Name: _____
BSAD 210—Montgomery College
David Youngberg

EXAM 4

Practice B

- 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. ___ Binominal distribution
2. ___ Correlation coefficient
3. ___ Expected value
4. ___ Hypergeometric distribution
5. ___ Kurtosis
6. ___ Margin of error
7. ___ Practically significant
8. ___ Standard deviation
9. ___ Standard normal distribution
10. ___ Statistically significant

A. A stronger correlation means this gets bigger
B. Conclusion if p-value is really small
C. Conclusion if p-value is really big
D. Describes the peakness of a normal distribution.
E. Forms the basis of critical z values
F. Mathematically shows that a rare and expensive event equals a common and cheap event
G. Never greater than 1 nor less than -1
H. Probability is constant
I. Probability is sometimes less than zero
J. Sample's value means you use the t distribution; population's value means you use the z distribution
K. Subtracted from and added to the mean
L. The number of successes is bounded by the number of trials
M. When a genuine difference is very large

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

3 points each.

11. Students with more psychological issues tend to have poorer nutrition. How should you apply this information?
    a. Eating healthy is never a bad idea.
    b. Better nutrition puts less stresses on your mind
    c. Causation is hard to determine
    d. The average is not the same as any particular observation.
    e. None of the above

12. Tyrone is looking for a job in information technology. Based on conversations from people in the industry, he estimates that his chance of getting a promotion is 60% and his chance of being hired is 80%. What is the chance that he's promoted <u>assuming</u> that he's hired?
    a. 20%
    b. 48%
    c. 60%
    d. 80%
    e. None of the above

13. Consider the previous question. If promotion in information technology is automatic—100% of hired people get promoted—then what should his chances of the nonconditional promotion?
    a. 20%
    b. 48%
    c. 60%
    d. 80%
    e. None of the above

14. People who had a job as a teenager are more likely to have a job as an adult. This correlation could be explained by a confounding variable. How?
    a. People who get a job as an adult cause them to get a job as a teenager.
    b. People with a lot of motivation are more likely to get a job, regardless of how old they are.
    c. People who get a job as a teenager get experience and that additional experience causes them to get a job as an adult.
    d. People who get a job as a teenager tend to be low income, which makes it easier for them to get a job as an adult.
    e. None of the above

15. If the coefficient of an independent variable in a regression analysis is -4.8 with a p-value of 0.094, what should you conclude at 95% confidence?
    a. The coefficient is not statistically significant.
    b. The coefficient is statistically significant but not practically significant.
    c. The coefficient is statistically significant and is practically significant.
    d. The coefficient is statistically significant; practical significance depends on what the dependent and independent variables are.
    e. It is impossible to conclude anything based on the information provided.

16. True or false: For two variables with an equal number of observations, if one variable has a larger range, that variable must also have a larger standard deviation.
    a. True because both range and standard deviation are measures of dispersion. It makes sense that they move in concert.
    b. True because the number of observations are equal; it would not be true otherwise.
    c. False because ranges are always larger than standard deviations.
    d. It is true, but not for a reason listed.
    e. It is false but not for a reason listed.

17. Which of the following is an example of someone being risk averse?
    a. Randy spending $60 on a coin flip that has a 50% chance of winning $120 and a 50% chance of winning zero.
    b. Amir learning to juggle using very sharp knives.
    c. Chloe purchasing a lottery ticket.
    d. B & C
    e. None of the above

18. Which of the following is an example of a question requiring a Poisson distribution?
    a. If a major storm hits the DC area, and an average of ten trees typically fall in a neighborhood, what is the chance that four trees will fall in a particular neighborhood?
    b. If there's an average of 1.3 flaws per 50 feet of rope, what are the chances that there will be two flaws on a particular 50 feet of rope?
    c. If Simon selects four people at random to be his bodyguards, what are the chances that two of his bodyguards will be traitors if there's a five percent chance of any one bodyguard being a traitor?
    d. A & B
    e. None of the above

