WEEK 0 TUTORIAL

Arithmetic revision

1. Suppose that $x_1 = 3$, $x_2 = -2.5$, $x_3 = \sqrt{2}$, $x_4 = \sqrt{5}$, and $y_1 = -2.6$, $y_2 = -1$, $y_3 = \sqrt{6}$, $y_4 = \sqrt{5}$. Calculate the following quantities:

a. $\sum_{i=1}^{i=3}(x_i + y_i) = 0.764$

b. $\sum_{i=2}^{i=4} x_i^2 y_3^2 = 79.5$

c. $x_1^3 + \sum_{i=1}^{i=2} \frac{y_i^3}{x_i^2} = 24.89$

2. Given

$x_1 = 3$, $x_2 = 1$, $x_3 = 4$, $x_4 = 6$ and $x_5 = 8$,
$p_1 = 1/4$, $p_2 = 1/8$, $p_3 = 1/8$, $P_4 = 1/3$, and $p_5 = 1/6$, find:

a. $\sum_{i=1}^{4} p_i x_i = 3.375$

b. $\sum_{i=3}^{5} p_2 x_i = 2.25$

c. $\sum_{i=2}^{5} p_i x_i^2 = 24.79$

d. $\prod_{i=1}^{5} x_i = 576$

Statistical revision

3. All data can be classified as one of the two general types: **quantitative (numerical)** and **qualitative (categorical).** Quantitative data are measurements that are recorded on a naturally occurring numerical scale. On the other hand, qualitative data are measurements that cannot be measured on a natural numerical scale; they can only be classified into one of a group of categories.
- **Qualitative data** are considered ***nominal*** if there is no order in the dataset and can be represented as labels (e.g. Chocolate chip colors in M&Ms). They are considered ***ordinal*** if the data can be classified on a relational ranking (e.g. ranking of a lecturer: excellent, good, average, bad).
- **Quantitative data** are classified as ***discrete*** if the number of possible outcomes are countable (e.g. number of students in the classroom) or ***continuous*** if the number of possible outcomes are uncountable (e.g. temperature, weight)

Place these variables in the following classification tables. Justify your answer briefly.

a. Deposit at a bank

b. Gender

c. Sales volume of iPhone 15s

d. Soft drink preference (Pepsi, Coke, Sprite)

e. Temperature in Celsius

f. IELTS scores

g. Student names

h. Rating of a Statistics professor (bad, average, good)

i. Number of home video screens

j. Age brackets of 18–30, 31–50, 51–70, 70+.

k. Passport number.

l. A country's inflation rate.

Qualitative (Categorical)		**Quantitative (numerical)**	
Nominal	**Ordinal**	**Discrete**	**Continuous**
b, d, g, k	h, j	c, f, i	a, e, l

4. Fill in the blanks with the following terms: sample, population, Statistics, Descriptive Statistics, Inferential Statistics, variable.

a. The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions is referred to as _ *Statistics* _.

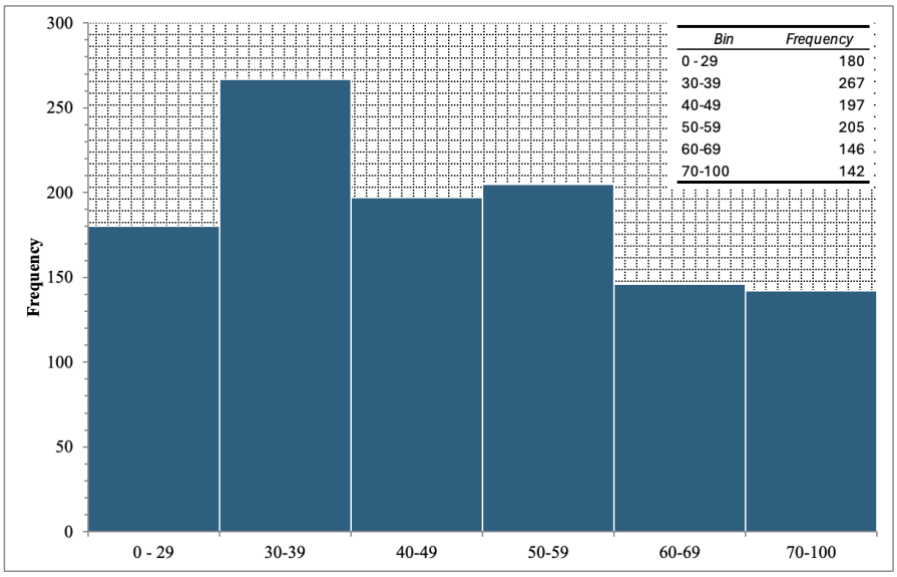
b. Methods of organizing, summarizing, and presenting data in an enlightening way are called _ *Descriptive Statistics* _.

c. The methods used to estimate a value of a population on the basis of a sample are called _ *Inferential Statistics* _.

d. A portion, or part, of the group of interest is referred to as a _*sample*_.

e. The entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest is known as a _*population*_.

