

SRI LANKA INSTITUTE OF ADVANCED TECHNOLOGICAL EDUCATION

Higher National Diploma in Information Technology

First Year, Second Semester Examination – 2017

HNDIT1214 / IT12142 – Statistics for IT

Instructions for Candidates:

Answer four (04) questions only.

Non programmable calculators are allowed.

No. of questions : 05

No. of pages : 04

Time : Two (02) hours

Marking Scheme

01.

- (i). a) What is meant by statistics?

Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting numerical data to assist in making more effective decisions.

----- (2 Marks)

- b) What are the importance of statistics? State two (02).

- ***Statistical techniques are used to make decisions that affect our daily lives***
- ***The knowledge of statistical methods will help to understand how decisions are made and give you a better understanding of how they affect you. etc.***

----- (2 = 2 Marks)

- (ii). A student carried out a survey to find out people's opinion on attending sports events in his local area. He stood outside the local sports stadium and asked a random sample of people their opinions on attending sports events as they entered the stadium.

- a) Is the data that the student collects primary or secondary? Justify your answer.

Primary data ----- (1 mark)

Student carried out the survey and collect the data himself ----- (1 Mark)

- b) Give two reasons why this sample may be biased.

Biased

- ***"Local stadium", should consider stadiums outside the locality***
- ***People attending a stadium probably not representative of the general Population.***
- ***The stadium may be holding non sporting event like music or dancing when the survey was carried out. etc.***

----- (2 = 2 Marks)

- c) Make two (02) suggestions to improve the accuracy of his survey.

- ***Increase the number of people surveyed.***
- ***Carryout a house to house survey.***
- ***Conduct a survey outside a supermarket (to increase the population)***

etc. ----- (2 = 2 Marks)

- (iii). Highest recorded temperatures (°F) for 50 selected cities for various periods are as follows:

118	115	114	113	112	111	110	110	109	108
108	107	107	107	106	106	106	106	105	105
105	105	104	104	104	104	103	103	103	102
102	102	102	102	102	102	102	101	101	100
100	99	99	99	98	98	98	97	93	90

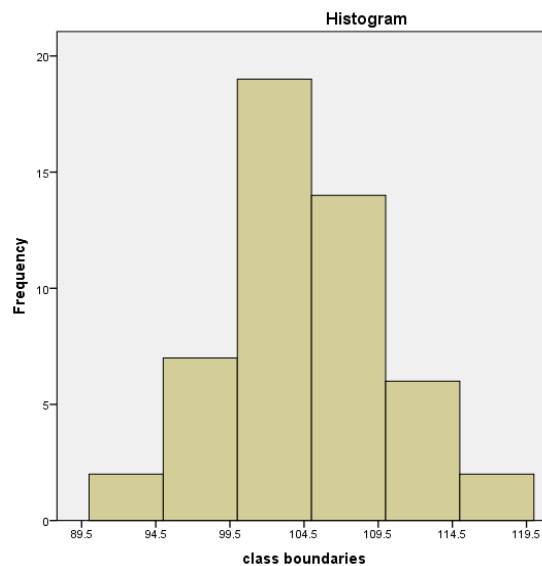
Using the above data,

- a) Construct a frequency distribution starting from multiple of 10 and by selecting class width as 5.

Class Interval	Tally marks	frequency
90 – 94	//	2
95 – 99	/// //	7
100 – 104	/// // // //	19
105 – 109	/// // //	14
110 – 114	/// /	6
115 – 119	//	2

----- (5 Marks)

- b) Represent the above data in a histogram.



----- (6 Marks)

- (iv). Express each of the following in sigma notation.

a) $2 + 5 + 8 + 11 + 14 =$

$$\sum_{n=0}^4 (2 + 3n)$$

----- (2 Marks)

b) $(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + (x_4 - \mu)^2 =$

$$\sum_{i=1}^4 (x_i - \mu)^2$$

----- (2 Marks)

(Total 25 Marks)

02.

- (i). The following table represents the monthly household income (x \$1000) for 20 interview respondents.

Income (x)	Relative frequency (f/N)
$0 < x \leq 5$	0.15
$5 < x \leq 10$	0.05
$10 < x \leq 15$	0.40
$15 < x \leq 20$	0.30
$20 < x \leq 25$	0.10

- a) Redraft the above data in the form of a grouped frequency distribution.

<i>Class Interval</i>	<i>f</i>
0-5	3
5 - 10	1
10 - 15	8
15 - 20	6
20 - 25	2

----- (4 Marks)

- b) Calculate the following for the above data.

