

Name : Tauqeer Ahmed  
ID : 13912  
Department : BS (RAD)  
Subject : Biostatistics  
Submitted to: Sir Anwar Shamim  
Semester : 6<sup>th</sup>

(1)

Question 1 (a):  
Ans:

$$\text{Overall mean consumption for men} = \frac{204 + 259 + 266 + 317}{4} \\ = 261.5$$

$$\text{Overall mean consumption for women} = \frac{178 + 235 + 266 + 304}{4} \\ = 245.75$$

$$\text{Standard deviation} = \text{Standard Error} \times \sqrt{\text{No. of men}} \\ \text{For Men} = 0.9 \times \sqrt{1308} = 32.5$$

$$\text{Standard deviation} = \text{Standard Error} \times \sqrt{\text{No. of women}} \\ \text{For Women} = 0.8 \times \sqrt{1540} = 31.4$$

Overall Mean Consumption

$$\frac{1308 \times 261.5 + 1540 \times 245.75}{2848} \\ = 253$$

(a)

Question 1 (b):

Ans:

The figures for milk, root, vegetables, wheat flour indicate that in region Q1 these are consumed the most by both men and women, while the least in region Q4.

Question 1 (c):

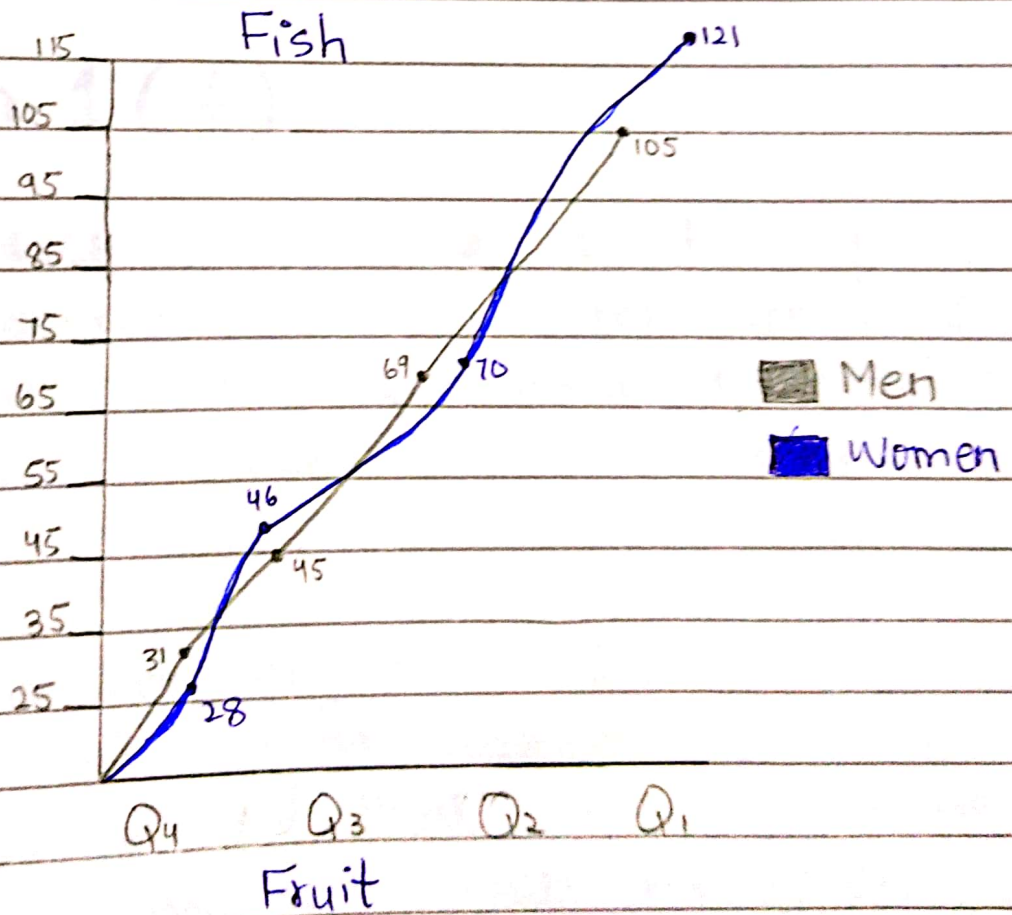
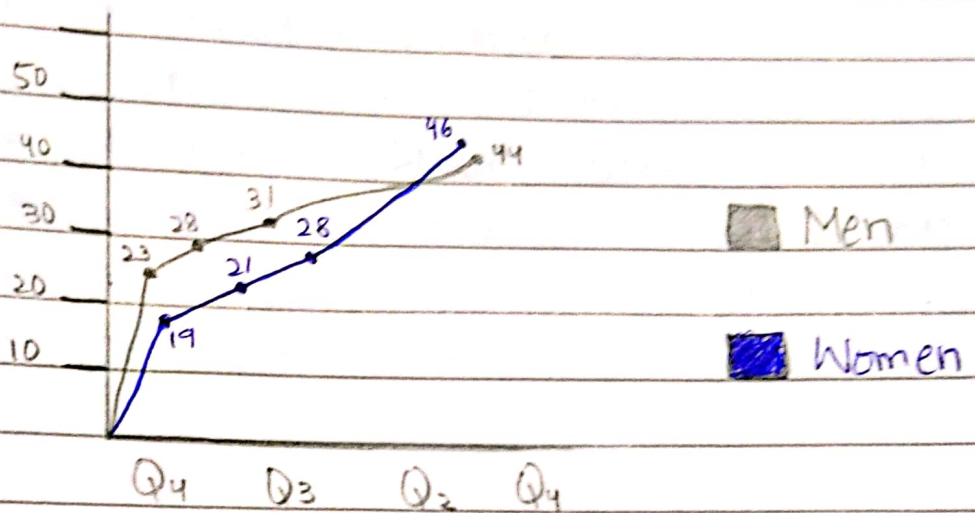
Ans:

