

Stat 104 Fall 2016: Exam 1 Solutions

True/ False#1

If X is a random variable, $E(X)$ must be less than 1.

a) True

b) False

$E(X)$ can take on any value

True/False #2

For any two random variables X and Y , if $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$, then the correlation between X and Y must be zero.

a) True

b) False

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(XY)$$

True/False #3

If X and Y are independent random variables, then the standard deviation of X+Y is the standard deviation of X plus the standard deviation of Y.

a) True

b) False

If two random variables are independent:

$$SD(X+Y) = \sqrt{Var(X) + Var(Y)}$$

True/False #4

If E and F are independent events, the $P(E)$ cannot be equal to $P(F)$

a) True

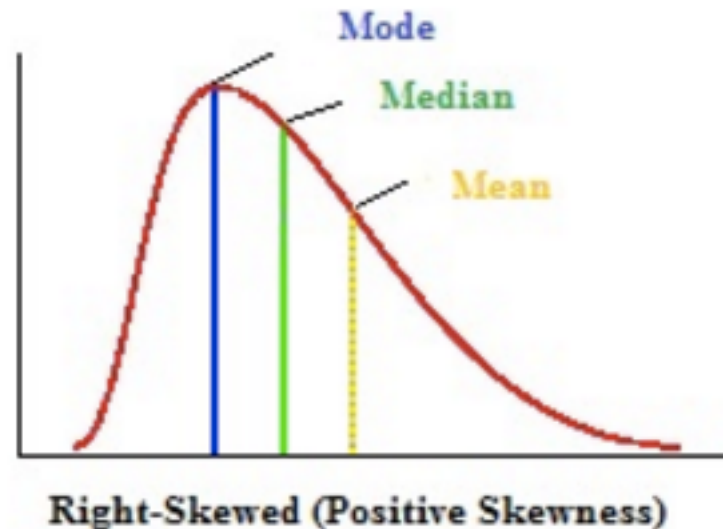
b) False

True/False #5

We know that the distribution of wealth in the United States is right skewed. Thus, the majority of people in the US have more than the average (i.e. mean) level of wealth.

a) True

b) False



True/ False #6

The statement “Of those surveyed, 59% said they favored the death penalty” is an example of the use of descriptive statistics.

a) True

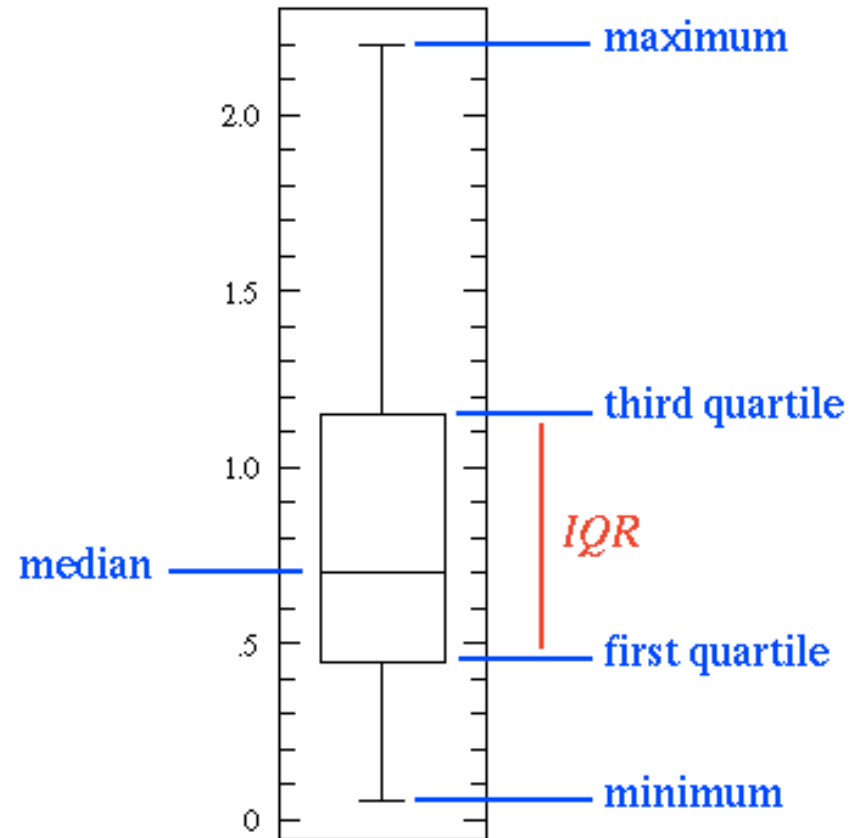
b) False

True/False #7

In a set of numerical data, the value of Q3 can never be smaller than the value of Q1.

