

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 I

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 $X_{electronic} \sim N(\bar{x}_{electronic}, s_{electronic})$ 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$p <- 2*pt(penguinHypTest$t,
                        penguinHypTest$dof
                        )

# 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	2

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 ($p = 2$). Therefore, the tag type has no appreciable influence on penguin foraging efficiency.

Part II

Download the dataset “diabetes.csv”. This dataset describes a diabetes study by Willems et al. 1997 in which blood glucose (stab.glu) was recorded for males and females (gender) in each of two locations (location).

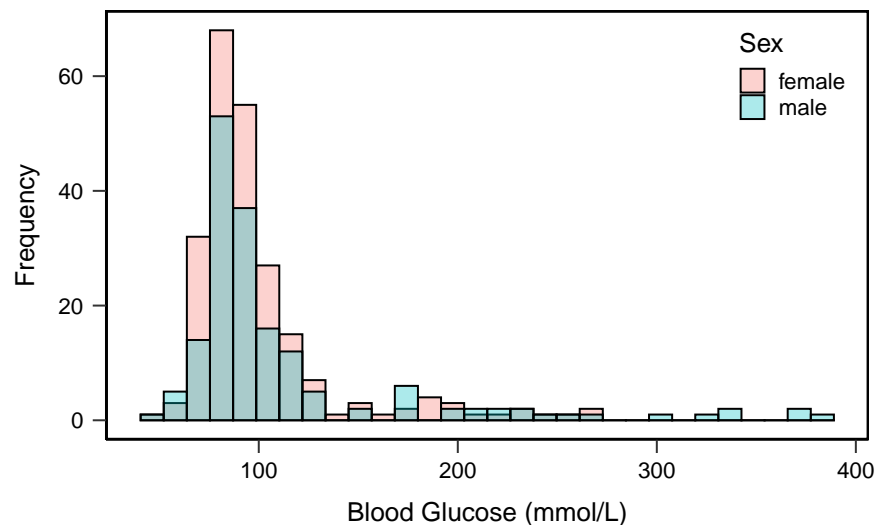
```
link <- "https://raw.githubusercontent.com/hlynch/Biometry2022/master/_data/diabetes.csv"
diabetes <- read_csv(link,
  show_col_types = FALSE
)

# Rename "gender" to "sex" for biological accuracy, as "male" and "female"
# are not genders
diabetes <- rename(diabetes, sex = gender)
```

Table 2: A selection of data from Willems et al. 1997.

Location	Sex	Blood Glucose
Louisa	male	74.17
Buckingham	female	145.35
Louisa	female	95.10
Buckingham	male	71.40
Louisa	male	83.42
Louisa	male	88.24
Louisa	female	90.15
Buckingham	female	75.22
Buckingham	male	80.03
Buckingham	male	104.12

Male and Female Blood Glucose Levels in England



Question 1 *Aggregating across locations, calculate the test statistic for the t-test to test the null hypothesis that males and females do not differ in their average blood glucose levels and compare it to its distribution under the null hypothesis.*

An unpaired two sample t-test will be used to determine if there is a difference between male and female blood glucose levels in England.

$$H_{null} : \bar{x}_{male} = \bar{x}_{female} \quad H_{alt} : \bar{x}_{male} \neq \bar{x}_{female}$$

```
# Males
males <- list(mean(diabetes$stab.glu[diabetes$sex == "male"]),
              sd(diabetes$stab.glu[diabetes$sex == "male"]),
              sum(diabetes$sex == "male"),
              var(diabetes$stab.glu[diabetes$sex == "male"])
            )
names(males) <- c("mean", "sd", "n", "var")

# Females
females <- list(mean(diabetes$stab.glu[diabetes$sex == "female"]),
                sd(diabetes$stab.glu[diabetes$sex == "female"]),
                sum(diabetes$sex == "female"),
                var(diabetes$stab.glu[diabetes$sex == "female"])
              )
names(females) <- c("mean", "sd", "n", "var")

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

# Calculate standard error of the means
diabetesHypTest$stErr <- sqrt(((males$var)/males$n)+((females$var)/females$n))

# Calculate the test statistic
diabetesHypTest$t <- (males$mean-females$mean)/diabetesHypTest$stErr

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

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

# Find the p value
diabetesHypTest$p <- 2*pt(diabetesHypTest$t,
                        diabetesHypTest$dof,
                        lower = FALSE
                      )

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

diabetesCI$LL <- males$mean-females$mean-1.96*diabetesHypTest$stErr
diabetesCI$UL <- males$mean-females$mean+1.96*diabetesHypTest$stErr
```

Table 3: Male and female blood glucose two-sample t-test results.

SE	T*	dof	p	95CI Lower	95CI Upper
5.471	2.318	251.076	0.021	1.959	23.406

Question 2 Compare your results with the R function “t.test”. (Copy and paste results below.) Explain what every single number reported in the output of t.test means.

```
diabetesttest <- t.test(diabetes$stab.glu[diabetes$sex == "male"],
                        diabetes$stab.glu[diabetes$sex == "female"]
                        )
```

Welch Two Sample t-test

```
data: Males and Females
t = 2.318, df = 251.08, p-value = 0.02125
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.906998 23.458207
sample estimates:
 Mean Male Glucose Mean Female Glucose
    112.49538         99.81278
```

The p-values produced by the t.test() calculations are the same as the p-values obtained from manual test statistic calculations.

