Course Title: Basic Statistics in Research

Course Code:

Credit Hours: 45

Credit Units: 3

Course Facilitator: Muwonge Emmanuel. MA (Econ), BA (Econ)

1.1 Course Description

This course deals with the study of the concepts and methods used to collect, present, analyze and interpret data concerning a given inquiry, and to draw conclusions in situations where uncertainties and variations occur. It introduces students to basic statistical concepts and methods used in social-economic analysis.

1.2 Course Objectives

This course is aimed at systematically equipping students with analytical skills in statistical methods and techniques in addressing various social-economic problems. The course will therefore be geared towards providing:

- A theoretical and applied statistical framework for the analysis of economic data using probabilistic methods.
- Competency in application of statistical theory to data collection, analysis and interpretation.
- Statistical tools essential for research process, objective decision-making and management through concepts and methods of inferential statistics.

1.3 Learning Outcomes

At the end of this course, students will be able to;

(i) Understand the various statistical studies performed in their field.

(ii) Design experiments, collect, organize, analyze and summarize data, and possibly make reliable predictions or forecasts for future use.

(iii) Demonstrate the ability to interpret and to employ descriptive statistics, indicators, questionnaires and other quantitative data in social-economic analysis.

1.4 Mode of Delivery

All in all, statistics is a science which involves the extraction of information from numerical data and it is used in making inferences (predictions and decisions) about the population from which the data is obtained(collected).

2.4 DESCRIPTIVE AND INFERENTIAL STATISTICS

Descriptive statistics

In this case, a statistician tries to describe a situation. Therefore descriptive statistics consists of data collection, organization, summation and presentation of data.

Inferential statistics

In this case, the statician tries to make inference from samples to populations. Inferential statistics uses probability i.e. the chance of an event occurring.

2.5 DATA COLLECTION AND SAMPLING TECHINQUES

Data is organized information. It can be numbers, words, measurements, observations or even just descriptions of things. Data can be collected in a variety of ways. The commonest method of data collection is through use of surveys. The surveys can be done by using a variety of methods namely: -

- Telephone surveys,
- Mailed questionnaire surveys,
- Personal interview surveys.

In order to obtain samples that are un biased i.e. that give each subject in the population a chance of being selected, staticians normally use the following four basic methods of sampling: -

(1) Simple random sampling.

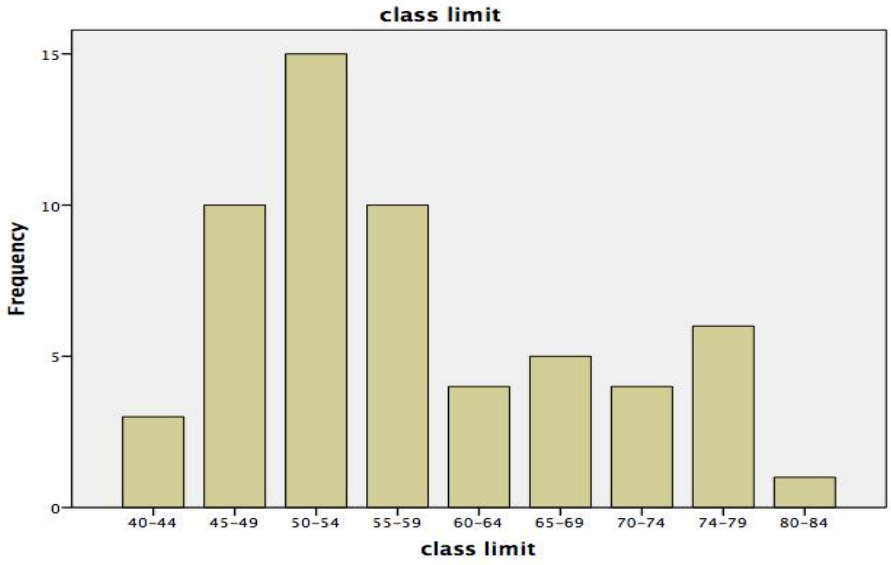
This method of sampling is also known as unrestricted random sampling or chance sampling. It is a method of sampling where every elementary unit for sampling has the same probability of being included or excluded in the sample as any other element of the target population.

(2) Stratified sampling method.

