

## ASSIGNMENT 1A – Due on Sunday 11:55pm of Week 4.

### TOPICS TESTED: DESCRIPTIVE STATISTICS & PROBABILITY

#### Do's & Don'ts:

\*\*\* All answer in this assignment can be done in any colour **except for red colour** as this is the colour your tutor will use to mark your assignment. Thus, answers in red colour will not be graded (even correct ones) \*\*\*

\*\*\* If you choose to handwrite your answer (scan and submit it electronically), make sure your **handwriting is readable** – this is a good practice for your exam. Failure to comply will result in a lack of marks awarded if your writing is unreadable and there will be no exception for this. \*\*\*

\*\*\* **Each sub-assignment will be worth 2.5%** towards your total score. Sub-assignments will have different point distributions within them as we aim to focus your attention to the important areas; however, the score for a sub-assignment remains 2.5%. \*\*\*

\*\*\* These questions are meant for you to solve **independently**, we encourage students to figure out the questions themselves as it would be good for their understanding of the topics; however, please feel free to consult your tutors if needed. **Plagiarism** (either from using online sources or copying the answers from your classmates) will be punished accordingly. \*\*\*

\*\*\* As this is considered to be an assignment (albeit a sub assignment), requests for special consideration or extension must be submitted at least **2 days BEFORE THE DEADLINE**. The due date is on Sunday, so the latest day you can ask for extension is on Friday (the last official working day of the week for the teaching team). Please follow Monash guideline to request for extensions (medical certificates, doctor or GP letter, etc). Emergencies are to be adjusted individually. \*\*\*

\*\*\* **No R or any other programming languages should be used in solving these questions.** All work for this assignment needs to be done manually, less the use of **non-programmable calculator** (this also applies to your Final Exam). Tutors are not required to answer questions in the difference between manual calculation and programmed calculation\*\*\*

\*\*\* **Late submission is 10% per day, after 5 days you will be given no marks.** Late submission is calculated as follows: If you get 70% on this assignment and you are late for 2 days (you submit on Tuesday), your score is now  $70\% - 20\% (2 \times 10\% \text{ per day}) = 50\%$ . This is done to ensure that the teaching team can release your result as soon as possible so that you can review on your mistakes and have a better study experience. \*\*\*

\*\*\* Please **show all working** in answering questions, your score will be **halved** if you don't comply\*\*\*

\*\*\* Assignments shall be marked completely in **two weeks' time** according to Monash Policies. If there are any changes to the marking time, we will duly inform you. Solutions **will not be released** for this assignment; you can come to the tutorial and ask for an explanation about how to solve the questions after scores are released. \*\*\*

\*\*\* Please do not send emails to tutors asking for suggestions, we have Moodle and consultations for that, In writing your inquiries on Moodle please try to be clear in your problem and not revealing your working to others as this might be counted as plagiarism on your part. A good format for inquiry topic would be "Assignment 1a – Tutorial 10 (your tutorial slot) – Question about median"\*\*\*

\*\*\* Assignments need to be submitted in **PDF** format. Failure to comply will result in 30% penalty\*\*\*

\*\*\* Filename format for submitting assignment "Assignment1A\_StudentId.pdf". File with wrong format incurs 30% penalty \*\*\*

## QUESTIONS:

### A. Descriptive Statistics: (2 Marks Total)

Consider the following set of numbers:

The times, in seconds, taken by 11 students to solve a puzzle are listed in order 7, 12, 13, 17, 18, 20, 22, 24, 25, 30, 45.

- 1) What is the median? **(0.5 Marks)**
- 2) What is the lower quartile and IQR? **(0.5 Marks)**
- 3) What is the sample standard deviation of the time data? **(0.5 Marks)** Are there any outliers? If there are, compute a new sample standard deviation with the outlier removed. **(0.5 Marks)**

### ANSWERS:

Descriptive statistics:  
Dataset: 7, 12, 13, 17, 18, 20, 22, 24, 25, 30, 45

1. Median = 20 → Middlemost value.

2. From the above set,

7, 12, 13, 17, 18, 20, 22, 24, 25, 30, 45

Q1                      Q2                      Q3

Lower quartile = Q1 = 13  
Upper quartile = Q3 = 25  
IQR = Inter-quartile range  
= 25 - 13  
= 12