Question 3 Interpret your results in words as you would in the results section of a manuscript (1-2 sentences)

An unpaired two sample t-test was used to test if there was a significant difference between the blood glucose levels of males (112.5 ± 63.61) and females (99.81 ± 37.12). At the critical value of $\alpha_C = 0.05$, the difference between male and female blood glucose levels is statistically significant ($p = 0.021$, $t^* = 2.318$, $df = 251.08$), and the null hypothesis is rejected. The 95th percentile confidence interval for the difference in mean blood glucose levels between males and females is (1.96, 23.41), suggesting that the females had significantly lower blood glucose levels than their male counterparts.

Question 4 How would the results differ if you used the median rather than the mean as the test statistic? Write a script to test the null hypothesis $H_0 : \text{median}_{\text{male}} = \text{median}_{\text{female}}$. Keep in mind that you will have to determine the distribution of the test statistic under the null hypothesis.

```
# This code was sourced in part from the original poster's updated code:
# https://stats.stackexchange.com/questions/176691/running-a-permutation-test-with-
# different-sample-sizes-in-r/176714

# Calculate the sample difference in medians
theta_hat <- median(diabetes$stab.glu[diabetes$sex == "male"]) -
  median(diabetes$stab.glu[diabetes$sex == "female"])

# Create vector capture the permutation test results
dif <- c()
```

```

# Create vector of the differences between the sample permutation medians
for (i in 1:1000) {
  samp <- sample(x = c(diabetes$stab.glu[diabetes$sex == "male"],
                      diabetes$stab.glu[diabetes$sex == "female"]),
                replace = FALSE)

  samp_male <- samp[1:sum(diabetes$sex == "male")]
  samp_female <- samp[sum(diabetes$sex == "male")+1:sum(diabetes$sex == "female")]

  dif[i] <- median(samp_male) - median(samp_female)
}

p_value <- sum(abs(dif) > abs(theta_hat))/length(dif)

```

The difference between the male and female permutation medians was 0.04 ± 2.2 . The quartiles of the permutations were -7.03(0%), -1.82(25%), 0(50%), 1.92(75%), 5.72(100%). The p-value of the original sample is 0.876. At $\alpha_c = 0.05$, the null hypothesis is rejected..

The difference in median blood glucose levels between males and females is (0.48), suggesting that the females had significantly lower blood glucose levels than their male counterparts.

Question 5 *If these were your data, under what circumstances would you feel comfortable lumping the data for each location together? Assuming you could lump the data together, what would be one advantage of doing so?*

In the case of this data, it would depend on what question I was asking. For something like blood glucose, which is dependent on so many socioeconomic and biological factors, my first instinct is to not lump the locations together to preserve potential differences in food quality, stress, health care availability, etc. If I were interested in the whole of England, I would probably lump them together, though I would question why I only sampled two cities to characterize a whole country. In the event I can lump the data together, I am gaining more power in my statistical tests due to the higher sample size and the assumption that the variance/mean is similar across the samples.

Extra Work

Part II Questions 2-4, Locations Separated

Question 1 *Aggregating across locations, calculate the test statistic for the t-test to test the null hypothesis that males and females do not differ in their average blood glucose levels and compare it to its distribution under the null hypothesis.*

```
# Get unique location names
uniqueLoc <- unique(diabetes$location)
```

Louisa Buckingham

```
# Divide into two datasets based on location
Louisa <- diabetes[which(diabetes$location == uniqueLoc[1]), c("sex", "stab.glu")]

Bucky <- diabetes[which(diabetes$location == uniqueLoc[2]), c("sex", "stab.glu")]
```

Louisa

An unpaired two sample t-test will be used to determine if there is a difference between male and female blood glucose levels in Louisa, England.

$$H_{null} : \bar{x}_{male} = \bar{x}_{female} \quad H_{alt} : \bar{x}_{male} \neq \bar{x}_{female}$$

```
# Males
LMales <- list(mean(Louisa$stab.glu[Louisa$sex == "male"]),
               sd(Louisa$stab.glu[Louisa$sex == "male"]),
               sum(Louisa$sex == "male"),
               var(Louisa$stab.glu[Louisa$sex == "male"])
               )
names(LMales) <- c("mean", "sd", "n", "var")

# females
Lfemales <- list(mean(Louisa$stab.glu[Louisa$sex == "female"]),
                 sd(Louisa$stab.glu[Louisa$sex == "female"]),
                 sum(Louisa$sex == "female"),
                 var(Louisa$stab.glu[Louisa$sex == "female"])
                 )
names(Lfemales) <- c("mean", "sd", "n", "var")

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

# Calculate standard error of the means
LouisaHypTest$stErr <- sqrt(((LMales$var)/LMales$n)+((Lfemales$var)/Lfemales$n))

# Calculate the test statistic
LouisaHypTest$t <- (LMales$mean-Lfemales$mean)/LouisaHypTest$stErr

# Calculate the degrees of freedom
top <- ((LMales$var/LMales$n)+(Lfemales$var/Lfemales$n))^2
```



```

bottom1 <- ((LMales$var/LMales$n)^2)/(LMales$n-1)
bottom2 <- ((Lfemales$var/Lfemales$n)^2)/(Lfemales$n-1)

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

# Find the p value
LouisaHypTest$p <- 2*pt(LouisaHypTest$t,
                        LouisaHypTest$dof
                        )

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

LouisaCI$LL <- LMales$mean-Lfemales$mean-1.96*LouisaHypTest$stErr
LouisaCI$UL <- LMales$mean-Lfemales$mean+1.96*LouisaHypTest$stErr

