Name: **KEY**
BSAD 210—Montgomery College

EXAM 2

- There are 110 possible points on this exam. The test is out of 100.

- You have one class period to complete this exam, but you should be able to complete it in less than that

- Please turn off all cell phones and other electronic equipment.

- Be sure to read all instructions and questions carefully.

- Remember to show all your work. You may print your formulas in Excel using the Show Formulas option in the Formulas tab. Printed versions of your work showing formulas ***and*** showing the results counts as showing your work. But you must include both with your test for "showing your work" to count this way. Write your name on both printouts.

- Note the last sheet lists some helpful Excel commands.

- *Please print clearly and neatly.*

Part I: Matching. *Write the letter from the column on the right which best matches each word or phrase in the column on the left. You will not use all the options on the right and you cannot use the same option more than once.*
2 points each.

1. **C** Binominal distribution
2. **F** Independent
3. **E** Mutually exclusive
4. **D** Poisson distribution
5. **I** Prediction market
6. **H** Risk averse
7. **G** Risk loving

A. Defined by if the sample size is less than 5% of the population
B. Defined by if the likelihood of a success changes with each trial
C. Used if wondering the probability of three machines (out of 50 identical ones) breaking
D. Used if wondering the probability of ten neighborhood burglar alarms going off in a week.
E. Means probabilities of events can be added together
F. Means probabilities of events can be multiplied together
G. State lotteries make money based on this idea
H. Warranties make money based on this idea
I. Prices can be interpreted as probabilities

1. *Each machine has an equal likelihood of breaking (as they are identical). Since the chance of one breaking doesn't change if another breaks, the probabilities are constant; you use binomial distribution.*
2. *If events are independent, probabilities are multiplied together to determine the chance that they both happen at the same time.*
3. *If both events can't happen at the same time (mutually exclusive), to determine the chance of either happens, the probabilities are added together.*
4. *Any number of alarms could go off in a week; since this is a defined interval (in this case of time), use the Poisson distribution to determine the probability that exactly ten alarms go off.*
5. *Prediction markets create prices: pay such and such amount for a ticket worth some amount if the event happens and $0 if it doesn't. That price is thus the expected value of the ticket; isolate the payoff and you get a probability.*
6. *People would rather give up a fixed amount of money (the cost of the warranty) to avoid the chance of a larger amount of money (the cost of replacement). Such individuals will pay even if the cost of the*

warranty is higher than the expected cost of replacement; firms pocket the difference; that's how they make money.

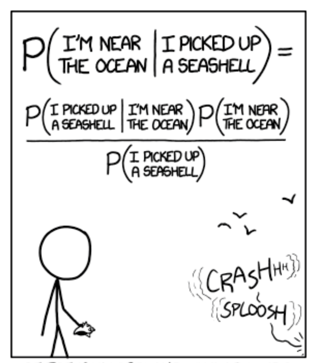
7. *People may choose between the cost of a lotto ticket or the ticket, which has a small chance of being worth a lot of money. Its expected value is less than the cost of the ticket (so payouts in winnings are less than total revenue from the tickets); only risk loving individuals will buy them.*

Part II: Multiple Choice. *Choose the best answer to the following.* 4 points each.

8. The ***most important*** difference between a binomial distribution and a hypergeometric distribution is based on what?
   a. If the size of the sample relative to the population is large or not.
   b. If the standard deviation of the population is constant or not.
   c. If the interval between events is constant or not.
   d. If the chance of success is high or not.
   e. None of the above

The most important difference is that in a hypergeometric, the probability of success is not constant while in a binomial distribution, it is constant (or close enough to being constant).

9. It's commonly suggested that if you hold a seashell to your ear, you can hear the ocean. This xkcd comic[1] suggests this is a poor conclusion. What statistical tool is used to make this point?
   a. Expected value
   b. Bayes' Theorem
   c. Discrete probability function
   d. Learned Hand Rule
   e. None of the above



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

The equation at the top of the comic is clearly Bayes' Rule: P(A|B) = [P(B|A)P(A)] / [P(B)]. In other words, the probability you're near the ocean, assuming you picked up a seashell, is very high. When else do you pick up a seashell? So of course you're hearing the ocean.

