BUS 310, Final Exam D, Part III, Spring 2019 Name KEY
Section

Instructions: This exam is in three parts: Part I is to be completed partly at home using the materials posted on Blackboard for Part I and you will answer questions about that work in class below; Part II is to be completed entirely in class using your computer. Part III is to be done entirely in class without your computer.

1. You may not use cell phones, and you may only access internet resources you are specifically directed to use: You may access your data file for Part I of the exam in Blackboard. You may access the data files posted to Blackboard for the Exam part II, but not for Part III.
2. It is a violation of the honor code to communicate with other students in or out of the class during the exam, by any means. Students whose exams show evidence of coordination will be reported.
3. Show all work to support your reasoning. Primarily, this can be done in Excel. Deletion of evidence of your logical process can result in loss of credit. A significant amount of credit goes toward process, reasoning and interpretation.
4. When rounding, do not over-round. In general, do not report dollar amounts beyond the penny. Means should be rounded to one digit more than the original data; standard deviations to two digits more. Do not report fractions rounded to single digit expressions: $\frac{131}{256} \neq \frac{1}{2}$, and do not round decimals or percents to a single digit: $0.57846 \ldots \neq 60\%$ *or* 0.6. Report a minimum of two digits, up to four, unless otherwise specified in the problem.
5. If a problem asks for an explanation, state the solution clearly, then interpret or explain in addition to stating the solution, not in place of. Explanations without solutions, just as solutions without explanations, will not be awarded full credit.

Part I: At Home

This part was completed at home. You can upload the Excel file for Part I to the Part I folder in Blackboard for use during the Exam period. However, this submission will **not** be graded in this location, it must be submitted to the "**to be graded** folder" to receive credit.

Part II: In Class (with computer)

Before completing Part III, complete Part II in class. Return the paper to your instructor and put away your computer. Then pick up Part III.

Part III: In Class (without computer)

1. You may use a handheld calculator for this portion of the exam. Any calculator is fine, as long as it is not on a device that connects to the Internet. That means, you may not use the calculator on your phone or smart watch. You may also not share calculators with another student taking the exam at the same time.
2. This is Exam D.
3. Answer the questions on the paper exam. Sign the honor code statement on the next page.
4. Turn in your paper copy of the exam to your instructor. Your instructor will attach this portion of the exam to the version of Part II that you submitted previously.

Honor Code Statement:

I, ______________________________ (print your name), agree to abide by the George Mason Honor Code and Academic Integrity Pledge: *To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University Community and with the desire for greater academic and personal achievement, I, a student member of the university community, pledge not to cheat, plagiarize, steal, or lie in matters related to academic work.* Furthermore, I have read and I agree to follow the guidelines laid out in the instructions for this exam above. I also agree not to participate in the efforts of other students to circumvent these guidelines, or to assist in their violations of the code, and will report such efforts in a timely manner.