It increases from Q4 to Q1 for fruit and fish, while decreases from Q4 to Q1 for rice.



(3)

Question 1 (d):  
Ans:





(4)

Question 1 (e):

Ans:

- Consumption of Fruit = Women > Men
- Consumption of Rice = Men > Women
- Consumption of Wheat Flour = Women > Men
- Consumption of Fresh Vegetables = Women > Men
- Consumption of Meat = Men > Women
- Consumption of Fish = Women > Men

Question 1 (f):

Ans:

The values of standard deviation of the above commodities for men and women shows that the standard deviation values are greater for men than women.

Standard deviation for men = 32.5

Standard deviation for women = 31.4

Hence proved that the standard deviation values for men are greater than women.

(5)

Question 2 (a):

Ans:

A census aims to count the entire population of a country, and at the location where each person actually lives. The census asks questions from people living in homes and group living situations, including how many people live or stay in each home and the sex, age and race of each person.

Question 2 (b):

Ans:

A census is a study of every unit everyone or everything in a population. It differs from a sample survey because the sample survey does not attempt to reach the whole population. Census also differs from a routine data collection because routine data is the information that you collect from your users habitually.



(6)

Question 2 (c):  
Ans:

Since the response rate is not 100% which means it is not full, this is a problem for the accuracy of the census. The missing 6% is an issue. These people are likely to be untypical of the 94% who participated.

E.g: estimating the homeless from the 94% reached would be inaccurate.

Question 2 (d):

Ans:

The people do not always take the census seriously because 'Jedi knight' is not in any real sense a religion. This casts doubt on the accuracy of the other responses. This may also show a distrust of government and the collection of data by government agencies. Therefore, the above discussion shows that all responses may not be taken seriously and there still may be value in asking the question.



(7)

Question 2(e):

Ans:

The potential problems in conducting the 2021 UK census online include, i) the additional information held by government agencies is unlikely to be complete. There may be concerns about security of information when it is supplied online. Conducting the census online may also be a problem for those living in backward areas having no access to internet or limited internet capability. The ONS (Office of National Statistics) has decided to put provisions in place to support households that may have difficulty accessing an online census. Households in some areas will receive a paper form to start with, and any household will be able to receive a paper form on request.

