Answers.

MAT 2375: Introduction to Statistics
Midterm 1
February 14, 2017

Professor: Kelly Burkett **Exam Duration: 80 minutes**

Student Number:________________________

Name: ______________________________

- Only non-graphing, non-programmable calculators are allowed.
- This is a closed book exam. Formulas have been provided on pages 7-9 of the exam booklet.
- The exam is out 50 points total. There are 7 questions.
- Questions 1 to 4 are multiple choice and worth 5 marks each. No part marks will be given for the multiple choice questions. Write your answer to questions 1 to 4 in the boxes below.
- Questions 5 to 7 are long answer questions. Be sure to show all of your work as part marks will be given.
- **Cellular phones, unauthorized electronic devices or course notes (unless an open-book exam) are not allowed during this exam. Phones and devices must be turned off and put away in your bag. Do not keep them in your possession, such as in your pockets. If caught with such a device or document, the following may occur: academic fraud allegations will be filed which may result in you obtaining a 0 (zero) for the exam.**

  By signing below, you acknowledge that you have ensured that you are complying with the above statement.

  Signature: ____________________________

Answers to questions 1 to 4 (5 points each) :

E	D	B	A
#1	#2	#3	#4

Multiple choice questions (1-4)

Question 1 *A manufacturer of photocopiers is studying the time that their devices spend out-of-service in a six month period. Eight clients are recruited and the number of hours out-of-service is recorded for each client's photocopier. Compute the IQR of the sample times given below:*

0 16 21 26 28 28 34 57

28 26 21 16 28 34 57 0

A) IQR=12.5
B) IQR=49.75
C) IQR=25
D) IQR=57
E) IQR=15.25

$n=8$

Q1: $m = 0.25(9) = 2.25$ $Q1 = 0.75(16) + 0.25(21) = 17.25$

Q3: $m = 0.75(9) = 6.75$ $Q3 = 0.25(28) + 0.75(34) = 32.5$

$IQR = 32.5 - 17.25 = 15.25$

Question 2 *Let Y be the average grade of a student in a calculus course and let x is the grade that student obtained on a math skills test administered before the beginning of the semester. The following gives summary statistics for the grades on both tests for 10 randomly sampled students enrolled in the class.*

$$\sum x_i = 460, \quad \sum y_i = 760, \quad \sum x_i^2 = 23634$$

$$\sum y_i^2 = 59816, \quad \sum x_i y_i = 36854.$$

Determine the equation for the linear regression of Y on x.

A) $y = -12.183 + 0.76556x$
B) $y = 46.501 + 0.64129x$
C) $y = -2.738 + 0.64129x$
D) $y = 40.784 + 0.76556x$
E) $y = 76 + 0.76556x$

$$\hat{\beta} = \frac{36854 - \frac{1}{10}(460)(760)}{23634 - \frac{1}{10}(460)^2} = \frac{1894}{2474} = 0.76556$$

$$\alpha = \frac{760}{10} - 0.76556\left(\frac{460}{10}\right) = 40.784$$

Question 3 *The tensile strength of an aluminum alloy is normally distributed with mean 10 gigapascals (GPa) and standard deviation 1.4 GPa. Find the probability that a randomly selected unit of the alloy has a tensile strength above 12 GPa.*

A) 0.9236 *B)* 0.0764 *C)* 0.3241 *D)* 0.0032 *E)* 0.0913

$$Pr(X > 12) = 1 - Pr(X < 12)$$
$$= 1 - Pr\left(\frac{X-10}{1.4} < \frac{12-10}{1.4}\right)$$
$$= 1 - \Phi(1.4286)$$
$$= 1 - 0.9236$$
$$= 0.0764$$

Question 4 *Let $X_1, ..., X_n$ be a random sample from a continuous distribution with probability density function*

$$f_X(x) = \theta x^{\theta-1}, 0 < x < 1, \; 0 < \theta < \infty.$$

Find a method of moments estimator for θ

A) $\frac{\overline{X}}{1-\overline{X}}$ *B)* $\min(X_i)$ *C)* $\overline{X}$ *D)* $\sum_{i=1}^{n} \ln(X_i)$ *E)* $\prod_{i=1}^{n} X_i$

$$E[X] = \int_0^1 x\theta x^{\theta-1}\,dx = \left.\frac{\theta x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

$$\overline{X} = \frac{\tilde{\theta}}{\tilde{\theta}+1} \Rightarrow \overline{X}\hat{\theta} + \overline{X} = \hat{\theta}$$
$$\overline{X} = \tilde{\theta}(1-\overline{X})$$
$$\tilde{\theta} = \frac{\overline{X}}{1-\overline{X}}$$