<i>Class Interval</i>	<i>f</i>	<i>x_m</i>	<i>fx</i>	<i>F</i>
0 - 5	3	2.5	7.5	3
5 - 10	1	7.5	7.5	4
10 - 15	8	12.5	100	12
15 - 20	6	17.5	105	18
20 - 25	2	22.5	45	20
			$\sum fx = 265$	

- a. Mean

$$\bar{x} = \frac{\sum fx}{\sum f} \quad \text{----- (1 Mark)}$$

$$\bar{x} = \frac{265}{20} = 13.25$$

$$\bar{x} = 13.25 \times 1000 = 13250 \text{----- (2 Marks)}$$

- b. Mode

$$Mode = L_{mo} + \left[\frac{D_a}{D_b + D_a} \right] (C) \quad \text{----- (1 Mark)}$$

Model class = (10 - 15)

$$Mode = 10 + \left[\frac{7}{2+7} \right] (5) = 13.89$$

Mode = 13890 ----- (2 Marks)

c. Quartile deviation

$$Q_1 = L + \frac{\left(\frac{n}{4} - CF \right)}{f} (c)$$

$$Q_1 = 10 + \frac{\left(\frac{20}{4} - 4 \right)}{8} (5) \rightarrow Q_1 = 10.625 \quad \text{----- (2 Marks)}$$

$$Q_3 = L + \frac{\left(\frac{3n}{4} - CF \right)}{f} (c)$$

$$Q_3 = 15 + \frac{\left(3 * \frac{20}{4} - 12 \right)}{6} (5) \rightarrow Q_3 = 17.5 \quad \text{----- (2 Marks)}$$

Quartile deviation = $(Q_3 - Q_1) / 2 = 14.0625$

Quartile deviation = 14062.5 ----- (2 Marks)

- (ii). a) Find x and y so that the ordered data set has a mean of 42 and a median of 35.

17, 22, 26, 29, 34, x , 42, 67, 70, y

$$Median = 35 = \frac{34+x}{2}$$

$x = 36$ ----- (2 Marks)

$$Mean = 42 = \frac{17 + 22 + 26 + 29 + 34 + x + 42 + 67 + 70 + y}{10}$$

$$x + y = 420 - 307 = 113$$

$$y = 113 - 36$$

$$y = 77 \quad \text{----- (2 Marks)}$$

- b) The mean and standard deviation of 20 items were calculated by a student as 40 and 10 respectively. But while calculating them an item 15 was misread as 50. Find the correct mean and standard deviation. (5 Marks)

$$\bar{x} = 40$$

$$\sigma = 10$$

$$40 = \frac{\sum x}{20} \rightarrow \sum x = 800$$

After correcting, $\sum x_{new} = 800 - 50 + 15 = 765$

$$\text{Correct } \bar{x} = \frac{765}{20} = 38.25 \quad \text{----- (2 Marks)}$$

$$\sigma^2 = \frac{\sum x_i^2}{N} - \bar{x}^2$$

$$\sum x_i^2 = 20(10^2 + 40^2) = 34000$$

After correcting, $\sum x_{new}^2 = 34000 - 50^2 + 15^2 = 31725$

$$\sigma_{new}^2 = \frac{31725}{20} - (38.25)^2 = 123.1875$$

$$\sigma_{new} = 11.09$$

Correct $\sigma = 11.09$ ----- (3 Marks)

(Total 25 Marks)

03.

(i). Define the following terms:

a) Permutations

A Permutation is an arrangement of n objects in a specific order using r objects at a time.

$${}_nP_r = \frac{n!}{(n-r)!} \text{ ----- (2 Marks)}$$

b) Combinations

A combination is the selection of r objects from n objects without regard to order.

$${}_nC_r = \frac{n!}{(n-r)!r!} \text{ ----- (2 Marks)}$$

(ii). Show that ${}_nC_r - {}_nC_{n-r} = 0$

$$\begin{aligned} \text{L.H.S.} &= \frac{n!}{(n-r)!r!} - \frac{n!}{(n-(n-r))!(n-r)!} \\ &= \frac{n!}{(n-r)!r!} - \frac{n!}{(n-r)!r!} = 0 \Rightarrow \text{R.H.S.} \text{ ----- (4 Marks)} \end{aligned}$$

(iii). Using the digits 1, 2, 3 and 5, how many 4 digit numbers can be formed if

a) The first digit must be 1 and repetition of the digits is allowed?

$$N = 1 * 4 * 4 * 4 = 64 \text{ ----- (2 Marks)}$$

b) The first digit must be 1 and repetition of the digits is not allowed?

$$N = 1 * 3 * 2 * 1 = 6 \text{ ----- (2 Marks)}$$

c) The number must be divisible by 2 and repetition is not allowed?

