Comparing the number of shoes men and women own

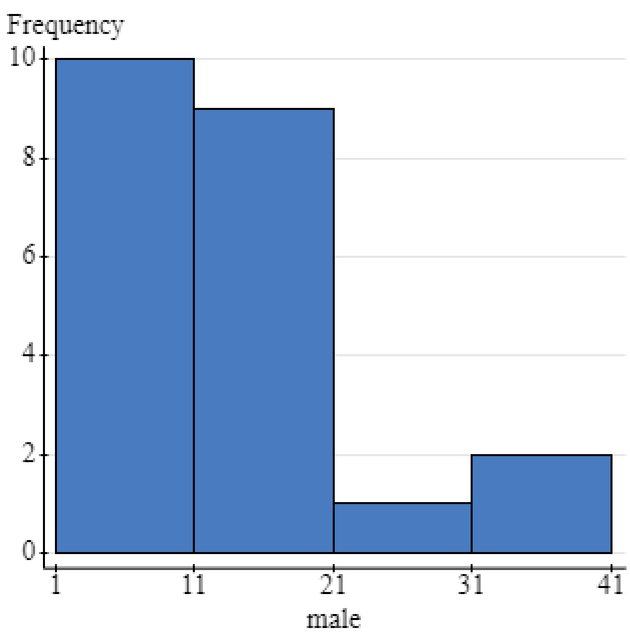
Eliza Juma

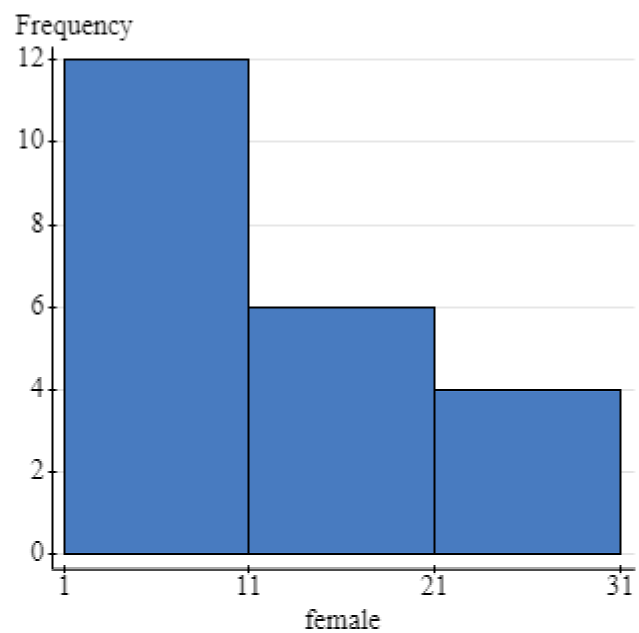
Nathalie Colindres

University of Central Missouri

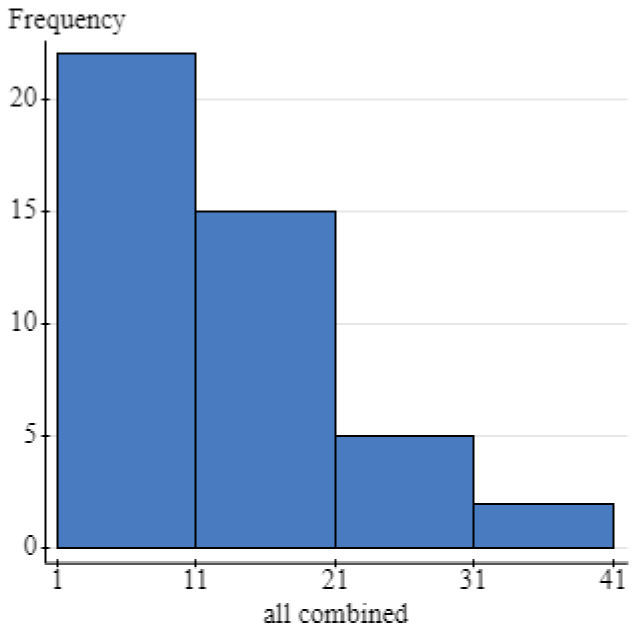
In this course,both me and Nathalie partnered together to collect data of UCM college students on campus. We wanted to know who between both genders owns the most shoes. Both me and Nathalie studied both males and females , the variable that was studied is the amount of shoes male and female own; For this data we walked around UCM campus collecting data from both male and female for each pair of shoes they own. Our population of interest was college students on campus.

Histogram for each category





Histogram for both data combined



When we observe the data.The histogram data for both males and females combined skewed to the right with no outlier. The histogram data of males skewed right and the female data both skewed to the right.

Find the Five-Number-Summary for the quantitative variable for each category.

