Hypothesis testing

Eduardo Alarcón & Alfonso Pineda

2022-12-03

Confidence Intervals

We will perform a unilateral test for the population mean μ with a large sample.

We are going to test if the mean (μ) of our population regarding the rithm of the song, the top songs of Spotify, match with the information from the entire dataset of Spotify, which for women is 650 and for men 600

Therefore:
$$H_0 = \mu(\text{female artists}) = 550 \quad H_1 = \mu(\text{female artists}) > 550$$
$$H_0 = \mu(\text{male artists}) = 600 \quad H_1 = \mu(\text{male artists}) > 600$$

We will start with the female artists:

```
rithmF <- SpotifySongs$rithmF
t.test(rithmF, mu=630, alternative="greater")
```

```
##
##  One Sample t-test
##
## data:  rithmF
## t = 2.192, df = 36, p-value = 0.01746
## alternative hypothesis: true mean is greater than 630
## 95 percent confidence interval:
##  647.0655      Inf
## sample estimates:
## mean of x
##  704.2647
```

Recalculate the P-value:

```
pnorm(2.192, lower.tail=FALSE)
```

```
## [1] 0.01418975
```

As the *p-value* is 0.01746, and even lower when recalculated (0.01418975), we can state that we have sufficient data and evidence to reject the initial hypothesis and take the second hypothesis, resulting in the songs created by female artiss that end up in the top charts on Spotify have a higher rithm than the rest of songs in spotify, including all songs, created by Male artists, Female, Bands and colaborations.

Now, we will calculate the hypothesis testing for the male artists, with the same mean, as this number is from the entire charts of spotify

```
rithmM <- SpotifySongs$rithmM
t.test(rithmM, mu=630, alternative="greater")
```

```
##
##  One Sample t-test
```

```
## 
## data:  rithmM
## t = -0.22364, df = 49, p-value = 0.588
## alternative hypothesis: true mean is greater than 630
## 95 percent confidence interval:
##  575.4498      Inf
## sample estimates:
## mean of x
##  623.5797
```

We obatain a p-value of 0.588, and after recalculating:

```
pnorm(-0.22364, lower.tail=FALSE)
```

```
## [1] 0.5884813
```

we obtain a p-value of 0.5884813, which is extremely high and result, tells us that there is enough evidence to keep the inicial hypotesis. Thus, the mean of the male artist's rithm of the top songs of spotify match with the rithm of the general songs in spotify.
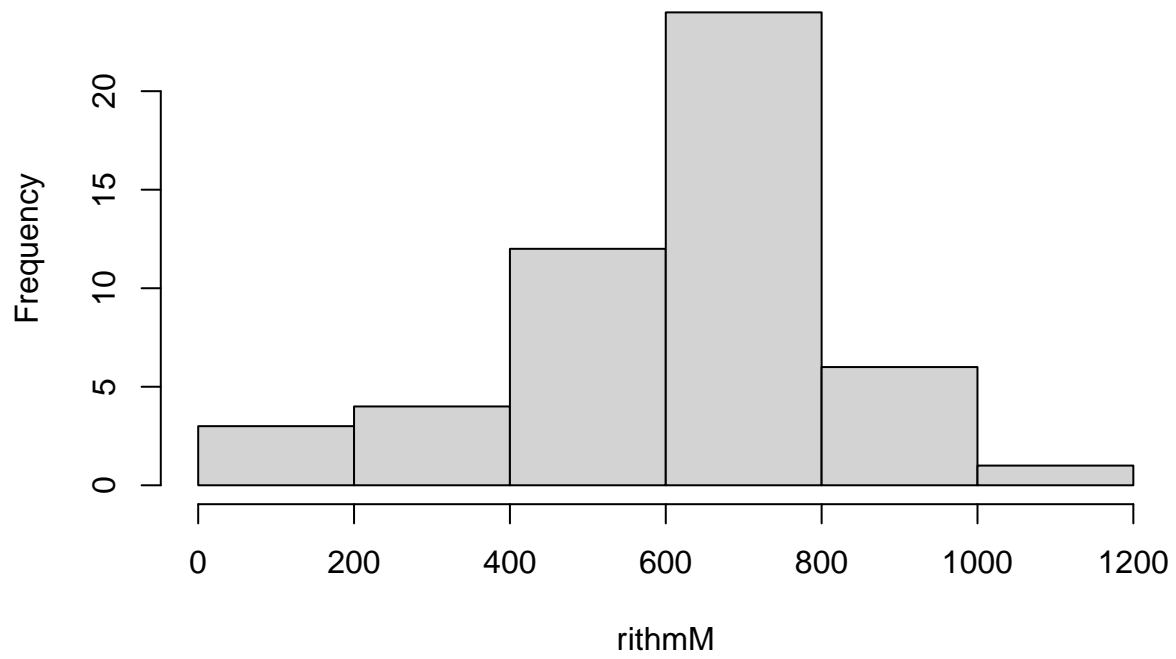
All in all, we can state that the songs created by females artists have a higher rithm than the rest of the songs in spotify with a confidence level of 95%. The oposite happens with the male-interpreted songs. We can state that the rithm of a male-interpreted song will be 630 units with a 95% confidence.
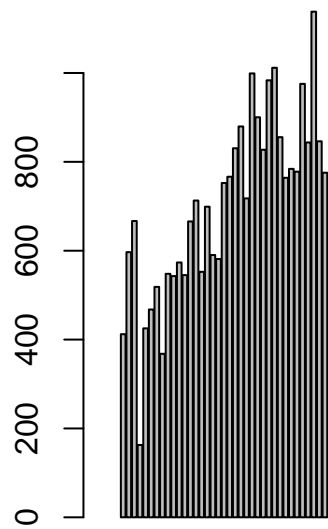
```
hist(rithmF)
```