10. Consider a Poisson distribution. If $\lambda=2$, how does the probability of getting two events in the time period compare with getting one event in that same time period?

[1] http://xkcd.com/1236/

a. More likely to get one event
b. **Equally likely**
c. More likely to get two events
d. It is impossible to tell with the information provided
e. None of the above

Interestingly, both scenarios are equally likely. $P(1) = (2^1)/[(e^2)(1!)] = 2/e^2$; $P(2) = (2^2)/[(e^2)(2!)] = 4/[(e^2)(2)] = 2/e^2$

You could also put =POISSON(1,2,0) and =POISSON(2,2,0) into Excel; you'll get the same result: 0.270671

11. An expected cost is an expected value but with a negative payoff. Therefore, one of the lessons of this unit is that a rare event that comes with a high cost can discourage bad behavior as well as a common event that comes with a low cost. Which of the following is a rare event that comes with a high cost?
    a. **Getting arrested because you were pulled over but didn't have your driver's license.**
    b. Being late for class and getting disappointing looks from your professor and classmates.
    c. Getting caught after murdering someone in full view of dozens of witnesses and video cameras and then being sent to jail for a long time.
    d. B & C
    e. None of the above

Being pulled over is a rare event; on any given day you are unlikely to be pulled over. (Yes, I'm aware that some ethnicities are disproportionally pulled over—"driving while black"—but it is still rare event even within that group.) But going to jail is a high cost. This is why people get their licenses even though it's unlikely they'll be caught driving without one.

The murdering option is not a low probability event. If you murder someone with a lot of witnesses and there's video of you committing your crime, you are very *likely to get caught.*

12. A test that's 100% specific and 0% sensitive:
    a. Will *never* have a false positive
    b. Will *never* have a false negative
    c. Will *never* have a true positive
    d. **A & C**
    e. B & C

Such a test will always be able to detect a negative state; therefore, all of its negatives will be true. There will be no false positives. Such a test will also never correctly detect a positive state. It will never get a true positive.

In fact, all results will come back negative. It will either correctly identify a negative state, calling it negative (thus no false positives), or it will incorrectly identify a positive state, still calling it negative (thus no true positives). It will always give back negative results. It's a stupid test; it tells you nothing.

13. It's conventional wisdom that prevention is always better than a cure because it's often cheaper to prevent a bad thing from happening than correcting a bad thing after it happened. But the Learned Hand Rule suggests that's not true. How?
    a. **A cost for non-problems must also be incurred; that's why p matters.**
    b. Burden only matters when there's negligence.
    c. The expected value is often negative.
    d. Sensitivity is greater than specificity.
    e. None of the above

When people argue how great prevention is, they often forget that the cost of prevention must be incurred even when no problem happens. Consider a train yard. Whenever a train moves, there is a chance that someone is underneath it doing maintenance work.

14. Which of the following pairs are pairs of independent probabilities?
    a. The chance of rain and the chance that there will be heavy traffic
    b. The likelihood of a child being bitten by a dog and the chance that the child has a hot dog in her pocket.
    c. The probability that an employee stole something from your stole and the probability that your security system is broken.
    d. A & C
    e. **None of the above**

In each of these examples, the probabilities are connected; there are conditional probabilities. For example, the probability of heavy traffic given it rains is not the same as the probability of heavy traffic given it does not rain. If $P(A|B) = P(A|\sim B)$, and vice versa, A and B would be independent.

15. If two events are mutually exclusive, then one can sum the probability of each event to find the chance that at least one of the events occur. If the events are ***not*** mutually exclusive, how could you determine the chance that at least one of the events occur?
    a. Multiply the probabilities together.

b. Add the probabilities together and then divide by the probability that both occur.
c. **Add the probabilities together and then subtract the probability that both occur.**
d. Add the probabilities together and then subtract 1.
e. You can't determine this.

