BEE552 Biometry Week 4

Maria Feiler

2/16/2022

My Learning Journey

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

- I asked / answered **5** questions posed in class.
- I asked **0** questions in Slack.
- I answered **1** questions posed by other students on Slack.
- I came to Heather's office hours: **No**
- I came to Jose's office hours: **No**
- 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?

> **6**

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

> **Doing fine for now**

Problem Set

Part 1

We will be using a recently published dataset on ancient rings and ribs that may have been used as early forms of money. These data come from the paper "The origins of money: Calculation of similarity indexes demonstrates the earliest development of commodity money in prehistoric Central Europe" by M.H.G. Kuijpers and C. Popa (PLoS ONE, January 20, 2021).

Download the data in the file "rings_and_ribs.csv".

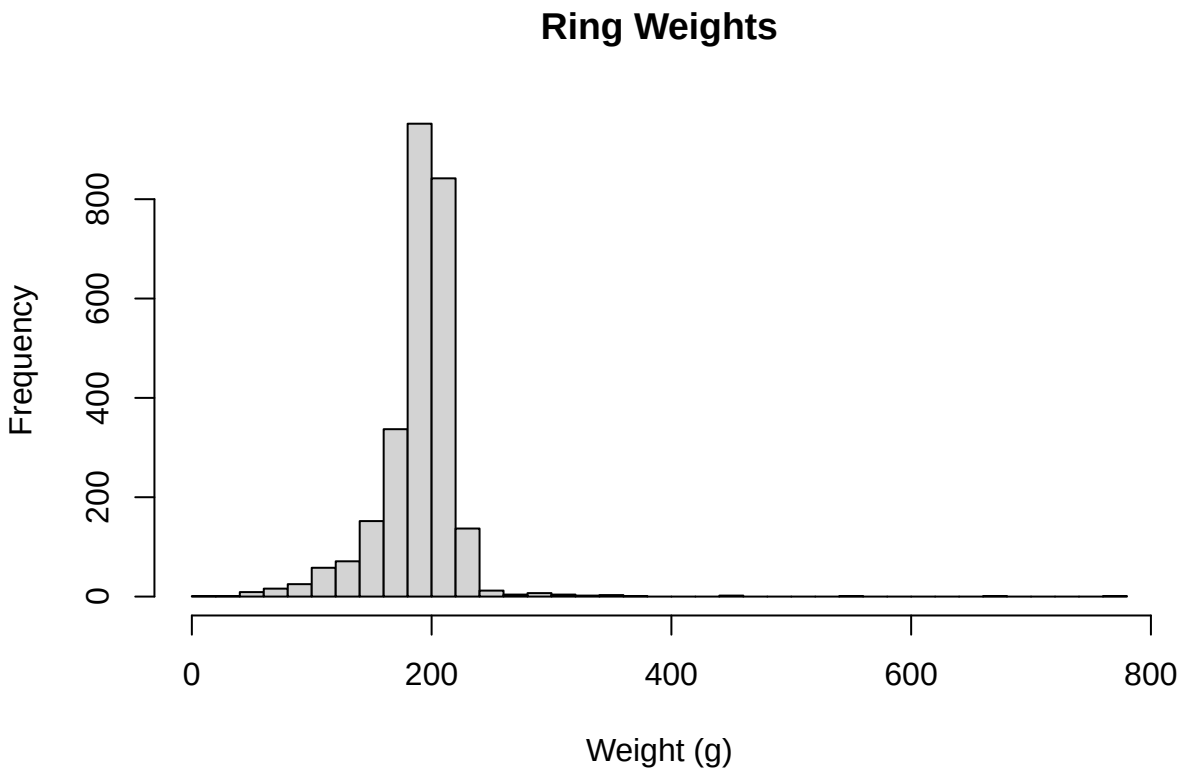
```
money <- read_excel("journal.pone.0240462.s002.xlsx",
                    sheet = 1
                    )

# Rename the weight column to make it tidy
money <- rename(money, Weight = `Weight (g)`)
```

Table 1: A selection of data from Kuijpers and Popa 2021.

