2019-11-04

Midterm Policies

You should have the following items on your desk (and nothing else) during the midterm:

- Pen and/or pencil and/or eraser - make sure you have enough writing utensils and they are functional. You should not try to open your bag to get another pen. Have everything ready prior to the midterm.
- Scientific calculator.
- UCI Student ID card - we will come around to check your ID during the midterm.

The following items are **not** permitted during the midterm

- Scrap paper
- Watches - remaining time will be displayed on the screen.
- Cell phones - turn your cell phone off completely and put it in your bag not in your pocket.
- Laptops / Ipads/Tablets
- Graphing calculator
- Notes of any kind, books etc.
- Any note taking device or any device with internet connectivity

Note that if you use an eraser (for pen or pencil) and do not erase well, you may not receive any points if previous answer and the latest answer are visible simultaneously.

Multiple choice questions have one correct answer. Only choose one of the choices. Write your answers in the boxes provided. You should only write a single letter in each of these boxes.

For questions that are **not** multiple choice, show your work. Write your final answers in the boxes provided. You should only write numbers in these boxes. For instance if your answer is $P(A) = 0.12$ only write 0.12 inside the box.

You will have 75 minutes. You may not leave the classroom if you finish early. Use the remaining time to review your answers.

You may not start the midterm until you are prompted to start. Just write your name and student ID number and wait for the prompt to start.

Name (as it appears on Canvas): **Answer key**

Student ID number:

Copyright © Dr. Mine Dogucu. All Rights Reserved.

The `avocado` data frame has information on avocado sales in 2017 as reported by the Hass Avocado Board. Answer the following three questions based on the output below.

```r
glimpse(avocado)
```

```
## Observations: 5,722
## Variables: 3
## $ Date          <chr> "12/31/17", "12/24/17", "12/17/17", "12/10/17", ...
## $ average_price <chr> "1.47", "1.45", "1.43", "1.29", "1.39", "1.5", "...
## $ type          <chr> "conventional", "conventional", "conventional", ...
```

```r
count(avocado, type)
```

```
## # A tibble: 2 x 2
##   type             n
##   <chr>        <int>
## 1 conventional  2862
## 2 organic       2860
```

1. The variable `average-price` shows the average sale price of certain type of avocado on a given day. R has read `average-price` as a character variable. If you were to mutate `average_price` as a different type of variable that would reflect its statistical characteristics, in other words what type of variable would you mutate to?

(A) character
(B) double
(C) integer
(D) factor

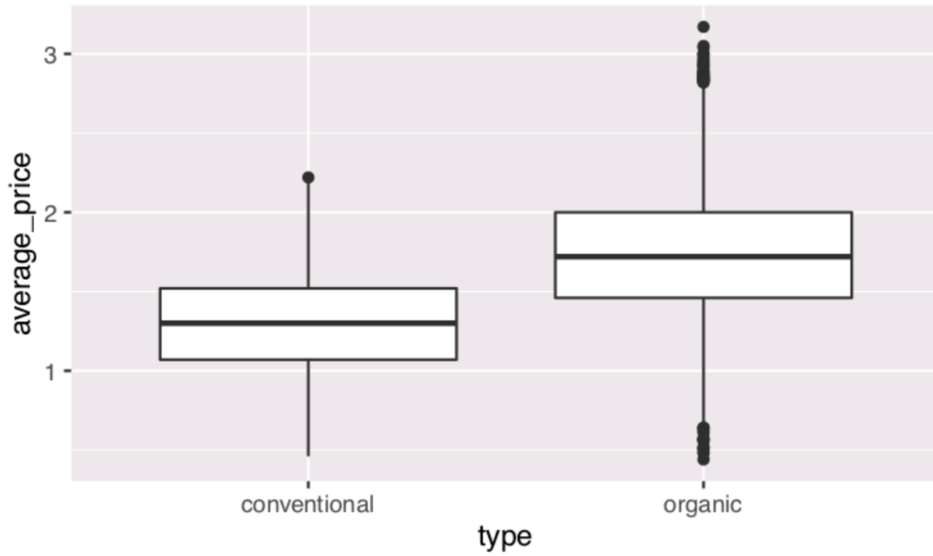
B

2. If you were to mutate `type` as a different type of variable that would reflect its statistical characteristics, in other words what type of variable would you mutate to?

