

COURSE CODE: STS 201

COURSE TITLE: APPLIED STATISTICS FOR NON MAJOR

NUMBER OF UNIT: 3 UNITS

COURSE DURATION: THREE HOURS PER WEEK.

COURSE COORDINATOR: DR. AGWUEGBO S.O.N NAMDI AND DR. (MRS) DANSU (agwuegbo_son@yahoo.com, bolanlebm@yahoo.ca).

LECTURER OFFICE LOCATION: ROOM A308; COLNAS BUILDING, AMREC

COURSE CONTENT:

Nature of statistics, Data presentation, Diagrammatic representation of data, Summarization of data, probability, Sampling distribution, Estimation of parameters, Test of hypothesis, Correlation and Others.

COURSE REQUIREMENTS:

This is a compulsory course for all statistics students. Students are expected to have a minimum of 75% attendance to be able to write the final examination.

READING LIST:

- 1.) Probability and statistics for engineers & scientists by Walpole and Myers.
- 2.) Applied statistics by department of statistics.
- 3.) Introduction to basic statistics by Adamu.

LECTURE NOTES

NATURE OF STATISTICS

1. Explain the following:
 - a) Data
 - b) Observation
 - c) Information
 - d) Phenomenon
 - e) Statistics
 - f) Population
 - g) Variable
 - h) Sample
 - i) Attribute
 - j) Census
 - k) Descriptive statistics
 - l) Statistical method
 - m) Discrete variable
 - n) Continuous variable

2. Mention the two types of data and illustrate with examples.
- 3a) Discuss the sources of data and the various methods of data collection.
 - b) What are the advantages and disadvantages of these methods?
4. What are the characteristics of a good questionnaire?
- 5a) What is measurement scale?
 - b) Explain the four basic types of measurement scale.

DATA PRESENTATION

1. Explain the following:
 - a) Raw data
 - (b) Array
 - c) Distribution
 - (d) Frequency distribution
 - e) Cumulative frequency distribution
 - f) Relative frequency distribution
2. Differentiate between a histogram and frequency polygon and illustrate each with an example.
3. From the following distribution:

Classes	No. of Object
10 – 14	19
15 – 19	24
20 – 24	37
25 – 29	81
30 – 34	43
35 – 39	30
40 – 44	16
	250

- a) Find the class interval
 - b) Find the class boundaries
 - c) Find the class mark
 - d) Find the class width or size of class
 - e) Find the cumulative frequency of the distribution
 - f) Find the relative frequency of the distribution
4. Consider the distribution given below

Age (Year to the next birthday)	Frequency
15 < 20	37
20 < 25	81
25 < 30	43
30 < 35	24
35 < 40	9
40 < 45	6
Total	200

- a) Construct a cumulative frequency graph
- b) Construct a histogram
- c) Construct a frequency polygon
- d) Construct a bar chart

5. A company administers an aptitude test to 100 applicants for a job with, the company. The following are the times taken to complete a simple task for each applicant, measured to the nearest second.

44 92 72 45 85 61 66 46 59 57 52 40 93 54
52 64 65 44 51 66 92 58 74 42 43 56 46 52
45 56 68 40 48 76 71 99 51 72 52 56 69 58
40 76 70 42 52 46 73 59 41 55 74 66 64 47
58 46 52 54 63 89 87 41 57 68 59 81 82 60
67 68 97 57 47 53 61 52 49 47 86 55 54 48
85 45 84 53 49 47 70 78 58 96 54 62 60 57
58

- a) Construct a frequency table for the above data using classes of 40 – 49, 50 – 59, 60 – 69, etc.
- b) Construct a cumulative frequency distribution.
- c) Construct a relative frequency distribution.
- d) Draw the histogram.
- e) Draw the Ogive.

