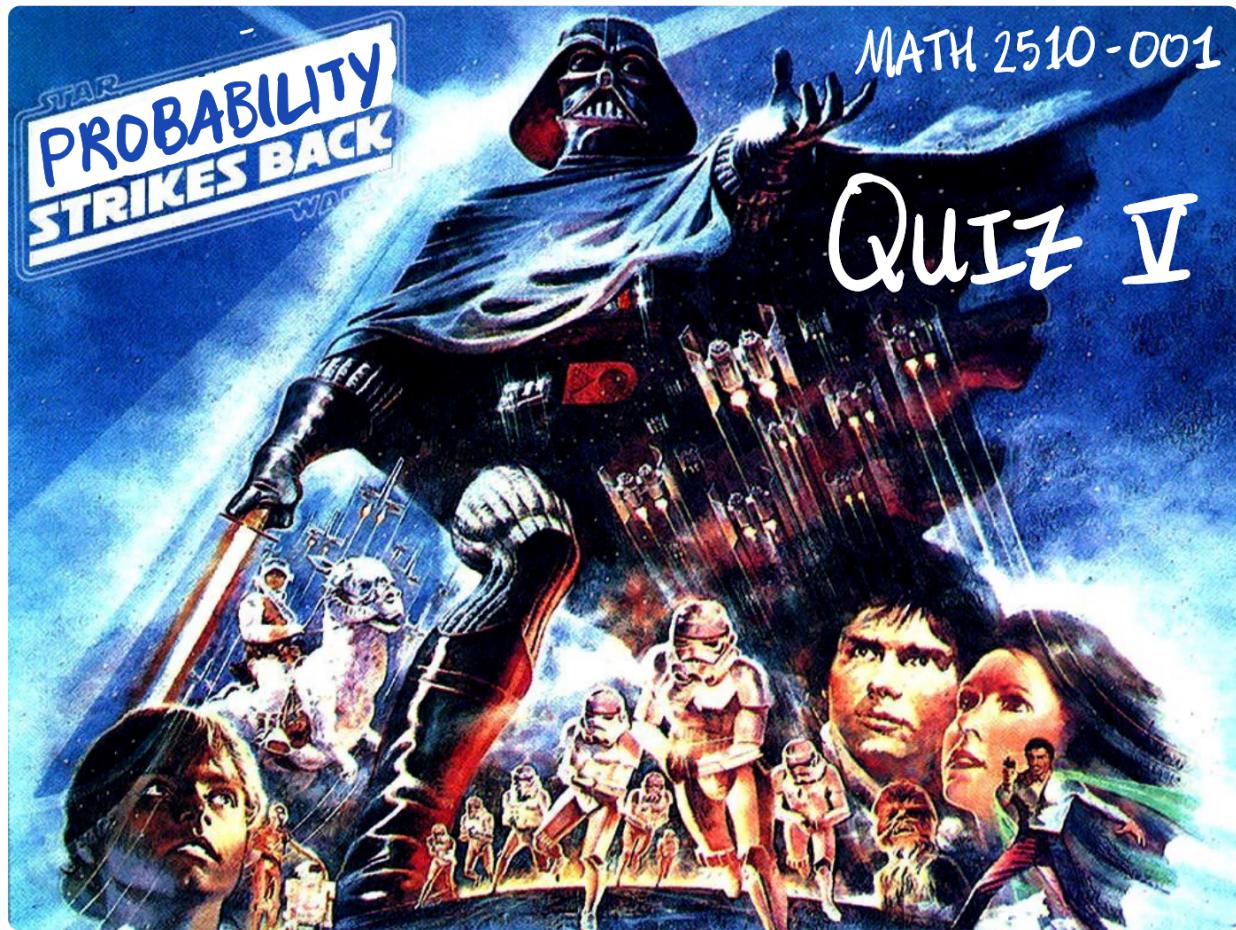


Your name (print clearly in capital letters): _____



Ground Rules

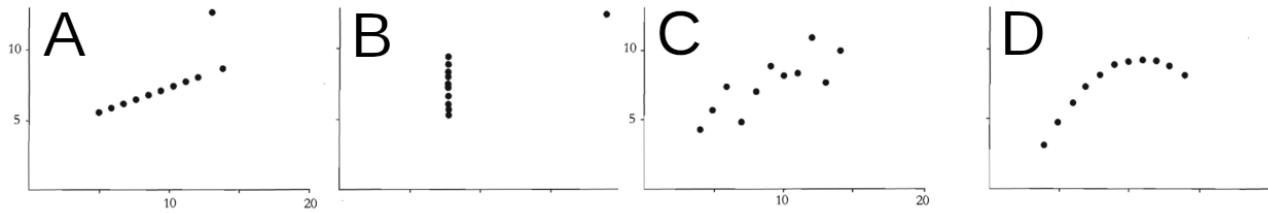
- This is an **graded** quiz. (Recall that quizzes make up about 1/10 of your final grade.)
- It is meant to give you some **low-stakes practice** for the upcoming midterm, which will be in-class on Wednesday, October 16th. (The first midterm is worth about 3/20 of your final grade.)
- I will grade the first **five out of six sections**, unless you explicitly tell me which section you'd like me **not to grade**.
- Which section do you want Colton **not to grade**? (Please circle one.) 1 2 3 4 5 6.
- You may not collaborate.
- You may use a statistical calculator if you wish, but each problem has been written so that it can be reasonable done with pencil and paper if necessary.
- You have until 8:55am, and you may leave class once you have finished. Have a great weekend!

1. ANSCOMBE'S QUARTET

Graphics *reveal* data. Indeed graphics can be more precise and revealing than conventional statical computations. Consider Anscombe's quartet: all four of these data sets are described by exactly the same linear model (at least until residuals are examined).

I	II	III	IV				
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

$N = 11$
 mean of X's = 9.0
 mean of Y's = 7.5
 equation of regression line: $Y = 3 + 0.5X$
 standard error of estimate of slope = 0.118
 $t = 4.24$
 sum of squares $\sum (X - \bar{X})^2 = 110.0$
 regression sum of squares = 27.50