Location	Country	Zone	Date	Weight	Type	Museum	Source
Mondsee	Austria	1	EBA	163.0	Ring	Vienna	Lenerz- de Wilde documentation
Radostice	Czech Republic	1	EBA	205.0	Ring	Ceské Budejovice	Moucha 2005
Mauthausen	Germany	1	EBA	211.0	Ring	Bad Reichenhall	Lenerz- de Wilde documentation; Menke 1987/1979
Bermatingen	Germany	1	EBA	78.0	Rib	Konstanz	Lenerz- de Wilde documentation
Stradonice	Czech Republic	1	EBA	64.0	Rib	Louny	Moucha 2005
Langesöd	Germany	1	EBA	187.0	Ring	Munich	Menke 1978/1979
Mikulov	Czech Republic	1	EBA	210.0	Ring	Brno	Lenerz- de Wilde documentation
Niederscheyern	Germany	1	EBA	70.0	Rib	Munich//Nuremberg	Lenerz- de Wilde documentation
Lesonice	Czech Republic	1	EBA	188.0	Ring	Vienna	Lenerz- de Wilde documentation
Lengfelden	Austria	1	EBA	206.0	Ring	Salzburg	Lenerz- de Wilde documentation
Bernhaupten	Germany	1	EBA	182.0	Ring	Munich	Lenerz- de Wilde documentation
Mauthausen	Germany	1	EBA	189.0	Ring	Bad Reichenhall	Lenerz- de Wilde documentation; Menke 1987/1979
Mauthausen	Germany	1	EBA	221.0	Ring	Bad Reichenhall	Lenerz- de Wilde documentation; Menke 1987/1979
Obereching	Austria	1	EBA	185.0	Rib	Salzburg	Moolsteiner and Maoesta 1988
Temelín	Czech Republic	1	EBA	88.0	Rib	Prague//Týn nad Vltavou//Prague ARÚ	Moucha 2005
Unknown (Wien)	Austria	1	EBA	185.0	Ring	Vienna	Lenerz- de Wilde documentation
Havalda	Czech Republic	1	EBA	220.0	Rib	Ceské Budejovice	Moucha 2005
St. Pölten - Spratzern	Austria	1	EBA	189.0	Ring	Vienna	Lenerz- de Wilde documentation
Bernhaupten	Germany	1	EBA	182.0	Ring	Munich	Lenerz- de Wilde documentation
Mauthausen	Germany	1	EBA	207.0	Ring	Bad Reichenhall	Lenerz- de Wilde documentation; Menke 1987/1979
Mauthausen	Germany	1	EBA	194.5	Ring	Bad Reichenhall	Lenerz- de Wilde documentation; Menke 1987/1979
Straubing	Germany	1	EBA	185.0	Ring	Straubing	Lenerz- de Wilde documentation
Mittermühle (Mainburg)	Germany	1	EBA	62.0	Rib	Munich	Lenerz- de Wilde documentation
Suché Vrbné	Czech Republic	1	EBA	189.0	Ring	Ceské Budejovice	Moucha 2005
Obereching	Austria	1	EBA	204.0	Rib	Salzburg	Moolsteiner and Maoesta 1988
Kostice	Czech Republic	1	EBA	194.0	Ring	Brno	Lenerz- de Wilde documentation
Vlasatice	Czech Republic	1	EBA	176.0	Ring	Brno	Lenerz- de Wilde documentation
München-Luitpoldpark	Germany	1	EBA	183.0	Rib	Munich	Lenerz- de Wilde documentation
Dušníky nad Vltavou	Czech Republic	1	EBA	200.0	Ring	Prague//Roudnice nad Labem	Moucha 2005
Blucina	Czech Republic	1	EBA	203.0	Ring	Brno	Lenerz- de Wilde documentation

Question 1 *For now, let's group all the locations together. Make a histogram of ring weight.*



Question 2 *What is the FORMULA for the sample mean and standard deviation (in other words, what is the formula you would want to use if you wanted to estimate the population mean [the μ parameter assuming a normal distribution] and the population variance [σ^2 if we assume a normal distribution] from a sample that represented a random subset of the entire population)?*

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

Question 3 *What is the population about which we are trying to make inference?*

The population is the total collection of rings and ribs ever used for currency in the context of Germany, Austria, the Czech Republic, and Poland during this time period.

Question 4 *Using R, what are the mean and the standard deviation of ring weight?*

The mean of the ring weights is 172.633. The standard deviation of ring weights is 48.296.

