

QSCI 482 Story 4

Simon Hans Edasi

2023-11-06

```
# quick function to make plotting faster
t_plot <- function(alpha=0.05, dof = 5, t_obs = 0){

  # define rejection regions
  rej_left <- qt(alpha/2, df = dof)
  rej_right <- qt(alpha/2, df = dof, lower.tail = F)
  #plot the curve with rejection regions
  x <- seq(-4, 4, length.out = 100)
  plot(
    x,
    y = dt(x, df = dof),
    type = 'l')
  ## x and y for the whole area
  xReject <- c(seq(-4, 4, by=0.01))
  yReject <- dt(xReject, df = dof)
  yReject[xReject > rej_left & xReject < rej_right] <- 0
  polygon(c(xReject, xReject[length(xReject)]), xReject[1]),
    c(yReject, 0, 0), col='red')

  abline(v = t_obs)
}
```

Question 1

Browse all of the sheets in the Excel workbook to familiarize yourself with the data, and look at the pictures of what each measurement represents. Create an R project for Assignment 4, start a new .R script file, and save it. Follow the instructions from the Lab 4 handout to:

a. Create a .csv file (suggestion “Sebastes_diaconus_data.csv”) from the appropriate sheet in Excel (just the data, not any accompanying analysis) and load it into R. Examine the read-in data to ensure that text columns remain as text, and that there are not any extra columns or rows in the data. Upload the .csv file to Canvas as part of your solution.

```
df <- read.csv('Sebastes_diaconus_data.csv')
# df
df <- df[!(is.na(df$FishNumber)), ]
write.csv(df, 'fish_data.csv', row.names = F)
```

b. Write R code to extract the data for the caudal peduncle of Deacon Rockfish from the .csv file and save it to a variable.

```
caud_ped_d <- df[df$FishID == 'S. diaconus',]$CaudalPeduncle
caud_ped_m <- df[df$FishID == 'S. mystinus',]$CaudalPeduncle
```

Question 2

The variance for the two putative species is quite similar for the caudal peduncle data. Run a statistical test to test if the means of caudal peduncle length (data are percentages of standard length) for the two samples of the putative species are equal, assuming that variances are equal.

a. Run the test the long way around (calculating the appropriate test statistic, the p-value, critical values of the test statistic, and reaching conclusions for $\alpha = 0.05$)

We want to run a two sample ttest of equal variance. We can use pooled variance as the best estimator of the variance.

