UNIVERSITI BRUNEI DARUSSALAM

SEMESTER II EXAMINATION, SESSION 2019/2020

for the Degree of Bachelor of Science
Fourth Year Course

SM-4331 ADVANCED STATISTICS

Time allowed: TWO (2) Hours

Wednesday, 13 May 2020. 2:00 pm.

Instructions to candidates:

1. Check that this paper has 3 pages (including this cover page).
2. There are four (4) questions in total. Answer **ALL** questions.
3. Write all your answers neatly, showing your working, in the answer booklets provided.
4. Write your name and registration number on the answer booklet(s).
5. You are supplied with a Formula Sheet and Statistical Tables.
6. Calculators may be used.

Question:	1	2	3	4	Total
Points:	11	21	18	10	60
Score:					

Question 1

(a) Show that the mean squared error of an estimator $\hat{\theta}$ for θ, given by $\text{MSE}(\hat{\theta}) = \text{E}\left[(\hat{\theta}-\theta)^2\right]$, can be decomposed into two parts: the bias of the estimator, and the variance of the estimator. **[3 marks]**

(b) Two independent random samples, of n_1 and n_2 observations, are drawn from normal distributions with the same variance σ^2. Let s_1^2 and s_2^2 be the sample variances of the first and the second sample, respectively. Show that

$$\hat{\sigma}^2 = \frac{1}{n_1+n_2-2}\left\{(n_1-1)s_1^2 + (n_2-1)s_2^2\right\}$$

is an unbiased estimator for σ. **[3 marks]**

(c) A public health researcher wants to know how two schools–one in the inner city and one in the suburbs–differ in the percentage of students who smoke. A random survey of students gives the following results: of 125 students in the inner city, 47 smoke, while of 153 students in the suburbs, 52 smoke. Derive an approximate 90% confidence interval for the difference between the smoking rates in the two schools. **[5 marks]**

Question 2

Let $\mathbf{X} := \{X_1, \ldots, X_n\}$ be a random sample independently drawn from a Poisson distribution with mean $\mu > 0$. The pmf of each X_i is $f_{X_i}(x|\mu) = e^{-\mu}\mu^x/x!$.

(a) Obtain the maximum likelihood estimate for μ. Is it unbiased? **[4 marks]**

(b) Find the Cramér-Rao lower bound for the variance of unbiased estimators of μ. **[4 marks]**

Let $\mathbf{Y} := \{Y_1, \ldots, Y_m\}$ be another random sample drawn independently of X_i from another Poisson distribution with mean $\lambda > 0$.

(c) Find the maximum likelihood estimator of μ when $\lambda = \mu$, i.e. the MLE of μ using **both** samples $\mathbf{X}$ and $\mathbf{Y}$. **[4 marks]**

(d) Derive the log-likelihood ratio test statistic for testing

$$H_0 : \mu = \lambda \quad \text{v.s.} \quad \mu \neq \lambda,$$

and specify its asymptotic distribution. **[6 marks]**

(e) Will you reject the null hypothesis H_0 if $n = 20$, $m = 15$, $\sum_i X_i = 21.4$ and $\sum_j Y_j = 14.6$? **[3 marks]**

Question 3

Due to its viral popularity, employees are worryingly found to be spending too much of company's time on the social media platform Kit Kot. In deciding whether or not to ban the platform during working hours, the company decides to estimate the total number of unproductive hours in a month that an employee spends at work creating video content for Kit Kot. Apparently, the use rate varies among different age groups, so the company decides to use stratified random sampling across three age groups: 1) less than 30; 2) 30 to 55; and 3) greater than 55 years old. The table below shoes the data obtained from a stratified random sample.

n_h	Stratum 1	Stratum 2	Stratum 3
1	8	4	1
2	24	0	8
3	0	8	
4	0	3	
5	16	1	
6	32	5	
7	6	24	
8	0	12	
9	16	2	
10	7	8	
11	4		
12	4		
13	9		
14	5		
15	8		
16	18		
17	2		
18	0		
σ_h^2	36	25	9
N_h	86	72	52

Table 1: Total unproductive hours spent on Kit Kot. Note that the true variances in each stratum $\left(\sigma_h^2 = N_h^{-1}\sum_i (X_i - \mu_h)^2\right)$ is assumed to be known.

(a) Calculate the sample mean and the **corrected** stratum variance, S_h^2, of the number of monthly unproductive hours within each strata. **[6 marks]**

(b) Calculate the estimate of the mean number of monthly unproductive hours from the stratified sample. State the variance of this estimator. **[4 marks]**

(c) Determine the minimum sample size required to achieve a bound of 0.01 on the estimation error of the mean with 95% confidence. **[4 marks]**

(d) Estimate the **total** number of monthly unproductive hours, and construct a suitable 95% confidence interval for this estimate. **[4 marks]**

Question 4

Aptitude test scores, x, and grades for a statistics test, y, for $n = 22$ students yield the following data:

$$\bar{x} = 80, \quad \bar{y} = 75, \quad \sum_{i=1}^{22}(x_i - \bar{x})(y_i - \bar{y}) = 800,$$

$$\sum_{i=1}^{22}(x_i - \bar{x})^2 = 1000, \quad \sum_{i=1}^{22}(y_i - \bar{y}) = 720.$$

(a) Estimate the regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where each $\epsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, and give an estimate for σ^2. **[5 marks]**

(b) Using a Wald test at the 5% significance level, test the hypothesis $H_0 : \beta_1 = 1 + 2\beta_0$ against the alternative $H_1 : \beta_1 \neq 1 + 2\beta_0$. **[5 marks]**

End of Paper