



KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2011/2012

INSTITUTIONAL BASED PROGRAMME (IBP) – AUGUST SESSION

EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE

SMA/SS/571: APPLIED CONTINGENCY TABLES ANALYSIS

DATE: Friday 30th December 2011

TIME: 2.00p.m – 5.00p.m

INSTRUCTIONS:

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

Question 1

- a) Define the meaning of “Contingency Table”?

Explain what is contained in this (i, i) -th cell of a contingency table.

[2marks]

- b) i) Consider a 2 x 2 contingency table with the entries as shown

a_{11}	a_{12}	a_{1+}
a_{21}	a_{22}	a_{2+}
a_{+1}	a_{+2}	n

For a null hypothesis of independence show that the Pearsonian chi-square test statistic is given

$$\chi^2 = \frac{n (a_{11}a_{22} - a_{12}a_{21})^2}{a_{1+}a_{2+}a_{+1}a_{+2}}$$

[12marks]

- i) As part of a study designed to identify possible subcultural, ethnic, and family influence of mobility after teacher training the following data was obtained from graduates of a Teacher Training College

<u>Labor force mobility</u>		
<i>Opportunity</i>	<i>Low</i>	<i>High</i>
<i>Low</i>	46	19
<i>High</i>	6	43

Test at $\alpha = 0.05$ level of significance, the null hypothesis that opportunity level is independent of labor force mobility. What is the *P-value*? [6marks]

- c) A hypothetical community has two ethnic groups classified according to whether one is a male (M) or female (F). The ratio of male to female in the two groups are 3:1 for group A and 2:1 for group B. The following data represent 120 people classified according to sex in the community

		Group A	
		M	F
Group B	M	60	15
	F	35	10

Test at $\alpha = 0.05$ whether the data is in agreement with the proposed model [carefully state the hypotheses under study]. What is the *P-value*? [10marks]

Question 2

- a) For testing independence of characteristics in a $r \times c$ contingency table we can use Akaike Information Criterion (AIC) to choose a model consistent for the null hypothesis (H_0) and alternative hypothesis (H_1). The following expression for AIC are available.

$$AIC(H_0) = 2(r + c - 2) - 2 \sum_i \sum_j a_{ij} \log \left(\frac{a_{i+} a_{+j}}{n^2} \right)$$

$$AIC(H_1) = 2(rc - 1) - 2 \sum_i \sum_j a_{ij} \log \left(\frac{a_{ij}}{n} \right)$$

Show that $AIC(H_1) - AIC(H_0)$

$$\rightarrow -\chi^2 + 2(r-1)(c-1) \text{ as } n \rightarrow \infty$$

What is the implication of this result with respect to the comparison of the models $AIC(H_1)$ and $AIC(H_0)$? [6marks]

- b) Consider the following contingency table on classification of therapy and diagnostic group for some 170 psychiatric patients.

<u>Diagnostic group</u>	<u>Type of therapy</u>	
	Psychotherapy	Drug therapy
Alcoholic psycho	48	23
Organic psycho	19	80

Use AIC to test for independence of diagnostic group and type of therapy. [14marks]

Question 3

The number of houses with a VCD player in 10 rows of “row houses” in a middle class city estate are:

Row	:	1	2	3	4	5	6	7	8	9	10
No. of VCD	:	4	8	16	3	15	10	24	20	18	17
No. of houses	:	21	43	39	40	38	40	43	40	35	37

- State the null hypothesis of homogeneity of proportions appropriate for this situation and the alternative hypothesis [3marks]
- Using likelihood ratio statistic test $\alpha = 0.05$ level of significance the hypothesis in part (i). Indicate the magnitude of the *p-value* in this case. [15marks]
- What does the result of part (ii) imply in terms of residents of this estate? [2marks]

Question 4

- In a 2-dimensional contingency table with r rows and c columns:
 - Write down an expression for the natural logarithm of estimated frequency of cell (i,j) under the null hypothesis of independence. [3marks]
 - Using the expression in part (i) obtain the log linear model, explaining clearly what each term in the model measures. [6marks]
- Using results of part (a) analogously, write down an expression for a

log linear model of a 3-dimensional contingency table with r rows, c columns and k layers under the null hypothesis of mutual independence.

[3marks]

- c) Given below is log linear model fitted to data on lizard species (Sagrei and angusticeps) and their habitat, the G^2 goodness-of-fit statistic and the degree of freedom (df)

1 = Perch height; 2 = perch diameter; 3 = species (sagrei or angusticeps).

Model	G^2	df
[1] [2] [3]	25.04	4
[1] [2 3]	12.43	3
[1 3] [2 3]	2.03	2
[1 2] [1 3] [2 3]	0.15	1

Using backward elimination procedure and the partitioning of G^2 , formulate the most parsimonious model and give interpretation with reference to the variables involved in this problem.

[9marks]