Question 5 (10 points) *Let $X_1, X_2, \ldots, X_n$ be a random sample (i.i.d.) of size n from some continuous distribution $F_X(x)$ and let $Y_1 < Y_2 < \ldots < Y_n$ be the order statistics $(Y_1 = X_{(1)}, \ldots, Y_n = X_{(n)})$.*

(a) *Use $F_{Y_1}(y_1) = Pr(Y_1 < y_1)$ to prove that the pdf of Y_1 is*

$$f_{Y_1}(y_1) = n(1 - F_X(y_1))^{n-1} f_X(y_1).$$

DO NOT use the formula for the rth order statistic.

$$F_{Y_1}(y_1) = Pr(Y_1 < y_1) = 1 - Pr(Y_1 > y_1) \; (1)$$
$$= 1 - Pr(X_1 > y_1, X_2 > y_1, \ldots (1), X_n > y_n)$$
$$= 1 - Pr(X > y_1)^n$$
$$= 1 - [1 - F_X(y_1)]^n \; (1)$$

$$f_{Y_1}(y_1) = (1) \frac{dF_{Y_1}(y_1)}{dy_1} = -n[1 - F_X(y_1)]^{n-1} \cdot (-f_X(y_1))$$
$$= n(1 - F_X(y_1))^{n-1} f_X(y_1) \; (1)$$

b) *Let $n = 6$ and assume that each X_i is Exponential with rate $\lambda = \frac{1}{2}$, that is $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Find $P(Y_1 > 4)$.*

$$Pr(Y_1 > 4) = 1 - Pr(Y_1 < 4) = [1 - \overset{(1)}{F_X}(4)]^6$$

$$F_X(x) = \int_0^x \overset{}{\underset{(1)}{\lambda}} e^{-\lambda t}\, dt = -e^{-\lambda t}\Big]_0^x = 1 - \underset{(1)}{e^{-\lambda x}}$$

So $Pr(Y_1 > 4) = [1 - (1 - \overset{(1)}{e^{-1/2(4)}})]^6 = (e^{-2})^6 = e^{\overset{(1)}{-12}}$

or // $Pr(Y_1 < 4) = \int_0^4 f_{Y_1}(y_1)\, dy_1$ where $f_{Y_1}(y_1) = 6[1 - (1 - e^{-\frac{1}{2}y_1} (1))]^5 \frac{1}{2} e^{-1/2 y_1}$

$$= 3e^{-1/2 y_1} \cdot e^{-5/2 y_1}$$
$$= 3e^{-3y_1} \; (1)$$

$$= \int_0^4 3e^{-3y_1}\, dy_1$$
$$= -e^{-3y_1} (1) \Big]_{y_1=0}^{y_1=4}$$
$$= 1 - e^{-12}$$

So $Pr(Y_1 > 4) = 1 - (1 - e^{-12}) = e^{-12}$ (1)

(1) $F_X(x)$.

Question 6 (10 points) *Let $X_1, ..., X_n$ be a random sample from a continuous distribution with probability density function*

$$f_X(x) = \frac{1}{\theta} x^{\left(\frac{1-\theta}{\theta}\right)}, 0 < x < 1,\ 0 < \theta < \infty.$$

Find the maximum likelihood estimator for θ and determine if the estimator is unbiased.

$$L(\theta) = \theta^{-n} \left(\prod_{i=1}^{n} x_i\right)^{1/\theta - 1} \quad 1$$

$$\ell(\theta) = -n \ln\theta + \left(\frac{1}{\theta} - 1\right) \sum_{i=1}^{n} \ln x_i \quad 1$$

$$\frac{\partial \ell}{\partial \theta} = \frac{-n}{\theta} + \left(-\frac{1}{\theta^2}\right) \sum_{i=1}^{n} \ln x_i \quad 1$$

$$0 = -\frac{n}{\hat{\theta}} - \frac{1}{\hat{\theta}^2} \sum \ln x_i \quad 1 \quad \Rightarrow \quad \hat{\theta} = -\frac{1}{n} \sum \ln x_i \quad 1$$