19. Which of the following ***clearly*** would be independent to the chance of a corn crop failing on Wu's small farm?
    a. The chance of a drought.
    b. The chance a corn crop failing on Aziz's nearby farm.
    c. The chance that the price of corn will fall next year.
    d. B & C
    e. None of the above

20. Use Practice Final Exam Data Set 2 for this question. It's made up of hypothetical data of fictional cities. What's the difference between the average Net Water for cities on the coast and the average Net Water for cities not on the coast?
    a. 2.1
    b. 4.8
    c. 5.9
    d. 7.0
    e. None of the above

21. Armadillo Business Consulting has 40 associates who are eligible for promotion; seven of those associates are accounting majors. If eight associates are promoted, what is the chance that at least two of those will be accounting majors? (Use hypergeometric distribution to determine the answer.)
    a. 0.1282
    b. 0.3024
    c. 0.4306
    d. 0.5694
    e. 0.8718

22. Kali's investing in three companies. Each company has an 80% chance of failing. What's the chance that exactly one company will succeed?
    a. 12.8%
    b. 20.0%
    c. 38.4%
    d. 48.8%
    e. There is not enough information to determine the answer.

23. Allen's testing a new user interface to understand if people who are color blind like it more than the previous user interface. Color blind users rated the old interface at 6.7 out of 10. Based on a sample of 12 color blind users, with a population standard deviation of 1.8 and a sample standard deviation of 2.9, the new average rating was 8.4 out of 10. Is the new user interface a statistically significant improvement for color blind users?
    a. At the 99% level, yes.
    b. At the 99.9% level, yes.
    c. A & B
    d. None of the above, but it would at the 95% level.
    e. None the above; it's not statistically significant at all.

24. We've discussed that statistics is the application of mathematics to sample data. At the heart of that application is what idea?
    a. Calculating important values like averages and standard deviations.
    b. Determining which equation to use.
    c. Evaluating if something is due to randomness or not.
    d. Estimating the line of best fit.
    e. Identifying sample biases.

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

25. Using Practice Final Exam Data Set 2, run a regression with Net Water as the dependent variable and Coast?, Pollution, Precipitation, Average Income, and Population Density as the independent variables. Then answer the following:
    a. Which variables are statistically significant at the 95% level?
    b. There is multicollinearity in this model; how should you fix the multicollinearity?
    c. Re-run the regression with the fix you suggest in part B. This is a better model than the original; what about the regression output shows you that?

    a. _____
    _____
    _____

    b. _____
    _____
    _____

    c. _____
    _____
    _____

26. Trinity is a lawyer attempting to determine which expert she wants to use in a trial. The expert must be able to explain the clearly explain the material to a jury while also being personable, cooperative, trustworthy, and able to withstand cross examination. Trinity does not have time to interview these candidates in depth so she relies on an agency's recommendation. Suppose 10% of potential experts are good in a trial. The agency's recommendation is 85% sensitive and 75% specific. What's the chance that the expert would be good at trial (G) if the agency recommended him (R)?

27. Jaya's company is thinking about investing in new blood testing technology. There's a 90% chance it will not work and Jaya's company will lose $1 million from the investment. But there's a 10% chance it will work and the profits for the company depend on market performance, as indicated by the table. What's the expected value of this investment?

Market Performance	*Probability*	*Profit*
Excellent	5%	$25 million
Good	15%	$15 million
Fair	50%	$6 million
Terrible	30%	$2 million

28. Highland Farms grows potatoes with an average yield of 30,900 pounds per acre. On 13 acres, they try a new fertilizer to try to increase yields. On these 13 acres, the average yield of 32,500 pounds and a sample standard deviation of 1300 pounds. Is this a statistically significant difference or not at 99.9% confidence? Remember to show your work and justify your answer.

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

$$\mu = \lambda$$

<u>Hypothesis testing</u>

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 <u>on</u>):

	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											