$$N = 3 * 2 * 1 * 1 = 6 \text{ ----- (3 Marks)}$$

(iv). In a certain country, the car number plate is formed by 4 digits from the digits 1, 2, 3, 4, 5, 6, 7, 8 and 9 followed by 3 letters from the alphabet. How many number plates can be formed if neither the digits nor the letters are repeated?

$${}_9P_4 * {}_{26}P_3 = 47,174,400 \text{ ----- (5 Marks)}$$

(v). To buy a computer system, a customer can choose one of 4 monitors, one of 2 keyboards, one of 4 computers and one of 3 printers. Determine the number of possible systems that a customer can choose from.

$${}_4C_1 * {}_2C_1 * {}_4C_1 * {}_3C_1 = 4 * 2 * 4 * 3 = 96 \text{ ----- (5 Marks)}$$

(Total 25 Marks)

04.

(i). Define the following terms:

a) Experiment

Any process that yields a result or an observation ----- (2 Marks)

b) Independent event

Two events are independent if the occurrence of the first event has no effect on the probability that the second event will occur. ----- (2 Marks)

c) Mutually exclusive event

**If two events are mutually exclusive if they cannot occur at the same time.
----- (2 Marks)**

(ii). A committee of 5 people is to be formed randomly from a group of 10 women and 6 men. Find the probability that the committee has

a) 3 women and 2 men.

There are ${}_{16}C_5$ ways to select 5 people (committee members) out of a total of 16 people (men and women)

There are ${}_{10}C_3$ ways to select 3 women out of 10.

There are ${}_6C_2$ ways to select 2 men out of 6.

There are ${}_{10}C_3 * {}_6C_2$ ways to select 3 women out of 10 AND 2 men out of 6.

$P(3 \text{ women AND } 2 \text{ men}) = {}_{10}C_3 * {}_6C_2 / {}_{16}C_5 = 0.412087$ ----- (3 Marks)

b) 5 women.

$P(5 \text{ women}) = {}_{10}C_5 * {}_6C_0 / {}_{16}C_5 = 0.0576923$ (in ${}_6C_0$ the 0 is for no men) --- (3 Marks)

c) at least 3 women.

$P(\text{at least 3 women}) = P(3 \text{ women or } 4 \text{ women or } 5 \text{ women})$

$= P(3 \text{ women}) + P(4 \text{ women}) + P(5 \text{ women})$

$= 0.412087 + 0.288461 + 0.0576923 = 0.758240$

----- (4 Marks)

(iii). In a factory, three machines, J , K and L , are used to make biscuits.

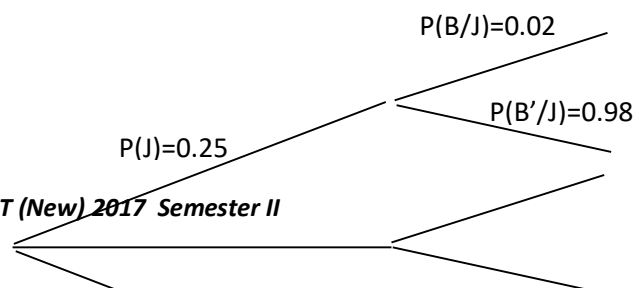
Machine J makes 25% of the biscuits.

Machine K makes 45% of the biscuits.

The rest of the biscuits are made by machine L .

It is known that 2% of the biscuits made by machine J are broken. 3% of the biscuits made by machine K are broken and 5% of the biscuits made by machine L are broken.

a) Draw a tree diagram to illustrate all the possible outcomes and associated probabilities.



$$P(B/K)=0.03$$

$$P(K)=0.45$$

$$P(B'/K)=0.97$$

$$P(B/L)=0.05$$

----- (2 Marks)

A biscuit is selected at random.

- b) Calculate the probability that the biscuit is made by machine J and is not broken.

$$P(J \cap B') = P(J)P(B'/J) = 0.25(0.98) = 0.245$$

----- (2 Marks)

- c) Calculate the probability that the biscuit is broken.

$$\begin{aligned}
 P(B) &= P(J \cap B) + P(K \cap B) + P(L \cap B) \\
 &= P(J)P(B/J) + P(K)P(B/K) + P(L)P(B/L) \\
 &= 0.25(0.02) + 0.45(0.03) + 0.3(0.05) \\
 &= 0.0335
 \end{aligned}$$

----- (2 Marks)

- d) Given that the biscuit is broken, find the probability that it was not made by machine K .

$$\begin{aligned}
 P(K'/B) &= \frac{P(K' \cap B)}{P(B)} = \frac{P(J \cap B) + P(L \cap B)}{P(B)} \\
 &= \frac{0.25(0.02) + 0.3(0.05)}{0.0335} \\
 &= 0.597
 \end{aligned}$$

----- (3 Marks)

(Total 25 Marks)

05.

