UTHM
Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
(ONLINE)
SEMESTER I
SESSION 2020/2021

COURSE NAME	:	STATISTICAL INFERENCE
COURSE CODE	:	BWB 20503
PROGRAMME CODE	:	BWQ
EXAMINATION DATE	:	JANUARY / FEBRUARY 2021
DURATION	:	3 HOURS 30 MINUTES
INSTRUCTION	:	ANSWER **ALL** QUESTIONS **OPEN BOOK EXAMINATION**

TERBUKA

THIS QUESTION PAPER CONSISTS OF **FIVE (5)** PAGES

Q1 (a) Given n independent binary variables $X_1, X_2, \ldots, X_n$, that can take the values of 0 or 1, are identically distributed as $X_i \sim Ber(\theta)$ where $i = 1,2,\ldots,n$. Consider the estimator $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(i) Describe the distribution of the random variable $S = n\hat{\theta}$.

(1 mark)

(ii) Identify the mean and variance of this distribution.

(2 marks)

(iii) By writing $\hat{\theta}$ in terms of S, compute the expectation $E(\hat{\theta})$ and the bias of this estimator. State whether this estimator unbiased.

(4 marks)

(iv) Based on distribution of $n\hat{\theta}$, examine $\text{var}(\hat{\theta})$ and the standard error of this estimator.

(4 marks)

(b) Assume that n independent count variables $\{X_1, X_2, \ldots, X_n\}$ are identically distributed as $X_i \sim Poi(\theta)$ where $i = 1,2,\ldots,n$ and to estimate $E(X_i) = \theta$, consider the sample mean estimator $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(i) Describe the distribution of the random variable $S = n\hat{\theta}$.

(1 mark)

(ii) Use the expectation of the distribution of $S = n\hat{\theta}$ to compute the expectation $E(\hat{\theta})$ and the bias of this estimator. State whether this estimator unbiased.

(4 marks)

(iii) Use the variance of the distribution of $S = n\hat{\theta}$ to examine $\text{var}(\hat{\theta})$ and the standard error of this estimator.

(4 marks)

(iv) Use the variance and the bias of $\hat{\theta}$ to compute the Mean Squared Error (MSE) of this estimator. Comment on the MSE behaviour in the asymptotic limit as $n \to \infty$. State whether this estimator is consistent or not.

(5 marks)

TERBUKA

Q2 (a) A study to test the null hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ by using a critical region $C = \{x : T(x) \leq c\}$, where c is the critical value. Here, $F_\theta(z) = \Pr(T \leq z; \theta)$ denotes the sampling distribution of $T(X)$ and $t = T(x)$ denotes the reported sample statistic evaluated for the data.

(i) Define hypothesis testing, type I error and type II error respectively.

(3 marks)

(ii) Demonstrate the size, $\alpha = F_{\theta_0}(c)$, power $1-\beta = F_{\theta_1}(c)$ and p value, $p = F_{\theta_0}(t)$.

(6 marks)

(iii) Discover, if any of these three probabilities in **Q2(a)(ii)** depend solely on the null distribution and the critical value that the size of the test equal the p-value.

(3 marks)

(iv) By rearranging the inequality $t < c_\alpha$ where $c_\alpha = F^{-1}{}_{\theta_0}(\alpha)$, prove that t being in the critical region is equivalent to the p value being less than or equal to the size of the test.

(3 marks)

(b) A company claims to be selling pints of milk that contain 4.7% plain water where the experiment group suspects that there is more plain water than this in the milk. A random sample of 15 pints was found to have a sample mean of 4.92% and a sample standard deviation of 0.53%. The percentage X of plain water in a pint can be considered to be independent and normally distributed, $X \sim N(\theta, \sigma^2)$.

(i) Outline the hypothesis for θ that represent the claims of the company and the experiment group.

(2 marks)

(ii) Using appropriate statistical tables such as percentage points of the t distribution, compute the critical value required for a test at the 0.1% level of significance.

(3 marks)

(iii) Calculate the value of the test statistic for the sample of data. Interpret the claims at the 0.1% level of significance.

(5 marks)

TERBUKA

Q3 (a) A simple experiment with a random sample of 10 mathematics students were found to have the following heights (in *cm*).

155, 185, 175, 175, 173, 170, 169, 164, 161, 185

Assume the heights of mathematics students are independently and normally distributed $X \sim N(\theta, \sigma^2)$ with a known variance of $\sigma^2 = 99.12\ cm^2$.

(i) Calculate a point estimate of the mean height θ and a probability that this estimate is exactly equals to the true value of θ.

(5 marks)

(ii) Compute the standard error, $\sigma/\sqrt{n}$ of the sample mean by using given variance.

(2 marks)

(iii) Estimate an equal-tailed 90% confidence interval for θ by using the sample mean and its standard error.

(3 marks)

(iv) Calculate an equal-tailed 50% confidence interval for θ.

(7 marks)

(b) Every month, ICM for the Guardian asks a sample of around 1000 people which political party they would vote for if there was an election. Result shows that $x = 425$ out of $m = 1000$ people said they would vote conservative. Assume that the count variable X can be modelled as $Bin(m, \theta)$ and that for large m this is well approximated by $N(m\theta, m\theta(1-\theta))$.

(i) Indicate an appropriate test statistic and critical region for testing hypothesis $H_0 : \theta = 0.5$ and $H_1 : \theta < 0.5$

(3 marks)

(ii) Calculate the test statistic value for the data, then use statistical tables to compute the *p*-value. Conclude whether the null hypothesis could be rejected at the 5% level of significance.

(5 marks)

TERBUKA

Q4 (a) Let $X_1,\ldots,X_n$ be independent random variables, each with the following probability density function

$$f(x) = \theta x^{\theta-1} \quad \text{for} \quad 0 < x < 1$$

where $\theta > 0$ is parameter to be estimated.

(i) Integrate the expectation of X_1 and derive a method of moments estimator for θ.

(6 marks)

(ii) Estimate the likelihood function, $L(\theta; X)$ for θ, where $X = (X_1,\ldots,X_n)'$.

(4 marks)

(b) Under standard growing conditions, the yields from plots sown with a particular crop are known to follow a normal distribution with expectation of 10 kg. When the crop is grown under new conditions in eight plots, the following independent yields (in kg) are recorded.

10.08, 9.79, 10.29, 10.41, 10.55, 11.02, 10.33, 10.49

The sum of these eight yields is 82.96 kg and the sum of the squared yields is 861.1946 kg.

(i) Conduct a test of size 1% of the null hypothesis that the expected yield, θ under the new growing conditions is $\theta = 10$ kg against the one-sided alternative hypothesis that $\theta > 10$ kg.

(12 marks)

(ii) Calculate an equal-tailed 99% confidence interval for θ.

(3 marks)

- END OF QUESTIONS -

TERBUKA