ASSIGNMENT 1 PART 2: CONTINIOUS DATA

Submitted By

Zajiba Sadia Islam

Student ID: 501279357

Group-09

Submitted for

Program: MEng in Biomedical Engineering

Course: Stats for the Health Sciences

BP8101, Winter 2024

Class Timing: Wednesday 3:00-6:00 PM

This assignment deals with the various statistical measurements calculated for continuous variable dataset. In this assignment, the weight of 34 friends and family members has been surveyed and various interpretations of the data has been made. The data has been represented in various ways by means of relevant graphs, and important data such as frequency, mean, standard deviation, percentiles etc has been calculated and shown with appropriate formula and calculations. The findings have also been presented in relevant chart forms.

The weight of 34 different people has been chosen randomly from a survey conducted among friends and relatives . The data is presented below.

|  |  |
| --- | --- |
| Person | Weight |
| 1 | 109.3 |
| 2 | 79.02 |
| 3 | 74.7 |
| 4 | 35.9 |
| 5 | 97.1 |
| 6 | 79.32 |
| 7 | 73.32 |
| 8 | 85.32 |
| 9 | 36 |
| 10 | 104.7 |
| 11 | 79.31 |
| 12 | 73.41 |
| 13 | 85.21 |
| 14 | 36 |
| 15 | 79.3 |
| 16 | 73.5 |
| 17 | 85.1 |
| 18 | 110 |
| 19 | 79.7 |
| 20 | 36 |
| 21 | 99.1 |
| 22 | 79.29 |
| 23 | 73.59 |
| 24 | 84.99 |
| 25 | 75.8 |
| 26 | 36.1 |
| 27 | 98.5 |
| 28 | 98 |
| 29 | 79.28 |
| 30 | 73.68 |
| 31 | 84.88 |
| 32 | 95 |
| 33 | 60 |
| 34 | 36.1 |

The summary of the given data can be shown in the following chart

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weight Range | Frequency, f | Cumulative frequency | Lower value of class | Upper value of class | Sum | Mid point,x | f\*x | x-x¯ | (x-x¯)^2 | f\*(x-x¯)^2 | Relative Frequency (f/n) | Cumulative Relative Frequency | Percentage(Relative frequency\*100) |
| 35-50 | 6 | 6 | 35 | 50 | 85 | 42.5 | 255 | -33.5 | 1122.25 | 6733.5 | 0.176471 | 0.176471 | 17.64706 |
| 51-65 | 1 | 7 | 51 | 65 | 116 | 58 | 58 | -18 | 324 | 324 | 0.029412 | 0.205882 | 2.941176 |
| 66-80 | 14 | 21 | 66 | 80 | 146 | 73 | 1022 | -3 | 9 | 126 | 0.411765 | 0.617647 | 41.17647 |
| 81-95 | 6 | 27 | 81 | 95 | 176 | 88 | 528 | 12 | 144 | 864 | 0.176471 | 0.794118 | 17.64706 |
| 96-110 | 7 | 34 | 96 | 110 | 206 | 103 | 721 | 27 | 729 | 5103 | 0.205882 | 1 | 20.58824 |
|  | 34 |  |  |  |  |  | 2584 |  |  | 13150.5 |  |  |  |

A survey was conducted among all my friends and relatives to disclose their weight. Among them, 34 different samples have been chosen at random.

The data that I chose here is continuous. This is because, the data of weight are not integers. It can have any value in between two consecutive values within a given range. The weight of a person can also be measured in great accuracy, for example 67.34 kg, 67.35 kg etc. Hence, the dataset containing the weight of a definite age group of people is continuous.

From the histogram of frequency vs Weight range, we see that the graph is roughly symmetrical with a central peak, representing a bell shape. Also, the dataset is that of continuous random variable data. Hence, it is a Normal Probability distribution.

The data can be presented graphically in various forms. These are shown below.