 residual sum of squares of Y = 13.75
 correlation coefficient = .82
 $r^2 = .67$



1. For each dataset of I, II, III, and IV (above) circle the letter of the graph A, B, C, or D corresponding to that data set. If no graph corresponds to a dataset, circle E.
2. Use *half open interval notation*¹ to roughly estimate the range of each dataset's X and Y values.
3. Circle whether (yes) or not (no) each dataset can be written as a *function of X* with output Y .

dataset	graph	range of X	range of Y	could be a function of X ?
I	A B C D E	[,)	[,)	yes no
II	A B C D E	[,)	[,)	yes no
III	A B C D E	[,)	[,)	yes no
IV	A B C D E	[,)	[,)	yes no

¹A half open interval is a set of real numbers $\{x \in \mathbf{R} : a \leq x < b\}$, with an closed endpoint at a and an open endpoint at b .

2. BASIC PROBABILITY COMPUTATIONS



Suppose A and B are events so that $P(A) = 0.40$ and $P(B) = 0.25$. What are the largest and smallest values that $P(A \cup B)$ could attain given only this information?²

- (A) $0 \leq P(A \cup B) \leq 0.4$
- (B) $0 \leq P(A \cup B) \leq 0.65$
- (C) $0.25 \leq P(A \cup B) \leq 0.65$
- (D) $0.25 \leq P(A \cup B) \leq 1$
- (E) $0.4 \leq P(A \cup B) \leq 0.65$

Your answer:

What additional information would you need in order to compute $P(A \cup B)$ exactly?

- (A) $P(A^c)$
- (B) $P(B^c)$
- (C) $P(\Omega)$
- (D) $P(A \cap B)$
- (E) $P(A)P(B)$

Your answer:

Suppose C and D are events so that $P(C) = 0.75$ and $P(D) = 0.60$. What are the largest and smallest values that $P(C \cup D)$ could attain given only this information?

