MATH 955

CHUKA   UNIVERSITY

UNIVERSITY EXAMINATIONS

FIRST YEAR EXAMINATION FOR THE DEGREE OF DOCTORATE OF PHILOSOPHY IN APPLIED STATISTICS

MATH 955: STATISTICAL ANALYSIS OF GENETIC DATA

STREAMS: PhD (STATS)   **TIME: 3 HOURS**

DAY/DATE: TUESDAY 06/04/2021   **8.30 A.M. – 11.30 A.M**

INSTRUCTIONS:
- Answer any **THREE** questions.
- Use of calculators and statistical tables is allowed.
- Do not write anything on the question paper.

QUESTION ONE (20 MARKS)

Simple Sequence Polymorphisms (SNPs) are usually considered binary random variable in genome-wide association studies (GWAs). Suppose SNP (X) is a binary random variable that takes value 0 with probability p and value 1 with probability 1-p. let $X_1, \ldots, X_n$ be independent and identically distributed (iid) samples of X.

(i) Compute a maximum likelihood estimation (MLE) estimate of p.   (5 marks)

(ii) Is $\hat{p}$ an unbiased estimate of p? Prove the answer.   (5 marks)

(iii) Compute the expected square error of $\hat{p}$ in terms of p.   (5 marks)

(iv) Prove that if you know that p lies in the interval $\left[\frac{1}{4}; \frac{3}{4}\right]$ and you are given only $n = 3$ samples of X, then $\hat{p}$ is an inadmissible estimator of p when minimising the expected square error of estimation.   (5 marks)

QUESTION TWO (20 MARKS)

(a) Using a hypothetical locus, derive the equations for estimating additive and dominance genetic variance.   (10 marks)

(b) The following data (Table 1) was obtained from a cross of two cultivars of rice.

Table 1: Mean and variance in number of spikelets in a cross of two cultivars rice

Genotype	Number of individuals analysed	Mean	Variance (σ^2)
P_1	20	24.5	79
P_2	20	25.9	62
F_1	20	24.3	67
F2	650	25.3	143
BC_1 (F_1 x P_1)	20	24.4	133
BC_2 (F_1 x P_2)	20	24.8	81

(i)   Using the equations derived in part (a) above compute the genetic variance and degree of dominance for number of spikelets in rice (Table 1).   (6 marks)

(ii)   Calculate heritability for the number of spikelets in rice (Table 1).   (4 marks)

QUESTION THREE (20 MARKS)

(a) Using appropriate statistical model(s), explain composite interval mapping as used in the QTL mapping.   (14 marks)

(b) Design a training populations for selective phenotyping in genomic prediction.   (6 marks)

QUESTION FOUR (20 MARKS)

(a) Given the following pedigree (Figure 1), construct a table showing the coefficients of relationship between each individual.   (10 marks)
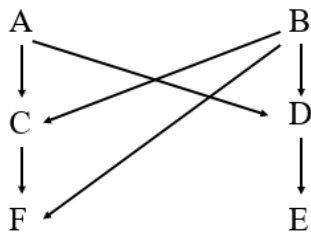

Figure 1: Relationship between different individuals

(c) With an aid of diagram, outline the steps in association mapping.   (10 marks)