(8)

Question 2 (f):

Ans:

Census itself means the study of every object under observation which, in real sense is a very difficult task. It is also very difficult for the government to go to every single person and collect all the data. A large server is required for incorporating additional data by the agency. A great number of expertises are required for the manipulation of the data. There may also be a doubt on the accuracy of the data.



(9)

Question 3 (a):  
Ans:

1) A.M = Arithmetic Mean

Rainfall	Number of years	Midpoints	C.F	F(x)
20-24	1	22	1	22
25-29	3	27	4	81
30-34	5	32	9	160
35-39	8	37	17	296
40-44	5	42	22	210
45-49	2	47	24	94
50-54	0	52	24	0
55-59	1	57	25	57
				<u>920</u>

$$A.M = \frac{\sum fx}{\sum n} = \frac{920}{25} = 36.8$$



(10)

2) G.M = Geometric Mean

Class	Mid Value (h)(z)	f	f log(x)
20-24	22	1	3.091
25-29	27	3	9.8875
30-34	32	5	17.3287
35-39	37	8	28.8873
40-44	42	5	18.6883
45-49	47	2	7.7003
50-54	52	0	0
55-59	57	1	4.0435
		n = 25	$\sum f \log(x) = 89.6263$

$$\text{G.M of } x = \text{antilog} \left( \frac{\sum f \log(x)}{n} \right)$$

$$= \text{antilog} \left( \frac{89.6263}{25} \right)$$

$$= \text{antilog} (3.5851)$$

$$= 38.552$$

(11)

3) H.M = Harmonic Mean

$f/x$
0.0455
0.1111
0.1562
0.2162
0.119
0.0426
0
0.0175

$$\sum \left( \frac{f}{x} \right) = 0.7082$$

$$n = 25$$

$$H = \frac{n}{\sum f_i \left( \frac{1}{x_i} \right)}$$

$$= \frac{25}{0.7082}$$

$$H.M = 35.30$$



(12)

$$4) \text{ Median} = \frac{\text{value of } (\frac{n}{2})^{\text{th}} \text{ observation}}{\text{value of } (\frac{25}{2})^{\text{th}} \text{ observation}}$$

$$= (12.5)^{\text{th}} \text{ which lies b/w the class}$$

The median class is 34.5 - 39.5

$L$  = lower boundary point of median class = 34.5

$$n = 25$$

$C$  = ~~the~~ cumulative frequency of the class preceding the median class = 9

$$h = 5$$

$$\text{Median} = L + \frac{h}{f} \left( \frac{n}{2} - C \right)$$

$$= 34.5 + \frac{5}{8} \left( \frac{12.5 - 9}{2} \right)$$

$$= 34.5 + \frac{5}{8} (3.50)$$

$$= 34.5 + 2.19$$

$$\text{Median} = 36.69$$



(13)

$$5) \text{ Mode} = \frac{L + f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$$

$L$  = lower class boundaries of the model class

$f_m$  = frequency of the model class

$f_1$  = frequency associated with the class following the model class

$h$  = width of class interval

So, the mode can be

$$\text{Mode} = 34.5 + \frac{(8-5)}{(8-5) + (8-5)} \times 5$$

$$= 34.5 + \frac{3}{3+3} \times 5$$

$$\text{Mode} = 37$$

## Quartiles:-

$$n = 25$$

Q1 class

class with  $(\frac{n}{4})^{\text{th}}$  value of the observation in cb column.

=  $(6.25)^{\text{th}}$  value of the observation in cb column and it lies in class 30-34

Q1 class : 29.5 - 34.5

(14)

The lower boundary point of 29.5-34.5 is 29.5

$$L = 29.5$$

$$Q_1 = L + \frac{\frac{n}{4} - cb}{b} \cdot c$$

$$= 29.5 + \frac{6.25 - 4}{5} \cdot 5$$

$$= 29.5 + \frac{2.25}{5} \cdot 5$$

$$= 29.5 + 2.25$$

$$= \boxed{31.75}$$

$\bar{Q}_3$  class

class with  $\left(\frac{3n}{4}\right)^{\text{th}}$  value of the observation in  $cb$  column.

$$= \left(\frac{3 \cdot 25}{4}\right)^{\text{th}} \text{ value of the observation}$$



(15)

in  $c_b$  column.

