Mid-term paper assignment

Submitted by:	Aiman Orakzai
Submitted to:	Sir Shamim Anwar
Subject :	Biostatistics
I’D number:	13613
Department:	Dental technology
Date:	16/04/2020

Rainfall	Mid points(x)	Number of years	Fx
20 – 24	22	1	22
25 -29	27	3	81
30 -34	32	5	160
35 -39	37	8	296
40 -44	42	5	210
45- 49	47	2	94
50 – 54	52	0	0
55 - 59	57	1	57

a. Find A.M,G.M,H.M,Median,Mode,Quartiles,Deciles,Percentiles,Range,M.D,Q.D,Variance,Standared Deviation, Coefficient of variation,Skewness for the following data.

Rainfall (inches)	Number of Years
20 - 24	1
25 - 29	3
30 - 34	5
35 - 39	8
40 - 44	5
45 - 49	2
50 - 54	0
55 - 59	1

A.M= ∑fx/∑f=920/25=36.8

Where x is the mid point and f is number of years

Median $M=L+\frac{n}{2}-cf}{f}\cdot c$

=34.5+12.5-98·5

=34.5+3.58·5

=34.5+2.1875

=36.6875

Details of median :

“””To find Median Class
= value of (n2)th observation

= value of (252)th observation

= value of 12th observation

From the column of cumulative frequency cf, we find that the 12th observation lies in the class 35-39.

∴ The median class is 34.5-39.5.

L=lower boundary point of median class =34.5

∴n=Total frequency =25

∴cf=Cumulative frequency of the class preceding the median class =9

∴f=Frequency of the median class =8

∴c=class length of median class =5

Median $M=L+\frac{n}{2}-cf\cdot c$

=34.5+12.5-98·5

=34.5+3.58·5

=34.5+2.1875

=36.6875”””

GM $=Anti\log\left(\sum f\log(x)/n\right)$

=*Anti*log(3.5851)

=36.0552

HM $=n/\sum\left(f/x\right)$

=250.7082

=35.3019

Mode

$=34.5+\left(8\text{-}5/(2\cdot 8\text{-}5\text{-}5)\right)\cdot 5$

$=34.5+\left(36\right)\cdot 5$

=34.5+2.5

Mode=37

For details

“”””

To find Mode Class
Here, maximum frequency is 8.

∴ The mode class is 34.5-39.5.

∴L=lower boundary point of mode class =34.5

∴f1= frequency of the mode class =8

∴f0= frequency of the preceding class =5

∴f2= frequency of the succedding class =5

∴c= class length of mode class =5

$Z=L+\left(f_1\text{-}f_02\cdot f_1\text{-}f_0\text{-}f_2\right)\cdot c$

$=34.5+\left(\frac{8-52\cdot 8-5}{-5}\right)\cdot 5$

$=34.5+\left(\frac{36}{}\right)\cdot 5$

$=34.5+2.5$

$=37$

“”

Quartiles

Q3 class : 39.5

For details

“”

Class with (3n4)th value of the observation in cf column

=(3·254)th value of the observation in cf column

=(18.75)th value of the observation in cf column

and 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.

“”

Decile

Here, n=25

D1 class : 27

For details

“”

Class with (n10)th value of the observation in cf column

=(2510)th value of the observation in cf column

=(2.5)th value of the observation in cf column

and it lies in the class 25-29.

∴D1 class : 24.5-29.5

The lower boundary point of 24.5-29.5 is 24.5.

∴L=24.5

$D_1=L+n_{10}\text{-}cff\cdot c$

=24.5+2.5-13·5

=24.5+1.53·5

=24.5+2.5

=27

“”

Percentile

Here, n=25
P10 class :27
for details

“”

P10 class :

Class with (10n100)th value of the observation in cf column

=(10·25100)th value of the observation in cf column

=(2.5)th value of the observation in cf column

and it lies in the class 25-29.

∴P10 class : 24.5-29.5

The lower boundary point of 24.5-29.5 is 24.5.