a) True

b) False



True/False #8

In a sample size of 40, the sample mean is 15. In this case the sum of all observations in the sample is 600.

a) True

b) False

$$\text{Mean} = \frac{\text{Sum of all observations}}{\text{Sample Size}} = \frac{600}{40} = 15$$

Multiple Choice #9

The standard deviation of a discrete random variable σ is the positive square root of its:

a) Variance

b) Mean

c) Range

d) IQR

e) Mode

$$SD = \sqrt{Variance}$$

Multiple Choice #10

Jim, Cathy, Alex, and Sue are thinking of going to a party. Jim has a 20% probability of going; for Cathy it's more like 70%. The probability of Alex going is 90%, and the probability of Sue going is 60%. They do not consult each other beforehand, so they are not affected by the other people's decisions of whether or not to go. What is the probability that they are all at the party?

- a) 0.188
- b) 0.091
- c) 0.372
- d) 0.065
- e) 0.076

Since all events are independent, we can multiple all together to get the probability of the 4 of them going to the party:

$$P(\text{All}) = 0.2 * 0.7 * 0.9 * 0.6 = 0.076$$

Multiple Choice #11

A construction company is bidding on the chance to do two different projects: a movie theater and a skating rink. The probability of getting the bid to do the movie theater is 20%. If the probability of getting the bid to do both is 1%, what should the probability of getting the bid to do the skating rink be for the two events to be independent?

a) 0.02

b) 0.03

c) 0.04

d) 0.05

e) 0.06

$$P(\text{Movie theater})=0.2$$

$$P(\text{Movie theater and Skating Rink})=0.01$$

$$\text{To be independent, } P(M)P(S)=P(M \text{ and } S)$$

$$0.2 * P(S)=0.01$$

$$P(S)= 0.05$$

Multiple Choice #12

Suppose X and Y are independent random variables, and suppose X is a binomial with $n=10$, and $p=0.4$; while Y is binomial with $n=12$ and $p=0.2$. Find the variance of $2X + 3Y$.

a) 26.88

b) 44.86

c) 32.51

d) 20.41

e) 13.95

$$\text{Var}(X) = 10 * 0.4 * (1 - 0.4) = 2.4$$

$$\text{Var}(Y) = 12 * 0.2 * (1 - 0.2) = 1.92$$

$$\text{Var}(2X + 3Y) = 2^2 * 2.4 + 3^2 * 1.92 = 26.88$$

Multiple Choice #13

The average amount earned by 110 juniors for a week was \$35, while during the same week 90 seniors averaged \$50. What were the average earnings for that week for the combined group?

a) \$41.75

b) \$42.50

c) \$43.50

d) \$47.55

$$E(X) = (110/200) * \$35 + (90/200) * \$50 = \$41.75$$

Multiple Choice #14

Find the expected value of the following random variable:

X=	0	2	3	5
probability	0.5	0.1	0.3	0.1

a) 2.7

b) 1.2

c) 1.6

d) 2.1

e) 1.8

$$E(X) = (0 \cdot 0.5) + (2 \cdot 0.1) + (3 \cdot 0.3) + (5 \cdot 0.1) = 1.6$$

Multiple Choice #15

The annual cost of owning a dog is a random variable with mean \$695 and standard deviation \$45. The annual cost of owning a cat is a normal random variable with mean \$705 and standard deviation \$35. What is the probability that the total annual cost of owning one dog and two cats exceeds \$2000? [assume dog and cat ownership is independent].

a) 0.6466

b) 0.7180

c) 0.8251

d) 0.8965

e) 0.9322

$$D \sim N(\$695, 45^2)$$

$$C \sim N(\$705, 35^2)$$

$$D + 2C \sim (695 + 2(705), 45^2 + 2^2 35^2) = (2105, 6925)$$

$$P(X > 2000):$$

$$P(Z > (2000 - 2105) / \sqrt{6925}) = P(Z > -1.26) = 1 - 0.1038 = 0.896$$

Multiple Choice #16

At a local community college, 90% of students take English, 80% of those who don't take English take art courses, while only 50% of those who do take English take art. What is the probability that someone takes art?