=  $(18.75)^{\text{th}}$  value of the observation  
in  $c_b$  column it lies in the  
class 40-44.

Q3 class: 39.5 - 44.5

The lower boundary point of 39.5 - 44.5  
is 39.5

$$L = 39.5$$

$$Q_3 = L + \frac{\frac{3n}{4} - c_b}{b} \cdot c$$

$$= 39.5 + \frac{18.75 - 17}{5} \cdot 5$$

$$= 39.5 + \frac{1.75}{5} \cdot 5$$

$$= 39.5 + 1.75$$

$$= 41.25$$



(16)

Q.D

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2} = \frac{41.25 - 31.75}{2} \\ = \frac{9.5}{2}$$

$$= \boxed{4.75}$$

Deciles:-

Which divide the distribution into ten equal parts, are called deciles. which is denoted by  $D_1, D_2, \dots, D_9$ . The calculation of each decile to be calculated is too large and time consuming. So, for the practice, we can calculate  $D_2$ .

$$D_2 = \left( \frac{2n}{10} \right)^{\text{th}}$$

$$= \left( \frac{2 \times 25}{10} \right)^{\text{th}}$$

$= 5^{\text{th}}$ , which corresponds in the class, (27.5 - 34.5) therefore

$$D_2 = l + \frac{h}{f} \left( \frac{2n}{10} - c.f \right)$$

(17)

$$= 29.5 + \frac{5}{5} (5 - 4)$$

$$= 29.5 + 1$$

$$\boxed{D_2 = 30.5}$$

**Percentile:-**

Here  $n = 25$

$P_{10}$  class:

class with  $\left(\frac{10n}{100}\right)^{\text{th}}$  value of the observation in cf column

$= \left(\frac{10 \cdot 25}{100}\right)^{\text{th}}$  value of the observation in

cf column.

$= (2.5)^{\text{th}}$  value of the observation in cf column and it lies in the class 25-29.

$P_{10}$  class : 24.5 - 29.5



(18)

The lower boundary point of 24.5-27.5 is 24.5

$$L = 24.5$$

$$P_{10} = L + \frac{100}{100} - cb \cdot c$$

$$= 24.5 + \frac{2.5 - 1}{3} \cdot 5$$

$$= 24.5 + 2.5 = 27$$

$P_{90}$  class: 44.5 - 49.5

The lower boundary point of 44.5-49.5 is 44.5

$$L = 44.5$$

$$P_{90} = L + \frac{90}{100} - cb \cdot c$$

$$= 44.5 + \frac{22.5 - 22}{2} \cdot 5$$

$$= 44.5 + \frac{0.5}{2} \cdot 5$$

(19)

$$= 44.5 + 1.25$$

$$= 45.75$$

percentile deviation

$$= \frac{P_{90} - P_{10}}{2} = \frac{45.75 - 27}{2} = \frac{18.75}{2} = \boxed{9.375}$$

Range =

Largest value - smallest value

$$= 57 - 22$$

$$\boxed{R = 35}$$

M.D =

$$M.D = \frac{\sum b(x - \bar{x})}{\sum b} = 5.456$$

Variance :-

$$s^2 = \left( \frac{\sum b \cdot d^2}{n-1} - \frac{(\sum b \cdot d)^2}{n} \right) \cdot h^2$$

$$= \left( \frac{82 - \frac{(-26)^2}{25}}{24} \right) \cdot 5^2$$



(20)

$$= \left( \frac{82 - 27.04}{24} \right) \cdot 25$$

$$= \frac{54.96}{24} \cdot 25$$

$$= 2.29 \cdot 25$$

$$= \boxed{57.25}$$

Standard deviation:-

$$\text{standard deviation} = s = \sqrt{\frac{\sum b \cdot d^2 - \frac{(\sum b \cdot d)^2}{n}}{n-1}} \cdot h$$

$$= \sqrt{\frac{82 - \frac{(-24)^2}{25}}{24}} \cdot 5$$

$$= \sqrt{\frac{82 - 27.04}{24}} \cdot 5$$

$$= \sqrt{\frac{54.96}{24}} \cdot 5$$

$$= \sqrt{2.29} \cdot 5$$

$$= \boxed{7.5664}$$

(21)

