ADM 2304

STATISTICS FOR MANAGEMENT 2

April 29, 2006, 19:00-22:00

Final Exam

NAME: _____

SIGNATURE: _____

STUDENT NUMBER: _____

SECTION:	A	(Wed. 13:00, Fri. 11:30)	Prof: YURI KHOROSHILOV
(Circle One)	B	(Tue. 16:00, Thu. 14:30)	Prof: YURI KHOROSHILOV
	C	(Tue. 16:00, Thu. 14:30)	Prof: JOHN C NASH
	D	(Wed. 10:00, Fri. 8:30)	Prof: LINDA SCHWEITZER
	E	(Mon. 11:30, Thu. 13:00)	Prof: SUREN PHANSALKER
	F	(Tue. 19:00)	Prof: SAAD KIRYAKOS
	G	(Mon. 19:00)	Prof: LINDA SCHWEITZER

Time allowed: 3 hours **Length: 11 pages.**

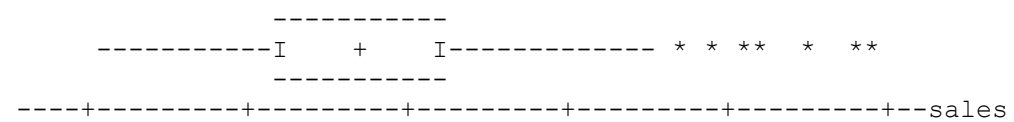
Instructions:

- Calculators, rulers, and one sheet of notes (8.5 x 14 in.) are allowed.
- **You must write your solutions and answers on the answer sheet provided (any work shown on the question booklet will NOT be marked)**
- Please write clearly and legibly.
- You may write your intermediate work on the exam question booklet (this work will not be marked)
- Statistical tables excerpts (z, t, F, chisquare) are provided
- You must hand everything in at the end of the exam (answer sheet, exam question booklet, stats tables, notes sheet).

Good luck and have a great summer!

Question	Value
1	14
2	15
3	19
4	19
Total	**67**

Question 1. Maxitronics has 544 stores across North America. A boxplot of the sales at the 544 stores is shown below. Use the .05 level of significance for all applicable parts of this question.



Descriptive statistics:

Variable	N	Mean	Median	StDev
sales	544	75.348	74.079	15.095

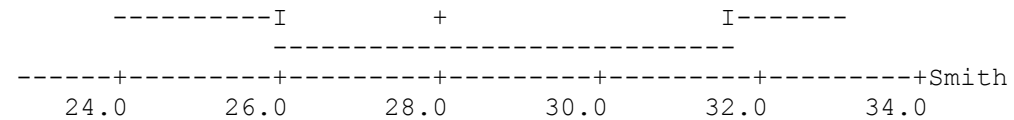
a) (2) Construct a 95% confidence interval for the mean sales per store.

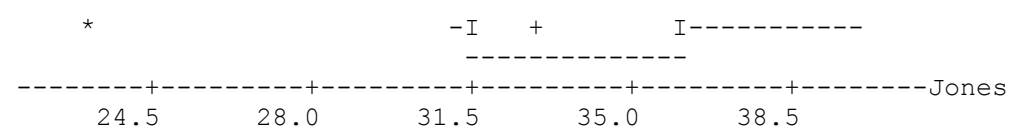
b) (2) What assumptions, if any, do you need to make to construct the interval in (a)? From the information supplied, are you satisfied that the calculation gives an appropriate interval?

c) (3) The CFO wants to use the median sales as his estimate of sales. Is it OK to use the median instead of the mean to estimate total sales? Explain why or why not using a hypothesis test.

d) (5) For eight weeks in a row, the Maxitronics stores in SmithVille and Jonestown have had the sales indicated below. Are the mean sales for Jonestown higher than those for Smithville? Perform an appropriate hypothesis test to answer this question. Be sure to indicate the type of test you are performing and justify any assumptions needed (if you can).

Row	Smith	Jones	Diff(J-s)
1	25	34	9
2	27	31	4
3	24	23	-1
4	32	32	0
5	33	38	5
6	29	40	11
7	27	33	6
8	31	33	2





e) (2) The manager in Smithville points out that, for one week out of the eight, his store did better than the Jonestown store and he claims that this implies that one cannot say his store is doing less well than Jonestown. Explain how the manager arrives at his claim. (Hint: If both stores have equal earnings in the long term, what is probability you observe 1 time when Smithville does better than Jonestown?)

Question 2. In 1995, research in Ottawa showed that 5% of all males under age 21 in the Ottawa – Carleton region suffered from asthma. Researchers now believe that increases in air pollution have led to increases in asthma. A 2005 study examined 768 males under 21 and found that 92 of them showed signs of some form of asthma. Use the .01 level of significance for all applicable parts of this question.

a) (3) Use this information to test the hypothesis that the level of asthma has increased since 1995.

b) (3) A parallel 2005 investigation in Sudbury involving 433 males under 21 showed 33 with signs of asthma. Using a hypothesis test, can you conclude that the rate of asthma in this group is different in the two locations?

c) (4) In developing countries, asthma is not a problem, but tuberculosis and COPD (obstructive lung disease) are. In fact, it is suggested that for men in the 22-49 age group, 2% will have asthma, 7% will have TB, and 4% will have COPD or lung cancer, with the rest "healthy" as far as their lungs are concerned. A study in an unnamed developing country (Country1) collected data on 1634 men of the appropriate age. The outcomes are below. Is this data for Country1 consistent with the suggested distribution?