a) 0.45

b) 0.90

c) 0.53

d) 0.10

e) 0.08

$$P(E)=0.9$$

$$P(A|E')=0.8$$

$$P(A|E)=0.5$$

$$P(A|E)P(E) + P(A|E')P(E') = (0.5*0.9) + (0.8*0.1) = 0.53$$

Multiple Choice # 17

In a certain game of chance, your chances of winning are 0.2. If you play the game five times and outcomes are independent, the probability that you win at least once is

a) 0.74

b) 0.33

c) 0.67

d) 0.92

e) 0.46

Binomial, $p=0.2$ and $n=5$

$$1 - P(X=0) = 1 - (1-p)^n = 1 - 0.8^5 = 0.67$$

Multiple Choice # 18

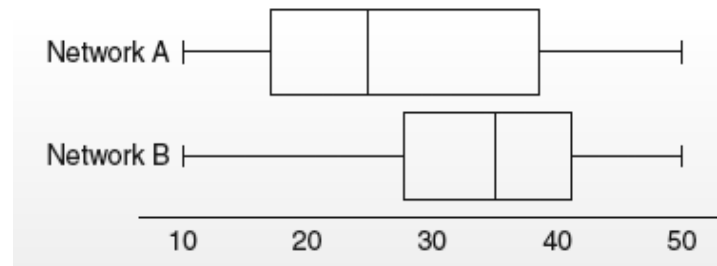
Which of the following statements is (are) correct?

- I. The area under a probability density curve for a continuous random variable is 1.
- II. A random variable is a numerical outcome of a random event.
- III. The sum of the probabilities for a discrete random variable is 1.

- a) I only
- b) II only
- c) I and III
- d) II and III
- e) I, II and III

Multiple Choice # 19

An advertiser is trying to decide which television station in town to use for his product. He gathers the ratings of all prime time shows on each network and constructs a boxplot of each. There are the same number of ratings for each network. The results are as follows. Based on these box plots, which of the following is a correct conclusion about the relative ratings for the two networks?



- a) The median rating for Network A is greater than for Network B.
- b) The range for Network A is greater than for Network B
- c) The IQR for the two networks is the same
- d) There are more ratings greater than 28 for Network A than for Network B
- e) The median rating for Network B is higher than for Network A.

Multiple Choice # 20

The length of songs played by a certain radio station is normally distributed with a mean of 210 seconds and a standard deviation of 25 seconds. If Angela turns on the radio just in time to catch the beginning of a new song, what is the probability the song will last between 170 and 225 seconds?

a) 0.05

b) 0.25

c) 0.67

d) 0.73

e) 0.90

$$L \sim N(210, 25^2)$$

$$P(170 < L < 225)$$

$$P((170-210)/25 < Z < (225-210)/25) = P(-1.6 < Z < 0.6) \\ = 0.7257 - 0.0548 = 0.67$$

Multiple Choice # 21

A factory utilizes two different production lines (A and B) for manufacturing gizmos. Since production line A is faster, it is able to produce 60% of the gizmos, with the remaining gizmos coming off production line B. It is known that 2% of the gizmos from production line A and 4% of the gizmos from production line B will contain some small defect. Based upon this, if a gizmo is randomly selected from among all gizmos produced at this factory and the gizmo is found to have a defect, what is the probability the gizmo came off production line B?

a) 0.667

b) 0.5

c) 0.571

d) 0.429

e) 0.016

$$P(A) = 0.6$$

$$P(B) = 0.4$$

$$P(D | A) = 0.02$$

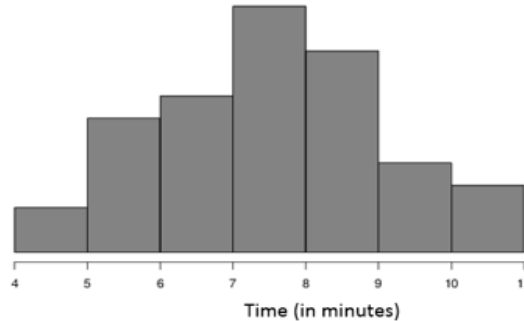
$$P(D | B) = 0.04$$

	P(D)	P(D')	
P(A)	=0.02*0.6 =0.012	=0.6-0.0012 =0.588	0.6
P(B)	=0.04*0.4 =0.016	=0.4-0.016 =0.384	0.4

$$P(B | D) = 0.016 / (0.012 + 0.016) = 0.57$$

Multiple Choice # 22

The amount of time it took students to run a mile in gym class was recorded. The histogram below shows the distribution of time (in minutes) for these students.



Which of the following is closest to the standard deviation of the one mile times?

a) 1

b) 2

c) 3

d) 4

e) 6

Multiple Choice # 23

On an EC10 midterm, the mean was 77, with a standard deviation of 10. If the professor adds 5 points to each score, what will be the new mean and new standard deviation?

- a) Mean 77, standard deviation 10
- b) Mean 77, standard deviation 15
- c) Mean 82, standard deviation 10
- d) Mean 82, standard deviation 13

New Mean = $77 + 5 = 82$
New SD = 10

Multiple Choice # 24

A local real estate agent is interested in the relationship between the cost of a house (y) in dollars and its area (x) in square feet. She randomly selects 50 houses around town and computes the least squares regression line: $\hat{y} = 14578.01 + 136.89x$. Which of the following is the correct interpretation of the slope of the least squares regression line?