DIAGRAMMATIC REPRESENTATION OF DATA

1. Explain the following:

- a) Bar chart
- (b) Component bar chart
- c) Multiple bar chart
- (d) Pie charts

2. Differentiate between a bar diagram and a pie chart

3. Draw a pie chart using the following information

Marital status	No. of Women
Single	670
Married	480
Separated	120
Divorce	330
Widow	400

4. The following data gives the enrolment of junior students in a secondary school.

Session	No. of Male	No. of Female	No. of Students
90/91	500	1,000	1,500
91/92	750	1,000	1,750
92/93	840	960	1,800
93/94	1,050	950	2,000

Present the information in:

- a) Simple bar diagram.
- b) Component bar.
- c) Percentage component bar.
- d) Multiple bar diagram.

SUMMARISATION OF DATA

1. Explain the followings:
 - a) Measures of Central Tendency
 - b) Skewness
 - c) Measures of Variability
 - d) Coefficient of Variation
2. What are the properties of a typical value or our central tendency?
3. Consider the following distribution

Age in year to the next birthday	No. of observation
15 < 20	37
20 < 25	81
25 < 30	43
30 < 35	24
35 < 40	9
40 < 45	6
	200

Calculate:

- a) The mean
- (b) median
- (c) mode
- (d) variance
- (e) Standard deviation
- (f) Quartile deviation
- (g) Coefficient of Variation
- (h) Skewness

PROBABILITY

1. Define the following:

- | | |
|------------------------|-----------------------------|
| (a) Experiment | (b) Sample Space |
| (c) Sample Point | (d) Event |
| (e) Permutation | (f) Combination |
| (g) Mutually Exclusive | (h) Independent event |
| (i) Dependent event | (j) Conditional Probability |

2a) List two types of probability distribution.

b) List four types of discrete probability distribution.

c) What are the properties of binomial distribution?

d) What are the properties of Normal distribution?

3a) Show that the letters of the word ANTICIPATION can be arranged in three times as many ways as the letters of the word COMMENCEMENT.

b) In the random experiment of tossing 5 coins, list the event that

i.) at least 3 heads occur

ii.) exactly 2 heads

iii.) no heads at all.

4a) Simplify the following:

i.) ${}^{10}P_4$ (ii) ${}^{10}P_4$ (iii) 5C_2 (iv) 5P_2

b) If ${}^nP_5 : {}^nP_3 = 2:1$, what is the value of n?

c) If ${}^nP_3 / {}^nC_4 = 6$, find n.

5.) Using normal tables, find the values of the following probabilities:

a) $P(Z < 0.20)$

(b) $P(Z < -1.62)$

c) $P(0.57 < Z < 1.62)$

(d) $P(-1.50 < Z < 2.50)$

SAMPLING DISTRIBUTION

1. Explain the following:

a) Sample distribution

(b) Population distribution

c) Sampling distribution

(d) Statistic

e) Parameter

(f) Central Limit Theorem

g) Sampling with replacement

(h) Sampling without replacement

2. A finite population consists of the numbers 1, 2, 3, 4, 5.

a) Construct the sampling distribution of the mean (\bar{X}) when samples of size are drawn.

i.) with replacement (ii.) without replacement

b) Verify that $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}}^2 = \sigma^2/n$

3. The distribution of student height of 168cm and the standard deviation of 4cm. What is the probability that the mean of a random sample of 64 students is greater than 162cm?

4. A normal population with unknown variance is believed to have an average of 20, is unlikely to obtain random sample of size 9 from his population which has a mean $\mu = 24$ and a standard deviation $\sigma = 4/1$? If not, what conclusion would you draw?

5. a) Find $t_{0.025}$ when $V = 14$

b) Find $t_{0.1}$ when $V = 10$

(c) Find $t_{0.995}$ when $V = 7$

d) Find $P(-t_{0.005} < t < t_{0.005})$

(e) Find $P(t > t_{0.025})$

6. Given a random sample of size 24 from a normal distribution. Find K such that

a) $P(-2.069 < t < k) = 0.965$

b) $P(k < t < 2.807) = 0.095$

c) $P(-k < t < k) = 0.90$

7. The UNAAB feed mills claims that the average content of each bag of feed is 50kg. if a random of 10 bags selected at random have average contents as 52, 48, 50, 53, 51, 50, 49, 47, 50, 52 kilogrammes, would you agree with the manufacturer's claim?

