**INTRODUCTION TO PROBABILITY USING R**

BEING A SHORT TRAINING PRESENTED AT OFFA R USERS GROUP MEETING ON 13TH FEBRUARY, 2024 AT STATISTICAL LABORATORY, STATISTICS DEPARTMENT, THE FEDERAL POLYTECHNIC OFFA, NIGERIA

BY

**UDOKANG, ANIETIE EDEM (OGANIZER, ORUG)**

CHIF LECTURER, STATISTICS DEPARTMENT,THE FEDERAL POLYTECHNIC OFFA, NIGERIA

**1. Introduction**

This topic is carefully chosen because of the important of probability in Statistics, most especially when it has to do with inferential Statistics.

We deal with probability in many fields of Statistics such as Econometrics, Time Series, Biostatistics, Medical Statistics, Sampling Theory, Inference and Business Statistics.

Welcome on board as we discuss the elementary part of Statistics that has to do with the Concepts of Probability.

Before then, let’s get ourselves familiar with R Software.

**2. What is R?**

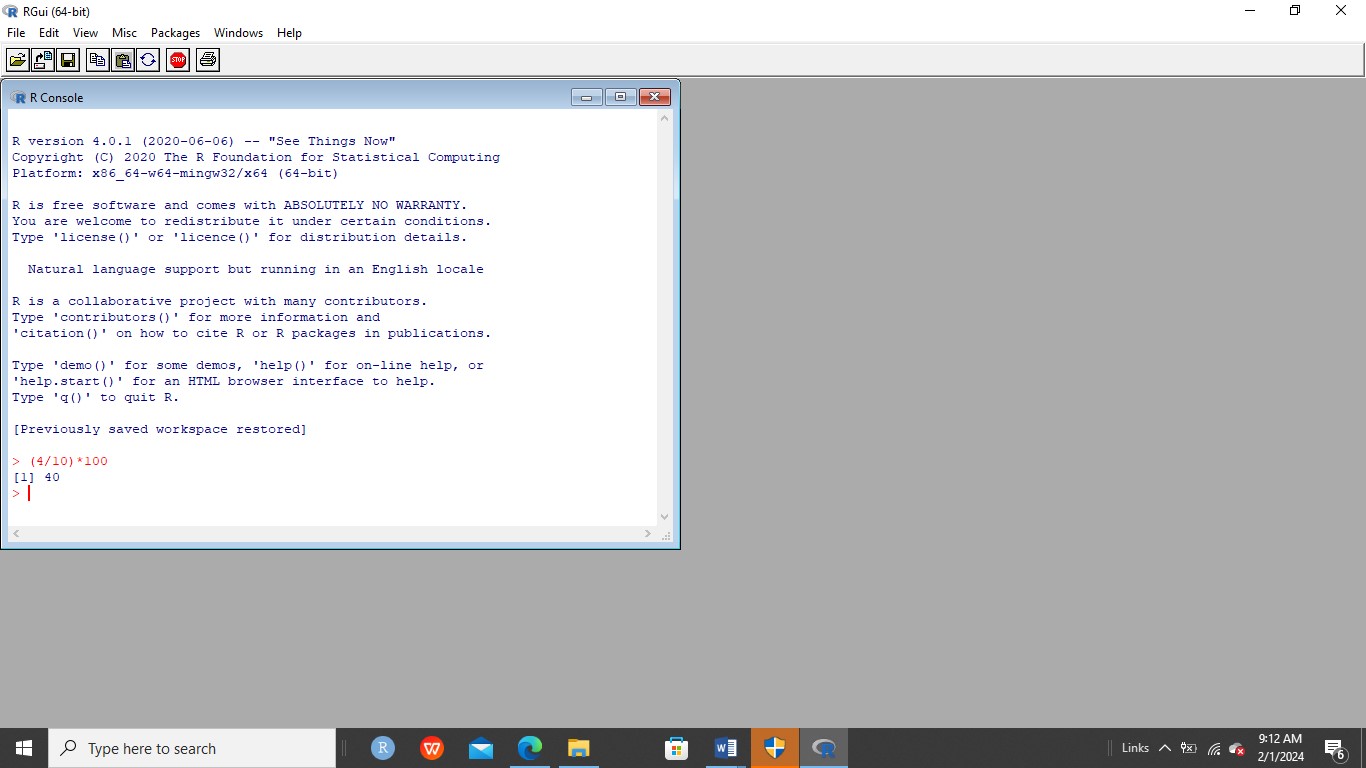
R is a free statistical software created by statisticians Ross Ihaka and Robert Gentleman and supported by R Core Team and the R Foundation for Statistical Computing.

**3. Where can I Find R?**

Visit: <https://cran.r-project.org/>

Follow the instructions for free download and install in your system.

**Display of the R Console (**A Tool to type command and see the result of the command**)**



**4. What is Probability?**

Probability can be defined as a quantity from 0 to 1used to measure the likelihood or the chance of an event occurring.

