

BEE552 Biometry Week 5

Maria Feiler

2/23/2022

My Learning Journey

Over the last week, I participated in Biometry in the following ways:

- I asked / answered **6** questions posed in class.
- I asked **2** questions in Slack.
- I answered **0** questions posed by other students on Slack.
- I came to Heather's office hours: **Yes**
- I came to Jose's office hours: **Yes**
- I met with Heather or Jose separately from office hours: **No**

Anything not falling into one of the above categories?

No

On a scale of 1 (no knowledge) to 10 (complete expert), how would I rate my comfort with R programming after this week?

7

Any topics from last week that you are still confused about?

Doing fine for now

Problem Set

Part 1

One concern about applying metal tags to penguins is that doing so may increase drag in the water. In a study of the impact of foraging tags on penguin foraging efficiency, the mean length of 344 foraging trips for penguins with a metal tag was 12.70 days with a standard deviation of 3.71 days. For those with an electronic tag (which presumably would not increase drag), the mean foraging trip length was 11.60 days with standard deviation of 4.53 days over 512 trips. Calculate the confidence interval for the difference in foraging trip length and calculate the p-value for the null hypothesis that there is no difference between tag types. (Do not use the R function 't.test'. I'm asking you to calculate these yourself.) Do these data provide evidence that mean foraging trips are longer for penguins with a metal tag?

An unpaired two sample t-test will be used to determine if there is a difference in foraging lengths of penguins with different tag lengths.

$$\begin{array}{lll} H_{null} : \bar{x}_{metal} = \bar{x}_{electronic} & \bar{x}_{metal} = 12.70 & \bar{x}_{electronic} = 11.60 \\ H_{alt} : \bar{x}_{metal} \neq \bar{x}_{electronic} & s_{metal} = 3.71 & s_{electronic} = 4.53 \\ & n_{metal} = 344 & n_{electronic} = 512 \end{array}$$

Assuming the distributions of penguin foraging lengths are normally distributed, then the distributions of foraging lengths are represented by $X_{metal} \sim N(\bar{x}_{metal}, s_{metal})$ and $X_{metal} \sim N(\bar{x}_{electronic}, s_{electronic})$ for metal and electronic tags, respectively.

Assuming that penguin foraging lengths between those with metal versus electronic tag are independent, then...

$$\bar{X}_{metal} - \bar{X}_{electronic} \sim N(\bar{x}_{metal} - \bar{x}_{electronic}, \frac{s_{metal}^2}{n_{metal}} + \frac{s_{electronic}^2}{n_{electronic}})$$

Therefore, the standard error of the difference of means is given by...

$$SE = \sqrt{\frac{s_{metal}^2}{n_{metal}} + \frac{s_{electronic}^2}{n_{electronic}}},$$

the test statistic is...

$$T^* = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}},$$

and the degrees of freedom are calculated using...

$$dof = \frac{\left(\frac{s_{metal}^2}{n_{metal}} + \frac{s_{electronic}^2}{n_{electronic}}\right)^2}{\frac{\left[\frac{s_{metal}^2}{n_{metal}}\right]^2}{n_{metal}-1} + \frac{\left[\frac{s_{electronic}^2}{n_{electronic}}\right]^2}{n_{electronic}-1}}.$$

Finally, the 95 percentile confidence interval of the difference in means is given by...

$$\bar{x}_{metal} - \bar{x}_{electronic} \pm z \sqrt{\frac{s_{metal}^2}{n_{metal}} + \frac{s_{electronic}^2}{n_{electronic}}}$$

... where z is 1.96, or the z-value corresponding with 95% confidence.

```

# Penguins with metal tags
metal <- list(12.70, 3.71, 344, 3.71^2)
names(metal) <- c("mean", "sd", "n", "var")

# Penguins with electronic tags
electronic <- list(11.60, 4.53, 512, 4.53^2)
names(electronic) <- c("mean", "sd", "n", "var")

# Create list for hypothesis testing results
penguinHypTest <- vector("list", length = 4)
names(penguinHypTest) <- c("stErr", "t", "dof", "p")

# Calculate standard error of the means
penguinHypTest$stErr <- sqrt(((metal$var)/metal$n)+((electronic$var)/electronic$n))

# Calculate the test statistic
penguinHypTest$t <- (metal$mean-electronic$mean)/penguinHypTest$stErr

# Calculate the degrees of freedom
top <- ((metal$var/metal$n)+(electronic$var/electronic$n))^2
bottom1 <- ((metal$var/metal$n)^2)/(metal$n-1)
bottom2 <- ((electronic$var/electronic$n)^2)/(electronic$n-1)

penguinHypTest$dof <- top/(bottom1 + bottom2)

# Find the p value
penguinHypTest$p <- pt(penguinHypTest$t,
                       penguinHypTest$dof,
                       lower.tail = TRUE
                       )

# Calculate the 95% confidence interval
penguinsCI <- vector("list", length = 2)
names(penguinsCI) <- c("LL", "UL")

penguinsCI$LL <- metal$mean-electronic$mean-1.96*penguinHypTest$stErr
penguinsCI$UL <- metal$mean-electronic$mean+1.96*penguinHypTest$stErr

```

Table 1: Two-Sample t-Test Results

SE	T*	dof	p
0.28	3.89	821.22	1

The 95% confidence interval of the difference of mean foraging length between tag types is (0.545, 1.655). The difference in foraging lengths between penguins with different tag types is not statistically significant ($p = 1$). Therefore, the tag type has no appreciable influence on penguin foraging efficiency.