Co-efficient of variation:-

$$\text{co-efficient of variation} = \frac{S}{\bar{X}} \cdot 100\%$$

$$= \frac{7.5664}{36.8} \cdot 100\%$$

$$= \boxed{20.56\%}$$

Skewness (SK) =

$$\text{Skewness} = \frac{\text{mean} - \text{mode}}{\text{S.D}}$$

where, mean = 36.80, Mode = 37

and S.D = 7.5666

$$\text{S.K} = \frac{36.80 - 37}{7.566}$$

$$= - \frac{0.20}{7.56}$$

$$= \boxed{-0.026}$$



(22)

(B)

$x$	$1/x$	$\log(x)$	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})$
22	0.045	1.34	-14.8	219.04	14.8
27	0.037	1.4313	-9.8	96.04	9.8
27	0.037	1.4313	-9.8	96.04	9.8
27	0.037	1.4313	-9.8	96.04	9.8
32	0.0312	1.5051	-4.8	23.04	4.8
32	0.0312	1.5051	-4.8	23.04	4.8
32	0.0312	1.5051	-4.8	23.04	4.8
32	0.0312	1.5051	-4.8	23.04	4.8
32	0.0312	1.5051	-4.8	23.04	4.8
32	0.0312	1.5051	-4.8	23.04	4.8
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
37	0.027	1.5682	0.2	0.04	0.2
42	0.0238	1.6232	5.2	27.04	5.2
42	0.0238	1.6232	5.2	27.04	5.2
42	0.0238	1.6232	5.2	27.04	5.2
42	0.0238	1.6232	5.2	27.04	5.2
42	0.0238	1.6232	5.2	27.04	5.2

(23)

47	0.0212	1.6720	1.76	3.10	1.76
47	0.0212	1.6720	1.76	3.10	1.76
57	0.019	1.7558	20.2	408.04	20.2
Total	0.73	38.92	55.31	1172.12	136.40

$$\text{Mean} = \frac{\sum x}{N} = \frac{900}{25} = 36.8$$

$$\text{G.M} = (22 \times 27 \times 27 \times \dots \times 57)^{\frac{1}{25}} = 36.0555$$

$$\text{H.M} = \frac{N}{\sum (\frac{1}{x})} = 35.302$$

$$\text{Median} = \text{size of } \frac{25^{\text{th}}}{2} \text{ item} = \frac{37+37}{2} = 37.$$

$$\text{Mode} = \text{Most frequent value} = 37$$

$$Q_1 = \text{Size of } \frac{25^{\text{th}}}{4} \text{ item} = 6^{\text{th}} \text{ item} = \dots = 32$$

$$Q_3 = \text{Size of } 3 \times \frac{25^{\text{th}}}{4} \text{ item} = 18^{\text{th}} \text{ item} = 42$$

$$D_1 = \text{Size of } \frac{25^{\text{th}}}{10} \text{ item} = \frac{25^{\text{th}}}{10} = 2.5^{\text{th}} = 27$$



(24)

$$P_1 = \text{Size of } \frac{25}{100} \text{th item} = 1^{\text{st}} \text{ item} = 22$$

$$\text{Range} = 57 - 22 = 35$$

$$\text{M.D} = \frac{\sum(x - \bar{x})}{n} = 5.456$$

$$\text{Q.D} = Q_3 - Q_1 = 42 - 32 = 10$$

$$\text{Variance} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = 57.25$$

$$\text{S.D} = \sqrt{57.25} = 7.5666 \quad \text{C.V} = \frac{S}{\bar{x}} \times 100 =$$

$$0.206$$

$$\text{Skewness} = \frac{3(\text{mean} - \text{median})}{\text{S.D}} =$$

$$\frac{3(36.8 - 37)}{7.5666} = -0.08$$

Did direct because shortness of time.