(A) character
(B) double
(C) integer
(D) factor

D

Copyright © Dr. Mine Dogucu. All Rights Reserved.

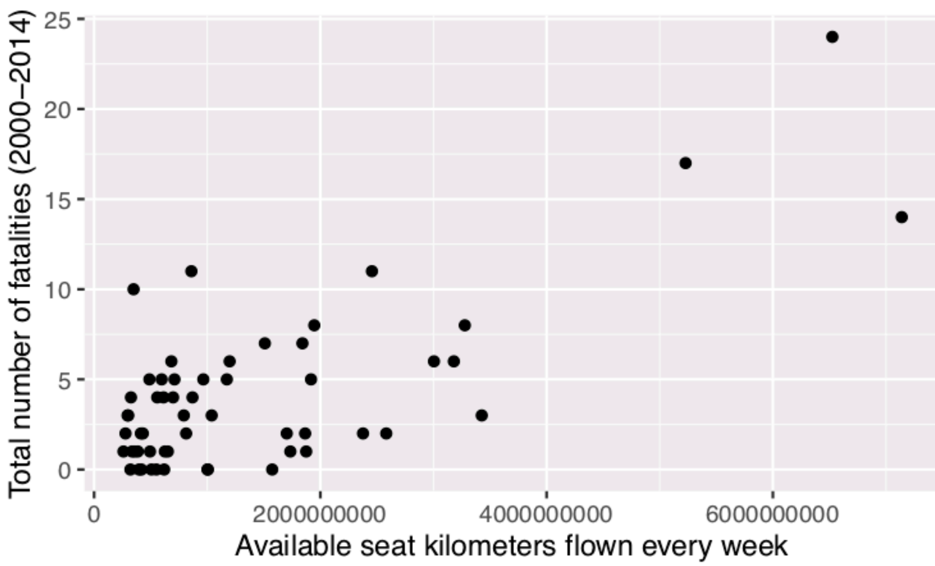


3. Which of the following statements is true based on the boxplot above?

(A) The minimum average price of conventional avocados is greater than the first quartile of organic avocados.
(B) The median of organic avocados is less than the third quartile of conventional avocados.
(C) Interquartile range of organic avocados is greater than interquartile range of conventional avocados.
(D) The maximum of conventional avocados is less than the third quartile of organic avocados.

C

Copyright © Dr. Mine Dogucu. All Rights Reserved.



4. Below is a scatterplot showing the relationship between the available seat kilometers flown every week by an airline, and the total number of fatalities an airline experienced between 2000 and 2014. Each point is a different airline. Which of the following is false? Write your choice (a, b, c, or d) in the box. Only one correct answer.

(A) Total number of fatalities is a discrete numerical variable.
(B) Total number of fatalities is the explanatory variable.
(C) There were airlines that did not have any fatality.
(D) There are three airlines that have more than 4000000000 available seat kilometers every flown every week.

B

5. Police expects 15 cars to pass through a particular intersection every hour. Which of the following calculates the probability that exactly 17 cars pass through that intersection in an hour?

(A) $\binom{17}{15}\frac{15}{17}^2(1-\frac{15}{17})^{15}$

(B) $\frac{15}{17}^2\frac{15}{17}^{15}$

(C) $\frac{15^{17}}{17!}e^-15$

(D) $\frac{15}{17}$

C

Copyright © Dr. Mine Dogucu. All Rights Reserved.

6. In 2009 there was a large outbreak in the U.S. of a flu strain clinically known as H1N1, but commonly known as the Swine Flu since the strain was similar to one found in pigs. That year, 20% of the population contracted Swine Flu. A test was developed at the time to diagnose whether a patient had the disease or not. If we knew a patient had Swine Flu, there was a 98% probability that they would get a positive test result. If we knew a patient did not have Swine Flu, there was a 96% probability that they would get a negative test result.

a. What is the probability that a patient did not have Swine Flu and got a positive test result?

