Write your name on the cover of the test booklet and nowhere else. Enclose this sheet with the booklet. Failure to follow these directions will cost you 1 point. The test has 240 points (to be scaled down to 180 points) and is scheduled to take 120 minutes. Therefore, expect to spend 1 minute for every 2 points. For example, a 12-point question should take 6 minutes. I cannot give some extra time but not a lot because some students have sports they need to get to.

The formulas in Excel © which you may need to know are written as such in Excel.
NORM.INV(probability, mean, standard_dev) FREQUENCY(data_arrays, bins_arrays)crtl-shift-enter
RAND() mean $\pm$Z*standard deviation/SQRT(n) where Z = 2.58 (99%), 1.96 (95%), 1.645 (90%)
n = (Z*standard deviation/desired error)^2 P$\pm$Z*SQRT(P*(1-P)/n) n = $\pi(1-\pi)*(Z/E)^2$
left critical value = NORM.S.INV(probability) two tail critical value = ABS(NORM.S.INV(prob/2))
P-value left tail test = NORM.S.DIST(Z,cumulative) P-value right tail test = 1- left tail P-value
P-value for the two-tail test = IF(Z>0,2*(1-NORM.S.DIST(Z,cumulative)),2*NORM.S.DIST(Z,cum.))
left critical value for t is T.INV(probabilty,df) right critical value = T.INV.RT(probability,df)
two tail critical value is T.INV.2T(probabilty,df) left P-value T.DIST(t,df)
right P-value T.DIST.RT(t,df) two tail P-value T.DIST.2T(t,df) $Z=(\bar{X}_1-\bar{X}_2)/SQRT(s_1^2/n_1+s_2^2/n_2)$
$Z = (P_1-P_2)/SQRT(P_C*(1-P_c)*(1/n_1+1/n_2)$ where $P_c = (P_1n_1+P_2n_2)/(n_1+n_2)$ $Z = (\bar{X}-\mu)/(\sigma/SQRT(n))$
$t = (\bar{X}-\mu)/(s/SQRT(n))$

1) (8 points each) Answer **THREE** of the following referring to the regression results on the last page. **Briefly explain your logic for each part.**
A) Are the overall regression results good?
B) Is the coefficient for the price of the good statistically significant?
C) Are the good being analyzed and the "other good" complements, likely complements, substitutes, likely substitutes, or likely unrelated?
D) Write the equation which you would use if you were to predict the sales if the price of the good was \$20/unit, the income was \$200, and the price of the other good was \$10. Do not solve it.
E) How did the computer get df of 3 and 32?

2) (12 points) Answer EITHER Part A OR Part B.
A) Why should recommendations be part of a report? Explain your logic.
B) What goes into an appendix? Explain your logic.

3) (12 points) Answer EITHER Part A OR Part B.
A) Pearson's Correlation Coefficient, r or ρ, assumes that the samples have a *bivariate normal distribution.* What does that mean?
B) The *method of least squares* adds up the squares of what? State what that means in English and using the variables $\hat{Y}$, $\bar{Y}$, and/or Y. (You only need two of them and it is probably best to name the ones you use in case you got two of them confused. That way I can give more partial credit.)

4) (16 points) Answer EITHER Part A OR Part B.
A) Explain when you would use NORM.INV(probability, mean, standard_dev) and when you would use NORM.S.INV(probability). Explain your logic.
B) When I typed NORM.INV(0.10, 20, 4), I got 14.9. What does that number tell me? Explain your

logic.

5) (16 points) Answer EITHER Part A OR Part B.
A) Explain the difference in audience and writing format between a technical report and a managerial report. Why is the writing different? Explain your logic.
B) Explain the difference in audience and writing format between a technical report and a managerial report. Which would you use if you were writing a press release? Explain your logic.

6) (18 points) Answer EITHER Part A OR Part B.
A) Suppose you were given statistics about aggregate consumption (all consumption in the economy) and GDP. You calculated Pearson's Correlation Coefficient, r or ρ, and got .9. What does that tell us? Does it give us a causation? Explain your logic for both parts.
B) Why is the confidence band bow tie shaped? Explain your logic.

7) (18 points) Answer EITHER Part A OR Part B.
A) Suppose you are doing a test which Type I errors cause major problems but Type II errors cause minor problems. What are <u>two</u> ways you can reduce the chance of the Type I errors? Explain why <u>both</u> of them reduce the chance of making a Type I error.
B) We said the standard is to have $\alpha = 0.05$. However, there are times we might want to have a larger α. Why might we want to have a larger α? Explain your logic.

8) (18 points) Answer EITHER Part A OR Part B.
A) What is the *Law of Large Numbers*? Explain the logic about how this affects the size of the confidence interval.
B) I gave you the equation $n = \pi(1-\pi)*(Z/E)^2$. When do you use this? What does each term mean?

9) (20 points) Answer EITHER Part A OR Part B.
A) Suppose you I gave an exam to 15 students and expected an average of 85, but the average was 82. I would want to know if the poor performance is statistically significant from 85. What is the null hypothesis? Would you use a Z-test or a t-test to test this? Explain your logic. Would you use a left-tail, right-tail, or two-tail test? Explain your logic and reference one of the equations at the top of the exam.
B) Suppose you had data on the high temperatures for this date for the past 76 years. Suppose today's high was 77 and the average high was 69. You want to know if that is statistically significantly different from the norm. What is the null hypothesis? Would you use a Z-test or a t-test to test this? Explain your logic. Would you use a left-tail, right-tail, or two-tail test? Explain your logic and reference one of the equations at the top of the exam.

