Name: _____
BSAD 210—Montgomery College
David Youngberg

EXAM 4

Practice #1

- There are 110 possible points on this exam. The test is out of 100.
- You have two hours 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 print outs.
- Try all questions! You get zero points for questions that are not attempted.
- Note the last sheet lists all the equations you will need for this exam.
- *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. ___ Binomial distribution
2. ___ Central Limit Theorem
3. ___ Median
4. ___ Poisson distribution
5. ___ Prediction market
6. ___ Self-selection bias
7. ___ Standard deviation
8. ___ Survivorship bias
9. ___ t score
10. ___ z score

A. Example: Comparing average newspaper revenues in 1990 and average revenues in 2015.
B. Example: Examining average test scores in a country where testing is voluntary
C. Example: Having all students in a remedial math class take a survey that's meant to represent the opinions of all students.
D. Example: Likelihood that two of fifty different buses are transporting one of five 1st grade classes (one class per bus)
E. Example: Probability that one of ten very different companies will perfect a new treatment for cancer this year
F. Prices can be interpreted as probabilities
G. If an odd number of observations is all either very high or very low, this will be either very high or very low.
H. If an odd number of observations is all either very high or very low, this will be very high.
I. If an odd number of observations is all either very high or very low, this will be somewhere between the observations.
J. Main reason for why we never say an analysis "proves" something
K. Use when you do know σ
L. Use when you don't know σ
M. Used if wondering the probability of ten neighborhood burglar alarms going off in a week.

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

3 points each.

11. 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

12. Which of the following causal claims is better described as reverse causation?
    a. Education causes income
    b. Sales cause advertising
    c. Size of house causes quality of car
    d. B & C
    e. None of the above

13. What does R^2 represent?
    a. The fraction of the variation the regression explains
    b. Regression Sum of Squares divided by Error Sum of Squares
    c. Regression Sum of Squares divided by Total Sum of Squares
    d. A & C
    e. None of the above

14. Integrated Systems manufactures internet modems. On average, the modems download one MB in 2.2 seconds with a standard deviation of 0.3 seconds. Assume modem performance follows a normal distribution. Suppose Integrated Systems wants to offer a warranty for underperforming modems. The minimum quality should result in replacing no more than 12% of modems. How many seconds should that minimum quality be?
    a. 0.6474 seconds
    b. 1.8474 seconds
    c. 2.5525 seconds
    d. 5.7250 seconds
    e. None of the above

15. Jason works quality control in a textiles factory. His job is to reject any fabric with more than one error per yard of fabric (the average under normal conditions). 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 or more errors in one yard. What is the chance that a yard of fabric would have three or more errors if the machine is working properly?
    a. 0.08
    b. 0.18
    c. 0.92
    d. 0.98
    e. None of the above

16. 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

17. 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

18. Suppose you ran a regression and found heteroscedasticity. What should you do first?
    a. Start over with new variables
    b. Drop as few variables as possible to remove the heteroscedasticity
    c. Check how significant your explanatory variables are
    d. A & B
    e. None of the above

19. 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

20. Use the Practice Final Exam Data Set to answer this question. It is hypothetical data on a hypothetical grocery store chain called The Happy Spud, with each observation a particular location. Using the coefficient of variance, determine which variable is most consistent.
    a. Annual profit
    b. Square feet
    c. Inventory value
    d. Advertising spent
    e. None of the above

21. Use the Practice Final Exam Data Set to answer this question. Create a dummy variable called East? to indicate which region the store is in (1=East, 0=West). Then run a regression with East? and Square feet predicting Annual profit. At 90% confidence, what is the result of the East? variable?
    a. It's statistically significant, with being located in the East reducing profits by 62.7%.
    b. It's statistically significant, with being located in the East increasing profits by 62.7%.
    c. It's statistically significant, with each additional dollar in profit reduces the chances of being located in the East by 62.7%.
    d. It's not statistically significant.
    e. None the above

22. Ursula works for CBS. Suppose CBS changed its time slot of a long running show, *The Amazing Race*, to Fridays. Suppose in previous seasons, the show average 8 million viewers with a population standard deviation of two million (based on over 200 episodes). The new time slot for the latest season—based on nine episodes—averaged 6.5 million episodes. Does changing the time slot result in a statistically significant change in viewership?
    a. At the 95% level, yes.
    b. At the 99% level, yes.
    c. A & B
    d. None of the above, but it would at the 90% level.
    e. None the above; it's not statistically significant at all.