P(SF - and test +) = P(SF -)*P(test +|SF -)
=0.8(0.04)
=0.032

$\boxed{0.032}$

b. What is the probability, in general, of getting a negative test result?

P(test -) = P(SF - and test -) + P(SF + and test -)
=0.8(0.96) + 0.2(0.02)
=0.768 + 0.004
= 0.772

$\boxed{0.772}$

c. What is the probability that a patient does not have Swine Flu, given they got a negative test result?

P(SF - | test -) = P(SF - and test -) / P(test -)
=[0.8(0.96)] / 0.772
=0.768 / 0.772
= 0.995

$\boxed{0.995}$

Copyright © Dr. Mine Dogucu. All Rights Reserved.

7. Everyone loves an orange kitty, but did you know that female orange tabby cats are rare? Studies have shown that only 2% of orange tabby cats are female. We'd like to see if this holds up in our area, so we find 14 orange tabby cats.

a. What is our random variable? What distribution does it follow?

X = number of female orange tabby cats
X ~ Binomial(n=14, p=0.02)

b. What is the probability that at least one but at most three of the cats we find are female?

P(1 <= X <= 3) = P(X=1) + P(X=2) + P(X=3)
= choose(14,1)(0.02)^1(0.98)^13
+ choose(14,2)(0.02)^2(0.98)^12
+ choose(14,3)(0.02)^3(0.98)^11
=0.246

0.246

c. What is the probability that at most two cats are female?

P(X<= 2) = P(X=0) + P(X=1) + P(X=2)
= choose(14,0)(0.02)^0(0.98)^14
+ choose(14,1)(0.02)^1(0.98)^13
+ choose(14,2)(0.02)^2(0.98)^12
=0.997

0.997

d. Now assume that we are going to keep looking for orange tabby cats until we find one that's female. How many cats do we expect we'll have to find to get our first female?

1/0.02 = 2/100 = 50 cats

50

Copyright © Dr. Mine Dogucu. All Rights Reserved.

8. There is a game where you can pick a rubber duck out of a pool and win the prize at the bottom of the duck. There are 30 ducks in the pool. 16 of the ducks have no prize, 6 have a prize of \$1, 4 have a prize of \$2, 3 have \$5 dollar prize, and 1 has a \$10 prize.

a. Assume the game is free, how much can you expect to win from the game?

x	f(x)
0	16/30
1	6/30
2	4/30
5	3/30
10	1/30

E(X) = 0(16/30) + 1(6/30) + 2 (4/30) + 5(3/30) + 10(1/30)
= 1.30

$\boxed{\$1.30}$

b. Now the game costs \$1.5, what is the expected winnings and variance of this game?

y	f(y)
-1.5	16/30
-0.5	6/30
0.5	4/30
3.5	3/30
8.5	1/30

Y = X-1.5
E(Y) = 1.3-1.50 = -0.2

E(Y^2) = (-1.5)^2(16/30) + (-0.5)^2(6/30)
+ (0.5)^2 (4/30) + (3.5)^2(3/30)
+ (8.5)^2(1/30)
= 4.917

Var(Y) = E(Y^2) - [E(Y)]^2
= 4.917 - [-0.2]^2
= 4.88

-\$0.20
4.88

c. Now assume that the first game is free and each additional game is \$1.5 to play. How many games does it take until you expect to lose money?

E(X + aY) < 0
E(X) + E(aY) < 0
E(X) + aE(Y) < 0
1.3 + a(-0.2) < 0
a(-0.2) < -1.3
a > -1.3 / -0.2
a > 6.5

