Probability and Statistics: Final Exam

Spring 2018

Name: ____

This exam is scheduled for 110 minutes. Notes and other outside materials are not permitted. Non graphing calculators are allowed; if you do not have any, numerical formulas are enough.
Show all work to receive full credit, except where specified. The exam is worth 60 points.

Problem Number	Problem Points	Points Earned
MC	10	
TF	5	
FR1	10	
FR2	10	
FR3	10	
FR4	8	
FR5	7	
Total	60	

Multiple Choice

$q_{0.25} = 2 + 0.25$ $= 2.25$ $q_{0.75} = 15 + 0.75 \cdot 3$ $= 17.25$

(2 points each) Circle the correct answer for each question. *You need not justify your answer, but one point of partial credit may be awarded.*

1, 2, 3, 3, 4, 7, 15, 18, 21, 25

D

1 Consider the dataset below:

$$1, 21, 3, 18, 4, 7, 15, 3, 18, 2, 25$$

What is the interquartile range for this dataset?

- (A) 17
- (B) 18
- (C) 14
- (D) 15
- (E) 16

$\Phi\left(\frac{85-90}{3}\right) = -1.6$ $\frac{2}{3} = 0.67$

B

2 Let X be a normally distributed random variable with mean $\mu = 90$ and variance $\sigma^2 = 9$. What is a good approximation of $P(85 \leq X \leq 92)$?

Note: Tables for the cumulative distribution function Φ of $N(0,1)$ are provided at the end of the exam

- (A) 0.45
- (B) 0.70
- (C) 0.62
- (D) 0.84
- (E) 0.77

0.748
0.0548

$E[X] = 20$

A

3 Let X be a binomial random variable with parameters $n = 200$ and $p = 0.1$. According to Markov's inequality, what is an upper bound for $P(X \geq 120)$?

- (A) $\frac{1}{6}$
- (B) 0.287
- (C) $\frac{1}{8}$
- (D) $\frac{2}{17}$
- (E) 0.1173

$20 \geq 120 P(X \geq 120)$

$\frac{20}{120} \geq P($

D

4 Consider two random variables X and Y with the following joint probability mass function

		X		
		0	1	2
Y	0	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{8}$
	1	$\frac{1}{8}$	$\frac{1}{6}$	$\frac{1}{6}$

What is $E[X^3Y]$?

$\frac{1}{6} + \frac{8}{6} = \frac{9}{6} = \frac{3}{2}$

(A) $\frac{1}{2}$

(B) $\frac{5}{2}$

(C) $\frac{9}{4}$

(D) $\frac{3}{2}$

(E) $\frac{7}{4}$

$\frac{1}{\lambda} = 22 \quad \lambda = \frac{1}{22}$

A

5 A continuous random variable X has an exponential distribution with a mean of 22. What is the variance of X?

(A) 484

(B) 22

(C) 0.0454

(D) 0.0020661157

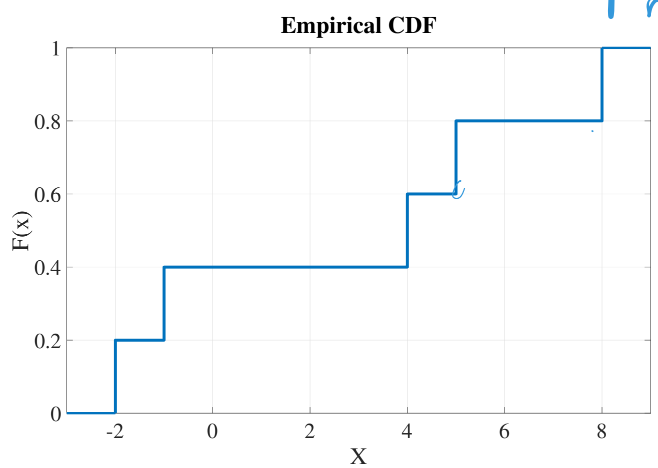
(E) 0.22

True or False

(1 point each) Indicate whether each statement is true or false. No partial credit will be given.