23. George works in a deli slicing meats using a special machine. This machine can slice meat very thinly (to maximize the flavor). Too thick and it wastes food. But it can never be too thin; thinner is always better. Suppose George wants to test if the machine is working properly. Normally it slices meat 0.5 millimeters thick. What is George's null hypothesis?
    a. $\mu < 0.5$ mm
    b. $\mu \leq 0.5$ mm
    c. $\mu = 0.5$ mm
    d. $\mu \geq 0.5$ mm
    e. $\mu > 0.5$ mm

24. If a distribution is positively skewed, which of the following is true about the central tendency?
    a. The median is higher than the mean
    b. The mean is higher than the median
    c. The mode is higher than the median
    d. B & C
    e. None of the above

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

25. Yolanda Jade is testing if any of her employees might sell company secrets to a competitor. She knows from industry analysis and historical evidence that 0.3% of employees will betray the company for money (we'll call them "traitors"). Suppose Yolanda develops a series of questions to test for treachery. It is 99% sensitive and 98% specific. If a result comes back positive, what is the chance that the employee is actually a traitor?

26. 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.

27. Justin is venture capitalist. He's thinking about backing a new video game with a radically unusual approach. Suppose there is a 60% chance of getting the game to work correctly. If it doesn't work, Justin would lose the $10 million he invested. If it works, the profit Justin can make on this investment is determined by how successful the game is (see table). Calculate Justin's expected value of this deal. Remember to show all work.

Market Performance	*Probability*	*Profit*
Good	50%	$20 million
Average	30%	$8 million
Bad	20%	$3 million

28. Pepper's testing a new workout routine he's developed to see if it's better than the standard one. The standard workout routine results in 5.1 pounds lost on average over the course of a week. It has a population standard deviation of 1.7 pounds. Using a sample of 25 people, Pepper finds that his results in 5.9 pounds of weight loss over a week with the same number of hours per day. Is his results statistically significant at the 90%, 95%, 99%, and/or 99.9% level? What is the p-value, to four decimal places, of this test? Make sure to show your work and report the relevant critical z-score and calculated z-score and how you found the p-value. (HINT: You will need Excel to find the p-value.)

Exam 4 Equation and Information Reference

Function	*Output*
ABS	The absolute value of an input
AVERAGE	Arithmetic mean of a dataset
BINOM.DIST	Binominal distribution for x number of successes
CONFIDENCE.NORM	Determines the margin of error to make a confidence interval (known σ)
CONFIDENCE.T	Determines the margin of error to make a confidence interval (unknown σ)
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
NORM.DIST	Returns the normal distribution for a specified mean and standard deviation.
NORM.INV	Returns the inverse of the normal cumulative distribution for a specified mean and standard deviation.
NORM.S.DIST	Returns the standard normal distribution. Can also be used to find the critical z scores.
NORM.S.INV	Returns the inverse of the standard normal cumulative distribution. Useful for finding critical z scores.
POISSON.DIST	Poisson distribution for x number of successes
QUARTILE	The 0^{th} to 4^{th} quartile of a dataset
SQRT	Finds the square root of the value in question.
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
T.INV	Finds area under a t distribution; useful for finding one-tailed critical t scores.
T.INV.2T	Finds area under a t distribution; useful for finding two-tailed critical t scores.
T.TEST	Various two population tests which use a t score.

Critical z scores

Confidence	α	$z_{\alpha/2}$	z_{α}
90%	0.1	1.645	1.280
95%	0.05	1.960	1.645
99%	0.01	2.576	2.330
99.9%	0.001	3.291	3.090

Critical t scores

Use the T.INV or T.INV.2T commands

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

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

Optimal Sample Size

$$n = \left(\frac{z_{\alpha/2}\sigma}{ME}\right)^2$$

z-test

$$z_{\bar{x}} = \left|\frac{\bar{x} - \mu_{H_0}}{\sigma/\sqrt{n}}\right|$$

Proportion

$$z_p = \left|\frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}\right|$$

t-test

$$t_{\bar{x}} = \left|\frac{\bar{x} - \mu_{H_0}}{s/\sqrt{n}}\right|$$

Adjusted R^2

$$R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-k-1}$$