- a) For each dollar increase, the estimated square footage of the house increases by 136.89.
- b) For each dollar increase, the estimated square footage of the house increases by 14578.01.
- c) For each additional square foot, the estimated cost of the house increases by \$136.89.
- d) For each additional square foot, the estimated cost of the house increases by \$14578.01.
- e) Since we are not given the data points, it is impossible to interpret the slope of the least squares regression line in this problem

Multiple Choice # 25

The U.S. Census Bureau reports that American spend an average of $\mu=24.3$ minutes commuting to work each day. Assuming that the distribution of commuting times is normal with a standard deviation of $\sigma=10$ minutes, how much time do you have to spend commuting each day to be in the highest 30% nationwide?

a) 29.5

b) 44.4

c) 26.6

d) 37.1

e) 41.8

Reverse look-up:

Z score for 30% in upper tail = 0.52

$$0.52 = \frac{X - 24.3}{10}$$

$$X = 29.5$$

Multiple Choice # 26

Which of the following statements are true for any continuous random variable

- a) The maximum value of $F(x)$ is 1
- b) The maximum value of $f(x)$ is 1
- c) Both a) and b)
- d) None of these

Multiple Choice # 27

Suppose $X \sim \text{Unif}(0,9)$. Find $E(X^2)$

a) 9

b) 3

c) 4.5

d) 27

e) 21

$$E(X^2) - [E(X)]^2 = \text{Var}(X)$$

$$E(X) = (0+9)/2 = 4.5$$

$$\text{Var}(X) = (9-0)^2/12 = 6.75$$

$$E(X^2) - 4.5^2 = 6.75$$

$$E(X^2) = 27$$

Multiple Choice # 28

Which of the following is true about any discrete random variable X ?

- a) The expected value of $X=np$.
- b) The sum of all possible values of X is equal to 1.
- c) The probabilities of all possible values of X must add to 1.
- d) The probability distribution is bell-shaped and symmetrical
- e) Approximately 95 percent of the values of X fall within 2 standard deviations of the mean.

Multiple Choice # 29

A random sample of records of electricity usage of homes in the month of July gives amount of electricity used and size (in square feet) of 135 homes. A regression was done to predict the amount of electricity used (in kilowatt-hours) from size. The resulting model is: $\text{usage} = 1204 + 0.6 \text{ size}$. How much electricity would you predict would be used in a house that is 2273 square feet?

a) 2567.8 kilowatt-hours

b) 3477.6 kilowatt-hours

c) 159.8 kilowatt-hours

d) 1363.8 kilowatt-hours

e) 1781.67 kilowatt-hours

$$1204 + 0.6(2273) = 2567.8$$

Multiple Choice # 30

Consider the same regression model as described in the problem above. What would a negative residual imply for people living in a house that is 2284 square feet?

$$\text{Residual} = y_i - \hat{y}$$

- a) Their house is bigger than expected.
- b) They are using the least amount of electricity of all of the houses sampled.
- c) Their house is smaller than expected.
- d) They are using less electricity than expected.
- e) They are using more electricity than expected.

Multiple Choice # 31

A study was done that measured the association between people's vertical jump ability (how high they can jump from a stationary position) and how much they can leg press. The resulting correlation was 0.98. Which of the following statements are true?

- i) If you leg press more than someone else, then you will have a higher vertical jump than him or her.
 - ii) If you leg press less than someone else, you will have a lower vertical jump than him or her.
 - iii) There is a very strong positive linear association between leg press ability and vertical jump.
- a) i only
 - b) ii only
 - c) iii only
 - d) i and ii only
 - e) i, ii and iii

Multiple Choice # 32

Based on the following output, what is the correlation between variables x and y?

a) -0.44

b) -0.63

c) -0.32

d) -0.58

e) -0.14

```
. summ x y
```

Variable	Obs	Mean	Std. Dev.	Min	Max
x	74	21.2973	5.785503	12	41
y	74	13.75676	4.277404	5	23

```
. corr x y, cov  
(obs=74)
```

	x	y
x	33.472	
y	-14.3924	18.2962

$$r_{xy} = \frac{Cov_{xy}}{s_x s_y} = \frac{-14.3924}{5.79 * 4.28} = -0.58$$

Multiple Choice # 33

Consider the following joint probability table:

		Y			
		0	5	10	15
X	0	0.02	0.06	0.02	0.1
	5	0.04	0.15	0.2	0.1
	10	0.01	0.15	0.14	0.01

Find $P(X=5|Y=5)$

a) 0.5

b) 0.42

c) 0.22

d) 0.08

e) 0.86

$$P(X=5|Y=5) = \frac{P(X=5 \text{ and } Y=5)}{P(Y=5)} = \frac{0.15}{(0.06+0.15+0.15)} = 0.42$$