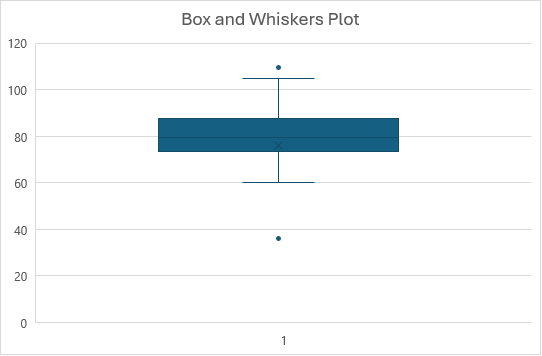
1. Histogram

From the histogram, it can be said that the graph is unimodal and roughly symmetrical. It is mound shaped as well.

1. Pie Chart
2. Line Chart
3. Dotplot
4. Stem and leaf plot

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Weight in kg | | |  |  |  |  |  |  | Key |  |  |
|  | 35 | 9 |  |  |  |  |  |  |  | 35 | 9 | =35.9 |
|  | 36 | 0 | 0 | 0 | 1 | 1 |  |  |  |  |  |  |
|  | 37 |  |  |  |  |  |  |  |  |  |  |  |
|  | 38 |  |  |  |  |  |  |  |  |  |  |  |
|  | 39 |  |  |  |  |  |  |  |  |  |  |  |
|  | 40 |  |  |  |  |  |  |  |  |  |  |  |
|  | 41 |  |  |  |  |  |  |  |  |  |  |  |
|  | 42 |  |  |  |  |  |  |  |  |  |  |  |
|  | 43 |  |  |  |  |  |  |  |  |  |  |  |
|  | 44 |  |  |  |  |  |  |  |  |  |  |  |
|  | 45 |  |  |  |  |  |  |  |  |  |  |  |
|  | 46 |  |  |  |  |  |  |  |  |  |  |  |
|  | 47 |  |  |  |  |  |  |  |  |  |  |  |
|  | 48 |  |  |  |  |  |  |  |  |  |  |  |
|  | 49 |  |  |  |  |  |  |  |  |  |  |  |
|  | 50 |  |  |  |  |  |  |  |  |  |  |  |
|  | 51 |  |  |  |  |  |  |  |  |  |  |  |
|  | 52 |  |  |  |  |  |  |  |  |  |  |  |
|  | 53 |  |  |  |  |  |  |  |  |  |  |  |
|  | 54 |  |  |  |  |  |  |  |  |  |  |  |
|  | 55 |  |  |  |  |  |  |  |  |  |  |  |
|  | 56 |  |  |  |  |  |  |  |  |  |  |  |
|  | 57 |  |  |  |  |  |  |  |  |  |  |  |
|  | 58 |  |  |  |  |  |  |  |  |  |  |  |
|  | 59 |  |  |  |  |  |  |  |  |  |  |  |
|  | 60 | 0 |  |  |  |  |  |  |  |  |  |  |
|  | 61 |  |  |  |  |  |  |  |  |  |  |  |
|  | 62 |  |  |  |  |  |  |  |  |  |  |  |
|  | 63 |  |  |  |  |  |  |  |  |  |  |  |
|  | 64 |  |  |  |  |  |  |  |  |  |  |  |
|  | 65 |  |  |  |  |  |  |  |  |  |  |  |
|  | 66 |  |  |  |  |  |  |  |  |  |  |  |
|  | 67 |  |  |  |  |  |  |  |  |  |  |  |
|  | 68 |  |  |  |  |  |  |  |  |  |  |  |
|  | 69 |  |  |  |  |  |  |  |  |  |  |  |
|  | 70 |  |  |  |  |  |  |  |  |  |  |  |
|  | 71 |  |  |  |  |  |  |  |  |  |  |  |
|  | 72 |  |  |  |  |  |  |  |  |  |  |  |
|  | 73 | 3 | 4 | 5 | 6 | 7 |  |  |  |  |  |  |
|  | 74 | 7 |  |  |  |  |  |  |  |  |  |  |
|  | 75 | 8 |  |  |  |  |  |  |  |  |  |  |
|  | 76 |  |  |  |  |  |  |  |  |  |  |  |
|  | 77 |  |  |  |  |  |  |  |  |  |  |  |
|  | 78 |  |  |  |  |  |  |  |  |  |  |  |
|  | 79 | 0 | 3 | 3 | 3 | 3 | 3 | 7 |  |  |  |  |
|  | 80 |  |  |  |  |  |  |  |  |  |  |  |
|  | 81 |  |  |  |  |  |  |  |  |  |  |  |
|  | 82 |  |  |  |  |  |  |  |  |  |  |  |
|  | 83 |  |  |  |  |  |  |  |  |  |  |  |
|  | 84 | 9 | 9 |  |  |  |  |  |  |  |  |  |
|  | 85 | 1 | 2 | 3 |  |  |  |  |  |  |  |  |
|  | 86 |  |  |  |  |  |  |  |  |  |  |  |
|  | 87 |  |  |  |  |  |  |  |  |  |  |  |
|  | 88 |  |  |  |  |  |  |  |  |  |  |  |
|  | 89 |  |  |  |  |  |  |  |  |  |  |  |
|  | 90 |  |  |  |  |  |  |  |  |  |  |  |
|  | 91 |  |  |  |  |  |  |  |  |  |  |  |
|  | 92 |  |  |  |  |  |  |  |  |  |  |  |
|  | 93 |  |  |  |  |  |  |  |  |  |  |  |
|  | 94 |  |  |  |  |  |  |  |  |  |  |  |
|  | 95 | 0 |  |  |  |  |  |  |  |  |  |  |
|  | 96 |  |  |  |  |  |  |  |  |  |  |  |
|  | 97 | 1 |  |  |  |  |  |  |  |  |  |  |
|  | 98 | 0 | 5 |  |  |  |  |  |  |  |  |  |
|  | 99 | 1 |  |  |  |  |  |  |  |  |  |  |
|  | 100 |  |  |  |  |  |  |  |  |  |  |  |
|  | 101 |  |  |  |  |  |  |  |  |  |  |  |
|  | 102 |  |  |  |  |  |  |  |  |  |  |  |
|  | 103 |  |  |  |  |  |  |  |  |  |  |  |
|  | 104 | 7 |  |  |  |  |  |  |  |  |  |  |
|  | 105 |  |  |  |  |  |  |  |  |  |  |  |
|  | 106 |  |  |  |  |  |  |  |  |  |  |  |
|  | 107 |  |  |  |  |  |  |  |  |  |  |  |
|  | 108 |  |  |  |  |  |  |  |  |  |  |  |
|  | 109 | 3 |  |  |  |  |  |  |  |  |  |  |
|  | 110 | 0 |  |  |  |  |  |  |  |  |  |  |