$$H_0 : \mu_d = \mu_m$$

$$H_A : \mu_d \neq \mu_m$$

```
alpha <- 0.05

n1 <- length(caud_ped_d)
X1 <- mean(caud_ped_d)
var_X1 <- var(caud_ped_d)

n2 <- length(caud_ped_m)
X2 <- mean(caud_ped_m)
var_X2 <- var(caud_ped_m)

dof <- n1 + n2 - 2
t_crit <- qt(alpha/2, df = dof, lower.tail = F)

s <- (var_X1 / n1) + (var_X2 / n2)

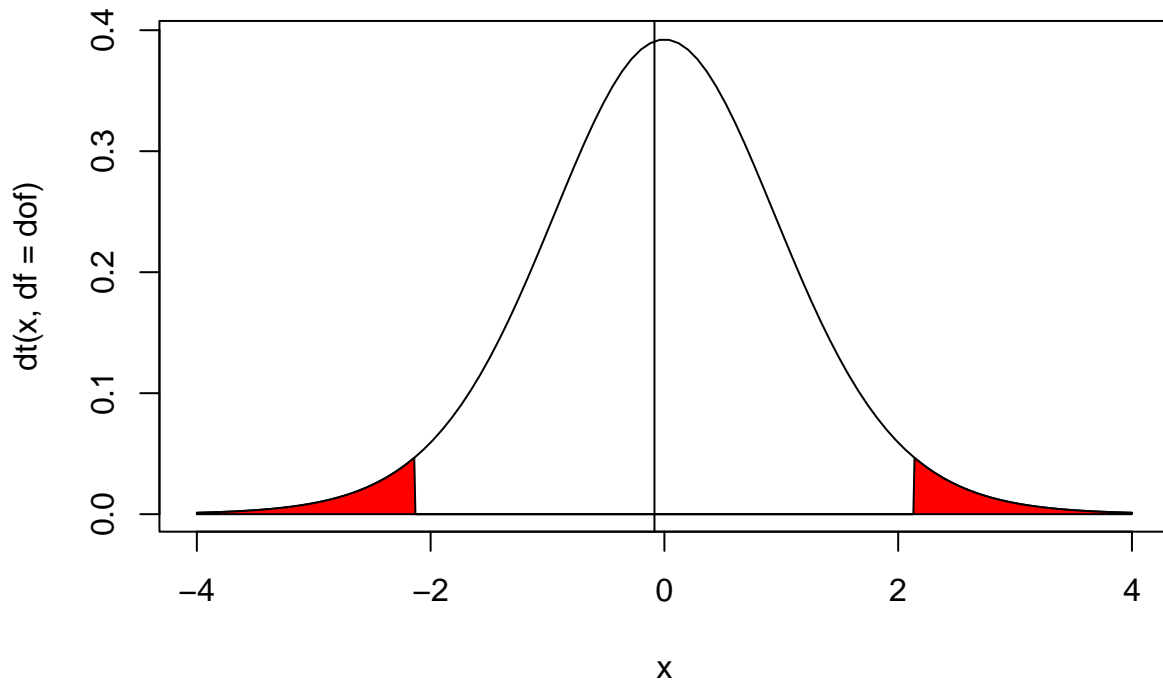
t_obs <- (X1 - X2) / sqrt(s)

p_value <- 2*pt(t_obs, df = dof)

library(glue)
glue('ratio of variances: {var(caud_ped_d) / var(caud_ped_m)}\ndiaconus mean: {X1}\nmystinus mean: {X2}\ndegrees of freedom: {dof}\nt statistic: {t_obs}\ncritical value: +/- {t_crit}\np-value: {p_value}\n')
```

```
## ratio of variances: 1.0562705300684
## diaconus mean: 18.30375
## mystinus mean: 18.3477777777778
## degrees of freedom: 15
## t statistic: -0.0851680399823201
## critical value: +/- 2.13144954555978
## p-value: 0.93325416949314
```

```
t_plot(alpha, dof, t_obs)
```



With a confidence $\alpha = 0.05$, we should not reject the null hypothesis there is no difference in mean caudal peduncle between species with a p-value of 0.933 and 15 degrees of freedom.

Run the test using a built-in statistical test in R, reporting the key results.

```
t.test(caud_ped_d, caud_ped_m, alternative = 'two.sided', var.equal = T)
```

```
##
## Two Sample t-test
##
## data:  caud_ped_d and caud_ped_m
## t = -0.085314, df = 15, p-value = 0.9331
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.143994  1.055939
## sample estimates:
## mean of x mean of y
##  18.30375  18.34778
```

The key result is the t-statistic and the p-value match what we calculated and the conclusion is still to not reject the null and conclude the mean caudal peduncle length is the same between species.

Question 3

The variance of anal spine I is much greater for Blue Rockfish than Deacon Rockfish. Run a statistical test to test if the means of anal spine I lengths (data are percentages of standard length) for the two samples of the putative species are equal, assuming that variances are not equal.

a. Run the test the long way around (calculating the appropriate test statistic, the p-value, critical values of the test statistic, and reaching conclusions for $\alpha = 0.05$)

```
as1_d <- df[df$FishID == 'S. diaconus',]$AnalSpineI
as1_m <- df[df$FishID == 'S. mystinus',]$AnalSpineI
```

Variances are not equal, so we will need to run a Welch's two sample t-test.