3. Sample standard deviation:

$$= \sqrt{\frac{\sum (x - \bar{x})^2}{(N-1)}}$$

$\bar{x} = \frac{7 + 12 + 13 + 17 + 18 + 20 + 22 + 24 + 25 + 30 + 45}{11}$   
 $= \frac{233}{11} = 21.18$

sample standard deviation

$$= \sqrt{\frac{\sum |x - u|^2}{(N-1)}}$$

$$= \sqrt{\frac{(201.0724) + (84.2724) + (66.9124) + (17.4724) + (10.1124) + (1.3924) + (0.6724) + (7.9624) + (14.5924) + (77.7924) + (567.3924)}{10}}$$

$$= \sqrt{\frac{1049.6364}{10}}$$

$$= 10.25 //$$

Now, Lower inner fence

$$= Q1 - 1.5 \times IQR$$

$$= 13 - 1.5 \times 12$$

$$= -5 //$$

Upper inner fence

$$= Q3 + 1.5 \times IQR$$

$$= 25 + 1.5 \times 12$$

$$= 43 //$$

True outlier = 45

Now,  $SD_{new} = \sqrt{\frac{\sum |x - u_{new}|^2}{(N-1)}}$

$$N = 10$$

$$u_{new} = \frac{7 + 12 + 13 + 17 + 18 + 20 + 22 + 24 + 25 + 30}{10}$$

$$= 18.8 //$$

Now,  $SD_{new} = \sqrt{\frac{2128}{45}} = 6.876 //$

**B. Probability: (8 Marks Total)**

1) Randomly pick two numbers from  $n$  numbers  $(1, 2, \dots, n)$ , what is the probability that one number is smaller than  $k$  ( $1 < k < n$ ), and the other one is larger than  $k$ ? **(1 Marks)**

**ANSWERS:**

B) PROBABILITY:

i. sample set =  $(1, 2, 3, \dots, n)$

Let us select two numbers  $n_1, n_2$  from this dataset.

$1, 2, 3, 4, \dots, n$

Let us assume this placement of  $k$ .

Now, numbers less than  $k \rightarrow (k-1)$

No. of ways of selecting a no.  $n_1$  from  $(k-1)$  numbers,

$= {}^{k-1}C_1 \quad \text{--- (1)}$

ii. Numbers greater than  $k$ , lesser than  $n \rightarrow (n-k)$

Thus, no. of ways of selecting a no.  $n_2$  from  $(n-k)$  numbers,

$= {}^{n-k}C_1 \quad \text{--- (2)}$

Now,

number of ways of selecting 2 nos from  $n$  numbers

$= {}^nC_2 \quad \text{--- (3)}$

Thus, required probability =  $P[n_1 < k < n_2]$

$= \frac{{}^{k-1}C_1 \times {}^{n-k}C_1}{{}^nC_2}$

$= \frac{(k-1)(n-k)}{{}^nC_2} = \frac{2(k-1)(n-k)}{n(n-1)/2}$

where  ${}^nC_2$

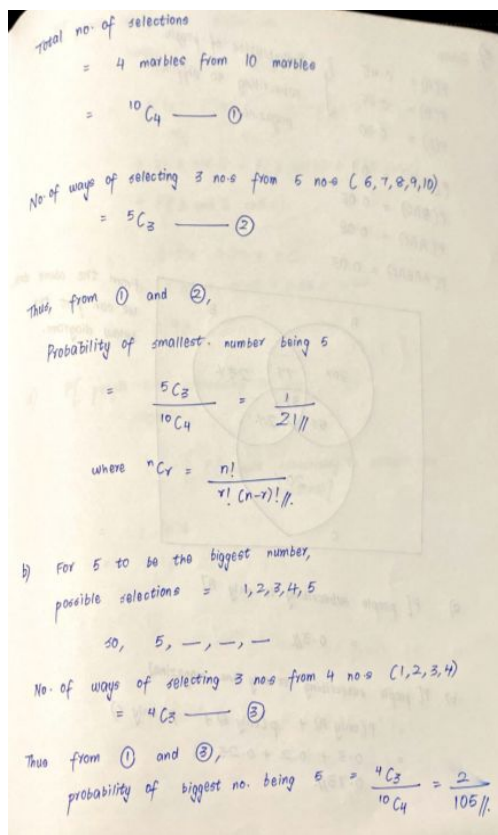
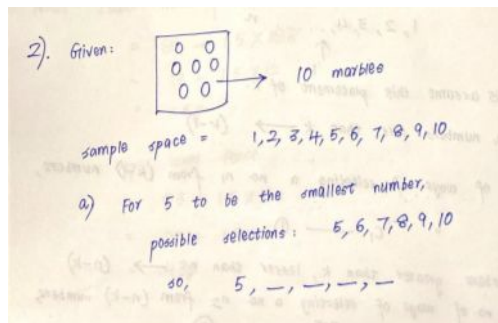
$= \frac{n!}{2!(n-2)!}$