1. Box Plot





Calculations:

Summary of data chart

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weight Range | Frequency, f | Cumulative frequency | Lower value of class | Upper value of class | Sum | Mid point,x | f\*x | x-x¯ | (x-x¯)^2 | f\*(x-x¯)^2 | Relative Frequency (f/n) | Cumulative Relative Frequency | Percentage(Relative frequency\*100) |
| 35-50 | 6 | 6 | 35 | 50 | 85 | 42.5 | 255 | -33.5 | 1122.25 | 6733.5 | 0.176471 | 0.176471 | 17.64706 |
| 51-65 | 1 | 7 | 51 | 65 | 116 | 58 | 58 | -18 | 324 | 324 | 0.029412 | 0.205882 | 2.941176 |
| 66-80 | 14 | 21 | 66 | 80 | 146 | 73 | 1022 | -3 | 9 | 126 | 0.411765 | 0.617647 | 41.17647 |
| 81-95 | 6 | 27 | 81 | 95 | 176 | 88 | 528 | 12 | 144 | 864 | 0.176471 | 0.794118 | 17.64706 |
| 96-110 | 7 | 34 | 96 | 110 | 206 | 103 | 721 | 27 | 729 | 5103 | 0.205882 | 1 | 20.58824 |
|  | 34 |  |  |  |  |  | 2584 |  |  | 13150.5 |  |  |  |

1. Mean= ∑fx/f

=2584/34

=76

1. Standard deviation

Variance, s2 = ∑ f (m − x̅)2 / n − 1

=13150.5/33

=398.5

Standard deviation, s= √s^2

=√398.5

=19.96

1. Range approximation of s= If range, R is about four times of standard deviation then standard deviation can be approximated as s ≈ R/4

≈(110-35.9)/4

≈18.525

Which is very similar to the calculated value of s, 19.96.