- (i). State two (02) properties of a discrete probability distribution.

$$\begin{aligned}
 0 \leq p(x) \leq 1 \text{ and} \\
 \sum p(x) = 1
 \end{aligned}$$

----- (1x2 = 2 Marks)

- (ii). A discrete random variable has the following probability distribution.

x	0	1	2	3	4	5
x^2	0	1	4	9	16	25
$P(X = x)$	0.1	0.1	0.2	0.2	0.3	0.1
$x \cdot P(x)$	0	0.1	0.4	0.6	1.2	0.5
$x^2 \cdot P(x)$	0.1	0.1	0.8	1.8	4.8	2.5

- a) Verify that $P(x)$ is a probability distribution.

$$\begin{aligned}
 \sum P(x) &= 0.1 + 0.1 + 0.2 + 0.2 + 0.3 + 0.1 = 1 \\
 \text{since } \sum P(x) &= 1, \text{ it is a probability distribution.}
 \end{aligned}$$

----- (2 Marks)

b) Find the following:

$$\begin{aligned} \text{a. } P(2 \leq x \leq 4) &= P(x=2) + P(x=3) + P(x=4) \\ &= 0.2 + 0.2 + 0.3 = 0.7 \quad \text{----- (2 Marks)} \end{aligned}$$

b. Expected value

$$\begin{aligned} E(x) &= \sum xp(x) \\ &= 0 + 0.1 + 0.4 + 0.6 + 1.2 + 0.5 = 2.8 \quad \text{----- (3 Marks)} \end{aligned}$$

c. Variance

$$\begin{aligned} \sigma^2 &= E(x^2) - (E(x))^2 \\ E(x^2) &= \sum x^2 P(X=x^2) = 0.1 + 0.1 + 0.8 + 1.8 + 4.8 + 2.5 = 10.1 \quad \text{---- (2 Marks)} \\ \sigma^2 &= 10.1 - 2.8 = 7.3 \quad \text{----- (2 Marks)} \end{aligned}$$

(iii). The average number of accidents at a level-crossing every year is 5. Calculate the probability that there are exactly 3 accidents there this year.

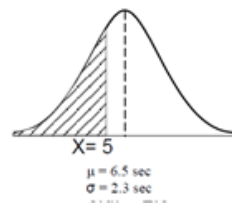
$$\mu = 5$$

$$P(x=k) = \frac{\mu^k e^{-\mu}}{k!}$$

$$P(x=3) = \frac{e^{-5} 5^3}{3!} = 0.14 \quad \text{----- (3 Marks)}$$

(iv). The download time of a resource web page is normally distributed with a mean of 6.5 seconds and a standard deviation of 2.3 seconds.

a) What proportion of page downloads take less than 5 seconds?

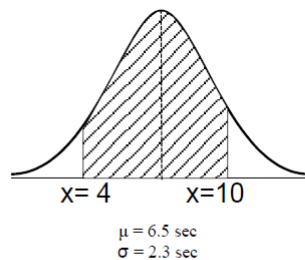


$$\begin{aligned} X=5 &\Rightarrow Z = \frac{5-6.5}{2.3} = -0.65 \\ P(Z < -0.65) &= 0.2578 \end{aligned}$$

i.e., approximately 25% of downloads take less than 5 secs

----- (3 Marks)

b) What is the probability that the download time will be between 4 and 10 seconds?



$$X=4 \Rightarrow Z = \frac{4-6.5}{2.3} = -1.09$$

$$P(Z < -1.09) = 0.1379$$

$$X=10 \Rightarrow Z = \frac{10-6.5}{2.3} = 1.52$$

$$P(Z < 1.52) = 0.9357$$

$$\begin{aligned} \text{Hence } P(4 < X < 10) &= P(-1.09 < Z < 1.52) \\ &= 0.9357 - 0.1379 = 0.7978 \end{aligned}$$

i.e. about 80% of downloads take between 4 and 10 seconds. (3 Marks)

c) How many seconds will it take for 35% of the download to be completed?

$$\text{For } p = 0.35, Z = -0.38$$

$$\text{Hence, } -0.38 = \frac{X-6.5}{2.3}$$

$$\Rightarrow X = -0.38 \times 2.3 + 6.5 = 5.63$$

i.e., 35% of downloads are completed in 5.6 seconds or less. (3 Marks)