Histogram of rithmF
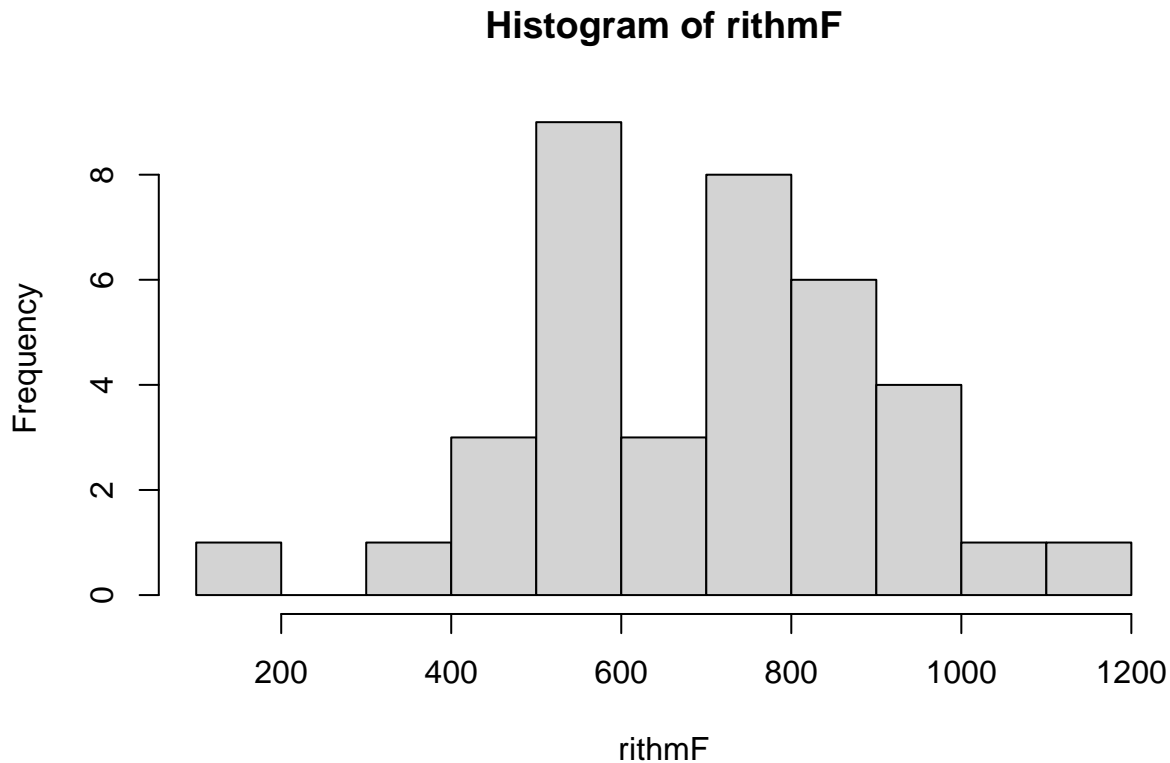

```
hist(rithmM)
```

Histogram of rithmM

barplot(rithmF)

barplot(rithmM)