1 Consider the empirical cumulative distribution function below.

$F_n(x) = 1 = \frac{\# \le 8}{n}$



It corresponds to a dataset with 10 data points.

Ⓣ True Ⓕ False

T

2 If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with finite mean μ and finite variance σ^2, then

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

where $\overline{X}_n$ is the sample mean, is ALWAYS an unbiased estimator for the variance σ^2.

Ⓣ True Ⓕ False

F

3 If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with finite mean μ and finite variance σ^2, then

$$S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}$$

where $\overline{X}_n$ is the sample mean, is ALWAYS an unbiased estimator for the standard deviation σ.

Ⓣ True neg Ⓕ False

$0 \leq p(x^2 Y4) \leq 1$

T

4 For any two random variables X and Y, $E[X^2Y^4]$ is always a positive number.

Ⓣ True Ⓕ False

F

5 The method of least squares only applies to linear regression.

Ⓣ True Ⓕ False

Free Response

Be sure to show all your work neatly and indicate your final answer where appropriate.

1 (10 points) The number N_1 of male customers entering a high end female clothing store per 1-hour time slot is a Poisson process with rate $\lambda_1 = 1$, and the number N_2 of female customers entering that same clothing store per 1-hour time slot is a Poisson process with rate $\lambda_2 = 2$. In this problem, we will assume that N_1 and N_2 are independent Poisson processes.

(A) (2 Points) In Problem 2 of Homework 9, you showed that the total number of customers $N = N_1 + N_2$ entering the store in any given 1-hour time slot is also a Poisson process. What is the rate of this Poisson process?

(B) (4 Points) Find the probability of the event "$N_{[0,1]} = 2$ and $N_{[0,2]} = 5$".

(C) (4 Points) If you know that $N_{[0,1]} = 2$, what is the probability that $N_{1\,[0,1]} = 1$?

A $\lambda_1 = 3$

B. $P_{N[0,1]}(2) \cdot P_{N[1,2]}(3)$

$$= \frac{3^2}{2!} e^{-3} \cdot \frac{3^3}{3!} e^{-3} = \frac{3^4}{4} e^{-6}$$

C. $P(N_1 = 1 \mid N = 2)$

$$= \frac{P(N_1 = 1) \cdot P_{N_2}(1)}{P_N(2)}$$

$$= \frac{e^{-1} \cdot 2 \cdot e^{-2}}{\frac{3^2}{2} e^{-3}} = \frac{2 \cancel{e^{-3}}}{\frac{9}{2} \cancel{e^{-3}}}$$

$\overline{}$ e 2
$2!$

$= \frac{4}{9}$

2 (10 Points) An airline uses an Airbus airplane with 150 seats for a flight between Cleveland and New York City. For this flight, data shows that the probability that a person who bought a ticket actually comes and checks in for the flight is only $p = 0.75$. The airline thus decides to sell more tickets n than seats on the airplane: $n > 150$. Let Y be the random variable corresponding to the number of people who bought a ticket and checked in for the flight.

A (3 Points) What is the exact probability mass function for Y? Please write it explicitly.

B (7 Points) Use the central limit theorem to compute an accurate approximation of the number n of tickets the airline can sell while being 95% sure that all the customers checking in will have a seat on the airplane.

