CE93 Review 1

Solutions to Quizzes 1 and 2

Quiz 1

Question 1 (20 points)

In the 10 years from 2000 to 2010, there were two major (M 5.0 or above) earthquakes in the San Francisco Bay Area, one in 2001 and one in 2007. Consider a data set of 10 observations, each of which is the number of major earthquakes in a year from 2000 to 2010.

(a) What is the sample mean of this data? [4]

(b) What is the sample median of this data? [4]

(c) What is the sample variance of this data? [4]

(d) What is the sample standard deviation of this data? [4]

(e) What is the sample coefficient of variation? [4]

Question 1

In the 10 years from 2000 to 2009, there were two major (M 5.0 or above) earthquakes in the San Francisco Bay Area, one in 2001 and one in 2007. Consider a data set of 10 observations, each of which is the number of major earthquakes in a year from 2000 to 2009.

Data Set is: [0,1,0,0,0,0,0,1,0,0]

(a) What is the sample mean of this data? [4] $\bar{x} = \frac{\sum_i x_i}{10} = 0.2$

(b) What is the sample median of this data? [4] *Median is average of 5th and 6th highest value=0*

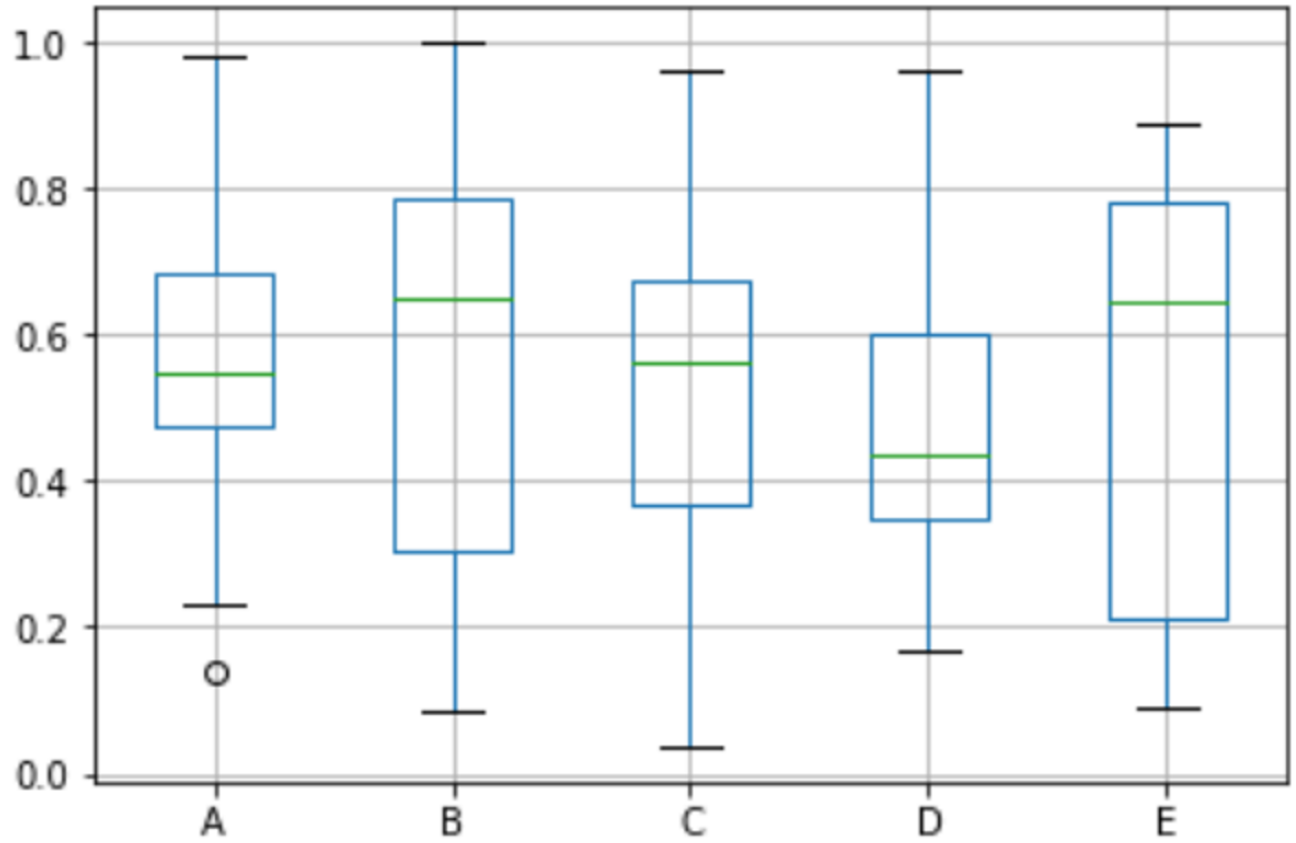
(c) What is the sample variance of this data? [4] $\text{var}(x) = \left(\frac{\sum_i x_i^2}{n} - \bar{x}^2 \right) \frac{n}{n-1} = \left(0.2 - 0.2^2\right) \frac{10}{9} = 0.178$

