Assignment – STAT8061

Kaushal Parmar (8890500)

Conestoga Collage

Course: - Data Visualization Tools and Techniques

Instructure's name: - Ifzal Bhatti

Due date: - 02-02-2024

Contents

INTRODUCTION

This is a statistics assignment that is about to pick own sample-data for student's weight(lbs) from two classes and answer some questions in a report form. The requirements for picking raw data are provided, and we need to explain the sampling method for each sample. We also need to include relevant graphs/methods /calculation during answering the question. There are different types of questions including finding the mean weight of male students in both samples, the difference between both samples weight's standard deviation, which sample has the higher standard deviation, drawing and interpreting pie-chart, bar graph and box plot.

1. RAW DATA

CLASS 1 (SAMPLE 1)		
SR.NO.	WEIGHT	Gender
1	110.01	M
2	111.02	F
3	111.08	M
4	112.08	M
5	113.06	M
6	116.01	M
7	116.05	F
8	119.04	M
9	120.04	F
10	121.03	M
11	122.07	M
12	123.06	M
13	124.04	M
14	125.04	M
15	127.04	M
16	128.02	F
17	129.01	M
18	131.04	M
19	132.08	F
20	133.06	M
21	135.04	M
22	136.08	M
23	137.06	F
24	139.03	M
25	139.05	F
26	142.01	M
27	143.08	F
28	145.03	F
29	145.07	M
30	146.03	M
31	149.07	F
32	150.04	F
33	151.04	M
34	155.09	M
35	156.01	F
36	157.08	M
37	158.01	M
38	159.02	M
39	159.03	M
40	160.06	M
41	161.03	M
42	161.06	F
43	162.07	M
44	162.08	F
45	163.05	M
46	164.04	M
47	164.08	F
48	168.06	M
49	169.07	M
50	171.07	M
51	175.09	F
52	178.03	F
53	178.04	M
54	179.01	F
55	179.08	M
standard deviation	20.75262	

CLASS 2 (SAMPLE 2)		
SR.NO.	WEIGHT	Gender
1	110.05	M
2	111.05	M
3	115.06	F
4	115.02	M
5	116.06	M
6	118.03	M
7	119.02	M
8	119.04	F
9	122.06	M
10	124.07	M
11	126.01	M
12	126.08	F
13	128.09	M
14	130.08	M
15	131.04	M
16	133.02	M
17	133.01	M
18	135.04	M
19	135.03	F
20	136.04	M
21	141.01	M
22	142.07	M
23	143.05	M
24	143.07	F
25	144.09	M
26	144.01	M
27	145.02	M
28	147.02	M
29	147.05	F
30	148.08	F
31	151.06	M
32	151.09	M
33	152.08	F
34	153.02	M
35	153.06	F
36	154.01	F
37	155.02	M
38	155.06	F
39	156.09	F
40	158.03	F
41	158.02	F
42	159.04	M
43	161.09	M
44	162.02	M
45	167.06	M
46	173.02	F
47	174.03	F
48	176.08	F
49	178.06	M
Standard deviation	**18.06012**	

i. Explain the sampling method (for sample-1 and sample-2) that you will use for such scenario.

Ans: - **Sample 1**

(Between 52 to 58)

I have selected **55.**

Female: - 32% (between 30% to 45 %) = 18 numbers

Male: - 68% (between 55 to 70 %) = 37 numbers

Sample 2

(between 46 to 50)

I have selected **49**.

Female: - 35% (between 30% to 45 %) = **17 numbers**

Male: - 65% (between 55% to 70 %) = 32 numbers

Explanation: -

➔ **I have used simple random sampling method for sample1 and sample2 .**

- **This method is a random selected subset of a population and in this sampling method each member of the population has an exactly equal chance of being selected.**

  Step 1 :- Define the population :- I have access to every individual students of the population(in order to collect data from all those who are selected for the sample). There are total 400 students in class 1 and 500 in class 2.

  Step 2 :- Decide on the sample size :- (one-gender's total between 30%-45% and the other's 55%-70%. Gender ratio cannot be same for both samples)

- 52 to 58 students from class 1
- 46 to 50 students from class 2

  Step 3:- **Randomly select sample** :- In the random number method , I assign every individual a number. By using a random number generator.