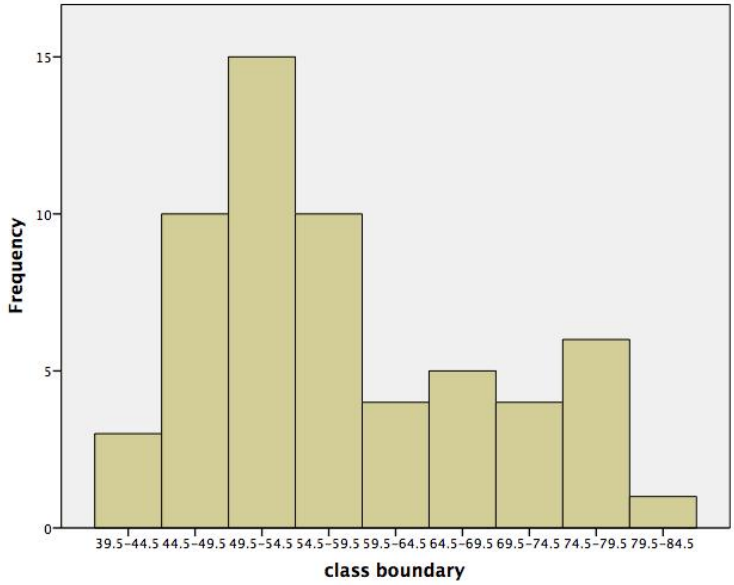
This is the method of sampling where the population is divided into non overlapping groups (or strata) according to different characteristics of the population. For example, suppose the headmaster of a given school wants to learn how students feel about their teachers, the headmaster would select the student from each group to use in the sample.

(3) Systematic sampling.

This is a sampling method that involves selection of sampling units at equal intervals after all these sampling units have been arranged in



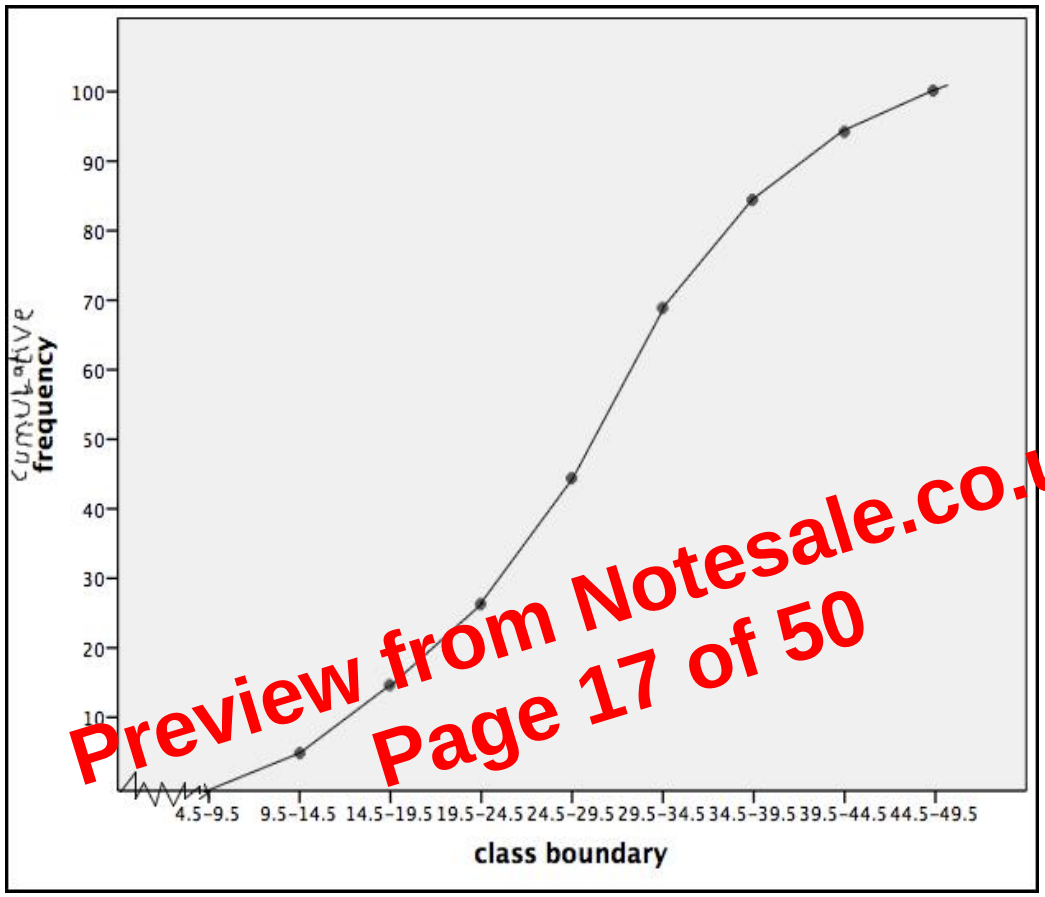
(ii) Histogram



(iii) Frequency Polygon

15-19	14.5-19.5	9	14
20-24	19.5-24.5	12	26
25-29	24.5-29.5	18	44
30-34	29.5-34.5	25	69
35-39	34.5-39.5	15	84
40-44	39.5-44.5	10	94
45-49	44.5-49.5	6	**100**
			Σf=100

O-GIVE



3.4 ADVANTAGES AND DISADVANTAGES OF HISTOGRAMS

Histograms are frequently used to display grouped frequency distributions graphically. They have the following advantages;-

1. They display clearly the comparative frequency of occurrence of data items within classes. This is seen by observing the relative heights of bars.
2. They indicate whether the range of values is wide (big) or narrow (small). The wider the histogram, the larger the range of values it's presenting and the narrower the histogram, the smaller is the range of the corresponding values.
3. They indicate the concentration of the values in the distribution (i.e. their skeweness) e.g.