(d) What is the sample standard deviation of this data? [4] $stdev(x) = \sqrt{\text{var}(x)} = 0.42$

(e) What is the sample coefficient of variation? [4] $\text{cov}(x) = \frac{stdev(x)}{\bar{x}} = \frac{.178}{.2} = .889$

Question 2 (15 points)

Consider the 5 boxplots below:



(a) Which box plot, A,B,C,D, or E, has the most outliers? [3]

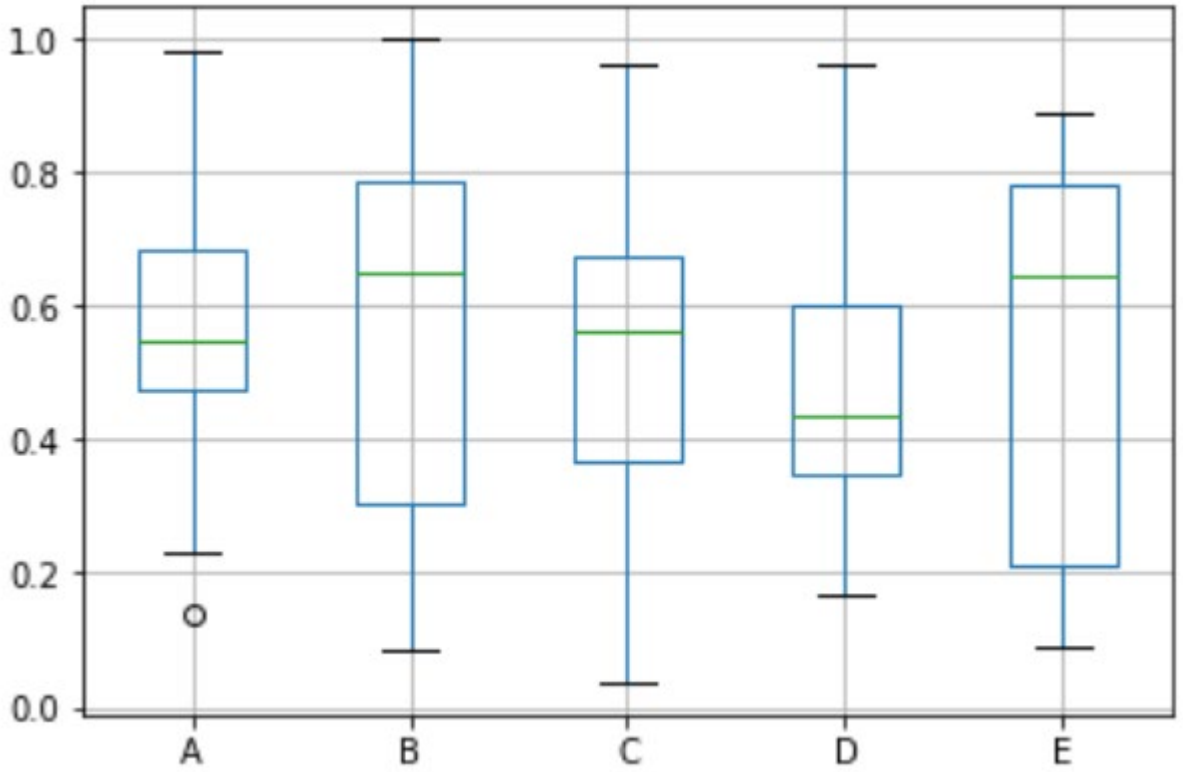
(b) Which box plot, A,B,C,D, or E, has the lowest median? [3]

(c) Which box plot, A,B,C,D, or E, has the greatest interquartile range? [3]