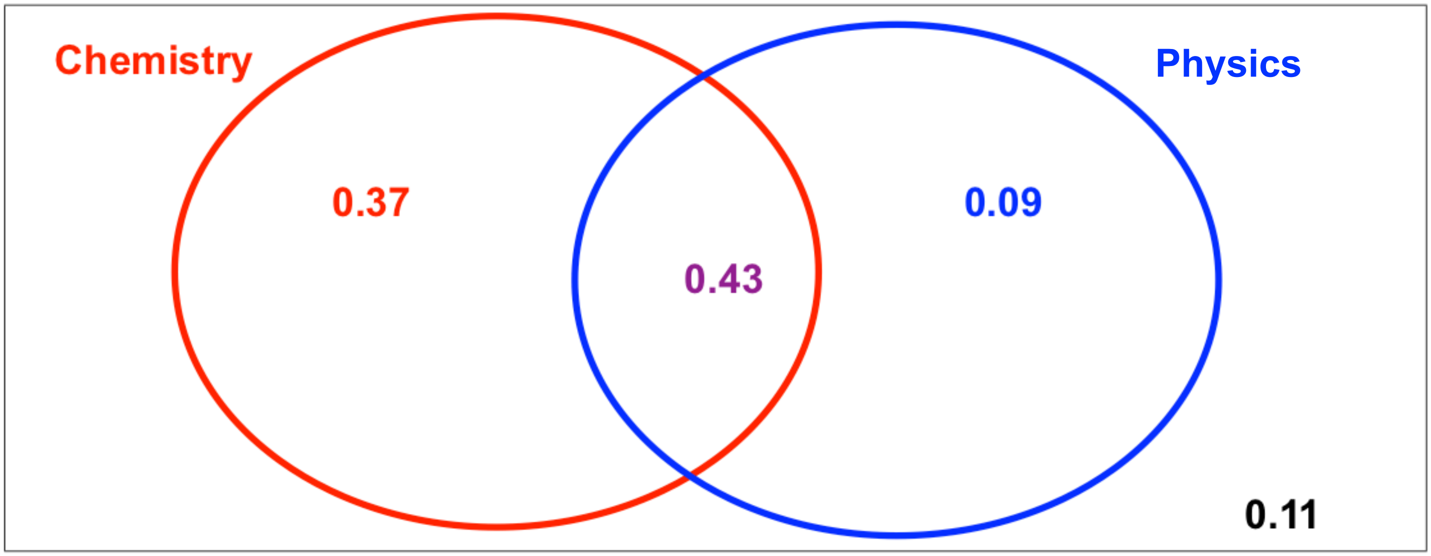
Will need to play 1 free game and 7 paid games (8 games total) to lose money

$\boxed{8}$

Copyright © Dr. Mine Dogucu. All Rights Reserved.

9. Imagine you are a student who is taking both intro chemistry and intro physics in the same quarter. Due to an unfortunate combination of schedules, you have an experiment for each class due on Monday. Because you have to split your time between the two experiments, and you don't have time to re-do either, there is a 80% chance that you will perform your chemistry experiment correctly, a 52% chance you will perform you physics experiment correctly, and a 43% chance that you will perform both your physics experiment and your chemistry experiment correctly.

a. Draw a Venn diagram of this scenario.



b. What is the probability that you perform your physics experiment or your chemistry experiment correctly?

P(P U C) = P(P) + P(C) - P(P and C)
= 0.52 + 0.8 - 0.43
= 0.89

$\boxed{0.89}$

c. What is the probability that you will perform your physics experiment correctly, but not your chemistry experiment?

P(P and C') = P(P) - P(P and C)
= 0.52 - 0.43
= 0.0.09

$\boxed{0.09}$

Copyright © Dr. Mine Dogucu. All Rights Reserved.

d. Are performing your chemistry experiment correctly and performing your physics experiment correctly mutually exclusive? Why? Give mathematical reasoning.

They are not mutually exclusive because P(P and C) = 0.043 != 0

e. Is performing your chemistry experiment correctly independent of you performing your physics experiment correctly? Why? Give mathematical reasoning.

P(P)*P(C) = 0.52(0.8) = 0.416 != 0.43

Thus, they are not independent

Copyright © Dr. Mine Dogucu. All Rights Reserved.

10. A research assistant has a job that requires them to count fireflies. Based on their experience, in a similar climate they had observed an average of 3 fireflies every five minutes. What is the probability that they will observe more than 8 fruitflies in the next 5 minutes?