∴L=24.5

P10=L+10n100-cff·c

=24.5+2.5-13·5

=24.5+1.53·5

=24.5+2.5

=27

“”””

Mean deviation(M.D) :

Mean $\bar{x}=\sum fx/\sum f$

=92025

=36.8

Variance = 54.96

Standard Deviation=7.5664

Skewness=0.428

Question 1 Solution:

In the United Kingdom there has been a national census every 10 years since 1801 (with the exception of 1941). At the time of the 2011 UK census, a government minister described the census as 'expensive, inaccurate and inefficient', and 'out of date almost before it's done'.

The minister also said that data held by the National Health Service, local councils, the postal service, the electoral register, tax returns, credit card firms and phone companies can do the job.

A proposal for the 2021 UK census is that it should be conducted online and that it should incorporate additional data held by government agencies.

a. Describe the purpose of a census.

b. Explaining how it differs from a sample survey and from routine collection of data by government agencies.

c. The 2011 UK census attracted a response rate of about 94% of the population. Discuss whether or not that is a problem for the accuracy of the census.

d. In the 2011 UK census, almost 170 000 people stated their religion as 'Jedi Knight'. (Jedi Knights are characters in the 'Star Wars' films.)
Discuss what responses of this type indicate about the attitudes of some members of the public to the census. Discuss also whether responses of this type invalidate asking a question about religion.

e. Discuss the potential problems in conducting the 2021 UK census online, and explain how these problems might be overcome.

f. Discuss the potential problems in incorporating additional data held by government agencies.

a. A **census** is the procedure of systematically acquiring and recording information about the members of a given population. This term is used mostly in connection with national population and housing **censuses**; other common **censuses** include traditional culture, business, supplies, agricultural, and traffic **censuses**.
b. **Census** covers more **data** items with less detail of every **data** item but **survey** covers less **data** items with more detail of every **data** item.
c. As everybody know 94% is a good rate, but still could not define whether it is a problem for the accuracy of census or not based on **"**There are two sources of **non-response error** in the **census**; person or household **non-response** and item **non- response**."
d. Such answers indicate that input of people was non serious during census and it may contain garbage data. Yes, this type of answers invalidates asking a question about religion.

e. **Coverage errors** that occur due to omission or duplication, but these could be resolved through data integrity methods. and **content errors** arise by incorrect reporting and these can be reduced by consistency checks.
f. Problems include cost pressures, concerns about intrusiveness, privacy and response burden, reduced cooperation, **difficulties** in accessing secure apartments and enumerating unsafe areas, more complex living arrangements, and timeliness concerns.

Given Data

	Men Sample					Women Sample				
	Q4	Q3	Q2	Q1		Q4	Q3	Q2	Q1	
	Mean	Mean	Mean	Mean	SE	Mean	Mean	Mean	Mean	SE
Fresh veg	204	259	266	317	0.9	178	235	266	304	0.8
fruits	31	45	69	105	0.5	28	46	70	121	0.4
rice	367	337	269	246	1	315	276	243	220	0.8
wheat flour	79	114	197	253	1	56	118	141	180	0.8
whole grain	2	2	6	27	0.1	1	3	6	22	0.1
root veg	7	11	16	29	0.1	6	12	17	28	0.1
meat	70	61	69	77	0.4	48	43	54	63	0.3
fish	23	28	31	44	0.2	19	21	28	46	0.2
milk	2	3	23	39	0.3	1	4	15	48	0.1

