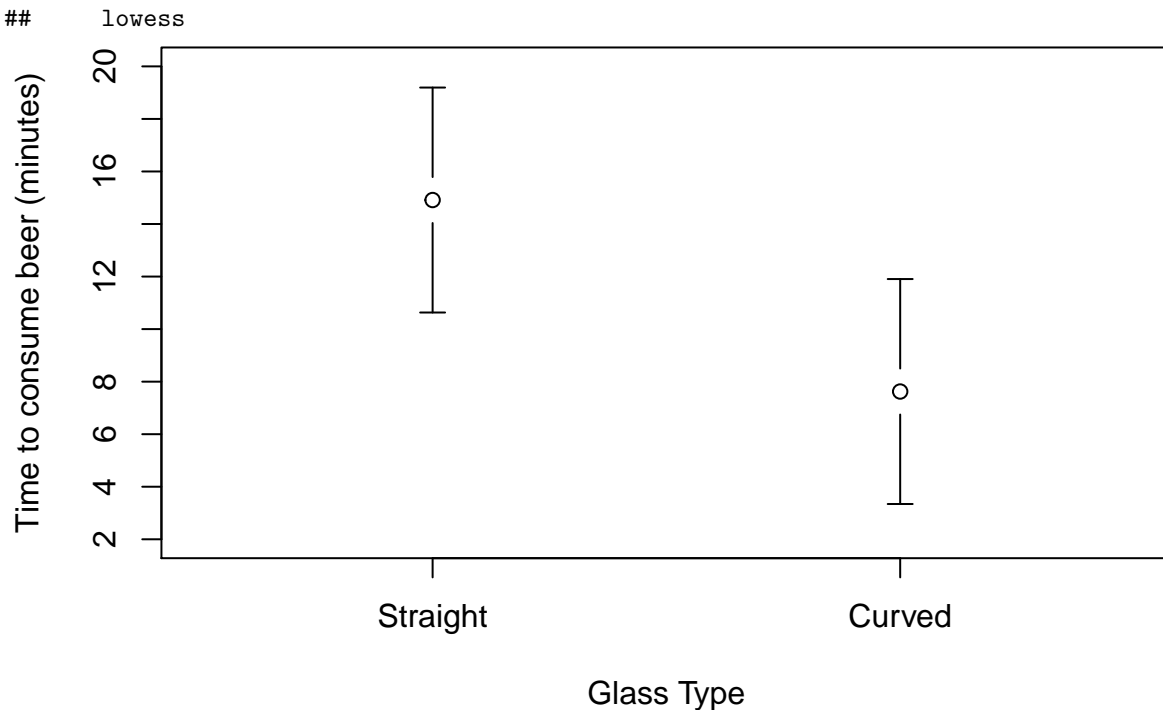
```
##
```

```
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
## lowess
```



```
#Based on a visual inspection, it seems that the mean drinking times are  
#different, though there is a slight overlap of the intervals.
```

```
# e. Now estimate the difference in mean drinking time between the two  
# treatments and calculate a 95% confidence interval for this difference.
```

```
## [1] 7.289
```

```
## [1] 3.007299 11.570701
```

```
# f. Carry out a t-test of the hypothesis that the mean drinking time differs  
# between the two groups. Be sure to clearly state null and alternative  
# hypotheses and to give a full statement of the conclusions of your test.
```

```
#step 1 - hypotheses  
#h0: xbar_straight = xbar_curved  
#ha: xbar_straight != xbar_curved
```

```
#step 2 - alpha  
#alpha = 0.05
```

```
#step 3 - define test statistic  
# t = xbar_diff / se_p
```

```
#step 4 - calculate test statistic
```

```
## [1] 3.576527
```

```
#step 5 - calculate the p-value and make a conclusion
```

```
## [1] 0.002156914
```

```
#We reject the null hypothesis that the mean minutes to finish a beer in a  
#straight glass is equal to the mean minutes to finish a beer in a curved glass  
#with an alpha of 0.05 (two-sample t-test: t = 3.577, df = 18, p = 0.0022)  
# g. Perform the same test using R's t.test function.
```

```
##  
## Two Sample t-test  
##  
## data: beer$Straight and beer$Curved  
## t = 3.5765, df = 18, p-value = 0.002157  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 3.007299 11.570701  
## sample estimates:  
## mean of x mean of y  
## 14.913 7.624
```

Answers