```

Table 4: Louisa male and female blood glucose two-sample t-test results.

SE	T*	dof	p	95CI Lower	95CI Upper
8.298	1.803	106.714	1.926	-1.302	31.226

Buckingham

An unpaired two sample t-test will be used to determine if there is a difference between male and female blood glucose levels in Buckingham, England.

$$H_{null} : \bar{x}_{male} = \bar{x}_{female} \quad H_{alt} : \bar{x}_{male} \neq \bar{x}_{female}$$

```
# Males
BMales <- list(mean(Bucky$stab.glu[Bucky$sex == "male"]),
               sd(Bucky$stab.glu[Bucky$sex == "male"]),
               sum(Bucky$sex == "male"),
               var(Bucky$stab.glu[Bucky$sex == "male"]))
names(BMales) <- c("mean", "sd", "n", "var")

# Females
Bfemales <- list(mean(Bucky$stab.glu[Bucky$sex == "female"]),
                 sd(Bucky$stab.glu[Bucky$sex == "female"]),
                 sum(Bucky$sex == "female"),
                 var(Bucky$stab.glu[Bucky$sex == "female"]))
names(Bfemales) <- c("mean", "sd", "n", "var")

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

# Calculate standard error of the means
BuckyHypTest$stErr <- sqrt(((BMales$var)/BMales$n)+((Bfemales$var)/Bfemales$n))

# Calculate the test statistic
BuckyHypTest$t <- (BMales$mean-Bfemales$mean)/BuckyHypTest$stErr

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

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

# Find the p value
BuckyHypTest$p <- 2*pt(BuckyHypTest$t,
                      BuckyHypTest$dof
                      )

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

BuckyCI$LL <- BMales$mean-Bfemales$mean-1.96*BuckyHypTest$stErr
BuckyCI$UL <- BMales$mean-Bfemales$mean+1.96*BuckyHypTest$stErr
```

Table 5: Buckingham male and female blood glucose two-sample t-test results.

SE	T*	dof	p	95CI Lower	95CI Upper
7.246	1.434	149.953	1.846	-3.814	24.591

Question 2 Compare your results with the R function “t.test”. (Copy and paste results below.) Explain what every single number reported in the output of t.test means.

```
ttestLouisa <- t.test(Louisa$stab.glu[Louisa$sex == "male"],
                     Louisa$stab.glu[Louisa$sex == "female"])

ttestBucky <- t.test(Bucky$stab.glu[Louisa$sex == "male"],
                     Bucky$stab.glu[Louisa$sex == "female"])
)
```

Welch Two Sample t-test

```
data: Louisa Males and females
t = 1.8031, df = 106.71, p-value = 0.07419
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.48783 31.41240
sample estimates:
 Mean Male Glucose Mean Female Glucose
      113.4417           98.4794
```

Welch Two Sample t-test

```
data: Buckingham Males and females
t = 1.0965, df = 134.02, p-value = 0.2748
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.580805 22.957418
sample estimates:
 Mean Male Glucose Mean Female Glucose
      110.4972           102.3089
```

The p-values produced by the t.test() calculations are the same as the p-values obtained from manual test statistic calculations.

Question 3 Interpret your results in words as you would in the results section of a manuscript (1-2 sentences)

An unpaired two sample t-test was used to test if there was a significant difference between the blood glucose levels of Louisa males (113.44 ± 70.5) and females (98.48 ± 32.4). At the critical value of $\alpha_C = 0.05$, the difference between Louisa male and female blood glucose levels is not statistically significant ($p = 1.926$, $t^* = 1.803$, $df = 106.71$), and the null hypothesis is not rejected. The 95th percentile confidence interval for the difference in mean blood glucose levels between Louisa males and females is (-1.3, 31.23), suggesting that the Louisa females had significantly lower blood glucose levels than their male counterparts.

An unpaired two sample t-test was used to test if there was a significant difference between the blood glucose levels of Buckingham (111.58 ± 56.58) and females (101.19 ± 41.56). At the critical value of $\alpha_C = 0.05$, the difference between Buckingham male and female blood glucose levels is not statistically significant ($p = 1.846$, $t^* = 1.434$, $df = 149.95$), and the null hypothesis is not rejected. The 95th percentile confidence interval for the difference in mean blood glucose levels between Buckingham males and females is $(-3.81, 24.59)$, suggesting that the Buckingham females had significantly lower blood glucose levels than their male counterparts.