(A) `ppois(q = 8, lambda = 3)`

(B) `dpois(x = 8, lambda = 3)`

(C) `1 - ppois(q = 8, lambda = 3)`

(D) `1 - dpois(x = 8, lambda = 3)`

C

Copyright © Dr. Mine Dogucu. All Rights Reserved.

Deferred Action for Childhood Arrival (DACA) is an immigration policy of the United States that provides a renewable two-year period of deferred action from deportation and work permit to individuals who were brought to the United States as children and who do not have a lawful presence. Note that DACA policy does not grant citizenship but grants two-year residency and work permit. On September 5, 2017 President Trump made a decision to end the DACA program. University of California was the first university in the nation to file suit in federal court against this decision. The oral arguments of this case will begin on November 12th in the US Supreme Court.

For the following five questions we will be working with public opinion on DACA policy and the US-Mexico border wall. In June 2018 Gallup, an internationally well-known American analytics company, indicated that

> 41% favor expanding construction of walls along U.S.-Mexico border
>
> 83% approve of allowing DACA immigrants to become citizens
>
> Republicans, Democrats agree on DACA, disagree on walls

Gallup also published some information about their survey methods as stated below:

> Results for this Gallup poll are based on telephone interviews conducted June 1-13, 2018, with a random sample of 1,520 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia.
>
> Each sample of national adults includes a minimum quota of 70% cellphone respondents and 30% landline respondents, with additional minimum quotas by time zone within region. Landline and cellular telephone numbers are selected using random-digit-dial methods.

11. Which of the following indicates the sample of the study the best?

(A) All residents of the United States aged 18 and older

(B) All citizens of the United States aged 18 and older

(C) All respondents of the survey

(D) All DACA recipients

C

12. Which of the following indicates the population of the study the best?

(A) All residents of the United States aged 18 and older

(B) All citizens of the United States aged 18 and older

(C) All respondents of the survey

(D) All DACA recipients

A

Copyright © Dr. Mine Dogucu. All Rights Reserved.

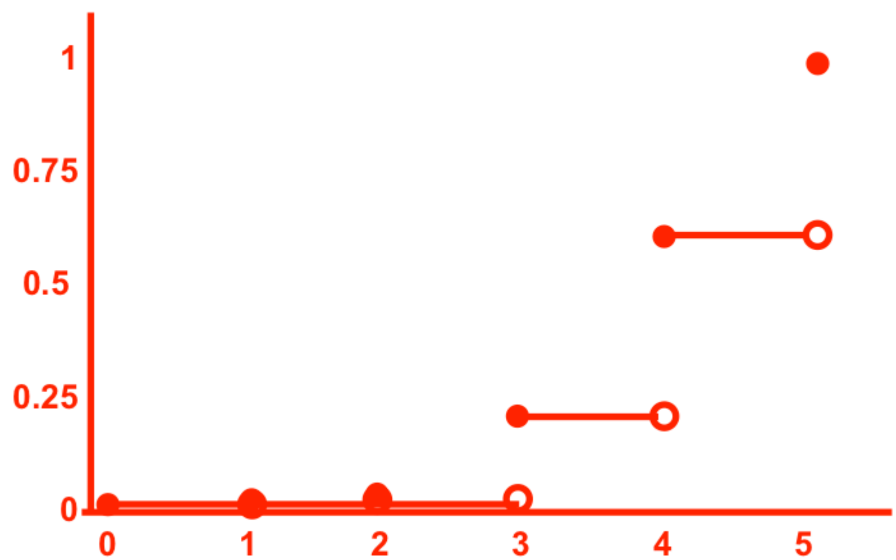
13. Assume that people either favor or oppose the expanding construction of walls along US - Mexico border. You randomly select five people from the population you defined above. What is the probability that none of them will favor the expanding construction of walls along US - Mexico border.