Question 5 *What is the formula for the standard error of the mean (heretofore s.e.)?*

$$SEM = \sqrt{\frac{S^2}{n}}$$

Question 6 *Describe the difference between the s.e. (of the mean) and the s.d.*

The standard deviation is a measure of the variance of your sample The standard error of the mean describes how close your sample mean approximates the population mean.")

Question 7 *Finish the sentence: The standard error is the standard deviation of...*

...the mean.

Question 8 *Use the MASS package's 'fitdistr' function to fit a normal distribution to the ring weight data. Do you get the roughly same answer as above?*

```
ringWeightFit <- fitdistr(money["Type" = "Ring",]$Weight,
                          "normal")

# Calculate difference in calculated and fitdistr() means
meandif <- unname(ringWeightFit$estimate[1]) - ringMean

# Calculate difference in calculated and fitdistr() standard deviations
sddif <- unname(ringWeightFit$estimate[2]) - ringSD
```

The difference between the calculated mean and the mean from fitdistr() is 0 The difference between the calculated standard deviation and the standard deviation from fitdistr() is -0.00546.

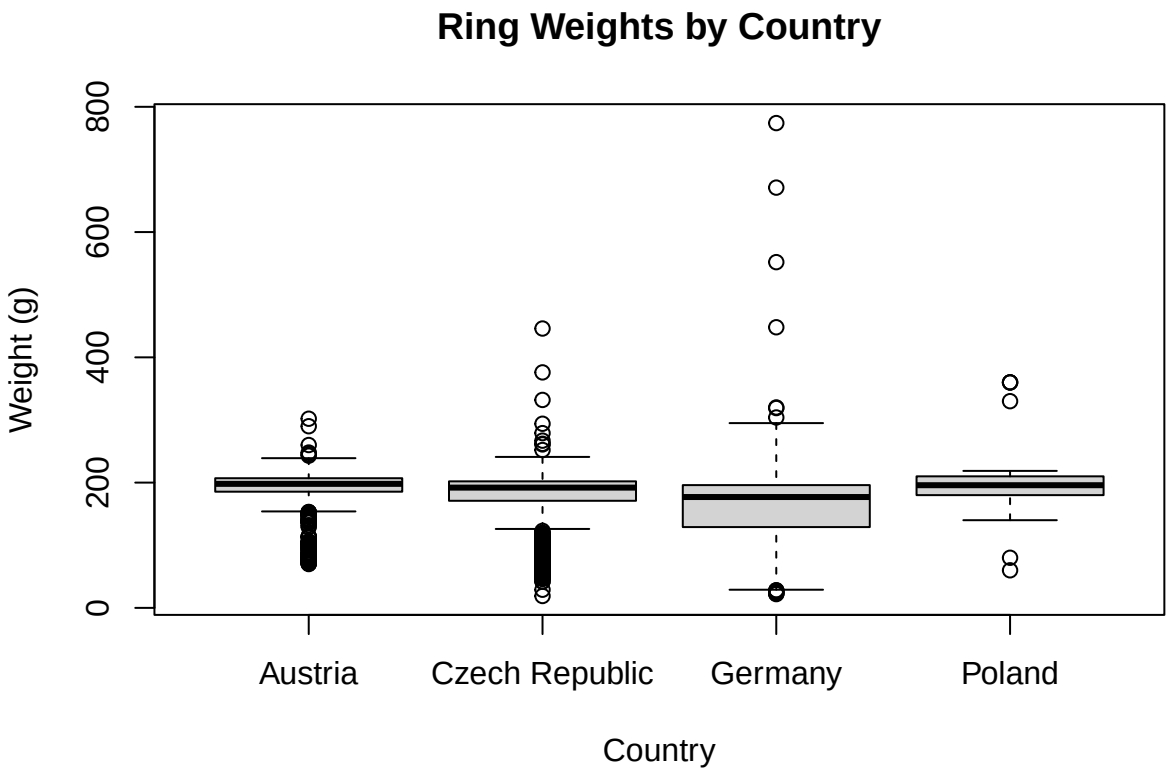
Question 9

Why might it not be valid to group the different locations when summarizing the data?

Though there might be differences between the rings used as currency in each country, all were considered to have moneytary value and might have changed hands outside of country/political entity lines. Since they are all from the same time period, if you want to characterize how the earliest money worked, you must use all the material that was used as money.