Summary statistics:

Column	Min	Q1	Median	Q3	Max	Mean	Std. dev.
male	1	8	12	16	40	13.909091	9.2885444
female	4	8	10	20	30	13.318182	7.3897378

Mean and IQR

Summary statistics:

Column	Mean	IQR
male	13.909091	8
female	13.944444	12

Relative Frequency table results for male:

Bins include the left endpoint
Count = 22

male	Relative Frequency
[1 to 11)	0.45454545
[11 to 21)	0.40909091
[21 to 31)	0.045454545
[31 to 41)	0.090909091

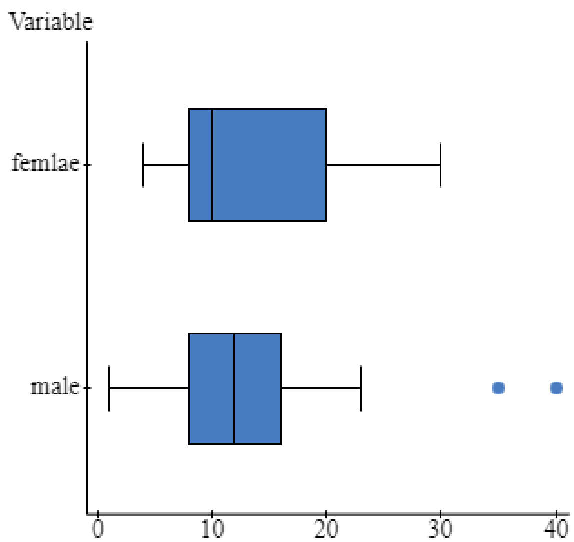
Relative Frequency table results for female:

Bins include the left endpoint
Count = 22

female	Relative Frequency
[1 to 11)	0.54545455
[11 to 21)	0.27272727
[21 to 31)	0.18181818

The difference between both the male and female frequency table.The male frequency seems to have a less relative frequency number between 1 to 11 but has the highest frequency than female.The male frequency has a higher frequency between 1 to 11 but its frequency is lower than the female data.

Box Plot of the quantitative data comparing the Box Plot for each category.



This box plot quantitative data for both male and female compares the box plot of each category. The female data is shorter on the left and so the data skewed to the right, the male seems symmetrical but there are two outliers for the male box plot , two male students owned 40 and 35 pairs of shoes which means the data is skewed right.

Mean and standard deviation for the data set.
Summary statistics:

Column	Mean	Std. dev.
male	**13.909091**	**9.2885444**
female	**13.944444**	**7.5729567**

Both means for males and females are close to one another Male (13.909) and Female (13.318), although their standard deviations have a noticeable difference. For Standard Deviation Males have a larger standard deviation (9.2885) and Females (7.3897). This difference in the standard deviations reflects the fact that in this study the quantitative values of shoes for males is larger and further away from the mean than the quantitative values of womens pairs of shoes when compared to the mean.

The hypothesis we are testing is if there is a difference in how many pairs of shoes people own based on gender, most specifically if men own more pairs of shoes than women. My Hypothesis is $H0$: $\mu1 = \mu2$ versus $H1$: $\mu1 > \mu2$ with $\alpha = 0.01$. I chose this level of significance because there is a low risk of making a type II error and having $\alpha = 0.01$ will help minimize the chance of making a type I error.

We are conducting the survey to figure out if on average men own more pairs of shoes than women.

H1 = H2
H1 > H2

Two sample T hypothesis test:
μ_1 : Mean of male
μ_2 : Mean of female
$\mu_1 - \mu_2$: Difference between two means

$H_0 : \mu_1 - \mu_2 = 0$
$H_A : \mu_1 - \mu_2 > 0$
(without pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	0.59090909	2.5305882	39.979893	0.23350662	0.4083

After finishing the hypothesis testing we concluded the mean pairs of shoes males own is not larger than the mean pair of shoes females own. For the hypothesis test we concluded a P-value of 0.4083. P-value is greater than the level of significance, so we do not reject the null. What we conclude about the population means is that males population mean is not greater than female population mean. There is no sufficient evidence to conclude that the mean number of shoes for males is larger than the mean number for females.

Two sample T confidence interval:
μ_1 : Mean of male
μ_2 : Mean of female
$\mu_1 - \mu_2$: Difference between two means
(without pooled variances)

95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	0.59090909	2.5305882	39.979893	-4.5236805	5.7054987

We calculated a 95% confidence interval of (-4.5236805, 5.7054987) when concluding the mean pairs of shoes between males and females. We are 95% confident that the mean number of shoes for males and females is between -4.5236805 and 5.7054987. Since the interval does contain 0 we do not reject the null hypothesis. Concluding the mean pairs of shoes males own is not greater then the mean pairs of shoes females own.

male	female
1	6
15	4
35	25
6	22
16	10
20	22
13	10
12	8
6	8
5	8
10	10
16	20
8	6
23	13
8	12
8	20
40	30
15	17
12	10
18	5
10	7
9	20