1. Formula for determining the quarters is, Qi = L + h/f\*(iN/4 – c.f).

The class containing Qi is the class whose cumulative frequency is just greater than (iN/4).

N = Total Frequency.

h = class size.

L = lower limit of the class containing Qi.

f = frequency of the class containing Qi.

c.f = cumulative frequency of the class interval preceding the interval containing Qi.

For the first quartile, Q1 i=1 in the above formula

 Since i=1 we have that (iN/4) = (1\*34/4) = 8.5

The value of cumulative frequency just greater than 8.5 is 21

Therefore, Q1 lies in the class interval 66-80. Here, L = 66, f = 14, h =15 and c.f = 7.

We substitute all these values in the above formula.

Q1 = L + h/f\*(3N/4 – c.f).

Q1 = 66 + 15/14\*(1\*34/4 – 7).

Q1 = 66 + 1.07\*(8.5 – 7) =67.6

1. For the third quartile, Q1 i=3 in the above formula

 Since i=3 we have that (iN/4) = (3\*34/4) = 25.5

The value of cumulative frequency just greater than 8.5 is 27

Therefore, Q3 lies in the class interval 81-95. Here, L = 81, f = 6, h =15 and c.f = 21.

We substitute all these values in the above formula.

Q3 = L + h/f\*(3N/4 – c.f).

Q3 = 81 + 15/6\*(3\*34/4 – 21).

Q3 = 92.25

1. Median = median of grouped data is l + [(n/2−c)/f] × h, where: l = lower limit of median class. n = total number of observations. c = cumulative frequency of the preceding (of median class) class, h=class interval.

Since n=34/2=17, the median is the 17th observation. It lies in the 66-80 range.

Hence l=66, frequency of median class, f= 14, c=7, h=15.

So Median= 76.71

1. Mode= From the frequency histogram, we obtain that the class interval of 66-80 contains the highest frequency, hence it is the modal class.

Mode = L + (f 1– f 0/2f 1– f 0– f 2 ) h.

L=66, f1=14, f0=1, f2=6,h=15

Mode=75.28

Here, l = lower limit of median class, h=class interval, f1= frequency of modal class, f0= frequency of class preceding modal class, f2= frequency of class following modal class.

1. 70th percentile =

Here, *n*=34  
  
  
*P*70 class :  
  
Class with (70*n/*100)*th* value of the observation in *cf* column  
  
=(70\*34/100)*th* value of the observation in *cf* column  
  
=(23.8)*th* value of the observation in *cf* column  
  
and it lies in the class 81-95.  
  
∴*P*70 class : 81-95  
  
The lower boundary is 81.  
  
∴*L*=81  
  
*P*70=*L*+(70*n/*100-*cf)/f*⋅*c*  
  
=81+(23.8-21)/6\*5  
  
=83.33

1. Z-score- Formula for calculation of z-score is (x-x̄)/ s

Previously we found, mean ̄x= 76.07

Standard deviation, s=21.87

The individual z-scores of each of these measurements are shown in the table below.