  Step 4 :- **Collect data from sample** :- Finally , I have collected data from sample1 and sample 2 (Capture continuous data between 109.99 lbs to 179.99 lbs)

2. Answer the below questions in a report form. Include all relevant graphs/methods/calculation/rationale etc. in a sequence with your answers.

a) Is the mean weight of male students in both samples same?

- **SAMPLE 1**

SR. NO.	MALE#1 weight (lbs)
1	110.01
2	111.08
3	112.08
4	113.06
5	116.01
6	119.04
7	121.03
8	122.07
9	123.06
10	124.04
11	125.04
12	127.04
13	129.01
14	131.04
15	133.06
16	135.04
17	136.08
18	139.03
19	142.01
20	145.07
21	146.03
22	151.04
23	155.09
24	157.08
25	158.01
26	159.02
27	159.03
28	160.06
29	161.03
30	162.07
31	163.05
32	164.04
33	168.06
34	169.07
35	171.07
36	178.04
37	179.08
TOTAL WEIGHT	5274.77
MEAN VALUE	142.5613514

Male students: - 68% (between 55 to 70 %) = **37** numbers

- Total weight of male students = 5274.77 (lbs) [=SUM (E6:E42)]

Sample MEAN formula = $\bar{x} = \frac{\sum x_i}{n}$ where, $\sum x_i$=sum , n = sample size

MEAN WEIGHT = total weight / male numbers [=E43/D42]
= 5274.77/37
= 142.56 lbs

- **SAMPLE 2**

Male students - 65% (between 55% to 70 %) **= 32** numbers

- **Total weight of male students = 4431.42 lbs [=SUM (K6:K37)**

SR.NO.	MALE#2weight (lbs)
1	110.05
2	111.05
3	115.02
4	116.06
5	118.03
6	119.02
7	122.06
8	124.07
9	126.01
10	128.09
11	130.08
12	131.04
13	133.02
14	133.01
15	135.04
16	136.04
17	141.01
18	142.07
19	143.05
20	144.09
21	144.01
22	145.02
23	147.02
24	151.06
25	151.09
26	153.02
27	155.02
28	159.04
29	161.09
30	162.02
31	167.06
32	178.06
Total weight	4431.42
Mean Value	138.481875

Sample MEAN formula = $\bar{x} = \frac{\sum x_i}{n}$ where, n = sample size Σx_i = sum

MEAN WEIGHT = total weight / male numbers [=k38/J37]

= 4431/32

= 138.481 lbs

-

- ✓ **Therefore, the mean weight of male students in both samples are not same.**

b) What is the difference between both samples' weight's standard deviation.

SAMPLE 1 (Total numbers of students 55)

- **Male = 37**
- **FEMALE=18**

CLASS 1 (SAMPLE 1)		
SR.NO.	WEIGHT	Gender
1	110.01	M
2	111.02	F
3	111.08	M
4	112.08	M
5	113.06	M
6	116.01	M
7	116.05	F
8	119.04	M
9	120.04	F
10	121.03	M
11	122.07	M
12	123.06	M
13	124.04	M
14	125.04	M
15	127.04	M
16	128.02	F
17	129.01	M
18	131.04	M
19	132.08	F
20	133.06	M
21	135.04	M
22	136.08	M
23	137.06	F
24	139.03	M
25	139.05	F
26	142.01	M
27	143.08	F
28	145.03	F
29	145.07	M
30	146.03	M
31	149.07	F
32	150.04	F
33	151.04	M
34	155.09	M
35	156.01	F
36	157.08	M
37	158.01	M
38	159.02	M
39	159.03	M
40	160.06	M
41	161.03	M
42	161.06	F
43	162.07	M
44	162.08	F
45	163.05	M
46	164.04	M
47	164.08	F
48	168.06	M
49	169.07	M
50	171.07	M
51	175.09	F
52	178.03	F
53	178.04	M
54	179.01	F
55	179.08	M
standard deviation	20.75262	

Total weight of both gender students = 7920.67 lbs

Standard deviation =

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}},$$ **were,**

s = Standard deviation

$\bar{x}$ = Sample mean

n = sample size.