ESTIMATION OF PARAMETERS

1. Explain the following:

a) Point estimate

(b) Interval estimate

c) Unbiased estimate

(d) Biased estimate

e) Estimator

(f) Estimate

g) Efficient estimator

(h) Confidence interval

2. A random sample of 25 UNAAB employees showed an average contribution of N12,500 to Ogun State Poverty Alleviation Programme, with a standard deviation of N225. Construct

- a) 90% Confidence interval
- b) 95% Confidence interval
- c) 99% Confidence interval

For the average contribution by all employees of UNAAB to Ogun State Poverty Alleviation Programme. Assume normally distributed.

3. A random sample of 400 smokers at Abeokuta is selected and 125 are found to have a preference for a brand called Benson and Hedges. Construct

- a) 90% confidence interval
- b) 95% confidence interval
- c) 99% confidence interval

for the proportion of the population of cigarettes smokers at Abeokuta who prefer Benson and Hedges.

4. A study is to be made by UNAAB students offering STS201. On Adolescent Sexuality and Fertility at Abeokuta. How large a sample is needed if they wish to be at least

- a) 90% confidence interval
- b) 95% confidence interval
- c) 99% confidence interval

the estimate differs from the true proportion by an amount not exceeding 0.01?

TEST OF HYPOTHESIS

1. Explain the following:

- a) Statistical hypothesis
- b) Null hypothesis
- c) Alternative hypothesis
- d) Type 1 error
- e) Type 2 error
- f) One tailed test
- g) Two tailed test
- h) Critical region
- i) Acceptance region
- j) Level of significance

2. State the null and alternate hypothesis to be used in the following claims and determine generally where the critical region is located:

- a) the mean rainfall at Abeokuta during the month of June is 68cm.
- b) On the average, students attend lectures within 2.5 kilometers of their homes at UNAAB, Abeokuta.

c) No more than 20% of the students of UNAAB contributed to Ogun State Poverty Alleviation Fund.

d) The proportion of voters favouring the new Student Union President is 0.63.

3. In Abeokuta, men have a mean height of 168cm, Standard deviation 8cm, and women have a mean height of 160cm, Standard deviation 5cm. In a random of 100 married couple, the average height difference between husband and wife was 5cm. Does this suggest that height of partner affects the decision to propose marriage?

4. Two drugs, A and B, were tested for a certain effect on UNAAB Laboratory mice. Two samples, each 50 mice were chosen randomly. One drug administered to each, and a measure, X of the effect obtained. The results were as follows:

Drug A: $\Sigma x = 1,750$

$\Sigma x^2 = 62,520$

Drug B: $\Sigma x = 1,980$

$\Sigma x^2 = 85,610$

Test at 1% and 5% levels of significance. The hypothesis that the two drugs have the same mean effect against the alternative hypothesis that drug B has a higher mean effect. You may assume that the effect on mice, of each drug is normally distributed.

CORRELATION AND REGRESSION

1. Explain the following:

- (a) Regression
- (b) Correlation
- (c) Scatter diagram
- (d) Regression coefficient
- (e) Correlation coefficient
- (f) Positive correlation
- (g) Negative correlation
- (h) Regression line
- (i) Coefficient of determination
- (j) Rank correlation

2. The table below gives the index figures for production and the price of an article over ten consecutive years

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Production	92	96	103	108	109	108	96	103	109	103
Price	109	111	94	93	89	84	100	106	87	97

- a) find an equation that expresses the index figure for production in terms of the price of an article.
- b) Calculate the Coefficient of Correlation (r).
- c) Use Spearman's Rank Correlation to calculate the correlation coefficient r_s
- d) Determine the percentage of observation that is explained by the regression equation.

3. A study was made by UNAAB Mills Ltd to determine the relationship between weekly advertising expenditures and sales. Assuming the coefficient of determination is $r^2 = 0.81$ and the prediction equation is $y = 0.957 - 0.032x$

- a) What is the relationship between advertising and sales.
- b) find the correlation coefficient r.