- (C) $0 \leq P(C \cup D) \leq 0.6$
- (D) $0 \leq P(C \cup D) \leq 0.75$
- (E) $0.6 \leq P(C \cup D) \leq 0.75$
- (F) $0.6 \leq P(C \cup D) \leq 1$
- (G) $0.75 \leq P(C \cup D) \leq 1$

Your answer:

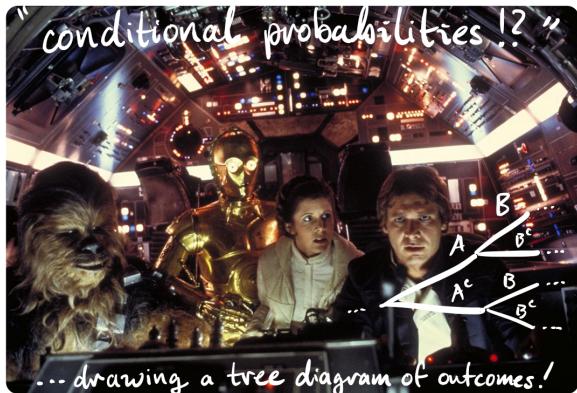
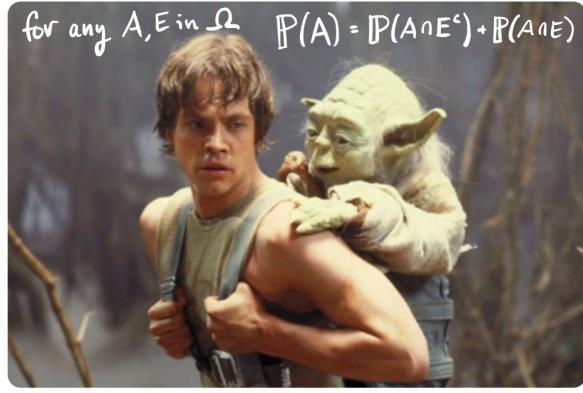
Would knowing $P(A^c \cap B^c)$ allow you to compute $P(A \cup B)$ exactly? (Choose the *best explanation*.)

- (A) No, because $P(A \cap B)$ remains unknown.
- (B) Yes, because $P(B^c)$ could be deduced.
- (C) No, because $P(A^c \cup B^c)$ remains unknown.
- (D) Yes, because $P(A^c \cap B^c) = 1 - P(A \cup B)$
- (E) No, because $P(A \cap B^c)$ remains unknown.

Your answer:

²In other words, choose the *highest possible bounds* for the value of the real number $P(A \cup B)$.

3. A TREE DIAGRAM FOR BAYES' RULE



Yoda is called to train Luke Skywalker. In his wisdom, Yoda knows that 90% of the time Luke is not trying hard enough (call this event E^c for “not enough”), and 10% of the time Luke is actually trying hard enough (call this event E for “enough”). You may assume E and E^c are complementary events.

Given that Luke is not trying hard enough, Yoda knows that Luke will fail to lift his X -wing (call this event X^c , for *no X-wing*) out of the swamp water in the Dagobah Cave³ with probability 0.95. On the other hand, given that Luke *is* trying hard enough, Yoda knows that Luke will fail to lift his X -wing out of the water with conditional probability 0.08.

Upon arriving in the Dagobah Cave, Yoda observes that Luke has failed to lift his X -wing out of the swamp. What is the probability that Luke was not trying hard enough?

Your answer (as a real number):

³In the movie *The Empire Strikes Back* (1980) Luke, travels with R2-D2 in his X -wing fighter to the swamp planet of Dagobah, where he crash-lands.

4. LIFE GOES ON: FURTHER TEDIOUS COMPUTATIONS

The value of Han Solo's portfolio (consisting mostly of smuggled goods and foreign investments) increases by 18% during a financial boom and by 9% during normal times. It decreases by 12% during a recession.

What is the expected return on this portfolio if each scenario is equally likely?

Your answer (as a percentage):

The Behavioral Risk Factor Surveillance System is an annual telephone survey designed to identify risk factors in the adult population of the Empire and report emerging health trends. The following table summarizes two variables for the respondents: health status and health coverage, which describes whether each respondent had health insurance.