5. The overall marks scored by 1137 level 4 students of Introduction to Statistics and Data Science module are shown in intervals in the following histogram.



Using the information from the histogram, fill in the following table:

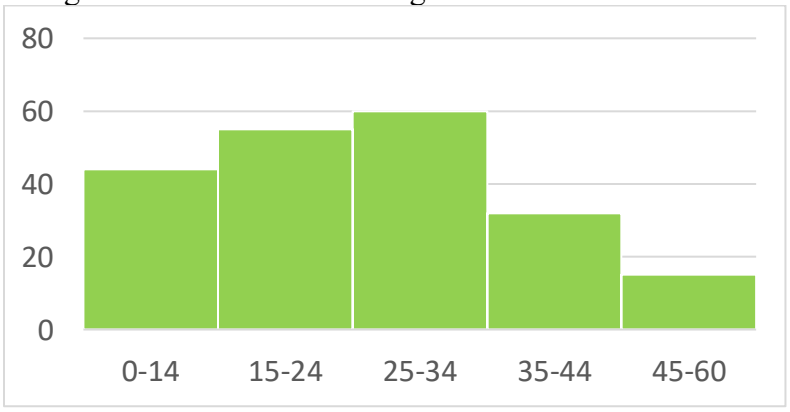
Mark Range	Frequency	Relative Frequency	Cumulative Frequency
$0 \leq \text{Mark} \leq 29$	180	180/1137 = 0.16	180
$30 \leq \text{Mark} \leq 39$	267	0.23	180 + 267 = 447
$40 \leq \text{Mark} \leq 49$	197	0.17	644
$50 \leq \text{Mark} \leq 59$	205	0.18	849
$60 \leq \text{Mark} \leq 69$	146	0.13	995
$70 \leq \text{Mark} \leq 100$	142	0.12	1137
Total	1137	1	1137

6. A financial advisor recorded the lengths of time that she spent in discussions with her clients during the last three months.

Length (minutes)	0-14	15-24	25-34	35-44	45-60
# of clients	44	55	60	32	15

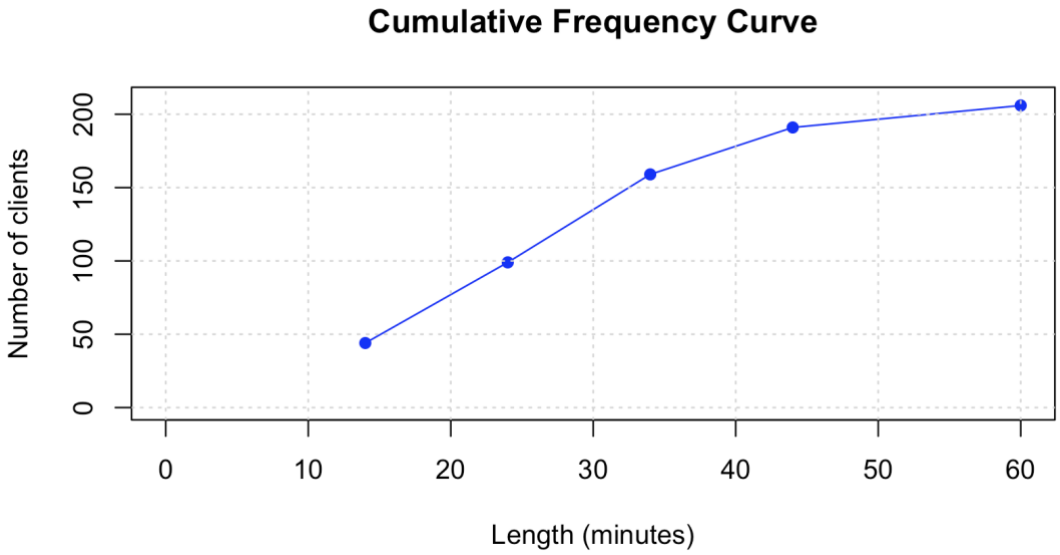
a. Represent the data given in the table in a histogram.

b.



b. Draw a cumulative frequency graph for the grouped data.

Length (minutes)	0-14	15-24	25-34	35-44	45-60
# of clients	44	55	60	32	15



c. Estimate the percentage of client who spent between 15 and 34 minutes.

(55+60)/206*100% = 55.8%

7. The 15 values in each of two sets of related data are denoted by x and y.

a. If y = 3x – 2, find $\bar{y}$, given that $\sum x = 90$

$\bar{y}$ = 3*90/15 – 2 = 16

b. If 5x + 3y = 13, find $\bar{x}$, given that $\sum x = -40$

$\bar{x} = \frac{-40}{15} = -2.67$

8. A set of n data values is denoted by y and it is given that $\sum\left(\frac{1}{2}y - 1\right) = 56 \ and \ \bar{y} = 18.$

Find the value of n.

n = 7