When there are two events, A and B, that are not mutually exclusive, there are four possibilities: neither of them happens, only A happens, only B happens, and both A and B happen.

If you want to determine what the chance of A or B (or both), adding the chance of each together would include some overlap. You would count times B comes up when you add in the chance of A and A comes up when you add in the chance of B. By subtracting the chance that both come up, you remove the double counting.

Note that when A and B are mutually exclusive, the chance that both of them occur is zero.

Consider two fair coins flipping at the same time. What's the chance at least one will come up heads?

		Coin 2	
		Heads	*Tails*
Coin 1	*Heads*	HH	HT
	Tails	TH	TT

*That's 0.5 for Coin 1 + 0.5 for Coin 2 = 1.00 Now subtract the chance both come up heads (0.25, or 0.5*0.5)*[2] *to get 0.75.*

Note that if you wanted the chance only one head comes up, you would again subtract the chance that both come up, resulting in 0.5.

16. Suppose Greg slipped on some spilled juice at a grocery store and he sued the store for damages because they didn't post a wet floor sign fast enough. According to the Learned Hand Rule, if Greg had slipped on some spilled juice at a hair salon, how would chances of the court ruling in Greg's favor changed?
    a. Better, because B is higher.
    b. Better, because p is lower.

[2] We can multiply because the coin flips are independent events.

c. Worse, because B is higher.

d. Worse, because p is lower.

e. Chances would be about the same / It is impossible to tell with the information provided.

Spilled liquids in a hair salon are far less likely than in a grocery store. The probability of an accident, p, would be lower and therefore it would be harder to blame the establishment for not taking proper precautions; pL would be lower and therefore, it is more likely $B > pL$.

Option C is also possible to argue. While wet floor signs cost the same regardless of who's buying them, posting it requires employee action. The employee must notice the problem, stop what they're doing, find the sign, and display it. One could argue that when employees at a hair salon are busy, they are busy with a directly helping a customer; stopping to do something else delays getting the job done and might irritate the customer, threatening a loss of business. Therefore, vigilance and precaution concerning spills is a greater burden in a hair salon than in a grocery store, were employee activities are more fluid and direct customer help (in the aisles, where spills are most likely to be) is more limited.

17. Which of the following products depend on people being risk averse?

a. Extended warranties

b. Car insurance

c. The lottery

d. A & B

e. None of the above

A warranty means you can return a defective product. Extending the warranty always cost more money; it's functionally the same as insurance. From a risk neutral standpoint, it's not a good deal: you are paying X amount to avoid a failure which will cost you Y. If $X < E(Y)$, companies would lose money on average. The lottery, on the other hand, is about paying X to ***get*** *Y. If $X > E(Y)$, states make money off their lottery (and because $X > E(Y)$, they do which is why they offer one). But it's not something a risk averse person would do.*

18. Jason works quality control in a textiles factory. His job is to reject any fabric with more than one error per yard of fabric. Suppose there's a 1% chance of getting an error. Jason wants to know if the textiles machinery is working properly and wonders how likely it would be to find three errors in one yard. If Jason pulls 100 yards of fabric, what probability function should he use?

a. Binominal

b. Hypergeometric
c. Poisson
d. It is impossible to tell given the information provided
e. None of the above

Poisson is about the number of events over a defined interval. While usually that interval is time, it doesn't have to be. In this case, the interval is a yard of fabric. Note a defining feature of Poisson—there number of events is infinite—applies here as well. There's no (obvious) limit to the number of errors on a yard of fabric.

19. A risk loving person _________.
    a. ***Always*** chooses the riskiest option.
    b. Would ***never*** pick the same option as a risk averse person.
    c. Prefers a 10% chance to win $50 than a five dollar bill.
    d. A & B
    e. All of the above

This is true by definition; when an expected value of an uncertain payoff equals the value of a certain payoff, a risk loving person prefers the uncertain payoff.