Note: Tables for the cumulative distribution function Φ *of* $N(0,1)$ *are provided at the end of the exam*

A

$$P_Y(Y) = \binom{n}{Y} Y^{0.75} (n-Y)^{0.25}$$

B

$$E[Y] = 0.75n$$

$$Var(Y) = 0.1825n$$

$$P(Y \leq 150) \geq 0.95$$

$$P\left(\frac{Y-\mu}{\sigma} \leq \frac{150-0.75n}{\sqrt{0.1825n}}\right) \geq 0.95$$

$$\Phi\left(\frac{150-0.75n}{\sqrt{0.1825n}}\right) \geq 1.65$$

$$150 - 0.75n = 1.65 \cdot \sqrt{0.1825n}$$

$$= 0.75 + 1.65\sqrt{0.1825}$$

$n = 187$

3 (10 points) A physics student measured the electrical conductivity in piece of copper as a function of temperature, and obtained the following data set, in appropriately nondimensionalized units:

Temperature	2	3	5	7
Electrical conductivity	3	6	7.5	12

(A) (2 Points) Draw the scatterplot for this data set.

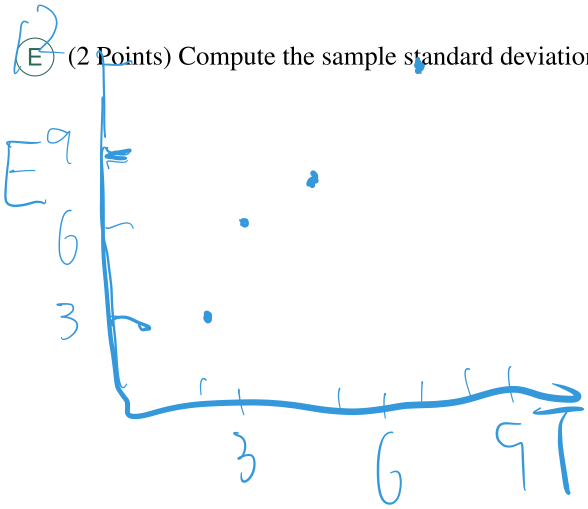
(B) (4 Points) Compute the best linear fit for the data set.

(C) (1 Points) Compute the residuals r_1, r_2, r_3, r_4 for this data set, measuring the difference between the data points and the best linear fit.

(D) (1 Points) Compute the sample mean of the residuals.

(E) (2 Points) Compute the sample standard deviation of the residuals.

(A)



(B)

(2,3) (3,6) (5,7.5) (7,12)

$$\beta = \frac{145.5 - 121.125}{87 - \frac{17^2}{4}} = 1.6525$$

$\bar{x}_n = \frac{17}{4}$

$\bar{y}_n = \frac{57}{8}$

$\alpha = 0.11$

$\bar{y}_n = 1.6525\,\bar{x}_n + 0.11$

(C) $r_1 = 3 - 1.65 \cdot 2 - 0.11 = -0.41$

$r_2 = 6 - 1.65 \cdot 3 - 0.11 = 0.94$

$r_3 = -0.86$

$r_4 = 0.34$

(D) $\bar{r}_n = \frac{57}{8} - \frac{17}{4} \cdot 1.65 - 0.11$

$= \frac{1}{400} \approx 0$

(E) $\sigma^2 = \qquad (0.41)^2 + \cdots (0.34)^2$

$= 1.91$

$Var(\bar{\sigma}^2) = \frac{1.91}{4} = 0.477$

$\bar{\sigma} = 0.69$

4 (8 points) Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with probability density function

$$f_X(x) = \begin{cases} \frac{1+\alpha x}{2} & \text{if } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where α is the parameter of interest, with $-1 \le \alpha \le 1$.

(A) (2 Points) Show that for any integer i between 1 and n, $E[X_i] = \frac{\alpha}{3}$

(B) (6 Points) Consider the estimator T for α given by $T = 3\frac{X_1 + X_2 + \ldots X_n}{n}$. Compute the bias, variance, and mean squared error of the estimator T.