$= \frac{n(n-1)}{2}$

2) A jar contains ten marbles (marked from 1 to 10). Four marbles are drawn without replacement, the number noted. Find the probability of:

- Smallest number is 5? (1 Marks)
- Biggest number is 5? (1 Marks)

**ANSWERS:**



3) There are 3 different magazines A, B and C offered for subscription for a community. 45% of the people in that community subscribe to A, 35% subscribe to B, 30% subscribe to C. 10% subscribe to A & B, 8% subscribe to A & C, 5% subscribe to B & C, 3% subscribe to all the three. Find the probability of:

- only subscribe to A; **(0.5 Marks)**
- only subscribe to one magazine; **(0.5 Marks)**
- subscribe to at least one magazine; **(1 Marks)**
- not subscribe to any magazine. **(1 Marks)**

**ANSWERS:**

3) Given:

$P(A) = 0.45$   
 $P(B) = 0.35$   
 $P(C) = 0.30$

$P(A \cap B) = 0.1$   
 $P(B \cap C) = 0.05$   
 $P(A \cap C) = 0.08$   
 $P(A \cap B \cap C) = 0.03$

Probabilities of people subscribing to different magazines:

From the above data, we can plot the below diagram.

a)  $P[\text{people subscribing to only A}]$   
 $= 0.30$

b)  $P[\text{people subscribing to only one magazine}]$   
 $= P(\text{only A}) + P(\text{only B}) + P(\text{only C})$   
 $= 0.30 + 0.23 + 0.20$   
 $= 0.73$

c)  $P[\text{people subscribing to at least one magazine}]$

$$\begin{aligned}
 &= P(A) + P(\text{only } B) + P(\text{only } C) \\
 &\quad + P(A \text{ and } B) + P(B \text{ and } C) + P(A \text{ and } C) \\
 &\quad + P(A \text{ and } B \text{ and } C) \\
 &= 0.3 + 0.23 + 0.2 \\
 &\quad + 0.07 + 0.03 + 0.05 + 0.02 \\
 &= 0.911.
 \end{aligned}$$

d)  $P[\text{people not subscribing to any magazine}]$

$$\begin{aligned}
 &= 1 - [P(\text{people subscribing to at least one magazine})] \\
 &= 1 - 0.9 \\
 &= 0.111.
 \end{aligned}$$



4) A toy factory has two machines, A and B. Machine A manufactures 70% of toys and machine B manufactures 30%. 80% of the toys from machine A and 90% of the toys from machine B are rated of standard quality.

- A toy is chosen at random and is found to be of standard quality. What is the probability that it has come from machine A? **(1 Marks)**
- If two toys are chosen at random and one of them is found to be of standard quality. What is the probability that the other one is defective? **(1 Marks)**

**ANSWERS:**

4).

Machine A: 70% (0.7) of toys. 80% (0.8) of toys from A are standard quality (SQ). 20% (0.2) of toys from A are non-standard quality (NSQ).

Machine B: 30% (0.3) of toys. 90% (0.9) of toys from B are standard quality (SQ). 10% (0.1) of toys from B are non-standard quality (NSQ).

Probabilities:

- $P(A \cap SQ) = 0.56$
- $P(A \cap NSQ) = 0.14$
- $P(B \cap SQ) = 0.27$
- $P(B \cap NSQ) = 0.03$

Legend: SQ: standard quality, NSQ: Non-standard quality

9) 
$$P(A|SQ) = \frac{P(A \cap SQ)}{P(SQ)}$$
  
[Apply conditional probability].

$$= \frac{0.56}{0.56 + 0.27}$$

$$= \frac{56}{83}$$

$$= 0.674711$$

4b)

$P[\text{Finding a defective toy, when one of them is of standard quality}]$

= 2

Probability of finding a non-standard quality toy =  $P(S)$

$$= (0.8 \times 0.7) + (0.9 \times 0.8)$$

$$= 0.56 + 0.72$$

$$= 0.83 \text{ --- (1)}$$

- Probability of finding a non-standard quality toy =  $P(NS)$

$$= (0.7 \times 0.2) + (0.3 \times 0.1)$$

$$= 0.14 + 0.03$$

$$= 0.17 \text{ --- (2)}$$

Probability of finding a defective toy, when one of them is of standard quality

$$= \frac{P(S) P(NS) + P(NS) P(S)}{P(S) P(NS) + P(S) P(NS) + P(NS) P(S)}$$

$$= \frac{(0.83)(0.17) + (0.17)(0.83)}{(0.83)(0.17) + (0.17)(0.83) + (0.83)(0.83)}$$

$$= 0.2906\%$$