10) (20 points) Answer EITHER Part A OR Part B.
A) Suppose you had data on 200 members of Greek organizations on campus and 150 non-Greek students. You found the former group is 45% Republican while the latter group is 38% Republican. You want to find out if the difference is statistically significant. What is the null hypothesis? Would you use a Z-test or a t-test to test this? Explain your logic. Would you use a left-tail, right-tail, or two-tail test? Explain your logic and reference one of the equations at the top of the exam.
B) This semester, I had 27 students start ECON 162. The average grade on Exam #1 was higher than the average grade on Exam #2 for the same students in the same class. I want to know if the difference is statistically significant. What is the null hypothesis? Would you use a Z-test or a t-test to test this?

Explain your logic. Would you use a left-tail, right-tail, or two-tail test? Explain your logic and reference one of the equations at the top of the exam.

11) (20 points) Answer EITHER Part A OR Part B.
A) The table to the right is showing how much the sales of cheese is at three different locations. What does "Anova: Single Factor" mean. What is the null hypothesis? Do we reject it? Explain your logic.
B) What does the "F-Test Two-Sample for Variances" mean? The table is looking at the time to get to a Pittsburgh suburb via two paths. What is the null hypothesis? Do we reject it? Explain your logic.

Question 11A Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
near milk	5	67	13.4	7.3
near eggs	5	69	13.8	3.7
near meat	5	56	11.2	2.7

ANOVA

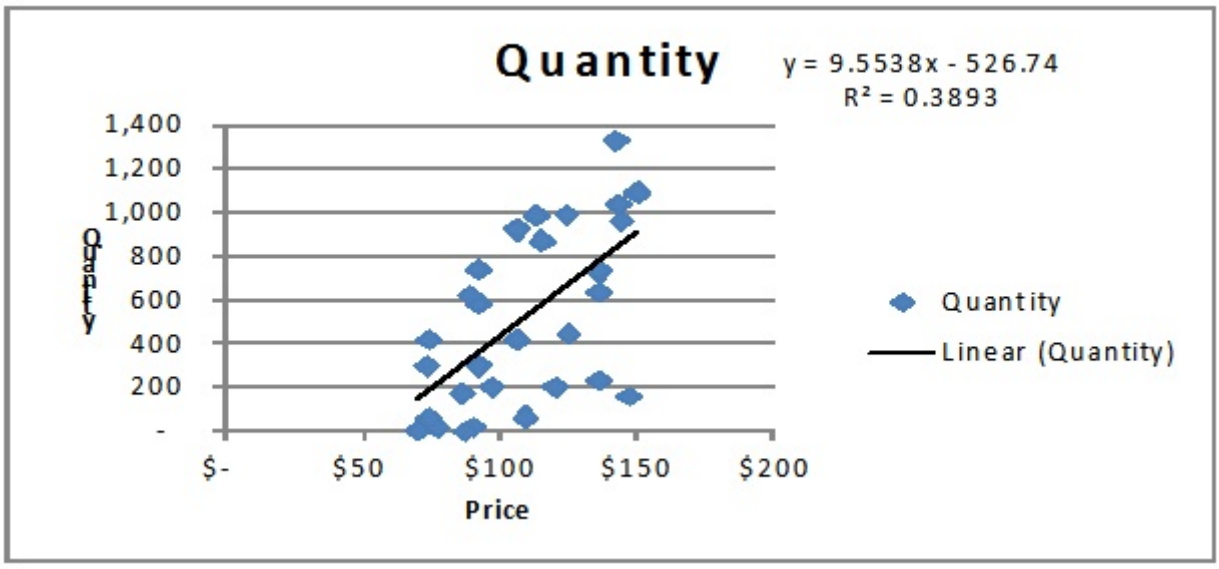
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	19.6	2	9.8	2.145985	0.159678	3.885294
Within Groups	54.8	12	4.566667			
Total	74.4	14				

Question 11B F-Test Two-Sample for Variances

	Route 50	*Wash Pike*
Mean	47.11111	48
Variance	46.36111	7.1428571
Observations	9	8
df	8	7
F	6.490556	
P(F<=f) one-tail	0.011602	
F Critical one-tail	3.725725	

12) (20 points) For the scatter diagram on the next page, are we using quantity to predict price or vice versa? Explain your logic. Would you say these results are good? Explain your logic. What would you estimate Pearson's Coefficient of Correlation (r or ρ) to be? Explain your logic.

13) (26 points) Answer EITHER Part A OR Part B.
A) Suppose you theorized that warmer cities would have more homeless people than colder cities. You have data on the number homeless people per 1000 residents and the average low temperature in January for 30 cities. What would your null hypothesis and alternative hypothesis be? Explain why you chose those. What would your Type I error be and what would your Type II error be? Explain your logic.
B) Suppose you theorized smaller colleges had less cheating than big universities. You have cheating statistics and school size for 50 small colleges and 20 big universities What would your null hypothesis and alternative hypothesis be? Explain why you chose those. What would your Type I error be and what would your Type II error be? Explain your logic.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.474445
R Square	0.225098
Adjusted R Square	0.152451
Standard Error	226.9103
Observations	36

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	478612.5013	159537.5	3.098522	0.040496773
Residual	32	1647624.44	51488.26		
Total	35	2126236.941			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	73.79738	274.201101	0.269136	0.789553	-484.7319878	632.326738
Price	-8.8336	6.012667796	-1.46916	0.151551	-21.08100357	3.41380335
Income	3.031799	7.451691618	0.406861	0.686819	-12.14679985	18.2103983
Other P	16.76455	8.313499869	2.016546	0.0522	-0.169490796	33.6985992