Bonus - 2nd derivative

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{n}{\theta^2} + \frac{2}{\theta^3} \sum \ln x_i$$

$$\left.\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\hat{\theta}} = \frac{n}{\left(-\frac{1}{n}\sum \ln x_i\right)^2} + \frac{2 \sum \ln x_i}{\left(-\frac{1}{n}\sum \ln x_i\right)^3}$$

$$= \frac{n^3}{(\sum \ln x_i)^2} - \frac{2n^3}{(\sum \ln x_i)^2}$$

$$= \frac{-n^3}{(\sum \ln x_i)^2} < 0$$

(2 marks bonus)

The estimator is unbiased if $E[\hat{\theta}] = \theta$ 1

$$E[\hat{\theta}] = -\frac{1}{n} \sum_{i=1}^{n} E[\ln x_i] \quad 1$$

$$E[\ln X] = \int_0^1 \ln x \cdot \frac{1}{\theta} x^{\frac{1}{\theta} - 1}\, dx \quad 1$$

$$= \underbrace{x^{1/\theta} \ln x \Big]_0^1}_{(*)} - \int_0^1 x^{1/\theta - 1}\, dx \quad 1$$

$$= 0 - \theta$$

$$= -\theta$$

So

$$E[\hat{\theta}] = -\frac{1}{n} \sum_{i=1}^{n} (-\theta) \overset{1}{=} \theta$$

$$(*) \quad 0 - \lim_{x \to 0^+} \frac{\ln x}{x^{-1/\theta}}$$

$$= 0 - \lim_{x \to 0^+} \frac{1/x}{-1/\theta\, x^{-1/\theta - 1}}$$

$$= 0 - \lim_{x \to 0^+} \frac{x^{1/\theta}}{-1/\theta}$$

$$= 0$$

Question 7 (10 points) *Consider the following linear regression model*

$$Y_i = \alpha + \beta x_i + \epsilon_i$$

where ϵ_i are independent and identically distributed with mean 0 and variance σ^2, for $i = 1, \ldots n$. Let $\hat{\alpha}$ and $\hat{\beta}$ be the usual least squares estimators of α and β, respectively. Show that

$$E(\hat{\beta}) = \beta, \quad Var(\hat{\beta}) = \frac{\sigma^2}{SS_X},$$

Note that all relevant formulas are provided on the formula sheets.

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{SS_x} = \frac{1}{SS_x}\left\{\sum (x_i - \bar{x})y_i - \bar{y} \overset{0}{\cancel{\sum (x_i - \bar{x})}}\right\}$$

$$E[\hat{\beta}] = \frac{1}{SS_X} \sum_{i=1}^{n} (x_i - \bar{x}) E[Y_i]$$ ① taking $E[\,]$ into sum

$$= \frac{1}{SS_X} \sum_{i=1}^{n} (x_i - \bar{x})(\alpha + \beta x_i)$$ ① substitute $\alpha + \beta x_i$

$$= \frac{1}{SS_x}\left\{ \overset{0}{\cancel{\sum_{i=1}^{n} (x_i - \bar{x})\alpha}} + \sum_{i=1}^{n} (x_i - \bar{x})\beta x_i \right\}$$ ① expand sum.

$$= \frac{\beta}{SS_X} \sum_{i=1}^{n} (x_i - \bar{x}) x_i$$ ① first term is 0.

(add 0)

$$= \frac{\beta}{SS_x} \sum_{i=1}^{n} (x_i - \bar{x}) x_i - \bar{x} \sum_{i=1}^{n} (x_i - \bar{x})$$

$$= \frac{\beta}{SS_x} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \beta$$ ① get SS_x in numerator ① answer

> Note: Some students may show this with $\hat{\beta} = \frac{\sum x_i y_i - \bar{x}\bar{y}}{SS_x}$. this is fine.

$$V[\hat{\beta}] = \left(\frac{1}{SS_X}\right)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \, V[Y_i]$$ ① squaring ① ~~expanding~~ $V[\,]$ in sum.

$$= \left(\frac{1}{SS_X}\right)^2 \cdot \sum_{i=1}^{n} \sigma^2 (x_i - \bar{x})^2$$ ① plug in $V(Y_i) = \sigma^2$

$$= \frac{\sigma^2}{SS_X}$$ ① answer.