But keep in mind that "risk loving" is a spectrum. If the certain payoff is greater than the expected value of an uncertain payoff (say, it was a 1% chance of winning $50 instead of a 10% chance), a risk loving person may still decide to go with the certain payoff. Only when the expected value equals the certain payoff can we truly distinguish the risk loving from the risk averse and the risk neutral.

Part III: Short Answer. *Answer the following.*
16 points each.

20. The 2014 "Weird Al" Yankovic song *Word Crimes* describes the proper use of grammar. It begins:

    If you can't write in the proper way
    If you don't know how to conjugate
    Maybe you flunked that class

    A class in English is a sort of test on how well you know grammar (among other things). Suppose 3% of people do not know proper grammar. Also suppose that the "test" of an English class is 95% sensitive and 99% specific. If someone flunks English (they failed the test), what is the probability they do not know proper grammar?

Time to pull out Bayes' Theorem. We're testing for ignorance of grammar (~G), with P(~G) being the probability you don't know grammar:

$$\mathrm{P}(\sim G|-) = \frac{\mathrm{P}(-|\sim G)\,\mathrm{P}(\sim G)}{\mathrm{P}(-|\sim G)\,\mathrm{P}(\sim G) + \mathrm{P}(-|G)\,\mathrm{P}(G)}$$

$$\mathrm{P}(\sim G|-) = \frac{(0.99)(0.03)}{(0.99)(0.03) + (0.05)(0.97)} = \frac{0.0297}{0.0297 + 0.0485} = \frac{0.0297}{0.0782} = 0.3798$$

There is a 37.98% chance that someone who fails English is poor at grammar.

21. Mary would like to grow vegetables in a small plot of land in her backyard and sell the crops to restaurants. She buys some bell pepper seeds online for $20. They cannot be returned. She later finds out that the seeds are bad 25% of the time; nothing will grow. But if they do grow, she'll make money based on how many peppers her crop yields and how good the crop is. She plants one hundred square feet of bell peppers.

Quality	*Probability*	*Price Per Pepper*
Good	80%	$0.25
Excellent	20%	$0.60

Quantity Per Square Foot	*Probability*
1	5%
2	35%
3	50%
4	10%

The first table indicates the chance each possible quality will occur and the price per pepper for that quality. The second table indicates the how likely various quantities of peppers would grow per square foot. Assume yield probabilities are independent with respect to quality probabilities.

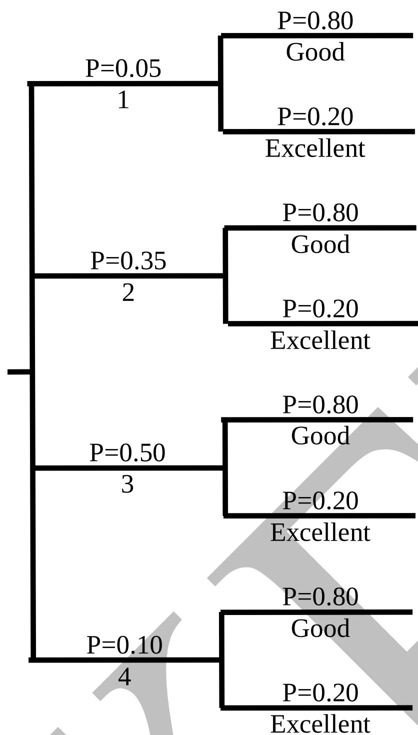
What is the expected profit (value) of this venture? Assume seeds are the only cost. (Remember to show your work. If using Excel, you may do this by selecting Show Formulas, under the Formulas tab, and printing off a copy to include with your exam along with a copy of the results.)