$$H_0 : \mu_{Diaconus} = \mu_{Mystinus}$$

$$H_A : \mu_{Diaconus} \neq \mu_{Mystinus}$$

```
alpha <- 0.05
n1 <- length(as1_d)
X1 <- mean(as1_d)
s_X1 <- var(as1_d) / n1

n2 <- length(as1_m)
X2 <- mean(as1_m)
s_X2 <- var(as1_m) / n2

t_prime <- (X1 - X2) / sqrt(s_X1 + s_X2)

dof_prime <- (s_X1 + s_X2)^2 / ( ((s_X1)^2 / (n1 - 1)) + ((s_X2)^2 / (n2 - 1)) )

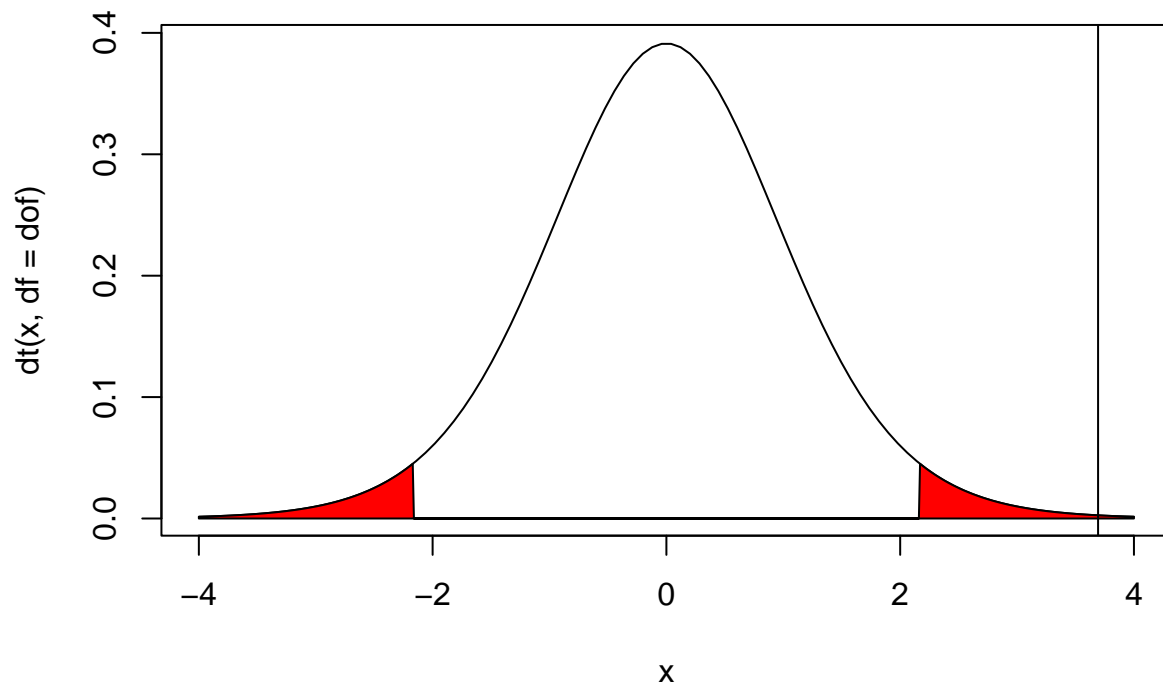
t_crit <- qt(alpha/2, df = dof_prime, lower.tail = F)

p_value <- 2*pt(t_prime, df = dof_prime, lower.tail = F)

glue('ratio of variances: {var(as1_d) / var(as1_m)}\ndiaconus mean: {X1}
mystinus mean: {X2}\ndegrees of freedom: {dof_prime}\nt statistic: {t_prime}
critical value: +/- {t_crit}\np-value: {p_value}
')
```

```
## ratio of variances: 0.319823293545337
## diaconus mean: 5
## mystinus mean: 3.797777777777778
## degrees of freedom: 12.8859835901609
## t statistic: 3.69313237279812
## critical value: +/- 2.16231337875695
## p-value: 0.00274316369210779
```

```
t_plot(alpha, dof_prime, t_prime)
```



