INTRODUCTION

1.0 STATISTICS

Statistics: This is concerned with the collection, ordering and analysis of data

Data: This consists of sets of recorded obeservations or values.

Statistic or Sample Statistic: Any quantity obtained from a sample for the purpose of estimating a population parameter.

Variable: Any quantity that can have a number of values. It may be discrete or continuous. A variable is any characteristic, number, or quantity that can be measured or counted. A variable may also be called a data item.

Discrete Variable: This is a variable that can be counted, or for which there is a fixed set of values. For example, the number of components in a machine.

Continuous Variable: This is a variable that can be measured on a continuous scale, the result depending on the precision of the measuring instrument, or the accuracy of the observer. E.g. the speed of rotation of a shaft, temperature of a coolant etc.

A statistical exercise normally consists of four stages:
1. Collecting of data, by counting or measurement [my addition: or webscraping or interviews etc.]
2. Ordering and presentation of data in a convenient form.
3. Analysis of the collected data.
4. Interpretation of the results and conclusions formulated

1.2 SAMPLING THEORY

In practice, we are interested in drawing valid conclusions about a large group of individuals or objects. Instead of examining the entire group, called the population, which may be difficult or impossible to do, we may examine only a small part of this population, which is called a sample.

CURRICULUM
PART A:
1. Measures of Central Tendency
2. Measures of Dispersion
3. Measures of Partition
4. Data Representation

PART B
1. Random Variables
2. Probability Distribution
3. Expectation of Random Variables
4. Moment Generating Functions
5. Discrete Distribution
6. Continuous Distribution
7. Joint Probability

PART C
1. Probability

PART D
1. R Programming.

NOTES
1.0 RANDOM VARIABLES
A random variable is a quantity that can be assigned a numerical value. A random variable is a variable that takes on numerical values according to a chance process. Random variables are ways to map outcomes of random processes to numbers. If you have a random process (like rolling a dice), you're mapping outcomes of that to numbers (quantifying the outcomes).

Suppose we are about to roll a die 4 times and record the number of sixes. The number of sixes in 4 rolls is random variable that will eventually take on a value. One of these values could be 0,1,2,3,4

There are two types of random variables
1. The Discrete Random Variable (DRV)
2. The Continuous Random Variable (CRV)

1. Discrete Random Variable: This is a random variable that can be assigned (distinct) whole number values i.e. They can take on a countable number of possible values. e.g. given a

random variable X, its value is 1 when the tossed coin is heads and its value is 0 when the tossed coin is tails. Number of lottery tickets purchased until the first winning ticket. Number of courses a randomly selected university student is taking.

2. Continuous Random Variables: These are random variables that exist between intervals. Decimal numbers exist here. e.g. Y = exact mass of a random animal selected at the VI zoo. The time until a newly released website gets its first hit. Height of a randomly selected adult Canadian male.

QUESTIONS
Approximately 3% of the US adult population is under correctional supervision. Suppose we randomly sample 2 US adults. Let X represent the number of adults in our sample that are under correctional supervision.

List the possible values of X and their probability of occuring.
The probability of someone in correctional supervision is0.03 and the probability of not is 0.97

Possibilities: NN, NC, CN, CC
Values of X : 0   1   1   2
Probabilities: 0.97 times 0.97, 0.97 times 0.03, 0.03 times 0.97, 0.03 times 0.03
Proababilities: 0.9409, 0.0291, 0.0291, 0.0009
P(X=0) = 0.9409
P(X=x) ⇒ P(x)

PROBABILITY DISTRIBUTION

The probability distribution for a random variable X is a listing of all possible values of X and their probabilities of occuring. This could be a table or some formula with a graphical representation.

1. Discreted Distribution
2. Continuous Distribution

Let X be the number of “heads” after 3 flips of a fair coin

HHH
HHT

HTH
HTT
THH
THT
TTH
TTT

P(X=0) = 1 over 8 // Means probability of getting 0 heads
P(X=1) = 3 over 8
P(X=2) = 3 over 8
P(X=3) = 1 over 8

Now how do we distribute it?
On our graph, the vertical is the probability (from 0 to 1)
On the horizontal, we have the outcomes.
Then for each value, we draw a bar (like a histogram)

1. Discrete Distribution: These are distributions that are characterised by discrete random variables. Examples include:
a. Bernoulli Distribution
b. Binomial Distribution
c. Poisson Distribution
Multinomial distribution
Negative binomial distribution
Geometric distribution
Hypergeometric distribution

All discrete probability distributions must satisfy
1. $0 \leqslant p(x) \leqslant 1$, for all x
2. $\sum p(x) = 1$

PROBABILITY DENSITY FUNCTION
This is a function that helps in calculating probabilities
The PDF of a discrete distribution is a table.

x	1	3	4
P(x)	1/6	1/3	1/2

P(x=1)=1/6
P(x=3)=1/3
P(x=4)=1/2

AXIOMS (RULES)
$p(x) \geq 0$
$p(x) \leq 1$
$\sum p(x) = 1$

EXPECTATION OF A RANDOM VARIABLE
The expected value of a random variable is the theoretical mean of the random variable. It is not based on sample data it is based on distribution.
$E(X)=\mu=\sum X\,P(X)$
It is the expected value from a discrete distribution. This is also the mean

x	1	2	3
P(x)	1/6	1/3	1/2

$\sum x\,P(x)=1\left(\frac{1}{6}\right)+2\left(\frac{1}{3}\right)+3\left(\frac{1}{2}\right)$
$\sum x\,P(x)=\frac{1}{6}+\frac{4}{6}+\frac{9}{6}$
$\sum x\,P(x)=\frac{14}{6}$
$mean(\mu)=\frac{7}{3}$

EXPECTATION OF A FUNCTION g(X)
$E[g(X)]=\sum g(X)\,p(X)$

PROPERTIES OF EXPECTATION
The expectation of a random variable (x) is:
1. $E(x)=\mu$
2. $E(x+y)=E(x)+(y)$
3. $E(ax)=aE(x)$ where a is a constant
4. $E(a)=a\,e.g.E(6)=6$
5. $E(xy)=E(x)\cdot E(y)$. x and y must be independent.

