MASENO UNIVERSITY

UNIVERSITY EXAMINATIONS 2013/2014

SECOND YEAR FIRST SEMESTER EXAMINATIONS FOR THE DEGREE OF BACHELOR OF SCIENCE (MATHEMATICS), BACHELOR OF SCIENCE IN ACTURIAL SCIENCE AND BACHELOR OF SCIENCE IN APPLIED STATISTICS WITH INFORMATION TECHNOLOGY

(MAIN CAMPUS)

MIT 201: STATISTICAL COMPUTING

Date: 25th November, 2013 *Time: 11.00 - 1.00 p.m.*

INSTRUCTIONS:

- **Attempt QUESTION ONE and ALL QUESTIONS in Section B.**

MIT: 201 – Statistical computing

SECTION A: Attempt all questions in this section (10 marks)

1. State whether **true** or **false**
   a. Inferential statistical methods deal with statements about unknown population parameters whereas descriptive statistical methods deal with summarizing and exploring data in the sample. (1 mark)
   b. Nominal categorical variables have natural ordering while ordinal categorical variables do not have natural ordering. (1 mark)
   c. A variable has a population and a sample distribution. (1 mark)
   d. The $median = 0.5X(n+1)$ observation. (1 mark)
   e. In measuring standard deviation of a sample, reason of dividing by n-1 instead of n is so that s^2 is on average equal to σ^2 which is biased when the sample is repeated many times. (1 mark)
   f. To compute the sample i^{th} percentile: order all observations in increasing order, then take the i*(n+1)100th observation. (1 mark)
   g. Probability theory is the basis of descriptive statistics. (1 mark)
   h. Sample mean and prevalence are examples of random variables. (1 mark)
   i. Expected value of the sum of independent random variables is the sum of their expected values (1 mark)
   j. For two random variables it holds that $Var(X \pm Y) = var(X) + var(Y) \mp 2covar(X, Y)$ (1 mark)

SECTION B: Attempt all questions in this section (10 marks)

2. Attempt question (a) to (f) in this section
   a. List 2 measures of location. (1 mark)
   b. Differentiate between measures of dispersion and measures of precision. (1 mark)
   c. In a family of 7 children, what is the probability of selecting 3 boys? (1 mark)
      i. What is the expected value of the number of boys in this family? (1 mark)
   d. In a sample of size n from population with prevalence π of disease D

      $$X_i = 1 \; if \; i^{\text{th}} \; individual \; in \; sample \; has \; D \; and \; X_i = 0 \; if \; not$$

      $X = X_1 + X_2 + \cdots . + X_n$ is number of individual with D in sample

      i. What is the expected value of number of individuals with the disease? (1 mark)
      ii. What is the estimated sample prevalence of the disease? (1 mark)
      iii. What is the standard deviation of the estimated sample prevalence of the disease? (1 mark)
   e. Suppose that age in a population being considered has a normal distribution with a mean of 60 years and a standard deviation of 10. What proportion of the sample has age greater than 70 years? (1 mark)

f. A sample of 18 people were drawn from a population of persons with Angina pectoris with a mean total cholesterol of 5.81mmol/l and standard deviation of 1.2, Calculate the standard error of the mean and approximate 95% confidence interval of the mean. (2 marks)

SECTION C: Attempt any 2 questions out of the 4 questions (20 marks each)

3. Dichotomous outcome: Sixty-five pregnant women at high risk of pregnancy-induced hypertension participated in a randomized controlled clinical trial comparing 100mg of aspirin daily and a matching placebo during the 3rd trimester of pregnancy. The observed numbers with hypertension are shown in the following table.

	Hypertension		
	yes	no	Total
Aspirin	5	29	34
placebo	10	21	31
Group Total	15	50	65

i. Is the risk of hypertension in aspirin treated women significantly lower than in placebo treated women? (1 marks)
   a. Which test do you use? (1 marks)
   b. What is the P-value? (3 marks)

ii. Give the **estimate** and **approximate 95% confidence interval** for the following of hypertension between aspirin and placebo treated women
   a. Difference in risk (3 marks)
   b. Risk ratio (3 marks)
   c. Odds ratio (3 marks)

iii. Suppose a new study is planned. What sample size is approximately needed in order to have a power of 90% if the risk of hypertension is 0.1 lower in aspirin treated women ($\alpha = 0.5$)? (6 marks)

4. A two-period two-treatment cross-over trial was carried out in 67 patients with cerebrovascular deficiency. Each patient was treated with either active treatment (A) in the first and with placebo (B) in the second period, or with placebo in the first and active treatment in the second period. The order in which the treatment were applied was randomized. The outcome variable was the assessment of ECG, scored as normal or abnormal. The table below gives the data. The variables are defined as: PATNO=patient number; ORDER= order in which the treatment were applied (AB or BA); ECG_A=assessment of ECG under treatment A (0=abnormal, 1=normal); ECG_B= assessment of ECG under treatment B (0=abnormal, 1=normal); DIF=ECG_A-ECG_B.

Quick summary of the data

	0	1		B	0	1	
A B	0	0	1	A 0	1	0	1
Total	6	6	22	9	2	4	18

Some descriptive statistics for the whole group are given in the table below.

	N	Std. Deviation		
	Statistic	Statistic	Std. Error	Statistic
ecg_A	67	.75	.054	.438
ecg_B	67	.63	.060	.487
Diff	67	.12	.050	.409
Valid N (listwise)	67			

The data are analyzed under the assumption that the treatment effect does not depend on the order in which the placebo and active treatment were given (i.e. no "carry-over" effect). Furthermore it is assumed that other is no difference between the two periods with respect to the chance of response (no "period" effect).

i. Give the cross table that adequately describe the data. (3 marks)
   a. Give the percentage response under the placebo and active treatment. (2 marks)

ii. Give an estimate of treatment effect (i. e. percentage response under the active minus placebo treatment). (2 marks)
   a. What is the corresponding standard error? (1 marks)
   b. Compute also an approximate 95% confidence interval (2 marks)

iii. Test with an approximate test the hypothesis that there is no treatment effect (provide the value of the test statistic and the corresponding p-value) (3 marks)
   a. Use also an exact test for this hypothesis (provide the bounds for the p-value) (3 marks)

Some statistics per treatment order group are given below.

AB

	N	Mean		
	Statistic	Statistic	Std. Error	Std
ecg_A	34	.82	.066	.387
ecg_B	34	.65	.083	.485
Diff	34	.18	.066	.387
Valid N (listwise)	34			

BA

	N	Mean		
	Statistic	Statistic	Std. Error	Std
ecg_A	33	.67	.083	.479
ecg_B	33	.61	.086	.496
diff	33	.06	.075	.429
Valid N (listwise)	33			

One assumption underlying the above analysis was that the treatment effect did not depend on the order in which placebo and active treatment were given.

i. Describe a simple approximate way to test this hypothesis. (1 marks)
ii. Carry out this test, give the p-value and state your conclusion. (3 marks)