**5. Two Approaches of Probability**

1. **Classical Approach** -This probability approach assumes equally likely outcomes based on prior knowledge.
2. **Relative Frequency** – This approach is based on observed occurrences over a large number of trials.

**Formula**

P(A) = n(A)/n(S)

Or

P(A) = f/n

Where

P(A) = number of event A

n(S) = number all possible events = sample space

f = frequency of a sub group

n = Total frequency of all the groups = sample size

**Illustrative Examples (R-4.0.1)**

**Example 1 (Classical Approach)**

An unbiased coin has two sides Head (H) and Tail (T). This implies that each side has an equal chance of occurrence when toss or flip. What is the probability of having a head when the coin is tossed once?

**R code**

outcomes <- c("Head", "Tail")

total\_outcomes <- length(outcomes)

head.outcome <- length(outcomes[outcomes == "Head"])

prob.head <- head.outcome / total\_outcomes

prob.head

**Example 2 (Classical Approach)**

Consider a situation that an unbiased coin is tossed twice or two of the coins are tossed twice. The sample space is HH, HT, TH and TT.

What is the probability of having two heads?

**R code**

outcomes <- c("HH", "HT", "TH", "TT")

total\_outcomes <- length(outcomes)

head.outcomes <- length(outcomes[outcomes == "HH"])

prob.head <- head.outcomes / total\_outcomes

prob.head

**Example 3 (Relative Frequency Approach)**

The scores and grades students in an examination are 75 (A), 75 (A), 70 (AB), 70 (AB), 70 (AB), 70 (AB), 70(AB), 66(B), 66 (B) and 66(B).

What is the probability of a student having AB in the examination?

There are ten students in the sample space, out which three have AB.

**R code**

grades <- c("A", "A", "AB", "AB","AB","AB","AB","B","B","B")

total.grades <- length(grades)

AB.grades <- length(grades[grades == "AB"])

prob.AB <- AB.grades / total.grades

prob.AB

We can as well put up a frequency distribution table and get the probabilities of 75 (A), AB (70) and B (66).

**R code**

library ('data.table')

Score<-c(75,75,70,70,70,70,70,66,66,66)

factor(Score)

table(Score)

prop.table(table(Score))

**6. Probability Properties with Illustrations and Examples**

**1. 0<P(A)<1,** where **A** is an event.

**2. P(A) = 1 – P(AI)**, where **AI** is the complement of **A**.

**3. P(**Φ) = 0, where Φ is the null or empty set.

**4. P(S) = 1**, where S is the sample space.

**Illustrative Examples of the properties**

Let a set S be a sample space with all integers from 1 to 10 inclusive.

Then S = {1,2,3,4,5,6,7,8,9,10}

Let A be all the elements in S less than 5.

A = {1,2,3,4,}

Let B be all the elements in S more than 4.

B = {5,6,7,8,9,10}

This example will be use to explain properties 1 to 4.

**Property 1 0<P(A)<1**

Property one will be seen while illustrating propertie 2to 3.

**Property 2 P(A) = 1 – P(AI)**

**R Code (the step by step codes how written to aid in the illustration)**

S <- c(1,2,3,4,5,6,7,8,9,10)

A<- c(1,2,3,4)

B<- c(5,6,7,8,9,10)

A.Complement=res <- S[is.na(pmatch(S,A))]

A.Complement

length(S)

length(A)

length(A.Complement)

prob.A <- sum(length(A)) / length(S)

prob.A

prob.A.Complement <- sum(length(A.Complement)) / length(S)

prob.A.Complement

prob.A=1- prob.A.Complement

prob.A

**Property 3 P(**Φ) = 0

In the example let’s find the intersection between A and B with its probability.

**R Code**

S <- c(1,2,3,4,5,6,7,8,9,10)

length(S)

A<-c(1,2,3,4)

B<-c(5,6,7,8,9,10)

phi=intersect(A, B)

length(phi)

prob\_to\_find <- c(phi)

prob.phi<- sum(length(phi)) / length(S)

prob.phi

**Property 4 P(S) = 1**

**R Code**

S <- c(1,2,3,4,5,6,7,8,9,10)

S=prob\_to\_find <- c(1,2,3,4,5,6,7,8,9,10)

length(S)

length(S)

prob.S <- sum(length(S)) / length(S)

prob.S

**Rules or Laws of Probability**

An extension to the properties above are addition, multiplication and conditional laws of probability.

1. **Addition Law of Probability**

If A, B and C are subsets of S, then

* P(A U B) = P(A) + P(B) if A and B are independent events also P(A U C) = P(A) + P(C) if A and C are independent events.
* P(A U B) = P(A) + P(B) – P(A ∩ B) if A and B are dependent events.
* P(A U B U C) = P(A) + P(B) + P(C) - P(A∩B) – P(A∩C) – P(B∩C) + P(A∩B∩C) if A, B and C are dependent events.