EXAMPLES
1. Solve $E(x+\mu)^2$
$E(x^2+2\mu+\mu^2)$
$E(x^2)+E(2x\mu)+E(\mu^2)$
$E(x^2)+2\mu E(x)+\mu^2$
$E(x)=\mu$
$E(x^2)+2\mu\cdot\mu+\mu^2$
$E(x^2)+3\mu^2$

2. $E(x-\mu)^3$
Answer: $E(x^3)-3\mu E(x^2)+2\mu^3$

VARIANCE
Variance is the average squared distance from the mean

The general formula for variance is:
$Var(x)=E(x-\mu)^2$
$Var=E(x^2)-\mu^2$
$Var=\sum x^2 P(x)-\mu^2$

Generally, the variance of X is
$Var(x)=E[X-\mu^2]=\sum (x-\mu)^2 p(X)$
Variance Var(X) can also be represented as σ^2

Practice Questions
1. Find the variance of the discrete distribution

x	4	8	12
P(x)	1/4	1/4	1/2

$Var(x)=E(x-\mu)^2$
$Var(x)=\sum (x-\mu)^2 P(x)$
$\mu=24/3$
$\mu=8$
$Var(x)=(4-8)^2\left(\frac{1}{4}\right)+(8-8)^2\left(\frac{1}{4}\right)+(12-8)^2\left(\frac{1}{4}\right)$
$Var(x)=\frac{16}{4}+0+\frac{16}{4}$
$Var(x)=8$

PROPERTIES OF VARIANCE
1. $Var(x)=E(x-\mu)^2$
2. When simplified, $Var(x)=E(x^2)-\mu^2$
3. Variance of a constant is 0, Var(a) = 0, Var(5)=0
Independent Events
4. Var(x+y)
Var(x)+Var(y)
E.g.
Var(x+2)
Var(x)+Var(2)
Var(x)+0
5. Var(x+a) = Var(x)
Dependent Events
6. Var(x+y)
Var(x)+Var(y)+2Cov(x,y)

QUESTIONS
1.Suppose 60% of American adults approve of the way the president is handling his job. Randomly sample 2 American adults. Let X represent the number that approve.

INTRODUCTION TO DISCRETE RANDOM VARIABLES.

CONTINUOUS DISTRIBUTION
The values are characterised by intervals
e.g. $P(x<2)$
$P(x>2)$
$P(0<x<2)$

PROBABILITY DENSITY FUNCTION
Ask for help here…
$$P(x)=\frac{e^{2x}}{10}$$
$$P(x)=\frac{x^{2-4x}}{100}$$
Find the probability from $P(x)=(0<x<2)$
$$P(0<x<2)=\int_0^2 \frac{x^{2-4x}}{100}\,dx$$

AXIOMS OF PROBABILITY
$0 \leqslant P(x) \leqslant 1$
That means $0\leq\int_a^b P(x)\,dx\leq 1$
PIECEWISE
$$P(x)=\begin{cases}\frac{81-t^2}{100}:0<t<9\\ 0:otherwise\end{cases}$$
Example 1. Find C if
$$P(x)=\begin{cases}c\,x^2:|x|\leq 1\\ 0:otherwise\end{cases}$$
Solution
Convert the absolute value
$|x|\leq 1$
$-1 \leqslant x \leqslant 1$
Applying the axiom
$$\int_a^b P(x)\,dx=1$$
$$\int_{-1}^1 C\,x^2\,dx=1$$
$$C=\frac{3}{2}$$

Use this to solve questions 2 and 3
$$P(x)=\begin{cases}\frac{3x^2}{2}:-1\leq x\leq 1\\ 0:otherwise\end{cases}$$
2.

$P\left(x \geq \frac{1}{2}\right)$

The boundary of the above probability is $\frac{1}{2} \rightarrow \infty$

$$\int_{\frac{1}{2}}^{1} \frac{3x^2}{2} dx = 0.4375$$

3. P(x ⩽ 0.6)

The range of the probability is $-\infty \rightarrow 0.6$

$$\int_{-\infty}^{0.6} \frac{3x^2}{2} dx$$

EXPECTATION (MEAN)

$E(x) = \mu$

$$E(x) = \int_{-\infty}^{\infty} x\, P(x)\, dx$$

VARIANCE

$Var(x) = E(x-\mu)^2 = E(x)^2 - \mu^2$

$$Var(x) = \int_{-\infty}^{\infty} x^2 P(x)\, dx - \mu^2$$

MOMENT GENERATING FUNCTION (FOR PIECEWISE FUNCTION)

$$m_T = E\left(e^{tx}\right) = \int_{-\infty}^{\infty} e^{tx} P(x)\, dx$$

4. Find the mean, variance and moment generating function of

$$P(x) = \begin{cases} \frac{3x^2}{2} : -1 \leq x \leq 1 \\ 0 : otherwise \end{cases}$$

i. E(x) = 0

ii. Var(x) = 0.6

iii. m_T

CUMULATIVE DISTRIBUTION FUNCTION

Step 1: $x < -\infty$, $f = 0$

Step 2: $x > \infty$, $f = 1$

Step 3: $-\infty \rightarrow x$