There's a lot going on in this question. Some things to remember:

- *One hundred square feet of crops were planted.*
- *There's a 75% chance the crops will grow and a 25% chance nothing will.*
- *The seeds cost $20, regardless if they grow or not.*

- *The dollars per square foot depend on the quality of bell peppers and the quantity of bell peppers. (These are independent probabilities.)*

On the last issue, it's useful to make a tree diagram:



(0.05)(0.80) = 0.04	(0.04)(1)($0.25) = $0.010
(0.05)(0.20) = 0.01	(0.01)(1)($0.60) = $0.006
(0.35)(0.80) = 0.28	(0.28)(2)($0.25) = $0.140
(0.35)(0.20) = 0.07	(0.07)(2)($0.60) = $0.084
(0.50)(0.80) = 0.40	(0.40)(3)($0.25) = $0.300
(0.50)(0.20) = 0.10	(0.10)(3)($0.60) = $0.180
(0.10)(0.80) = 0.08	(0.08)(4)($0.25) = $0.080
(0.10)(0.20) = 0.02	(0.02)(4)($0.60) = $0.048

So: the chance of getting one good bell pepper in a square foot is (0.05)(0.80) = 0.04. The expected value of that square foot is (0.04)(1)($0.25), or the chance it'll happen times the number of bell peppers you have to sell times the value of the bell peppers.

Add up all of these possibilities and you get: $0.848; each square foot is worth just under 85 cents.

Since there are 100 square feet, the expected value of the garden is ($0.848)(100) = $84.80.

But we're not done yet; remember that $84.80 is conditional on the seed being good. There's a 25% chance nothing will grow at all and a 75% chance the seeds are good. The expected revenue is ($84.80)(0.75) = $63.60. Subtract off the cost of the seeds and her expected profit is $63.60 - $20 = $43.60.

Note that I left out the 25% chance that nothing will grow. If nothing grows, she has no peppers to sell; her revenue from these seeds is zero. Thus there is a 25% chance she'll earn no money which becomes zero (25% times zero is zero). Adding zero to anything gets you the same thing again so I just left it out.

The most time consuming part of this question was filling out the tree diagram but you can do the same thing in far less time with Excel.

This is what I put into Excel:

	A	B	C	D	E	F
1	*Quality*	*Probability*	*Price Per Pepper*		*Quantity Per Square Foot*	*Probability*
2	Good	80%	$0.25		1	5%
3	Excellent	20%	$0.60		2	35%
4					3	50%
5					4	10%
6						
7						
8						
9		good	excellent			
10	1	4.00%	1.00%			
11	2	28.00%	7.00%			
12	3	40.00%	10.00%			
13	4	8.00%	2.00%			
14						
15		good	excellent			
16	1	$0.010	$0.006			
17	2	$0.140	$0.084			
18	3	$0.300	$0.180			
19	4	$0.080	$0.048			
20				**$0.848**		
21						

This is the same thing, but now you can see the formulas. (You can find this option under the Formulas tab. It'll automatically expand all your columns so I had to shrink them back manually to keep it all on one page.)

	A	B	C	D	E	F
1	*Quality*	*Probability*	*Price Per Pepper*		*Quantity Per Square Foot*	*Probability*
2	Good	0.8	0.25		1	0.05
3	Excellent	0.2	0.6		2	0.35
4					3	0.5
5					4	0.1
6						
7						
8						
9		good	excellent			
10	1	=F2*B2	=F2*B3			
11	2	=F3*B2	=F3*B3			
12	3	=F4*B2	=F4*B3			
13	4	=F5*B2	=F5*B3			
14						
15		good	excellent			
16	1	=B10*A10*C2	=C10*A10*C3			
17	2	=B11*A11*C2	=C11*A11*C3			
18	3	=B12*A12*C2	=C12*A12*C3			
19	4	=B13*A13*C2	=C13*A13*C3			
20				**=SUM(B16:C19)**		

Note that I locked certain cell references (as indicated with the dollar signs) to speed up making these calculations. When I copied and pasted the cells, some references updated and some did not. Consider B10. By selecting F2 without locking it, when I copied and pasted it to B11, F2 became F3, as I wanted. But B2 was locked; when I copied and pasted B10 to B11, B2 stayed B2, as I wanted (because this was the "good" column as opposed to the "excellent" column).

