University of Waterloo
Stat 372 – W16
Term Test III

Date: Thursday, March 31, 2016. Duration: 60 minutes

(Non-programmed) calculator permitted.

Family Name:______________________ First Name:______________________ I.D. #:____________

Signature:______________________

Instructor: P. Balka

Instructions:

- This exam has 5 pages including this cover page. The marks for each question are indicated (total of 30). **Show your work. Your grade will be influenced by how clearly you express your ideas, and how well you organize your solutions.**
- No questions will be permitted.
- Unless otherwise stated, include the *fpc* in all relevant calculations.

solutions

A mathematics test was given to 486 students prior to entering a certain college who then took a calculus class. A simple random sampling of 10 students are selected and their calculus grade and math test score are recorded.

It is known that the mean math test score for the 486 students was 52.

The sample mean and sample standard deviation of the math test score are 46 and 16.6, respectively.

The sample mean and standard deviation of the calculus grade are 76 and 15.1, respectively.

A scatterplot of the sample data indicated a linear relationship between calculus grade, y, and math test score, x.

The simple linear regression model

$$Y_i = \alpha + \beta(x_i - \bar{x}) + R_i \quad R_i \sim G(0,\sigma)$$

was fit to the data, yielding $\hat{\beta} = 0.7656$, with the residual sum of squares = $\sum \hat{r}_i^2 = 606$.

(Note that $\hat{\alpha}$ can be obtained from the information provided).

1) **[2]** Provide a ratio estimate of the mean calculus grade, $\mu(y)$ of these 486 students.

$$\left.\begin{array}{l} \mu(x) = 52 \\ \hat{\mu}(x) = 46 \\ \hat{\mu}(y) = 76 \end{array}\right\} \qquad \hat{\mu}(y)_{ratio} = \hat{\mu}(y)\frac{\mu(x)}{\mu(\hat{x})}$$

$$= 76\left(\frac{52}{46}\right) = 85.9$$

2) **[3]** Provide a regression estimate of the mean calculus grade. Show your work.

$$\hat{\mu}(y)_{reg} = \hat{\alpha} + \hat{\beta}(\mu(x) - \hat{\mu}(x))$$

$$= \hat{\mu}(y) + \hat{\beta}(\mu(x) - \hat{\mu}(x))$$

$$= 76 + 0.7656(52 - 46)$$

$$= 80.6$$

3) **[5]** Provide a 95% confidence interval for the mean calculus grade, based on the regression estimate in b). Show your work.

$\hat{Var}(\hat{\mu}(y)_{reg}) = \left(1-\frac{n}{N}\right)\frac{\hat{\sigma}_r^2}{n}$, where $\hat{\sigma}_r^2 = \sum \hat{r}_i^2/(n-1)$

$= 606/9$

$= \left(1-\frac{10}{486}\right)\frac{606/9}{10}$

$= 6.6$

$\Rightarrow SE(\hat{\mu}(y)_{reg}) = \sqrt{6.6} = 2.57$

95% Confidence Interval for $\mu(y)$:

$\hat{\mu}(y)_{reg} \pm 1.96\, SE(\hat{\mu}(y)_{reg})$

$= 80.6 \pm 1.96\ 2.57 = 80.6 \pm 5.0 = (75.6, 85.6)$

4) **[4]** For this data, estimate the percentage increase in precision (in terms of the variance of the estimator) of the regression estimator compared to that of the sample mean estimator, $\hat{\mu}(y)$.

$\hat{Var}(\hat{\mu}(y)_{reg}) = 6.6$

$\hat{Var}(\hat{\mu}(y)) = \left(1-\frac{n}{N}\right)\frac{\hat{\sigma}^2}{n}$

$= \left(1-\frac{10}{486}\right)\frac{(15.1)^2}{10}$

$= 22.3$

$22.3/6.6 = 3.38 \Rightarrow$ There is a 238% increase in precision from using $\hat{\mu}(y)_{reg}$ vs $\hat{\mu}(y)$

5) Suppose (exactly) 3 of the 10 students in the sample received a calculus grade of 90 or higher.

a) **[5]** Provide a 95% confidence interval for the **total number** of students in the frame who received a grade of 90 or higher.

$$\hat{\tau} = N\hat{\pi} = 486\left(\tfrac{3}{10}\right) = 145.8$$

$$SE(\hat{\pi}) = \sqrt{\hat{Var}(\hat{\pi})}$$

$$= \sqrt{\left(1-\tfrac{n}{N}\right)\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

$$= \sqrt{\left(1-\tfrac{10}{486}\right)\left(\tfrac{.3(.7)}{10}\right)} = 0.1434$$

$$\Rightarrow SE(\hat{\tau}) = N\,SE(\hat{\pi}) = 69.7$$

95% C.I for τ:

$$\hat{\tau} \pm 1.96\,SE(\hat{\tau}) = 145.8 \pm 136.6 = \underline{\underline{(8.8,\ 282.4)}}$$

b) **[5]** Calculate the sample size required to estimate the total number of students who received a grade of 90 or higher to within 50, 19 times out of 20. Ignore the *fpc*.

Find n such that $1.96 \cdot N\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} < 50$

$$\Rightarrow n \geq \left(1.96^2 N^2 \hat{\pi}(1-\hat{\pi})\right)/50^2$$

$$n \geq \left(1.96^2\, 486^2\, 0.3(0.7)\right)/50^2$$

$n \geq 76.2 \Rightarrow n \geq \underline{\underline{77}}$ students

c) **[2]** Would including the *fpc* in your calculations in b) yield a larger or smaller required sample size than that obtained without including the *fpc*? Briefly explain. No calculations are required.

Smaller, since fpc reduces the variance of the estimator, thus increasing the precision, so that you would require a smaller sample size for a given margin of error than if you did not include the fpc.

d) **[4]** Consider the SRS proportion estimate, $\hat{\pi} = \frac{\sum_{i=1}^{n} y_i}{n}$, where $y = \{1 \text{ if grade 90 or higher; 0 otherwise}\}$.

Derive $E(\hat{\pi})$ by defining an appropriate indicator random variable, I_i. Show your work.

$$I_i = \begin{cases} 1 & \text{if } i^{th} \text{ student selected in sample} \\ 0 & \text{o/w} \end{cases}$$

Then $\hat{\pi} = \sum_{i=1}^{n} y_i / n$

$$= \sum_{i=1}^{N} I_i y_i / n$$

$$E(\hat{\pi}) = \frac{1}{n} \sum_{i=1}^{N} y_i E(I_i)$$

$$\left(\begin{aligned} E(I_i) &= 0P(I_i=0) + 1P(I_i=1) \\ &= P(I_i=1) = P(\text{selected}) \\ &= \frac{n}{N} \end{aligned} \right)$$

$$= \frac{1}{\cancel{n}} \sum_{i=1}^{N} y_i \cancel{n}/N$$

$$= \sum_{i=1}^{N} y_i / N = \pi$$