Where, sample mean for sample1 = 144.012

And sample size is = 55

Therefore, Standard deviation for sample 1 = STDEV.S(O6:O60)

- **= 20.75262301**

- **SAMPLE 2 (Total numbers of students 49)**
  - **Male=32**
  - **Female=17**

CLASS 2 (SAMPLE 2)		
SR.NO.	WEIGHT	Gender
1	110.05	M
2	111.05	M
3	115.06	F
4	115.02	M
5	116.06	M
6	118.03	M
7	119.02	M
8	119.04	F
9	122.06	M
10	124.07	M
11	126.01	M
12	126.08	F
13	128.09	M
14	130.08	M
15	131.04	M
16	133.02	M
17	133.01	M
18	135.04	M
19	135.03	F
20	136.04	M
21	141.01	M
22	142.07	M
23	143.05	M
24	143.07	F
25	144.09	M
26	144.01	M
27	145.02	M
28	147.02	M
29	147.05	F
30	148.08	F
31	151.06	M
32	151.09	M
33	152.08	F
34	153.02	M
35	153.06	F
36	154.01	F
37	155.02	M
38	155.06	F
39	156.09	F
40	158.03	F
41	158.02	F
42	159.04	M
43	161.09	M
44	162.02	M
45	167.06	M
46	173.02	F
47	174.03	F
48	176.08	F
49	178.06	M
Standard deviation	18.06012	

Total wight of both gender students = 6974.31 lbs

Standard deviation =

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}},$$ **were,**

s = Standard deviation

$\bar{x}$ **= Sample mean,**

n = sample size.

Where sample mean = 142.3328571

And sample size = 49

Therefore, Standard deviation for sample 2 = STDEV.S(S6:S60)

- **= 18.06011559**

So, finally difference between both samples weight's standard deviation.

= **Sample 1 weight's standard deviation** – **Sample 2 weight's standard deviation**

= 20.75- 18.06

= 2.69

- ✓ **This is the difference between both samples' weight's standard deviation.**

c) Which sample has the higher standard deviation of female students' weights.

Standard deviation = $$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}},$$

SAMPLE 1 (Female)

SR.NO.	FEMALE#1 weight (lbs)
1	111.02
2	116.05
3	120.04
4	128.02
5	132.08
6	137.06
7	139.05
8	143.08
9	145.03
10	149.07
11	150.04
12	156.01
13	161.06
14	162.08
15	164.08
16	175.09
17	178.03
18	179.01
Standard deviation	20.76688009

SAMPLE 2 (Female)

SR.NO.	FEMALE#2 weight (lbs)
1	115.06
2	119.04
3	126.08
4	135.03
5	143.07
6	147.05
7	148.08
8	152.08
9	153.06
10	154.01
11	155.06
12	156.09
13	158.03
14	158.02
15	173.02
16	174.03
17	176.08
Standard deviation	17.75144967

- Standard deviation for SAMPLE 1(**Female students**) = **20.76688009 [STDEV.S(B6:B23)]**
- Standard deviation for SAMPLE 2(**Female students**) = **17.75144967 [=STDEV.S(H6:H22)]**

So, **therefore SAMPLE 1 has the higher standard deviation** of female students' weights.

d) What is the ratio between male and female students for each sample.

- ➔ For sample 1 I have selected 55(sample) between 52 and 58 students .where, MALE#1 = 37 and FEMALE#1 = 18
- ➔ For sample 2 I have selected 49(sample) between 52 and 58 students.where, MALE#2 = 32 and FEMALE#2 =17

So ,

For sample 1 Male#1 =37

Female#1 = 18

So, the ratio for male#1 =37/55 AND for Female#1=18/555

=0.672 *100 **=0.32 *100**

=67.27% **=32.72%**

For sample 2 Male#2 =32

Female#2 = 17

So, the ratio for male#2 =32/49 AND for Female#2 =17/49

=0.653*100 **=0.346*100**

=65.4% **=34.6%**

e) What percentage of all students have weights between 121 and 157 pounds.

- **For sample 1**

CLASS 1 (SAMPLE 1)		
SR.NO.	WEIGH	Gende
10	121.03	M
11	122.07	M
12	123.06	M
13	124.04	M
14	125.04	M
15	127.04	M
16	128.02	F
17	129.01	M
18	131.04	M
19	132.08	F
20	133.06	M
21	135.04	M
22	136.08	M
23	137.06	F
24	139.03	M
25	139.05	F
26	142.01	M
27	143.08	F
28	145.03	F
29	145.07	M
30	146.03	M
31	149.07	F
32	150.04	F
33	151.04	M
34	155.09	M
35	156.01	F

There is total **26** students who have weights between **121 and 157** pounds for sample 1.