(d) Which box plot, A,B,C,D, or E, has the lowest first quartile? [3]

Question 2 (15 points)

Consider the 5 boxplots below:



(a) Which box plot, A,B,C,D, or E, has the most outliers? [3] Outliers are dots--A

(b) Which box plot, A,B,C,D, or E, has the lowest median? [3] Median is green line--D

(c) Which box plot, A,B,C,D, or E, has the greatest interquartile range? [3] IQR is the box height-E

(d) Which box plot, A,B,C,D, or E, has the lowest first quartile? [3] First quartile is bottom on box--E

Question 3 (15 points)

On any given day, the probability that there is a dense fog is 0.25 and the probability that it rains is 0.15. The probability that it rains and there is dense fog is 0.1. Let R be the event that it rains and F be the event that there is dense fog.

(a) What is $P(F)$? [3]

(b) What is $P(R|F)$? [3]

(c) What is $P(F|R)$? [3]

(d) What is $P(F \cup R)$? [3]

(e) What is $P(FR)$? [3]D

Question 3 (15 points)

On any given day, the probability that there is a dense fog is 0.25 and the probability that it rains is 0.15. The probability that it rains and there is dense fog is 0.1. Let R be the event that it rains and F be the event that there is dense fog.

What is P(F)? [3] *0.25 (given)*

What is P(R|F)? [3] $P(R \mid F) = \frac{P(RF)}{P(F)} = \frac{0.1}{0.25} = 0.4$

What is P(F|R)? [3] $P(F \mid R) = \frac{P(RF)}{P(R)} = \frac{0.1}{0.25} = \frac{2}{3}$

What is P(F∪R) ? [3] $P(F \cup R) = P(F) + P(R) - P(FR) = 0.25 + 0.15 - 0.1 = 0.3$

What is P(FR)? [3] *0.1 (given)*

Quiz 2

Question 1 (20 points)

A survey of cars on a certain stretch of highway showed that 70% had only occupant, 15% had 2, 10% had 3, 3% had 4, and 2% had 5. Let X be the number of occupants in a randomly selected car.

(a) Let $p(x)$ be the PMF for X.

i. What is $p(3)$? [2]

ii. What is $p(100)$? [2]

(b) Let $F(x)$ be the CDF for X.

i. What is $F(3)$? [2]

ii. What is $F(100)$? [2]

(c) What is $E(X)$? [4]

(d) What is $E(X^2)$? [4]

(e) What is $VAR(X)$? [4]

Question 1 (20 points)

A survey of cars on a certain stretch of highway showed that 70% had only occupant, 15% had 2, 10% had 3, 3% had 4, and 2% had 5. Let X be the number of occupants in a randomly selected car.

(a) Let $p(x)$ be the PMF for X.

i. What is $p(3)$? [2] $p(3) = P(X = 3) = 0.1$

ii. What is $p(100)$? [2] $p(100) = P(X = 100) = 0$

(b) Let $F(x)$ be the CDF for X.

i. What is $F(3)$? [2] $F(3) = P(X \le 3) = 0.7 + 0.15 + 0.1 = 0.95$

ii. What is $F(100)$? [2] $F(100) = P(X \le 100) = 1$

(c) What is $E(X)$? [4] $E(X) = \sum_x xp(x) = 1*0.7 + 2*0.15 + 3*0.1 + 4*0.03 + 5*0.02 = 0.7 + 0.3 + 0.3 + 0.12 + 0.1 = 1.52$

(d) What is $E(X^2)$? [4] $E(X^2) = \sum_x x^2 p(x) = 1*0.7 + 4*0.15 + 9*0.1 + 16*0.03 + 25*0.02 = 0.7 + 0.6 + 0.9 + 0.48 + 0.5 = 3.18$

(e) What is $VAR(X)$? [4] $VAR(X) = E(X^2) - \left(E(X)\right)^2 = 3.18 - 1.52^2 = 0.87$

Question 2 (15 points)

Random variables X and Y have the joint distribution:

$f(x, y) = x + y$ for $0 \leq x \leq 1$ *and* $0 \leq y \leq 1$

$f(x, y) = 0$ for all other values of x and y.

(a) What is the marginal distribution for X, $p_X(x)$?[5]

(b) What is the conditional distribution for Y, $p_{Y|X}(y|x)$?[5]

(c) What is the conditional expectation of X when $Y = 1$, $E(X|Y = 1)$? [5]

Question 2 (15 points)

Random variables X and Y have the joint distribution:

$f(x,y) = x + y$ for $0 \le x \le 1$ *and* $0 \le y \le 1$

$f(x,y) = 0$ for all other values of x and y.

