PALESTINE POLYTECHNIC UNIVERSITY

Department of Applied Mathematics & Physics

Regression Analysis

First Exam (40 points) **Wednesday 14/10/2020** **60 Minutes**

Instructor: *Dr. Monjed H. Samuh*

Name: Key Student ID:

Q1]...[22 points] For $i = 1, 2, \ldots, n$, let $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where the X_i's are fixed and the ϵ_i's are uncorrelated random variables with mean zero and common variance σ^2. β_0 & β_1 are unknown constants. Let $\hat{Y}_i = b_0 + b_1 X_i$ be the least squares line with

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \text{and} \quad b_0 = \bar{Y} - b_1 \bar{X}.$$

Show that the sampling distribution of b_1 is normal with mean β_1 and variance $\sigma^2 / \sum_{i=1}^n (X_i - \bar{X})^2$.

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

$$= \frac{\sum (X_i - \bar{X}) Y_i - \sum (X_i - \bar{X}) \bar{Y}}{\sum (X_i - \bar{X})^2} \quad +1$$

$$= \frac{\sum (X_i - \bar{X}) Y_i}{\sum (X_i - \bar{X})^2} \quad +1 \quad = \sum_{i=1}^{n} \frac{(X_i - \bar{X})}{\sum (X_i - \bar{X})^2} Y_i \quad +1$$

$$= \sum K_i Y_i \quad +1 \quad \text{where } K_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$$

Thus, b_1 is a linear combination of Y_i; +1

If $Y_i \sim$ Normal $\Rightarrow$ b_1 is normal. +1

$$E(b_1) = E\left(\sum K_i Y_i\right) = \sum \left(E(K_i Y_i)\right) \quad +1$$

$$= \sum K_i E(Y_i) \quad +1 \quad = \sum K_i (\beta_0 + \beta_1 X_i) \quad +1$$

$$= \beta_0 \sum K_i + \beta_1 \sum K_i X_i \quad +1 \quad \Rightarrow$$

~~This page is intentionally left blank.~~

But $\sum k_i = \sum \frac{(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \text{Zero}.$ (+1) (+1)

$$\sum k_i x_i = \sum \frac{(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} x_i = \sum \frac{(x_i - \bar{x}) x_i}{\sum(x_i - \bar{x})^2}$$ (+1)

$$= \frac{\sum (x_i - \bar{x}) x_i}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = 1.$$ (+1) (+1)

Therefore, $E(b_1) = \beta_1$ #

$$var(b_1) = var\left(\sum k_i y_i\right)$$

$= \sum var(k_i y_i)$ (+1) $\quad \because y_i$ indep. (+1)

$= \sum k_i^2 \, var(y_i)$ (+1)

$= \sigma^2 \sum k_i^2$ (+1)

$= \dfrac{\sigma^2}{\sum (x_i - \bar{x})^2}$ (+1)

#

$$\left\{ \begin{aligned} \sum k_i^2 &= \sum \left(\frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2} \right)^2 \\ &= \frac{\sum (x_i - \bar{x})^2}{\left[\sum (x_i - \bar{x})^2\right]^2} \\ &= \frac{1}{\sum (x_i - \bar{x})^2} \end{aligned} \right.$$ (+1) (+1)

Q2]...[18 points] A substance used in biological and medical research is shipped by airfreight to users in cartons of 1000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route X and the number of ampules found to be broken upon arrival Y. Assume that first-order regression model is appropriate. The relation $Y = \beta_0 + \beta_1 X + \epsilon$ is estimated using this data and the least square line is found to be:

i	1	2	3	4	5	6	7	8	9	10
X_i	1	0	2	0	3	1	0	1	2	0
Y_i	16	9	17	12	22	13	8	15	19	11

$$\hat{Y} = 10.2 + 4X.$$

You are also given that:

$$\sum_{i=1}^{10}(\hat{Y}_i - Y_i)^2 = 17.6, \quad \sum_{i=1}^{10}(Y_i - \bar{Y})^2 = 177.6, \quad \sum_{i=1}^{10}(X_i - \bar{X})^2 = 10.$$

1. (**2 points**) Obtain a point estimate of the expected number of ampules when $X = 1$ transfer is made.

$$\hat{Y}\Big|_{X=1} = 10.2 + (4)(1) = \underline{\underline{14.2}}.$$

(+1) (+1)

2. (**1 points**) Estimate the increase in the expected number of ampules broken when there are 2 transfers as compared to 1 transfer.

$$b_1 = \underline{\underline{4}}.$$ (+1)

3. (**2 points**) Obtain the residual for the first case.

$$e_1 = Y_1 - \hat{Y}_1 = 16 - 14.2 = \underline{\underline{1.8}}$$

(+1) (+1)

4. (**7 points**) Estimate β_1 with a 95% confidence interval. Interpret your interval estimate.

$b_1 \pm t_{(0.975, 8)} S_{b_1}$ (+1)

$4 \pm (2.306)(0.469)$ (+1)

$\Rightarrow (2.918, 5.082)$ (+1)

We are 95% confident that the true value of β_1 lies between $(2.918, 5.082)$. (+1)

$t_{(0.975, 8)} = 2.306$ (+2)

$S_{b_1}^2 = \frac{MSE}{S_{xx}}$

$= \frac{17.6/8}{10}$

$= 0.22$ (+1)

$S_{b_1} \doteq 0.4690$

5. (**6 points**) Conduct a t-test to decide whether or not there is a linear association between number of times a carton is transferred X and number of broken ampules Y. Use $\alpha = 0.05$. State the alternatives, decision rule, and conclusion. What is the p-value of the test.

$H_0: \beta_1 = 0$ vs $H_A: \beta_1 \neq 0$ (+1)

Test statistic : $t_{cal} = \frac{b_1}{S_{b_1}} = \frac{4}{0.469} \simeq 8.529$ (+2)

Critical values ± 2.306.

We reject H_0 (+1)

p-value $= 2P(T_{(8)} > 8.529)$ (+1)

$\ll 0.0005$. (+1)

GOOD LUCK