Q1: Calculate the overall mean consumption of ..... For men and women combined

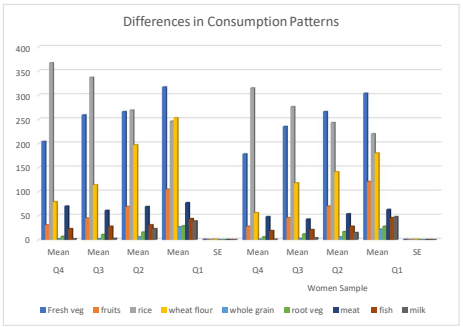
	Men Sample					Women Sample					
	Q4	Q3	Q2	Q1		Q4	Q3	Q2	Q1		
	Mean	Mean	Mean	Mean	Overall mean consumption	Mean	Mean	Mean	Mean	Overall mean consumption	Combined
Fresh veg	204	259	266	317	261.5	178	235	266	304	245.75	245.75
fruits	31	45	69	105	62.5	28	46	70	121	66.25	66.25
rice	367	337	269	246	304.75	315	276	243	220	263.5	263.5
wheat flour	79	114	197	253	160.75	56	118	141	180	123.75	123.75
whole grain	2	2	6	27	9.25	1	3	6	22	8	8
root veg	7	11	16	29	15.75	6	12	17	28	15.75	15.75
meat	70	61	69	77	69.25	48	43	54	63	52	52
fish	23	28	31	44	31.5	19	21	28	46	28.5	28.5
milk	2	3	23	39	16.75	1	4	15	48	17	17

Q2: Describe in words ..... Consumption indicate?

Ans: b. The standard error tells you how accurate the mean of any given sample from that population is likely to be compared to the true population mean. When the standard error increases, i.e. the means are more spread out, it becomes more likely that any given mean is an inaccurate representation of the true population mean.
Milk: low standard deviation value, good accuracy of value
Root Veg Very near to population mean value, good accuracy
Wheat Floor High value inaccurate representation of the true population mean.

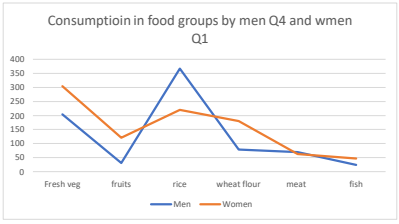
Q4: Draw a suitable diagram .....between men and women?

	Men Sample					Women Sample				
	Q4	Q3	Q2	Q1		Q4	Q3	Q2	Q1	
	Mean	Mean	Mean	Mean	SE	Mean	Mean	Mean	Mean	SE
Fresh veg	204	259	266	317	0.9	178	235	266	304	0.8
fruits	31	45	69	105	0.5	28	46	70	121	0.4
rice	367	337	269	246	1	315	276	243	220	0.8
wheat flour	79	114	197	253	1	56	118	141	180	0.8
whole grain	2	2	6	27	0.1	1	3	6	22	0.1
root veg	7	11	16	29	0.1	6	12	17	28	0.1
meat	70	61	69	77	0.4	48	43	54	63	0.3
fish	23	28	31	44	0.2	19	21	28	46	0.2
milk	2	3	23	39	0.3	1	4	15	48	0.1



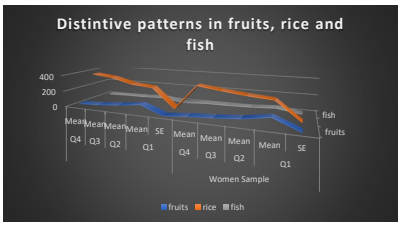
Q5: Men require on average .....vegetables or milk?

	Q4	Q1
	Men	Women
Fresh veg	204	304
fruits	31	121
rice	367	220
wheat flour	79	180
meat	70	63
fish	23	46



Q3:What distintive pattern is there ..... Q4 to Q1?

	Men Sample					Women Sample				
	Q4	Q3	Q2	Q1		Q4	Q3	Q2	Q1	
	Mean	Mean	Mean	Mean	SE	Mean	Mean	Mean	Mean	SE
fruits	31	45	69	105	0.5	28	46	70	121	0.4
rice	367	337	269	246	1	315	276	243	220	0.8
fish	23	28	31	44	0.2	19	21	28	46	0.2



Q6: Describe in your word.....show better result?

Based on the standard deviatioins, it indicates that men men consumed 20% or more food to maintain energy level and accuracy of women S.E is better than that of Men.