22. Alfonso works for a fruit company. He's in charge of quality control for bananas. It's too expensive to test every banana bunch in a crate so he requires his fellow workers to select a sample. Suppose he has them selects three banana bunches from each crate containing six banana bunches. Suppose, in one instance, the chosen box has three bad banana bunches. What is the probability that the sample from that crate will have exactly two bad banana bunches? Be sure to include any commands you put into Excel.

Here, "success" is defined as getting a bad bunch. Since the chance of success changes with each trial (three trials but only six bananas), we use hypergeometric. Let's make a list of what we know.

$N = 6$ (bunches in the crate)
$R = 3$ (bad bunches)
$n = 3$ (number of bunches being pulled)
$x = 2$ (successes in question)

Here's what you should have put in Excel:

=HYPGEOM.DIST(2,3,3,6)

For a result of 0.45

If you want to write out the equation by hand, here's what you would have wrote (we didn't get into the actual equations behind the Excel commands but just so you appreciate all that the program is doing for you):

$$\mathrm{P}(2,3) = \frac{\left(\frac{(6-3)!}{\left((6-3)-(3-2)\right)!\,(3-2)!}\right)\left(\frac{3!}{(3-2)!\,2!}\right)}{\left(\frac{6!}{(6-3)!\,3!}\right)} = \frac{\left(\frac{3!}{(3-1)!\,1!}\right)\left(\frac{3!}{1!\,2!}\right)}{\left(\frac{6!}{3!\,3!}\right)}$$

$$= \frac{\left(\frac{3!}{2!}\right)\left(\frac{3!}{2!}\right)}{\left(\frac{6!}{3!\,3!}\right)} = \frac{(3)(3)}{\left(\frac{(6)(5)(4)}{(3)(2)(1)}\right)} = \frac{9}{(2)(5)(2)} = \frac{9}{20} = 0.45$$

Regardless, you have a 45% of getting exactly two bad bunches.

Equation and Information Sheet

Function or Command	*Output*
ABS	The absolute value of an input
AVERAGE	Arithmetic mean of a dataset
BINOM.DIST	Binominal distribution for x number of successes
CORREL	Correlation coefficient of two variables
CTRL + `	Show formulas
CTRL + F	Find
CTRL + P	Print
CTRL + X	Cut highlighted area
CTRL + C	Copy highlighted area
CTRL + V	Paste highlighted area
CTRL + Z	Undo
F4	Makes cell reference absolute
GEOMEAN	Geometric mean of a dataset (adjustments must be added manually)
HYPGEOM.DIST	Hypergeometric distribution for x number of successes
LARGE	Larger values of a dataset (k=1 is largest, k=2 is second largest, k=3 is third largest…)
MAX	Maximum value of a dataset
MEDIAN	Median of a dataset
MIN	Minimum value of a dataset
MODE	Mode of a dataset
POISSON	Poisson distribution for x number of successes
QUARTILE	The 0^{th} to 4^{th} quartile of a dataset
SMALL	Smaller values of a dataset (k=1 is smallest, k=2 is second smallest, k=3 is third smallest…)
STDEV.S	Standard deviation of a sample

Coefficient of Variation

$$CV_{sample} = \frac{s}{\bar{x}}(100)$$

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|{\sim}A)P({\sim}A)}$$

Learned Hand Formula

$$B < pL$$

Binominal Distribution

$$\mu = np, \sigma = \sqrt{npq}$$

Hypergeometric Distribution

$$\mu = \frac{nR}{N}, \sigma = \sqrt{\frac{nR(N-R)}{N^2}}\sqrt{\frac{N-n}{N-1}}$$

Poisson

$$\mu = \lambda, \sigma = \sqrt{\lambda}$$