Name: _____
BSAD 210—Montgomery College
David Youngberg

FINAL EXAM

Practice A

- 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. Writing down what you put into Excel is sufficient to show your work.
- To access Data Analysis on Excel, select File (top left), then Options, then Add-ins, then Go… (for Excel Add-ins), then select Analysis ToolPak.
- 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. ___ Population
6. ___ Self-selection bias
7. ___ Standard deviation
8. ___ Survivorship bias
9. ___ t score
10. ___ z score

A. All possible objects of interest
B. Example: Comparing average newspaper revenues in 1990 and average revenues in 2015.
C. Example: Examining average test scores in a country where testing is voluntary
D. Example: Having all students in a remedial math class take a survey that's meant to represent the opinions of all students.
E. Example: Likelihood that 10 out of 50 randomly chosen buses, all in different cities, will be late; assume 400 out of 10,000 buses are late.
F. Example: Probability that one of ten very different companies will perfect a new treatment for cancer this year
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. Consider a hypothetical regression line of retail stores monthly profit (PROFIT) predicted by the percent of employees who quit each month (TURNOVER), with TURNOVER being statistically significant:

$$PROFIT = 650{,}000 - 42{,}000 * TURNOVER$$

    What is the punchline of TURNOVER?
    a. For every additional percent increase in turnover, monthly profit decreases by $42,000.
    b. For every additional percentage point increase in turnover, monthly profit decreases by $42,000.
    c. For every additional percent increase in turnover, monthly profit decreases by $420.
    d. For every additional percentage point increase in turnover, monthly profit decreases by $420.
    e. For every additional percentage point increase in turnover, monthly profit decreases by $4,200,000.

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. Political betting markets allow participants to bet real money on political events, including if someone becomes U.S. president and if someone becomes their party's nominee for

president. One can interpret the current price in these markets as probabilities. If Candidate X going price of becoming president is 35 cents, there there's a 35 percent chance that Candidate X will become president. Suppose Roger uses this information to calculate a president's "electability:" assuming they become the nominee, what is the chance that they become president? How could Roger reasonably calculate this?

a. Probability of becoming president divided by the probability of becoming the nominee.
b. Probability of becoming the nominee divided by the probability of becoming president.
c. Probability of becoming president times the probability of becoming the nominee.
d. Probability of becoming the nominee minus the probability of becoming president.
e. Roger doesn't have enough information to calculate this.

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 Razi invested 60% of her money in stocks, 30% in U.S. government bonds, and 10% in real estate. Suppose the rate of return for stocks is -4%, for bond is 1%, and for real estate is 5%. What is Razi's average rate of return?
    a. 0.67%
    b. 1.5%
    c. 2%
    d. 3%
    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 1 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 1 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 Number of competing stores in district predicting Annual profit. At 95% confidence, what is the result of the East? variable?
    a. It's statistically significant, with being located in the East reducing profits by 89.2%.
    b. It's statistically significant, with being located in the East increasing profits by 89.2%.
    c. It's statistically significant, with each additional dollar in profit reduces the chances of being located in the East by 89.2%.
    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. There is not enough information to determine an answer.
    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 95%, 99%, and/or 99.9% level?

    If you want to stretch yourself, determine the p-value, to four decimal places. 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.

Geometric Mean

$$Geometric\ Mean = \sqrt[n]{\prod_{i=1}^{n}(1 + x_i)} - 1$$

Weighted Average

$$Weighted\ Average = \frac{\sum_i^n (w_i x_i)}{\sum_i^n w_i}$$

Coefficient of Variation

$$CV = \frac{s}{\bar{x}}$$

Confidence interval for proportion

$$\widehat{CI}_{\bar{p}} = \bar{p} \mp z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Adjusted R^2

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

Bayes' Theorem

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

Binominal Distribution

$$\mu = np$$

Hypergeometric Distribution

$$\mu = \frac{nR}{N}$$

Poisson

$$\mu = \lambda$$

Hypothesis testing

z-test

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

t-test

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

z-test (proportion)

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

Critical z scores

Use =NORM.S.INV command

Confidence	*α*	$z_{\alpha/2}$	z_{α}
95%	0.05	1.960	1.645
99%	0.01	2.576	2.326
99.9%	0.001	3.291	3.090

Critical t scores

Use T.INV or T.INV.2T commands or see the table on the last page

p-values

Make your calculated value negative and then use one of the following (make sure cumulative is turned on):

	1 tail	*2 tails*
z	NORM.S.DIST	Multiply 1 tail result by 2
t	T.DIST	

Table B t distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											