Question 10 *Using the R command 'boxplot', create a boxplot to compare ring weight across different countries.*

```
ringsByCountry <- melt(data = as.data.table(money["Type" = "Ring",]),
                       id = "Country",
                       measure = "Weight"
                       )
```



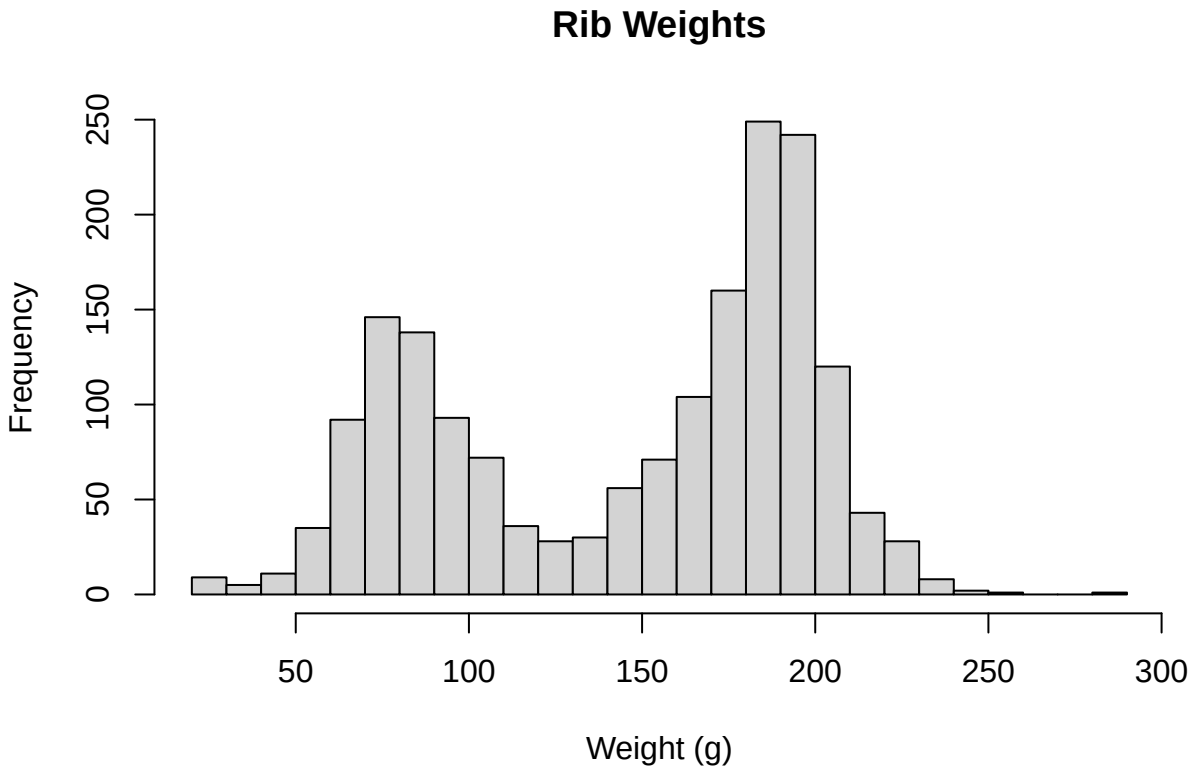
Question 11 *Make the case (in words and/or mathematically) that average ring weight in Germany is or is not statistically different from the average ring weight in Austria.*

I will test whether or not the mean weight of Austrian and German rings are statistically different using a critical value of $\alpha_c = 0.05$. The null hypothesis ($H_0 : \overline{x}_{German} = \overline{x}_{Austrian}$) will be tested using a student's t-test and will be rejected if $p > 0.05$.

```
GermanVSAustrian <- t.test(ringsByCountry$value[ringsByCountry$Country == "Germany"],
                           ringsByCountry$value[ringsByCountry$Country == "Austria"]
                           )
```

The average weight of Austrian rings (192.01 grams) is statistically greater than the average weight of German rings (161.91 grams) ($p < 0.001$).

Question 12 *So far, we've only been focused on ring weight. Let's go back and make a histogram of rib weight. Why would testing a hypothesis about average rib weight be harder?*



Making predictions or testing hypotheses about rib weight would be difficult because they appear to follow a bimodal distribution.