____

Student Signature and G#

____

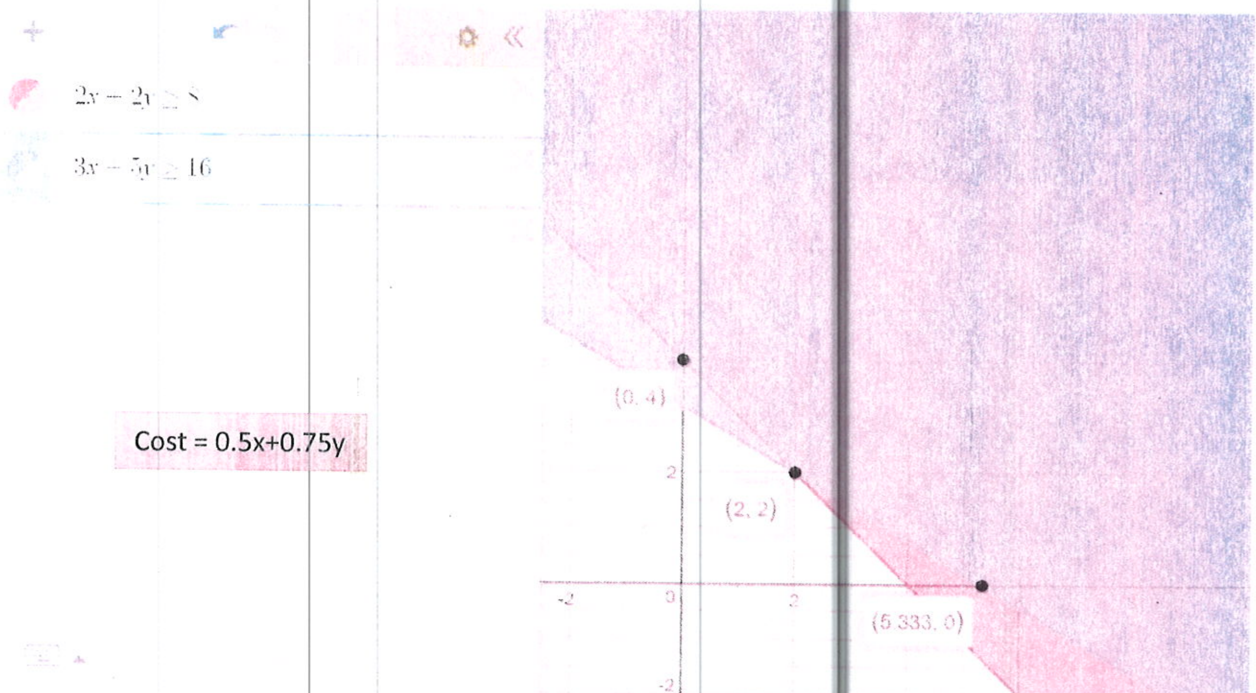
Today's Date

Part III:

1. When a hypothesis test is conducted, there are four possible combinations of outcomes: The null can be true, the null can be false, our conclusion can agree with the true state of nature, or it may not. A table of these situations is shown below. Two of these combinations are correct and two produce errors. Label all four possibilities as correct, or, if an error, which kind of error it is. (4 points)

	Conclusion: H_0 True	Conclusion: H_0 False
Nature: H_0 True	Correct	Type I
Nature: H_0 False	Type II	Correct

Below you will find the Desmos plot for the constraints for a linear programming problem. The graph shows the feasible region and its cornerpoints, the constraints graphed, and the cost function to be minimized. Not shown are the non-negativity constraints. Use this information to answer the questions that follow.



2. Based on the cost equation and the cornerpoints, find the minimum cost, and the values of the two variables produce this cost. (8 points)

$0.5(0) + 0.75(4) = 3$

$0.5(2) + 0.75(2) = 2.5$ ← Minimum @ (2,2)

$0.5(5.333) + 0.75(0) = 2.\overline{66}$

Below you will find a boxplot and output for a two-sample t-test of two brands of light bulbs. Use this information to answer the questions that follow.

t-Test: Paired Two Sample for Means

	Variable 1	*Variable 2*
Mean	100.0898	99.5594
Variance	7.958329253	39.76949661
Observations	100	100
Pearson Correlation	-0.120363277	
Hypothesized Mean Difference	0	
df	99	
t Stat	0.735458562	
P(T<=t) one-tail	0.231899543	
t Critical one-tail	1.660391156	
P(T<=t) two-tail	0.463799087	
t Critical two-tail	1.984216952	

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	100.0898	99.5594
Variance	7.958329253	39.76949661
Observations	100	100
Pooled Variance	23.86391293	
Hypothesized Mean Difference	0	
df	198	
t Stat	0.767746225	
P(T<=t) one-tail	0.221776286	
t Critical one-tail	1.652585784	
P(T<=t) two-tail	0.443552573	
t Critical two-tail	1.972017478	