Disease	Number
Asthma	40
TB	110
COPD/cancer	50
"healthy" lungs	1434

d) (5) Data for another oceanic country were also gathered. Based on the data given above and below, are the two developing countries different from the perspective of their distribution of lung disease? Partial MiniTab output is provided in **Appendix A.**

Disease	Number
Asthma	55
TB	80
COPD/cancer	80
"healthy" lungs	1350

Question 3. You would like to estimate the daily gasoline sales at the Shell gas station on Eagleson Road in Kanata. You have collected the daily average prices for the Shell station (in cents per liter) as well as the daily sales data (in liters) for the last 100 days. The corresponding MiniTab output is presented in **Appendix B**, MODEL 1.

a) (2) Referring only to the corresponding graphs, do you think the underlying assumptions of the regression analysis are satisfied? Explain briefly.

b) (2) Calculate R-squared for Model 1.

c) (2) Are the daily gas prices at Shell significant in predicting the daily sales? Use the .05 level of significance.

d) (1) You realize that the sales of gas at this Shell station may depend also on how competitive their gas price is. To better estimate the sales of gas, you include 2 additional variables in the model: 1) the daily gas price at the neighboring Mobil gas station; and 2) Shell's price advantage over the average daily price in the Ottawa region (calculated as the average daily price in Ottawa minus Shell's daily price - in cents per liter). The results of the corresponding Minitab analysis are given in Appendix A, MODEL 2. Which of the two models is better? Give 2 reasons.

e) (1) Is Model 2 Significant? Use the .05 level of significance. Support your answer.

f) (2) Which of the slopes in Model 2 are significant? Use the .05 level of significance. Support your answer.

g) (4) Regardless of your answers above, provide a point estimate as well as 90% prediction interval estimate for the daily gasoline sales at the Shell gas station on a weekday when their gas price is 90 cents/liter, the price at Mobil is 90.2 and they have a 0.4 cent advantage over their Ottawa area competitors. There are marks for showing your work briefly.

h) (3) Regardless of your answers above, do you think a 1-cent increase in the Price Advantage at the Shell station will decrease the sales by more than 30,000 liters (use a 5% level of significance)

i) (2) Compare the two models with respect to the explanatory variable "Shell". What reason might there be for the changes in the slope and pvalue?

Question 4. Hydroponics Inc grows tomatoes in hothouses even in the middle of ferocious Canadian winters. They have kept detailed records of the 'Dozens' of tomatoes produced in their facility with two factors: 'Light' and 'Temperature'. 'Light' keeps track of 3 different intensities of light and 'Temperature' keeps track of 4 different settings of temperature. **Appendix C** presents the data as well as 4 ANOVA models.

a) (2) Referring to Model 1, is the production of tomatoes related to the intensity of light? Give your answer in the form of a hypothesis test. Use the .05 level of significance.

b) (3) Referring to Model 1, compute the multiple comparisons (Bonferroni) margin of error to determine which of the Light intensities are significantly different. Use the .03 level of significance. Remember to state which are different.

c) (2) By looking **only** at the interaction plot, does it appear that there is an interaction between the factors 'Light' and 'Temperature' in terms of their effects on the 'Dozens' of tomatoes being produced? Explain VERY briefly.

d) (2) Based on the ANOVA analysis for Model 3, test whether there is an interaction between 'Light' and 'Temperature' in terms of their effects on the 'Dozens' of tomatoes produced. Use the .05 level of significance.

e) (3) Of the four models given, which model is the best model? Explain with evidence from the MINITAB output as well as your reasoning.

f) (2) Regardless of your answers above, your boss tells you to use Model 4. Use this model and specify which factor explains the most **variation** in the 'Dozens' of tomatoes being produced. Justify with evidence from the MINITAB output. Test this factor and determine if it is significant in the production of tomatoes. Use the .05 level of significance.

g) (1) Model 4 has two factors, so test whether the other factor (i.e. the factor not tested in f) is significant in the production of tomatoes. Use the .05 level of significance.

h) (2) Based on the 'Residual Plots' given for Model 4, argue why Model 4 can be taken as appropriate.