4. The following are estimated average of employment percentages, 1990 – 2000, for 10 states in Nigeria.

State	Male (x)	Female (y)
Lagos	4.12	3.29
Abeokuta	2.60	1.61
Anambra	3.24	3.58
Imo	4.32	4.99
Benue	2.88	3.02
Plateau	3.26	3.19
Kano	2.97	3.19
Kaduna	3.21	2.56
Rivers	1.12	0.19
Delta	1.04	0.99

Fit regression lines: (i) for y on x

(ii) for x on y and find the coefficient of correlation between x and y.

OTHERS STATISTICAL METHODS

1a) What is ANOVA?

b) Write an assumption for analysis of variance.

The data in the table below represent 5 random samples each of 5 from independent normal distribution with $\mu_1, \mu_2, \dots, \mu_5$ and common variance σ^2 . Test the hypothesis at $\alpha = 5\%$ that $\mu_1 = \mu_2 = \dots = \mu_5$

	A	B	C	D	E
	5	9	3	2	7
	4	7	5	3	6
	8	8	2	4	9
	6	6	3	1	4
	3	9	7	4	7
TOTAL	26	36	20	14	33

2. Four different test were used in the treatment of a course and the final grades of the students were recorded below

1	2	3	4
60	80	97	67
80	81	84	84
69	73	93	90
65	69	79	78
75	92	61	72

Test the hypothesis at $\alpha = 5\%$ that there is no differences in the final grades from the four different tests.

3a) What is a run?

b) From the following arrangement of M and F, is their evidence randomness = 5%

MM F MMMM FF MMMM FM FF MMMM F
M FF MMMMM F MMM F M F MMM

4) Consider the following measurements, which are weights of some people in kilograms.

163, 165, 160, 189, 161, 171, 158, 151, 169, 162, 151, 169,
162, 163, 139, 172, 165, 148, 166, 172, 163, 187, 173

Test the null hypothesis $\mu = 163$ against the alternative $\mu > 163$ at $\alpha = 5$.

5) Consider the following three samples.

Sample I	Sample II	Sample III
29	36	24
36	17	18
37	19	20
36	21	24
36	26	25
35	29	28
39	27	31
38	21	34
40	32	30
23	33	22
	27	21
		16

Use Kruskal – Wallis test at 0.05 to test

$H_0: \mu_1 = \mu_2 = \mu_3$

6) The following are the weights in kilograms before and after of 16 persons who stayed on a certain reducing diet for four weeks.

Before	After
147.0	137.9
183.5	176.2
232.1	219.0
161.6	163.8
197.5	193.5
206.3	201.4
177.0	180.6
215.4	203.2
147.7	149.0
208.1	195.4
166.8	158.5
131.9	134.4
150.3	149.3
297.2	189.1
159.8	159.1
171.7	173.2

Use Wilcoxon Signed – Rank test to test at $\alpha = 0.05$ whether the weight reducing diet is effective.

7) Define:

- a) Statistical Quality Control (SQC)
- b) Process Control
- c) What is Sampling Inspection Plan?
- d) What are the causes of variation?
- e) What is Control Chart?
- f) List the types of control charts and their examples.

8) The table below shows the thickness of some materials produced by a machine.

Draw control charts for mean and range and comment on the state of control.

Sample	1	2	3	4	5	6	7	8	9	10
	13	14	11	14	12	10	10	8	13	5
	8	12	10	10	10	12	16	12	8	8
	15	19	16	9	13	10	8	8	14	5
	8	10	8	13	7	8	10	10	7	10

9a) Define Time Series.

b) List the Components of a Time Series.

c) The following gives the volume of passengers ('000) of an Airline over a period of time.

Year	Jan – Mar	Apr – Jun	Jul – Sept	Oct – Dec
1991	24	35	55	30
1994	20	42	70	26
1995	26	37	82	38
1996	25	38	90	40

Estimate the trend

ii) using the least square method

iii) using the method of moving averages smoothing out the trend.

10a) What is an Index Number?

b) What are the uses of index numbers?

c) What are the limitations of index numbers?

Given the following table

Community	1970		1980	
	Price	Quantity	Price	Quantity
A	4	10	8	16
B	5	12	6	12
C	2	5	4	10
D	3	4	5	8
E	4	12	7	10

Find Laspeyre and Paasche quantity Indices.

Find Fishers' ideal price index