	A	B	C	D
1	Brand 1		Brand 2	
2	Battery	Lifetime	Battery	Lifetime
3	1	99.11	1	110.65
4	2	99.45	2	92.24
5	3	98.39	3	96.63
6	4	97.07	4	99.45
7	5	99.97	5	102.55
8	6	100.06	6	109.60
9	7	98.20	7	96.53
10	8	98.13	8	104.64
11	9	107.73	9	88.03
12	10	95.58	10	96.87
13	11	96.98	11	96.02
14	12	100.47	12	97.33
15	13	101.23	13	105.14
16	14	100.39	14	99.86
17	15	106.07	15	93.81
18	16	98.02	16	92.26
19	17	100.26	17	103.67
20	18	102.48	18	93.61
21	19	97.88	19	97.28

3. Based on the snapshot of the data, would you conclude that that the lightbulb data collected for the two brands are dependent or independent? (4 points)

independent

4. Based on your decision above, should you use the paired two-sample t-test, or the pooled (equal variance) two-sample t-test? (4 points)

pooled equal variance

5. If the company wants to know whether Brand 2 has a different lifespan than Brand 1, does that correspond to a one-tailed or two-tailed t-test (4 points)

2-tailed

6. State the null and alternative hypotheses, the test-statistic and p-value for the test described in the previous questions. State your conclusion in language a person not familiar with statistics can understand. (8 points)

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$t = 0.7677$

p-value $= 0.44 > .05$

fail to reject null

There is not sufficient evidence to conclude the light bulbs are different

Use the data in the correlation table to answer the question that follows.

	Three-month	*Six-month*	*One-year*	*Five-year*	*Seven-year*	*Ten-year*	*Thirty-year*
Three-month	1						
Six-month	0.995018236	1					
One-year	0.980587641	0.994322863	1				
Five-year	0.824150317	0.861942602	0.896782787	1			
Seven-year	0.764325584	0.806788198	0.847048391	0.993284711	1		
Ten-year	0.708407903	0.754748746	0.799680553	0.979556259	0.995836885	1	
Thirty-year	0.581770384	0.631833095	0.682138948	0.925093011	0.960802599	0.980922012	1

7. Based on the correlation table, which variable appears to be the best predictor of 30-year market averages? Explain your reasoning. (5 points)

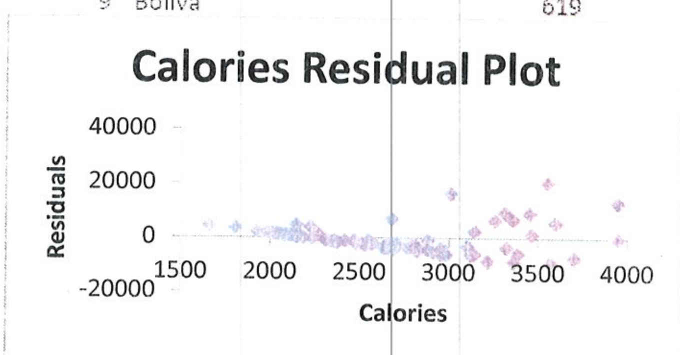
The 10-year is best predictor
highest correlation

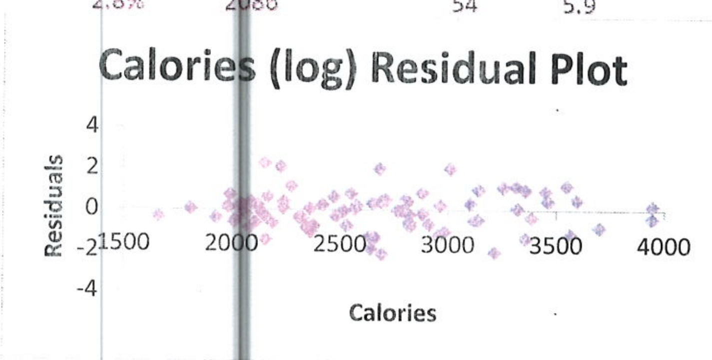
8. Why does it appear that the highest correlations for each variable on just off the diagonal? Why does this make sense in the real world? (6 points)