We find a statistical difference in anal spine I with p-value = 0.002743, 12.886 degrees of freedom.

b. Run the test using a built-in statistical test in R, reporting the key results

```
t.test(as1_d, as1_m)
```

```
##
##  Welch Two Sample t-test
##
## data:  as1_d and as1_m
## t = 3.6931, df = 12.886, p-value = 0.002743
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4983262 1.9061182
## sample estimates:
## mean of x mean of y
##  5.000000  3.797778
```

We find a statistical difference in anal spine I with p-value = 0.002743, 12.886 degrees of freedom.

Question 4

Now run statistical tests on the other measurements using the appropriate built-in statistical tests in R. You will need to decide if the variances are similar or not. Report the key results for differences in the means for:

a. The symphyseal knob

```
symp_d <- df[df$FishID == 'S. diaconus',]$SymphysealKnob
symp_m <- df[df$FishID == 'S. mystinus',]$SymphysealKnob
var(symp_m) / var(symp_d)
```

```
## [1] 1.979925
```

```
glue('ratio of variances: {var(symp_d) / var(symp_m)}')
```

```
## ratio of variances: 0.505069612590799
```

Our ratio of variances is between $\frac{1}{3}$ and 3, so we're close enough to consider them equal. We can use a simple t-test.

```
t.test(symp_d,symp_m,var.equal = T)
```

```
##
```

```
## Two Sample t-test
```

```
##
```

```
## data: symp_d and symp_m
```

```
## t = 0.88527, df = 15, p-value = 0.39
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -0.1824109 0.4415776
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 1.756250 1.626667
```

There is no statistically significant difference in mean length of symphyseal knob between the two rockfish with a p-value 0.39 and 15 degrees of freedom.

b. The anal spine II

```
as2_d <- df[df$FishID == 'S. diaconus',]$AnalSpineII
as2_m <- df[df$FishID == 'S. mystinus',]$AnalSpineII
glue('ratio of variances: {var(as2_d) / var(as2_m)}')
```

```
## ratio of variances: 0.61807808751356
```

Our ratio of variances is between $\frac{1}{3}$ and 3, so we're close enough to consider them equal. We can use a simple t-test.

```
t.test(as2_d,as2_m,var.equal = T)
```

```
##
```

```
## Two Sample t-test
```

```
##
```

```
## data: as2_d and as2_m
```

```
## t = 4.542, df = 15, p-value = 0.0003893
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## 0.9501438 2.6304117
## sample estimates:
## mean of x mean of y
## 10.912500 9.122222
```

There is a statistically significant difference in mean length of anal spine 2 between the two rockfish with a p-value 0.00039 and 15 degrees of freedom.

c .The anal spine III

```
as3_d <- df[df$FishID == 'S. diaconus',]$AnalSpineIII
as3_m <- df[df$FishID == 'S. mystinus',]$AnalSpineIII
glue('ratio of variances: {var(as3_d) / var(as3_m)}')
```

```
## ratio of variances: 0.95236464378979
```

Our ratio of variances is between $\frac{1}{3}$ and 3, so we're close enough to consider them equal. We can use a simple t-test.

```
t.test(as3_d,as3_m,var.equal = T)
```

```
##
## Two Sample t-test
##
## data: as3_d and as3_m
## t = 2.2757, df = 15, p-value = 0.03797
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.07488034 2.28845300
## sample estimates:
## mean of x mean of y
## 10.825000 9.643333
```

There is a statistically significant difference in mean length of anal spine 3 between the two rockfish with a p-value 0.038 and 15 degrees of freedom.

Question 5

What do you conclude from these tests and the data: should blue rockfish have been separated into two species or not? Explain your reasoning.

From the gathered data, we see that the fish have similar caudal peduncle lengths, however the three anal spines are different lengths with a statistical significance of 0.05. Considering both genetic and morphometric considerations, this fish should be divided into two sub species.