|  |  |  |
| --- | --- | --- |
| Weigh, w | wi-x¯ | Z-score (wi-x̄)/19.96 |
| 35.9 | -40.1 | -2.00877 |
| 36 | -40 | -2.00376 |
| 36 | -40 | -2.00376 |
| 36 | -40 | -2.00376 |
| 36.1 | -39.9 | -1.99875 |
| 36.1 | -39.9 | -1.99875 |
| 60 | -16 | -0.8015 |
| 73.32 | -2.68 | -0.13425 |
| 73.41 | -2.59 | -0.12974 |
| 73.5 | -2.5 | -0.12524 |
| 73.59 | -2.41 | -0.12073 |
| 73.68 | -2.32 | -0.11622 |
| 74.7 | -1.3 | -0.06512 |
| 75.8 | -0.2 | -0.01002 |
| 79.02 | 3.02 | 0.151284 |
| 79.28 | 3.28 | 0.164308 |
| 79.29 | 3.29 | 0.164809 |
| 79.3 | 3.3 | 0.16531 |
| 79.31 | 3.31 | 0.165811 |
| 79.32 | 3.32 | 0.166312 |
| 79.7 | 3.7 | 0.185348 |
| 84.88 | 8.88 | 0.444835 |
| 84.99 | 8.99 | 0.450345 |
| 85.1 | 9.1 | 0.455856 |
| 85.21 | 9.21 | 0.461366 |
| 85.32 | 9.32 | 0.466876 |
| 95 | 19 | 0.951786 |
| 97.1 | 21.1 | 1.056984 |
| 98 | 22 | 1.102068 |
| 98.5 | 22.5 | 1.127115 |
| 99.1 | 23.1 | 1.157172 |
| 104.7 | 28.7 | 1.437698 |
| 109.3 | 33.3 | 1.668131 |
| 110 | 34 | 1.703196 |

The value that is two standard deviations above the mean is

(x̄+2s)= 76+ 2\*19.96

=115.92

The value that is 1.5 standard deviations below the mean is

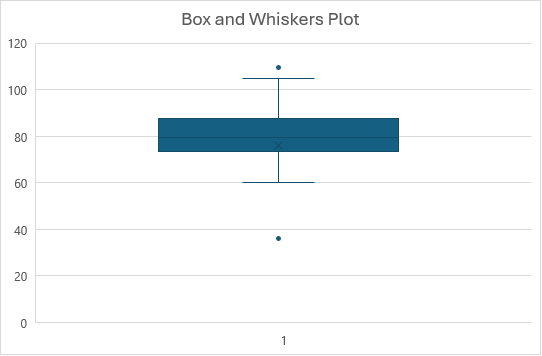
(x̄-1.5s)= 76- 1.5\*19.96

=46.06

From the previous calculations, the Five-Number Summary is shown in the chart below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Minimum value | First Quartile  Q1 | Median | Third Quartile  Q3 | Maximum value |
| 35.9 | 67.6 | 76.71 | 92.25 | 110 |

The box and whiskers plot is shown below:





From the chart, we find the lower fence as Q1-1.5(IQR)

= 67.6-1.5(92.25-67.6)

= 30.62

From the chart, we find the upper fence as Q3+1.5(IQR)

= 92.25+1.5(92.25-67.6)

= 129.22

Since all values fall within this range, there are no outliers.

Since the interquartile range is very small (24.65), and from the box and whiskers plot we see that the width of the box is very narrow as well, it appears that the middle 50% of the data is concentrated together.

Histogram showing mean and various standard deviations:

Find *P*(−.5 ≤ *z* ≤ 1.0); *P*(−1 ≤ *z* ≤ 1); *P*(−2 ≤ *z* ≤ 2); a value of *z* having area .025 to its right (*z*.025). Please, describe step by step, showing snapshot of the tables used and highlighting the values in the tables.

A table with numbers and letters

Description automatically generated

A table with numbers and letters

Description automatically generated

From the tables, we found the area to the left of *z =* −.5 as.3085 and the area to the left of z=1 as.84134

Hence For P(−.5 ≤ *z* ≤ 1.0),

By subtracting the two entries we get

=(.84134-.30854)

=0.5328

For P(−1 ≤ *z* ≤ 1)

From the tables, we found the area to the left of *z =* −1 as.15866 and the area to the left of z=1 as.84134

By subtracting the two entries we obtain

P(.84134-.15866)

=0.68268

For P(−2 ≤ *z* ≤ 2)

From the tables, we found the area to the left of *z =* −2 as.02275 and the area to the left of z=2 as.97725

By subtracting the two entries we obtain

P(.97725-.02275)

=0.9545

For a value of *z* having area .025 to its right (*z*.025)

P(1-0.025)

=1.96

A common notation for a value of *z* having area .025 to its right is *z(*.025) = 1.96.