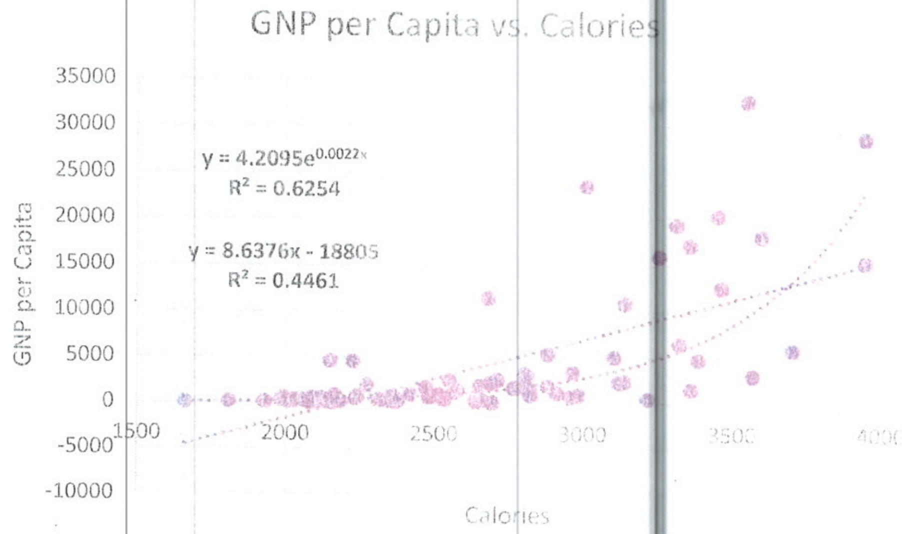
it means that generally the most recent behavior
is best predictor of what comes next.
This makes sense

Use the data in the residual graphs, scatterplots, and regression output on the pages that follow to answer the remaining questions. Included elements are a snapshot sample of all available variables used in the complete model (the first regression output). Two more regression outputs are included: Calories vs. GNP per Capita as a linear model, and Calories vs. LN(GNP per Capital) which analyses the exponential model shown in the scatterplot. The residual plots correspond to the linear and log models for the last two regression outputs respectively.

	A	B	C	D	E	F
1	Country	GNP per Capita	Population Growth	Calories	Life Expectancy	Fertility
2	Antigua_&_Barbud	4595	0.5%	2222	74	1.9
3	Argentina	2369	1.4%	3118	71	2.8
4	Bahamas	11514	1.9%	2678	69	2.2
5	Bangladesh	199	2.6%	1925	52	4.8
6	Belgium	15444	0.1%	3942	76	1.6
7	Belieze	1974	2.8%	2649	68	4.7
8	Benin	362	3.2%	2145	51	6.3
9	Boliva	619	2.8%	2086	54	5.9







9. Based on the first model containing all the variables, conduct a hypothesis test on the coefficient for Population Growth. State the null and alternative hypotheses, test-statistic and p-value. Does this test explain why the variable was removed from the final model? (8 points)

Ho: $\beta i = 0$
Ha: $\beta i \neq 0$
$t = -1.12$
p-value = 0.2656 > 0.05
fail to reject

yes. while it could change as variables are removed, this p-value is quite high and so by parsimony, it's better to assume coeff is zero.

10. Using the scatterplot, does the data appear to be linear or non-linear? (4 points)

the data appears to be either nonlinear or unequal variance
the residual plot suggests nonlinear is more likely

11. Model 2 is the linear model, and Model 3 the non-linear model. In which model is the R^2 value higher? State the higher value. Does this confirm your answer from above? (4 points)

Model 3 higher
$R^2 = 0.625$
yes

12. In the third Model, what is the proportion of variability in GNP per capita can be explained by the relationship with Calories? (4 points)

62.5%

13. Use Model 2 to predict the GNP for capita of a country with a population growth rate of 0.7%, consumes 2200 calories per day, has a life expectancy of 74 years and a fertility rate of 1.2. If the variable does not appear in the model, ignore it. (6 points)