		Health Status					
		Excellent	Very good	Good	Fair	Poor	Total
Health Coverage	No	459	727	854	385	99	2,524
	Yes	4,198	6,245	4,821	1,634	578	17,476
Total		4,657	6,972	5,675	2,019	677	20,000

If we draw one individual at random, what is the probability that the respondent has excellent health and doesn't have health coverage?

Your answer:

If we draw one individual at random, what is the probability that the respondent has excellent health or doesn't have health coverage?

Your answer:

The random variable X represents the number of credit cards that adults in the Empire have along with the corresponding probabilities (of the adults having X many cards).

X	0	1	2	3	4
$P(X)$	0.05	0.67	0.22	0.04	0.02

What is the probability that an adult has **at least** two credit cards?

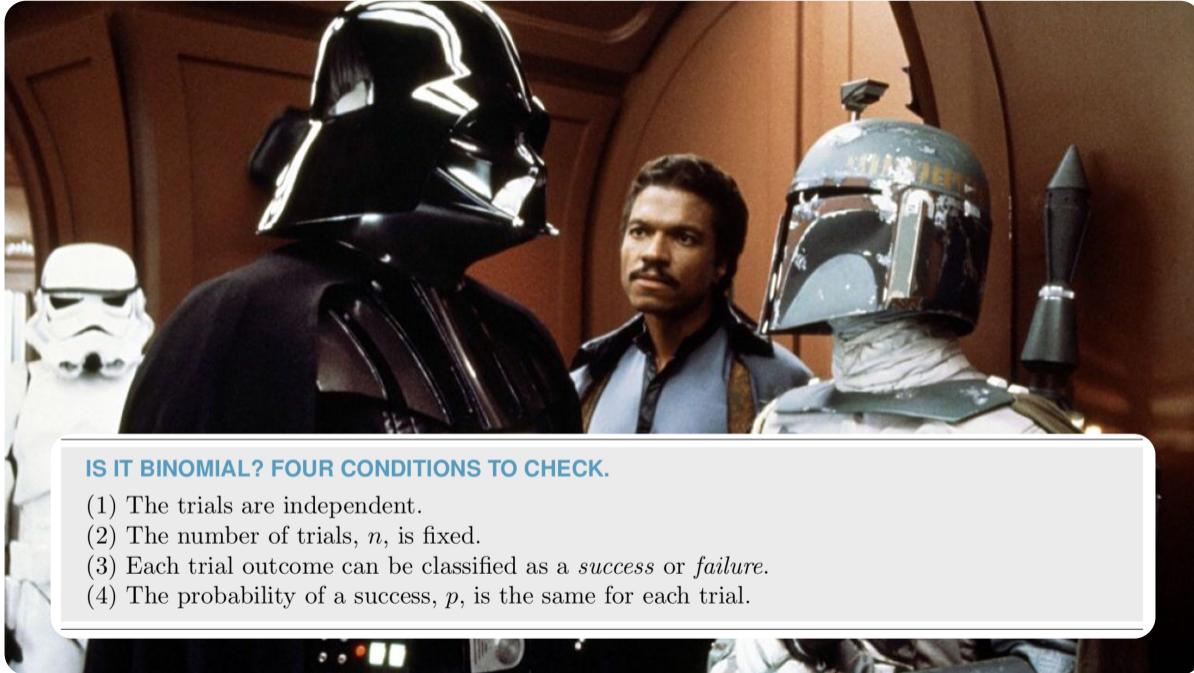
Your answer:

What is the mean⁴ of this probability distribution?

Your answer:

⁴The mean of a probability distribution $P(X)$ for X is the expected value $E(X)$ of the random variable X .

5. WOULD THE REAL BINOMIAL DISTRIBUTION PLEASE STAND UP?



IS IT BINOMIAL? FOUR CONDITIONS TO CHECK.

- (1) The trials are independent.
- (2) The number of trials, n , is fixed.
- (3) Each trial outcome can be classified as a *success* or *failure*.
- (4) The probability of a success, p , is the same for each trial.