Whereas,

- **For Sample 2**

CLASS 2 (SAMPLE 2)		
SR.NO.	WEIGI	Gend
9	122.06	M
10	124.07	M
11	126.01	M
12	126.08	F
13	128.09	M
14	130.08	M
15	131.04	M
16	133.02	M
17	133.01	M
18	135.04	M
19	135.03	F
20	136.04	M
21	141.01	M
22	142.07	M
23	143.05	M
24	143.07	F
25	144.09	M
26	144.01	M
27	145.02	M
28	147.02	M
29	147.05	F
30	148.08	F
31	151.06	M
32	151.09	M
33	152.08	F
34	153.02	M
35	153.06	F
36	154.01	F
37	155.02	M
38	155.06	F
39	156.09	F

There is total **31** students who have weights between 121 and 157 pounds for sample 2.

Then,

SUM of sample 1 and sample 2
=26 + 31
=57 (all students have weights between 121 and 157)

Now, Total weights of all students = 55 (sample 1) + 49 (sample 2)

=**104 students** (from class 1 and class 2)

- **% of all students = 57/104**

= **0.548 *100**

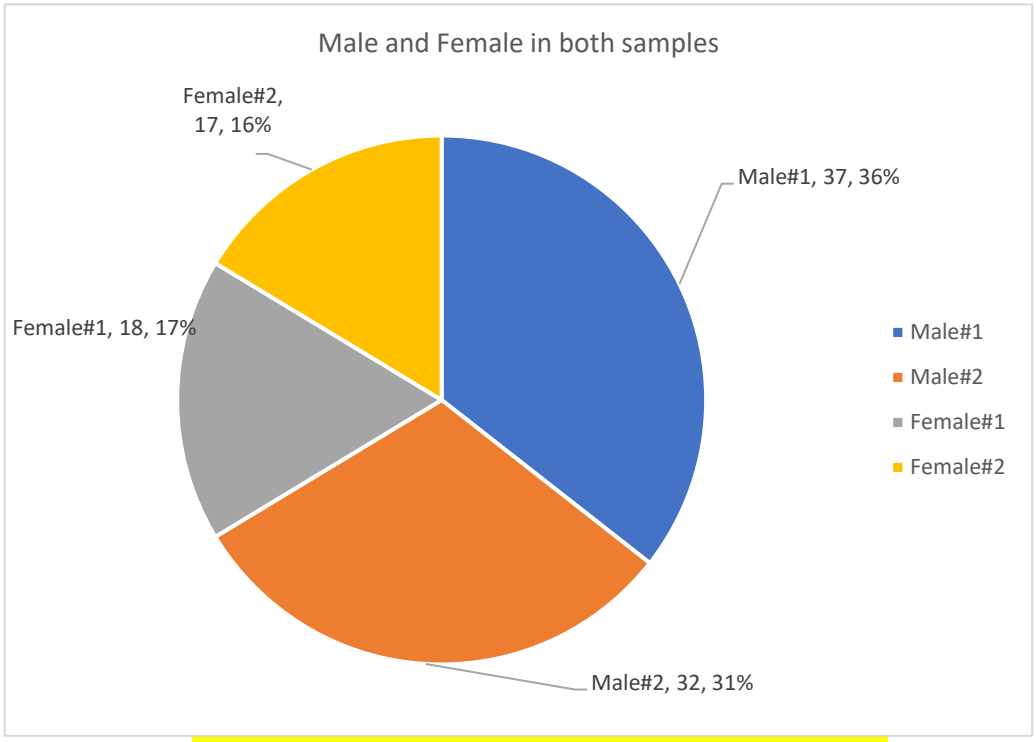
= 54.80%

✓ **so, 54.80% of all students have weights between 121 and 157 pounds.**

3. Draw and interpret pie-chart and bar-graph of male#1, male#2, female#1, and female#2 (#1 and #2 denote sample-1 and sample-2).