P(X=0) = choose(5,0) 0.41^0(0.59)^5
= 0.072

$\boxed{0.072}$

14. You randomly select five people from the population you defined above. Let X be the random variable that represents the number of people (out of 5) who approve of allowing DACA immigrants to become citizens. Plot the cdf of X.

x	f(x)	CDF
0	0.0001	0.001
1	0.003	0.0031
2	0.034	0.037
3	0.165	0.202
4	0.403	0.605
5	0.394	1



15. Assume that there are only two political affiliations (Democractic and Republican). Is political affiliation and agreement on DACA independent or not? Why or why not?

Because Republicans and Democrats both agree on DACA citizenship, political affiliation and agreement on DACA are independent since belonging to one party or another will not sway your opinion on this issue.

Copyright © Dr. Mine Dogucu. All Rights Reserved.

16. Let X be modeled by the function $f(x) = \frac{2}{x^3}$ for $x \geq 1$.

a. Show that this is a valid probability density function (pdf).

1) Show f(x) integrates to 1:

$$\int_1^\infty 2x^{-3}dx = [-x^{-2}|_1^\infty$$
$$= -\infty^{-2} + 1^{-2}$$
$$= 0 + 1$$
$$= 1$$

2) Show f(x) >= 0:

$$P(1 < X < 2) = \int_1^2 2x^{-3}dx$$
$$= [-x^{-2}|_1^2$$
$$= 0.75$$

$$P(10 < X < 20) = \int_{10}^{20} 2x^{-3}dx$$
$$= [-x^{-2}|_{10}^{20}$$
$$= 0.0075$$

$$P(100 < X < 200) = \int_{100}^{200} 2x^{-3}dx$$
$$= [-x^{-2}|_{100}^{200}$$
$$= 0.000075$$

b. What is the expected value of this distribution?

$$E(X) = \int_1^\infty 2XX^{-3}dx$$
$$= \int_1^\infty 2X^{-2}dx$$
$$= -2[x^{-1}|_1^\infty$$
$$= -2\left[\infty^{-1} - 1^{-1}\right]$$
$$= -2[0 - 1]$$
$$= 2$$

2

c. What is the variance of this distribution?

$$E(X^2) = \int_1^\infty 2x^2x^{-3}dx$$
$$= \int_1^\infty 2x^{-1}dx$$

DNE

The integral 1/x does not converge, so E(X^2) and the variance does not exist

Copyright © Dr. Mine Dogucu. All Rights Reserved.

d. What is the probability that X is at least 4?

$$P(X > 4) = \int_4^{\infty} 2X^{-3}dx$$
$$= [-x^{-2}|_4^{\infty}$$
$$= \left[-\infty^{-2} + 4^{-2}\right]$$
$$= 0.0625$$

$\boxed{0.0625}$

e. What is the probability that X is more than 3 but less than 10?

$$P(3 < X < 10) = \int_3^{10} 2X^{-3}dx$$
$$= [-x^{-2}|_3^{10}$$
$$= \left[-10^{-2} + 3^{-2}\right]$$
$$= 0.101$$

$\boxed{0.101}$

f. What is the probability that X is exactly 1?

$$P(X = 1) = \int_1^{1} 2X^{-3}dx$$
$$= [-x^{-2}|_1^{1}$$
$$= -1 + 1$$
$$= 0$$

$\boxed{0}$

Copyright © Dr. Mine Dogucu. All Rights Reserved.

$\bar{x} = \frac{\Sigma_{i=1}^{n} x_i}{n}$ $s^2 = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})}{n-1}$

$P(A^c) = 1 - P(A)$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$

Discrete Random Variables

Let X be a discrete random variable with a pmf of $p(x)$ then

$E[X] = \sum_x xp(x)$

Let X be a random variable, then

$Var(X) = E[X^2] - E[X]^2$

Linear combination of random variables X and Y and fixed numbers a and b:

$E[aX + bY] = aE[X] + bE[Y]$
$Var(aX + bY) = a^2Var(X) + b^2Var(Y)$

Distribution	**pmf**	$E(X)$	$Var(X)$
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Geometric	$(1-p)^{n-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Discrete Uniform	$\frac{1}{n}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
Poisson	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ

Continuous Random Variables

$E[X] = \int_{x \in \Omega_x} x f(x)\, dx$

Copyright © Dr. Mine Dogucu. All Rights Reserved.