COVENANT UNIVERSITY
NIGERIA

TUTORIAL KIT
OMEGA SEMESTER

PROGRAMME:
DEMOGRAPHY AND SOCIAL STATISTICS

COURSE: DSS 423

DISCLAIMER

The contents of this document are intended for practice and leaning purposes at the undergraduate level. The materials are from different sources including the internet and the contributors do not in any way claim authorship or ownership of them. The materials are also not to be used for any commercial purpose.

DSS 423
ADVANCED SOCIAL STATISTICS II
BY
FASINA, F.F

Tutorial Questions and Answers

1. Given a set of censuses data from 1990 to 2010 with over 40 thousands variables, how would derived a single data file to analyze your variables assuming you are to explore only the socio-demographic characteristics, access to media facilities and household living conditions.

2. A national survey is being planned for your country by an international NGO on Adolescent Reproductive Health Needs in Nigeria. As a Research Consultant, what methodological approach (es) would you recommend taking into cognizance that the results will be used for policy formulation? Note that the sample must be representative of all the states in the country.

3. The data in table 1 below represents the weight of newborn babies obtained from two different hospitals during the year 2011 in AMCO LGA. You are required to use Mann-Whitney U Test to confirm whether the average weight of the babies are the same, assuming your $\alpha = 0.01$.

Table 1: Weight of newborn babies

Observation	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Score A	0.3	0.43	1.34	0.18	0.41	0.47	0.15	0.38	0.82	0.49	0.45	0.1	0.16	0.33
Score B	1.09	0.87	0.99	0.9	0.37	0.94	1.11	0.56	1.05	0.46	0.51	0.86	0.24	0.52

4. An experiment was conducted to determine how a child's reading rate varies with his ability to write. 10 children were randomly selected from a class in one of the standard schools in the town and the result recorded in table 2. X are the scores on the ability to write and Y is a measure of reading rate. Assuming the First Order Regression model is appropriate, (i) Obtain the estimated regression function. (ii) Plot the estimated regression function and the data. (iii) Does your plot support the anticipation that writing skills increases with ability to read.

Table 2: Reading rates and ability to write

X	40	42	15	42	30	18	20	20	15	38
Y	85	80	120	70	105	116	94	90	120	90

5. The number of defective parts turned out by a machine is believed to be related to the speed at which the machine is operated. These data represent a random sample from the output of the machine.

Speed (X)	10	12	15	13	14	17	13	18
Defective (Y)	2	4	8	5	6	7	6	9

(i) Fit a least squares linear regression line to the data.
(ii) Test the significance of b at the 0.05 level of significance
(iii) Calculate the standard deviation of regression.

(iv) Estimate P when X = 15

6. The owner of a retailing organization is interested in the relationship between price at which a commodity is offered for sale and the quantity sold. The following sample data have been collected

Price (X)	25	45	30	50	35	40	65	75	70	60
Quantity Sold (Y)	118	105	112	100	111	108	95	88	91	96

(v) Fit a least squares linear regression line to the data
(vi) Test the significance of b at the 0.05 level of significance
(vii) Calculate the standard deviation of regression.
(viii) Estimate P when X -= 15

Q1. The data below shows the patterns of fertility in Gesco country before, during and after the implementation of National Health Insurance Scheme (NHIS) in 12 States in that country. Three patterns were identified as in table 1 below:

Table 1: Patterns of Fertility in Gesco

Before NHIS		**During NHIS**		**After NHIS**	
X	**Y**	**X**	**Y**	**X**	**Y**
190	177	252	226	206	226
261	222	228	196	239	229
194	167	240	198	217	215
217	176	246	206	177	189

Perform an analysis of covariance test to determine that the NHIS has any impact on fertility level in Gesco Country. Use $\alpha = 0.05$ level of significance. Derive and show your normal equations using Deviation approach. **(30 marks)**

Q2. Arising from the 30th Annual General Meeting of AMCO Technology is the consensus that splitting the share would likely increase the stock holding of individual shareholders in the company. The following observations were therefore made for over 14 months. The table 2 below shows that recent stock splits increase in value and at the same time decrease in value in some cases. As a consultant, you are required to employ Wilcoxon Signed-ranked Test to determine whether stock splits will continue to be beneficial for holders of AMCO stocks or not.

Table 2: Stock splits in AMCO Technology

Observation	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Score 1	20	18	24	14	5	26	15	29	15	9	25	31	35	12
Score 2	40	25	38	27	31	21	32	38	25	18	32	28	33	29

In addition the he management will like to have the following details

1. What are the null and alternative hypotheses? **(5 marks)**
2. With alpha = .05, what is the rejection rule and, **(5 marks)**

3. Your final advice (i.e. conclusion) **(10 marks)**

Q3. In a study of the relationship between incidence of adolescent fathers (Y) in various developing nations and adolescent living status (X_1) and delinquency (X_2), the following results were obtained (table 3):

Table 3: Adolescent fathers, living status and delinquency

Location	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y	64	73	61	76	72	80	71	83	83	89	86	93	88	95	94	100
X_1	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10
X_2	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4

It is assumed that the first order linear regression model with independent normal error terms is appropriate (i.e. the model with two independent X's and no interaction).

a) Using a matrix approach, find the estimated regression coefficients. State the estimated regression model and, how would you interpret b_1 and b_2? **(10 marks)**
b) You are to state your hypothesis, your decision rule and indicate how useful is the regression equation in explaining the relationship **(10 marks)**

Q4. A health officer is concerned with the cause of minor accidents in the soap factory of a large manufacturing plant. He would like to know whether or not there is a relationship between the number of accidents (Y_1), the number of workers in the factory (X_2) and the number of hours of overtime worked in a week (X_3). A small random sample of records at the factory yields this information in Table 4

Table 4: Workers numbers, hours worked and frequency of accident in the factory

No of Accidents (Y_1)	6	2	3	5	7	4	9	5	6	8
No of workers (X_1)	50	40	45	48	51	44	55	46	49	52
No of hours worked/week (X_2)	250	205	225	240	260	220	300	230	245	250

a. Determine the least square multiple regression equation using the 1st order regression equation and interpret your result **(10 marks)**
b. Compute the coefficient of determination and interpret your findings **(5 marks)**
c. Estimate the number of accidents when there are 50 workers and a total of 240 hours of overtime and comment on your observations **(5 marks)**

Q5. The data in table 5 below represents the weight of newborn babies obtained from two different hospitals during the year 2010 in AMCO LGA. You are required to use Mann-Whitney U Test to confirm whether the average weight of the babies are the same, assuming your $\alpha = 0.01$.

Table 5: Weight of newborn babies

Observati	1	2	3	4	5	6	7	8	9	10	11	12	13	14

on														
Score A	0.3	0.43	1.34	0.18	0.41	0.47	0.15	0.38	0.82	0.49	0.45	0.1	0.16	0.33
Score B	1.09	0.87	0.99	0.9	0.37	0.94	1.11	0.56	1.05	0.46	0.51	0.86	0.24	0.52