$y = 8.638x - 18{,}805$

$= 198.6$

14. Construct a 95% prediction interval for the prediction above. You may use 1.96 as the multiplier for the prediction margin. (6 points)

margin of error/prediction $1.96 * 5174.6 =$

$10{,}142.2$

$(-9{,}943.6,\ 10{,}340.8)$

15. Report the 99% confidence interval for the Calories coefficient from Model 1 (the full model). (4 points)

$(1.544,\ 10.477)$

SUMMARY OUTPUT (model 1)

Regression Statistics	
Multiple R	0.691027162
R Square	0.477518538
Adjusted R Square	0.44965286
Standard Error	5125.104196
Observations	80

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1800470511	450117627.7	17.13644071	5.0615E-10
Residual	75	1970001977	26266693.02		
Total	79	3770472487			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 99.0%*	*Upper 99.0%*
Intercept	-16075.80961	11578.79005	-1.388384239	0.169130089	-39141.94221	6990.322983	-46678.35565	14526.73642
Population Growth	-105278.3503	93853.82864	-1.121726751	0.265557203	-292244.7645	81688.06388	-353332.4302	142775.7296
Calories	6.010881434	1.68982288	3.557107379	0.000653986	2.644581635	9.377181233	1.544708177	10.47705469
Life Expectancy	92.55000493	132.8290119	0.696760471	0.488106706	-172.0589558	357.1589657	-258.5148243	443.6148342
Fertility	105.4697533	962.3370513	0.109597519	0.9130213	-1811.603959	2022.543466	-2437.970778	2648.910285

SUMMARY OUTPUT (model 2)

Regression Statistics	
Multiple R	0.66789146
R Square	0.446079002
Adjusted R Square	0.438977451
Standard Error	5174.572795
Observations	80

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1681928606	1.68E+09	62.8143	1.31146E-11
Residual	78	2088543882	26776204		
Total	79	3770472487			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	-18804.95149	2949.810357	-6.37497	1.19E-08	-24677.5728	-12932.33019	-26593.45382	-11016.44917
Calories	8.637590872	1.089841526	7.925548	1.31E-11	6.467883064	10.80729868	5.76003868	11.51514307

SUMMARY OUTPUT (model 3)

Regression Statistics	
Multiple R	0.79081175
R Square	0.625383224
Adjusted R Square	0.620580445
Standard Error	0.90822106
Observations	80

ANOVA

	df	SS	MS	F	Significance F
Regression	1	107.4080282	107.4080282	130.2127789	2.64932E-18
Residual	78	64.33950854	0.824865494		
Total	79	171.7475367			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99.0%	Upper 99.0%
Intercept	1.437340187	0.517739337	2.77618501	0.006883209	0.406600345	2.46808003	0.070332298	2.804348077
Calories	0.002182766	0.000191285	11.41108141	2.64932E-18	0.001801947	0.002563585	0.001677709	0.002687823

Standard errors: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

$s_{x_1-x_2} = s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Sample sizes: $n > \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$ $n > \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$ $m = n = \frac{4z_{\alpha/2}^2(\sigma_1^2+\sigma_2^2)}{w^2}$

Confidence intervals:

One sample: $\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two samples (independent): $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,n-1}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Test statistics:

One sample: $z\ or\ t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

Two samples: dependent: $z\ or\ t = \frac{\bar{d}_0-\delta}{\frac{s_d}{\sqrt{n}}}$

Independent: $z\ or\ t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$

Degrees of freedom (two samples, unpooled) $\nu = \frac{\left(\frac{s_1^2}{m}+\frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$

χ^2Tests: $\chi^2 = \sum_{all\ cells}\frac{(obs-\exp)^2}{exp}$

ANOVA: $MSE = \frac{\left(\sum_{j=1}^{J} n_j(\bar{Y}_j-\bar{\bar{Y}})^2\right)}{J-1}$ $MSS = \sum_{j=1}^{J}\frac{(n_j-1)s_j^2}{n-J}$ $F = \frac{MSE}{MSS}$