(a) What is the marginal distribution for X, $p_X(x)$?[5] $p_X(x) = \int_0^1 f(x,y)dy = xy + \frac{y^2}{2}\Big]_0^1 = x + \frac{1}{2}$

(b) What is the conditional distribution for Y, $p_{Y|X}(y|x)$?[5] $p_{Y|X}(y \mid x) = \frac{f(x,y)}{p_X(x)} = \frac{x+y}{x+\frac{1}{2}}$

(c) What is the conditional expectation of X when $Y = 1$, $E(X|Y=1)$? [5]

$p_{X|Y}(x \mid y) = \frac{f(x,y)}{p_Y(y)} = \frac{x+y}{y+\frac{1}{2}}$

$E(x \mid y=1) = \int_0^1 x p_{X|1}(x \mid 1)dx = \int_0^1 x \frac{x+1}{1+\frac{1}{2}}dx = \frac{2}{3}\left(\frac{x^3}{3} + \frac{x^2}{2}\right)\Big]_0^1 = \frac{5}{9}$

Question 3 (15 points)

A California license plate has seven characters, which may be either letters or digits. Any character is equally likely to be any of the 26 letters or 10 digits. Answer the following questions, without evaluating factorials, expressions with exponents, or products.

(a) What is the probability that a license plate contains only letters? [3]

(b) What is the probability that the number 7 appears once and only once on a license plate? [3]

(c) What is probability that the license plate is DONALDT? [3]

(d) Suppose the license plate contains only numbers, and each number on the plate appears only once. What is the probability that the numbers appear in order (lowest number first, second lowest number second, etc)? [3]

(e) What is the probability that no letter or digit appears more than once on the license plate? [3]

Question 3 (15 points)

A California license plate has seven characters, which may be either letters or digits. Any character is equally likely to be any of the 26 letters or 10 digits. Answer the following questions, without evaluating factorials, expressions with exponents, or products.

(a) What is the probability that a license plate contains only letters? [3]

$$P(only\ letters) = \frac{\#\ of\ sequences\ only\ letters}{\#\ of\ sequences\ total} = \frac{26^7}{36^7} = \left(\frac{13}{18}\right)^7$$

Question 3 (15 points)

A California license plate has seven characters, which may be either letters or digits. Any character is equally likely to be any of the 26 letters or 10 digits. Answer the following questions, without evaluating factorials, expressions with exponents, or products.

b. What is the probability that the number 7 appears once and only once on a license plate? [3]

$$P(7 \text{ once and only once}) = \frac{\#\text{ of places 7 could go*\# of sequences without 7 in other 6 places}}{\#\text{ of sequences total}}$$

$$= \frac{7 * 35^6}{36^7}$$

Question 3 (15 points)

A California license plate has seven characters, which may be either letters or digits. Any character is equally likely to be any of the 26 letters or 10 digits. Answer the following questions, without evaluating factorials, expressions with exponents, or products.

(c) What is probability that the license plate is DONALDT? [3]

$$P(DONALDT) = \frac{1}{\#\text{ of sequences total}} = \frac{1}{36^7}$$

Question 3 (15 points)

A California license plate has seven characters, which may be either letters or digits. Any character is equally likely to be any of the 26 letters or 10 digits. Answer the following questions, without evaluating factorials, expressions with exponents, or products.

(d) Suppose the license plate contains only numbers, and each number on the plate appears only once. What is the probability that the numbers appear in order (lowest number first, second lowest number second, etc)? [3]

$$P(\text{numbers in order}|\text{only numbers with no repeats}) = \frac{C_7^{10}}{P_7^{10}} = \frac{10!/7!3!}{10!/3!} = \frac{1}{7!}$$

Question 3 (15 points)

A California license plate has seven characters, which may be either letters or digits. Any character is equally likely to be any of the 26 letters or 10 digits. Answer the following questions, without evaluating factorials, expressions with exponents, or products.

(e) What is the probability that no letter or digit appears more than once on the license plate? [3]

$$P(\text{no letter or number appears more than once}) = \frac{\#\text{ of sequences with no repetitions}}{\#\text{ of sequences total}} = \frac{36!/29!}{36^7}$$