Gender	count
Male#1	37
Male#2	32
Female#1	18
Female#2	17

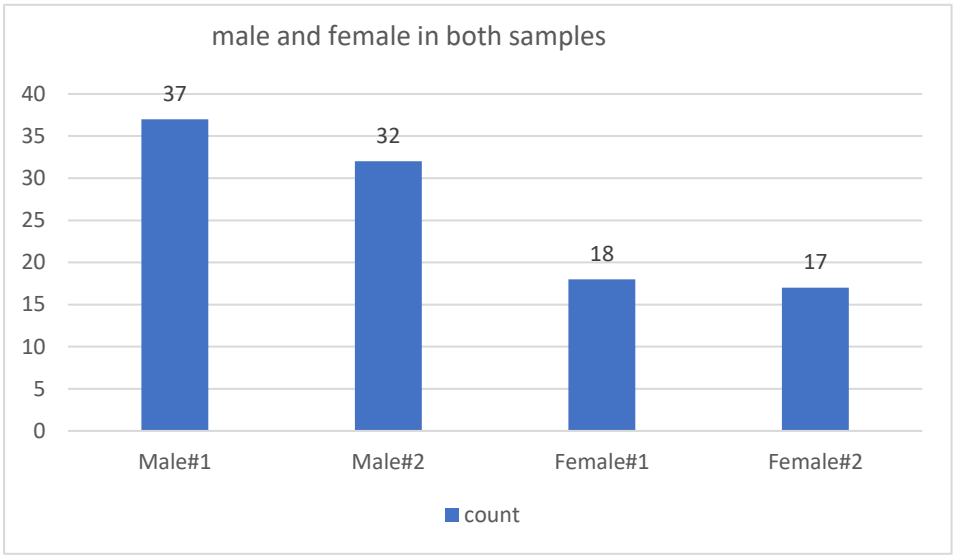
Pie chart: -



(*Note: - #1 and #2 represents sample1 and sample2)

- This pie chart represents the number of male and female students in two different class. The chart is divided up into **four sections**, each of which represents an individual student group classified by sample size and gender. The blue segment represents **37** male students from **Sample1**, accounting for **36%** of the total. The orange segment represents **32 male** students from **Sample2**, accounting for **31%** of the total. The grey segment represents 18 female students from **Sample1**, accounting for **17%** of the total. The yellow segment represents **17** female students from **Sample2,** accounting for **16%** of the total.

Bar graph: -

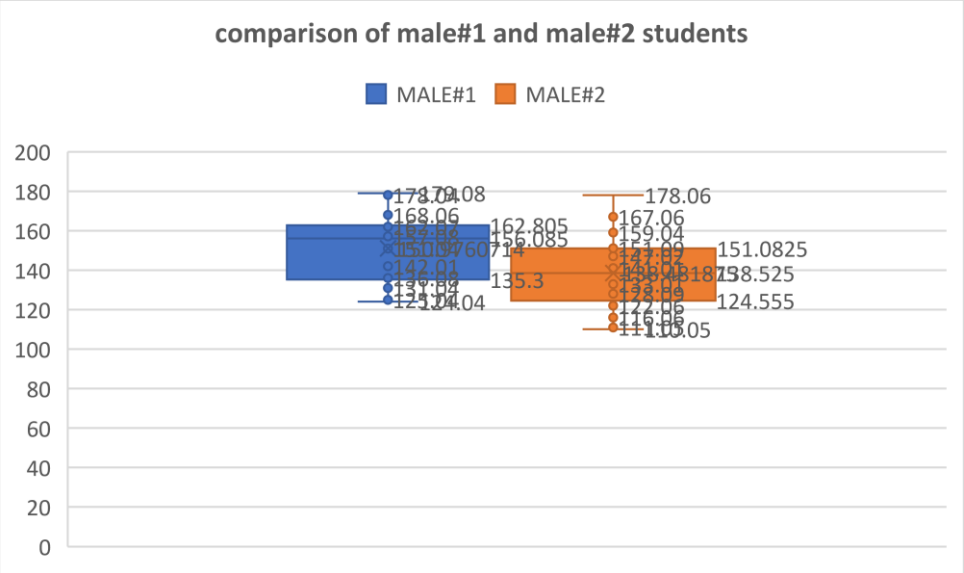


(*Note: - #1 and #2 represents sample1 and sample2)

- This bar graph compares the counts of male and female students in two different samples, labelled as #1 and #2. The blue bars represent males, with one is “Male#1” having a count of 37 and the other labelled “Male#2” having a count of 32. The orange bars represent females, with one labelled “Female#1” having a count of 18 and the other labelled “Female#2” having a count of 17.