Only one of the problems below require the binomial distribution. Which one? (Answer both regardless.)

Your answer:

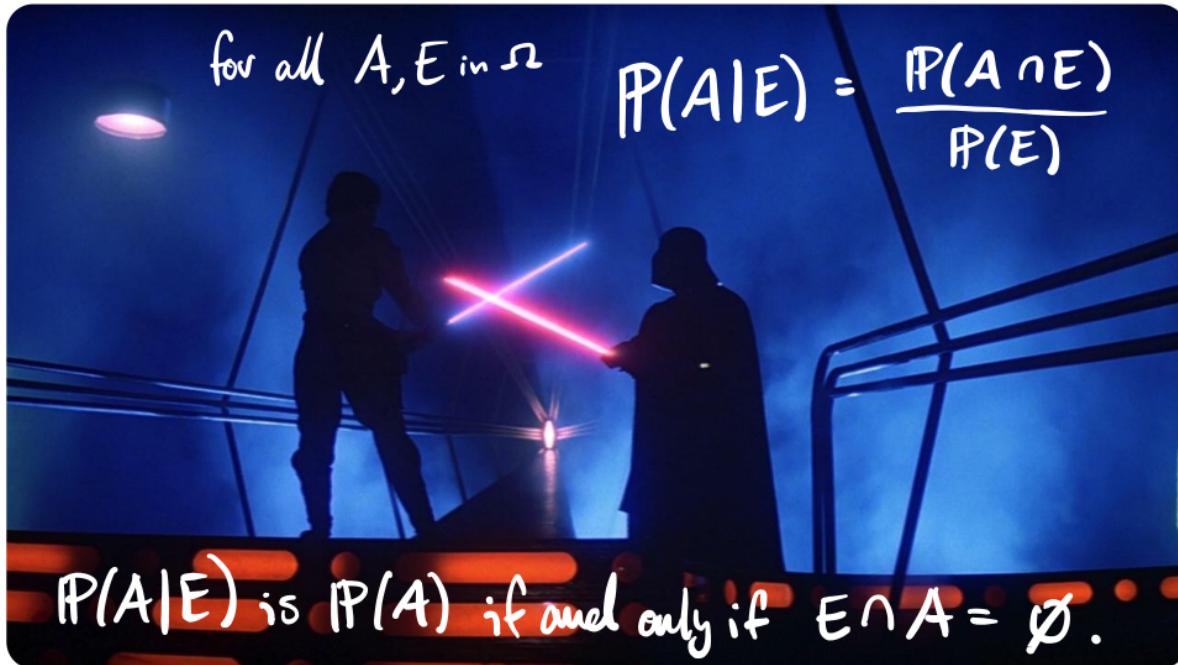
A. Chewie the Wookie has 4 identical boxes of snacks. Each box contains 40 snacks (18 purple snacks, 14 red snacks, and 8 green snacks). Chewie selects one snack at random from each box. What is the probability that exactly 1 of the 4 snacks is a red snack?

Your answer:

B. Chewie the Wookie has 4 identical boxes of snacks. Each box contains 40 snacks (18 purple snacks, 14 red snacks, and 8 green snacks). Chewie selects one snack at random from each box. What is the probability that exactly 1 of the 4 snacks is a red snack, and the rest (the remaining 3) are purple?

Your answer:

6. CONDITIONAL PROBABILITY, AGAIN!



A certain hereditary disease can be passed from a mother to her children (e.g., love of the dark side). Given that the mother has the disease (call this event M), her children independently will have it with probability $1/2$. Given that she doesn't have the disease (call this event M^c), her children won't have it either. A certain mother, who has probability $1/3$ of having the disease, has two children.

Find the probability that neither child has the disease. (Hint: give a uppercase letter to the events that the elder child does or does not have the disease. Do the same for the younger child. Then draw a tree diagram.)

Your answer:

Is whether the elder child has the disease independent of whether the younger child has the disease?

'Yes' or 'No', and briefly explain:

The elder child is found not to have the disease. A week later, the younger child is also found not to have the disease. Given this information, find the probability that the mother has the disease.

7. SCRATCH WORK