$Var[T] = 3\left(\frac{X_1}{n} + \cdots \frac{X_n}{n}\right)$

$= 9 Var\left[\frac{\cancel{x_1 \cdots x_n}}{\cancel{n}} x_i\right]$

A $E[X_i] = \int_{-\infty}^{\infty} x f(x)\,dx$

$= \frac{1}{2}\int_{-1}^{1} (x + \alpha x^2)$

$= \frac{1}{2}\left(\frac{x^2}{2} + \alpha \cdot \frac{x^3}{3}\right]_{-1}^{1}$

$= \frac{1}{2}\left(\frac{\alpha \cdot 2}{3}\right) = \frac{\alpha}{3}$

B $E[T] = E\left[3 \cdot \frac{X_1 + \cdots X_n}{n}\right]$

$= 3\left(E\left[\frac{X_1}{n}\right] + \cdots E\left[\frac{X_n}{n}\right]\right)$

$= 3 \frac{E[X_1] + \cdots E[X_n]}{n}$

$= \frac{3 \cdot n \cdot \frac{\alpha}{3}}{n}$

$= \alpha$

unbiased

$$Var(T) = \frac{9}{n^2} Var(X_i)$$

$$= \frac{9}{n^2}\left[\int_{-1}^{1}\left(x^2 \cdot \frac{1+\alpha x}{2}\right) - \frac{\alpha^2}{9}\right] \cdot n$$

$$= \frac{9}{n}\left[\frac{1}{2}\int_{-1}^{1}(x^2+\alpha x^3) - \frac{\alpha^2}{9}\right] = \frac{9}{n^2}\left[\frac{1}{2}\left[\frac{x^3}{3}+\frac{\alpha x^4}{4}\right]_{-1}^{1} - \frac{\alpha^2}{9}\right]$$

$$= \frac{9}{n}\left(\frac{1}{2}\left(\frac{1}{3}+\frac{\alpha}{4}+\frac{1}{3}-\frac{\alpha}{4}\right) - \frac{\alpha^2}{9}\right)$$

$$= \frac{9}{n}\left(\frac{3-\alpha^2}{9}\right)$$

$$= \frac{3-\alpha^2}{n}$$

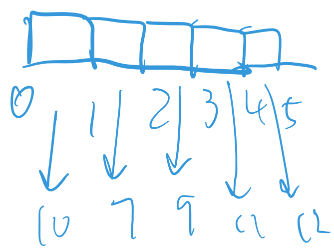
(c) $MSE = Var(T) + (E(T) - \theta)^2$

$$= \frac{3-\alpha^2}{n}$$

5 (7 points) A fast food restaurant is trying to predict the number of customers per hour it will have to serve. To do so, it assumes that this number is well modeled by a Poisson distribution, with unknown parameter λ. One day, in order to determine λ, the restaurant manager records the number of customers over a time span of 5 hours, split into separate 1-hour time slots, and finds the following numbers: 10, 7, 9, 11, 12.

What is the maximum likelihood estimate of λ?

$$P_{N_t}\, k = \frac{\lambda^k}{k!} e^{-\lambda}$$



$$L(\lambda) = \frac{\lambda^{10}}{10!} e^{-\lambda} \cdot \frac{\lambda^7}{7!} \cdot e^{-\lambda} \cdot \frac{\lambda^9}{9!} \cdot e^{-\lambda} \cdot \frac{\lambda^{11}}{11!} e^{-\lambda} \cdot \frac{\lambda^{12}}{12!} \cdot e^{-\lambda}$$

$$l(\lambda) = \ln(L(\lambda)) = \ln\left(\frac{\lambda^{10}}{10!}\right) - 5\lambda + \ln \frac{\lambda^7}{7!} \dots$$

$$= -5\lambda + \ln(\lambda^{10}) + \ln(10!^{-1})$$

$$= -5\lambda + 49 \ln \lambda + K \leftarrow \text{constant}$$

$$\frac{d\, l(\lambda)}{d\lambda} = -5 + \frac{49}{\lambda} = 0$$

$$\lambda = \frac{49}{5}$$

SCRAP PAPER – FEEL FREE TO DETACH

SCRAP PAPER – FEEL FREE TO DETACH

SCRAP PAPER – FEEL FREE TO DETACH