**Illustrative Example**

Set theory will be used to explain the laws of probability laws.

Let’s consider a sample space S with only three sets A and B.

Let S = {21,43, 22,35,50,60,20,45, 48, 57,64,67,82,33,44,80,90}

A = {21,43,22,50,48,,57,80}

B = {22,35,60,20,45,64,67,82,33,44}

C ={22,43,50,45,82,33,82,60,90}

What is the probability of AUB?

P(AUB) = n(AUB)/n(S)

P(A) = n(A)/n(S)

P(B) = n(B)/n(S)

P(A ∩ B) = n(A ∩ B)/n(S)

**R Code**

S <- c(21,43, 22,35,50,60,20,45, 48, 57,64,67,82,33,44,80,90)

length(S)

A<-c(21,43,22,50,48,57,80)

length(A)

prob.A<- sum(length(A)) / length(S)

prob.A

B<-c(22,35,60,20,45,64,67,82,33,44)

length(B)

prob.B<- sum(length(B)) / length(S)

prob.B

AUB=union(A,B)

AUB

length(AUB)

prob.AUB<- sum(length(AUB)) / length(S)

prob.AUB

AB=intersect(A, B)

AB

length(AB)

prob\_to\_find <- c(AB)

prob.AB<- sum(length(AB)) / length(S)

prob.AB

# Lets cross check the result

prob.AUB = 0.4117647+ 0.5882353- 0.05882353

prob.AUB

**Venn Diagram**

**R Code**

library("VennDiagram")

grid.newpage()

draw.pairwise.venn(area1=7, area2=10,cross.area=1,

category=c("A","B"),fill=c("Green","Blue"))

* **Try P(AUBUC)**

**2. Multiplication Law of Probability**

If A and B are subsets of S, then

* P(A ∩ B) = P(A). P(B) if A and B are independent events
* P (A ∩ B) = P(A).P(B/A) if A and B are idependent events

**Illustrative Example**

Let’s consider in a football field there are 100 footballs of two colours of 30 ash and 70 blue.

Let A represents ash colour footballs

B represents blue colour footballs

S represents ash and blue colour footballs

If two footballs are selected for training, one after the other with replacement, what is the probability that the first is ash and the second blue?

P(ash colour football) = P(A) = n(A)/n(S)

P(blue colour football) = P(B) = n(B)/n(S)

P(first ash and second blue) = P(A and B) = P(AB)

= P(A ∩ B) = P(A). P(B) [since the process of selection is with replacement]

**R Code**

#Let S represent all the football

n.S=100

#Let A represent ash colour football

n.A=30

#Let B represent blue colour football

n.B=70

prob.A=n.A/n.S

prob.A

prob.B=n.B/n.S

prob.B

prb.AB=prob.A\*prob.B

prb.AB

What is the probability that the first is ash and the second blue if the process of selection is without replacement?

P(blue given that it was ash) = P(B/A) = n(B)/(n(S)-1)

P(first ash and second blue) = P(A and B)

= P(AB) = P(A ∩ B) = P(A). P(B/A) [since the process of selection is without replacement]

**R Code**

prob.A=n.A/n.S

prob.BgivenA=n.B/(n.S-1)

prob.BgivenA

prb.AB=prob.A\*prob.BgivenA

prb.AB

**3. Conditional Law of Probability**

This probability explains how to get the probability of having an event base on the fact that another event occurred.

Probability of event A given that B has occurred.



**Illustrative Example**

Let’s consider the illustration in addition law of probability using sets.

Let’s consider a sample space S with only three sets A and B.

Let S = {21,43, 22,35,50,60,20,45, 48, 57,64,67,82,33,44,80,90}

A = {21,43,22,50,48,,57,80}

B = {22,35,60,20,45,64,67,82,33,44}

C ={22,43,50,45,82,33,82,60,90}

What is the probability of P(A/B)?

**R Code**

S <- c(21,43, 22,35,50,60,20,45, 48, 57,64,67,82,33,44,80,90)

length(S)

A<-c(21,43,22,50,48,57,80)

length(A)

prob.A<- sum(length(A)) / length(S)

prob.A

B<-c(22,35,60,20,45,64,67,82,33,44)

length(B)

prob.B<- sum(length(B)) / length(S)

prob.B

AB=intersect(A,B)

AB

length(AB)

prob.AB<- sum(length(AB)) / length(S)

prob.AB

prob.AgivenB= prob.AB/prob.B

prob.AgivenB

We have come to the end of the training today. The Offa-R-Users-Group (ORUG) is a place to learn and share knowledge in the use of R. I wish to see you next time. If you are a guest, find time to register as a member to actualize your goal in using R.

The ORUG (<https://www.meetup.com/fedpofa-r-users-group/> ) is sponsored by R Consortium and AniKem\_Consult. For any Enquiry Contact the Organizer (WhatsApp: +2349030912602, email: anietieeu@yahoo.com)

THANK YOU FOR PARTICIPATING IN THIS EVENT