46		3	13
47		3	16
48		0	16
49		6	22
50		4	26
51		4	30
		N=30	

3.5 GROUPED FREQUENCY DISTRIBUTION

In some cases, it is necessary to group the value of the data to summarize the data properly. If your data has more than 20 score values, you should create a grouped frequency distribution by grouping score values together into class intervals.

Example

Consider the following data of high temperatures for 50 days.

57	39	52	52	43
50	53	42	58	55
58	50	53	50	49
45	49	51	44	54
49	57	55	59	45
50	45	51	54	58
53	49	52	51	41
52	40	44	49	45
43	47	47	43	51
55	55	46	54	41

Required;

Create a grouped frequency distribution.

NB: In this case, the highest temperature is 59 and the lowest temperature is 39. Therefore, we would have 21 temperature values i.e. (59-39) +1=21.

This is greater than 20 values, so we should create a grouped frequency distribution.

STEPS

- Select an interval size so that you have7- 20 class intervals.
- Each interval must be the same size and they must not overlap.
- Create the various columns (i.e. tally column, mid point column and frequency column).

Question:

Compute the sample as well as the population covariance between the two variables and interpret the results.

Solution:

Number of cigarettes(X) per day	Saving($) per month (Y)	$x_i=X_i-\overline{X}$	$y_i=Y_i-\overline{Y}$	x_iy_i
0	45	-10	9	-90
5	42	-5	6	-30
10	33	0	-3	0
15	31	5	-5	-25
20	29	10	-7	-70
$\Sigma X_i=50$	**$\Sigma Y_i=180$**	**$\Sigma x_i=0$**	**$\Sigma y_i=0$**	**$\Sigma x_i,y_i=-215$**

From

$$\overline{Y} = \frac{\Sigma yi}{n}$$

$$= \frac{180}{5}$$

$$\overline{Y} = \mathbf{36}$$

$$\overline{X} = \frac{\Sigma xi}{n}$$

$$= \frac{50}{5}$$

$$X = \mathbf{10}$$

The population covariance

$$\Sigma xy = \frac{\Sigma xi\ yi}{N}$$

$$= \frac{-215}{5}$$

$$= \mathbf{-43}$$

Sample covariance

From; $$Sxy = \frac{\Sigma xi\ yi}{n-1}$$

$$= \frac{-215}{}$$

d. $A \cap A^1 = \{\}$
e. $A \cup A^1 = \{1,2,3,4,5,6\}$

Independent event

These are events where the occurrence or the non-occurrence of one of them does not affect the occurrence or the non-occurrence of the other. E.g. whether it rains in Luweero or not, it does not affect the flights at Entebbe airport

Conditional event

Events are said to be conditional if for one of them to occur, another one must have occurred already. E.g. for any student to sit the Ndejje university semester exams, he or she must have fully registered for that semester. Conditional events are written as E_1/E_2, which refers to the occurrence of event E_2 given that event E_1 has already occurred.

RULES/ LAWS OF PROBABILITY

(1) 0<_P<_1, i.e. If Ais an event in the sample space, the probability of A lies between zero and one.

(2) (a) P(S) =1. i.e. The probability of a sample space is one (1).
(b) P( ) =0. i.e. The probability of an empty set or null set = zero (0).

(3)) P (A) + P (A^1) =1 i.e. If A is an element in the sample space, then A and A^1 are complementary events. The sum of their probability is one (1).

NB;

The third law/rule holds, if A and A^1 are mutually exclusive

The table below shows the probabilities on the number of computer system failure in a week.

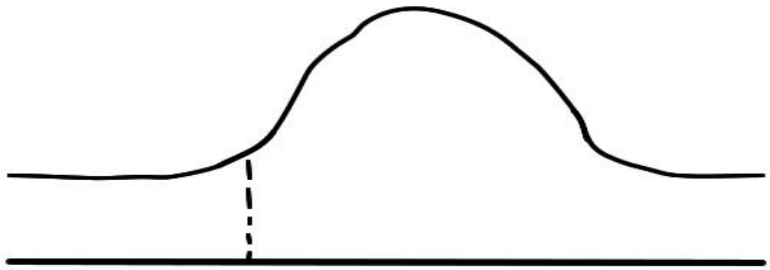
Number of failures	0	1	2	3	4
Probability	0.31	0.38	0.20	0.09	0.02

1) What is the probability that there will be less than two failures in a particular week?
2) What is the probability that there will be more than two failures is a particular week?
3) What is the probability that there will be atleast one failure in a particular week?

Steps involved in setting attest of hypothesis

1.Set the null hypothesis (H_o)
2. Set the alternative hypothesis (H_1)
3. State the level of significance ($\propto$)
4. State the critical region (Zc)

i.e



1. Perform the necessary computation and hence obtain the value of the statistic
2. Take the decisions, Accept the null test (H_o) if the value a of statistic is falling within the acceptance region and reject H_o if the value of a statistic is falling within the critical region (Zc).