4. Develop the following boxplot, interpret, and conclude (comparison should be based on all characteristics of boxplot discussed in the class):

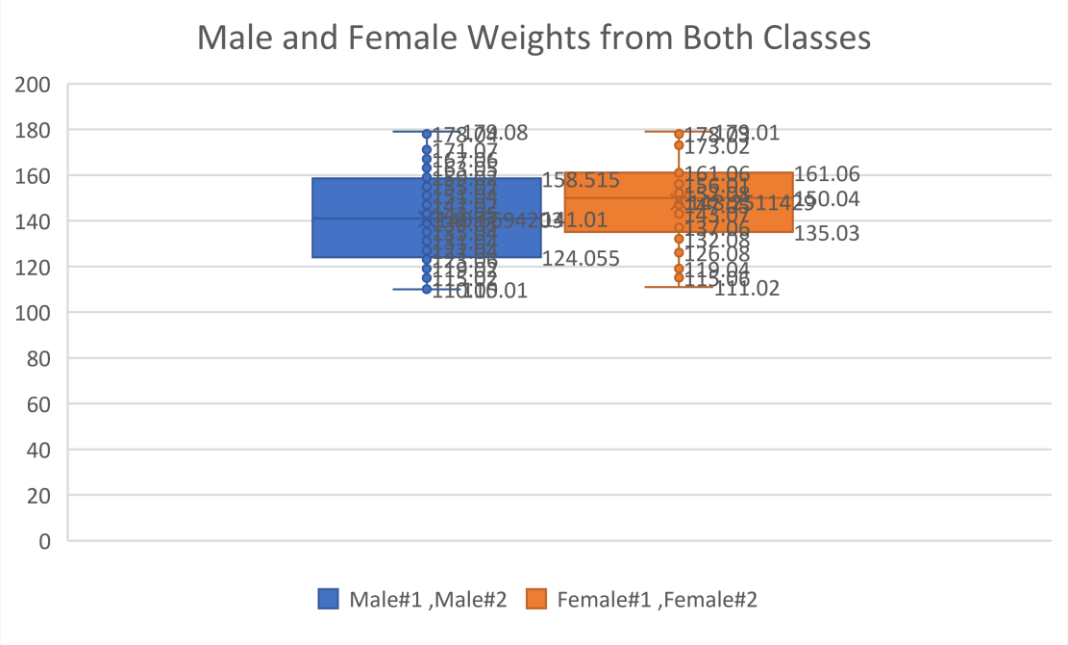
A. Male student weights in sample-1 Vs. male students' weight in sample-2.



(*where #1 and #2 represents sample1 and sample 2)

- This box plot compares two different samples, MALE#1 and MALE#2. MALE#1 is represented by the blue box, while MALE#2 is represented by the orange box. A form of statistics for comparing data points between these two categories is the box plot. The box plot elements for MALE#1's minimum value around 104 and maximum value is 177.5. Moreover, for male#1 Q1, median and Q3 values are approximately 131.5, roughly 151.5, close to168 respectively. Whereas, for male#2 minimum and maximum values are 116 roughly 116 and about 178 respectively. In addition to this, the Q1, median, Q3 values are 124.5,138 and around 157.

B. Male students' weight in both classes Vs. female students' weight in both classes



(*where #1 and #2 represents sample1 and sample 2)

- This box plot displays the distribution of male and female weights from two different classes or samples. The blue boxes represent MALE#1 and MALE#2, while the orange boxes represent FEMALE#1 and FEMALE#2. This illustrates that males in both samples generally have a wider range of weights compared to females. There's an overlap in the weights of males and females, indicating that some individuals from both genders have similar weights. Additionally, MALE#2 has a lower median value compared to MALE#1, indicating that the central tendency of MALE#2's data points is lower than that of MALE#1. However, both samples have similar maximum values but different minimum values, with MALE#2 having a higher minimum value than MALE#1

Conclusion

In summary, the review of the collected data provides illumination on the gender distribution, the weight distribution within a particular range, the variation of female students' weights, the gender distribution mean, and the features of the gender-based weight distribution. These results help to clarify how students' weights are distributed between the two classes.