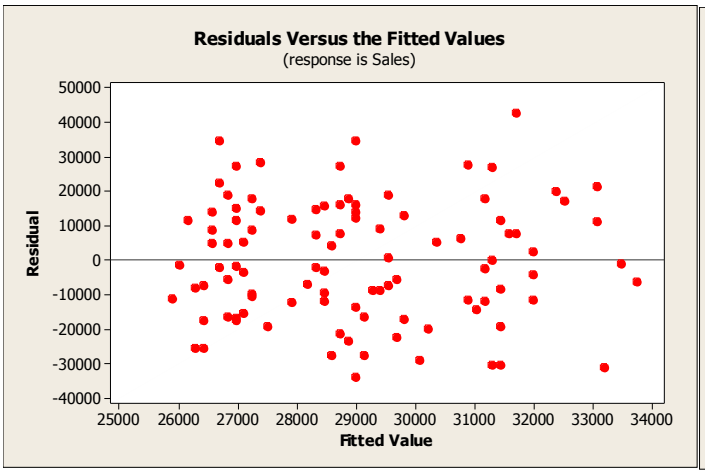
i) (2) Provide the Model 3 estimate (NOTE: MODEL 3) of the Dozens of tomatoes produced with Light Level 2 and Temperature setting 2.

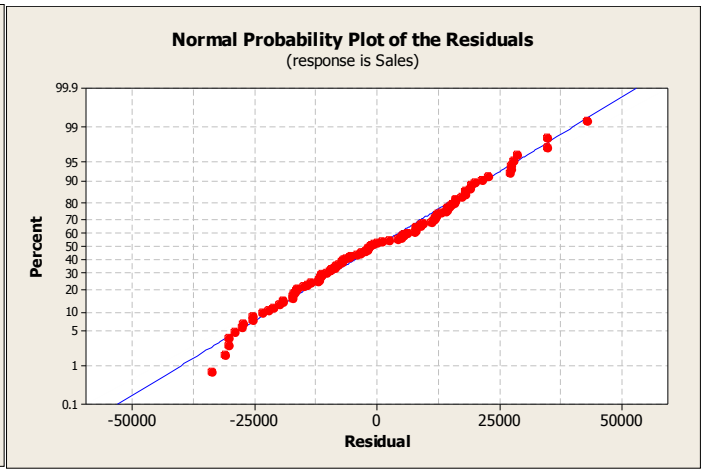
APPENDIX A (Question 2)

Expected counts are printed below observed counts

	COUNTRY1	COUNTRY2	Total
Asthma	40	55	95
	48.52		
TB	110	80	190
	97.05	92.95	
CODP	50	80	130
		63.60	
Healthy	1434	1350	2784
	1422.02		
Total	1634	1565	3199

APPENDIX B (Question 3)





Model 1:

Regression Analysis: Sales versus Shell

The regression equation is
Sales = 40599 - 135 Shell

Predictor	Coef	SE Coef	T	P
Constant	40599	9769	4.16	0.000
Shell	-135.4	113.3	_____	_____

S = 17076.9 R-Sq = _____ R-Sq(adj) = 0.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	416469543	416469543	_____	_____
Residual Error	98	28578868590	291621108		
Total	99	28995338133			

Model 2:

Correlations: Shell, Mobil, Price Advantage

	Shell	Mobil
Mobil	0.999 0.000	
Price Advant	0.101 0.316	0.140 0.164

Regression Analysis: Sales versus Shell, Mobil, Price Advantage

The regression equation is
Sales = 41892 - 50174 Shell + 50018 Mobil - 38184 Price Advantage

Predictor	Coef	SE Coef	T	P	VIF
Constant	41892	8266	5.07	0.000	
Shell	-50174	13473	-3.72	0.000	19828.3
Mobil	50018	13481	3.71	0.000	20018.2
Price Advantage	-38184	14056	-2.72	0.008	32.1

S = 14417.9 R-Sq = 31.2% R-Sq(adj) = 29.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	9039334671	3013111557	14.49	0.000
Residual Error	96	19956003462	207875036		
Total	99	28995338133			

Source	DF	Seq SS
Shell	1	416469543
Mobil	1	7088806802
Price Advantage	1	1534058327

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	_____	2894	(16839, 28328)	(-6607, 51773)

Values of Predictors for New Observations

New Obs	Shell	Mobil	Price Advantage
1	90.0	90.2	0.400

APPENDIX C (Question 4)

Data: Light, Temperature

		Temperature			
		1	2	3	4
Light	1	696	440	808	536
		624	536	784	696
		736	480	712	560
	2	536	480	720	432
		616	440	696	536
		464	384	784	448
	3	712	560	832	624
		624	536	880	648
		736	656	784	632

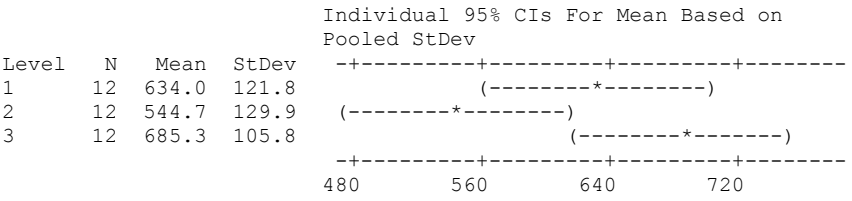
Cell Contents: Dozens

Model 1

One-way ANOVA: Dozens versus Light

Source	DF	SS	MS	F	P
Light	2	121611	60805	_____	_____
Error	33	471797	14297		
Total	35	593408			

S = 119.6 R-Sq = 20.49% R-Sq(adj) = 15.68%



Pooled StDev = 119.6

Model 2

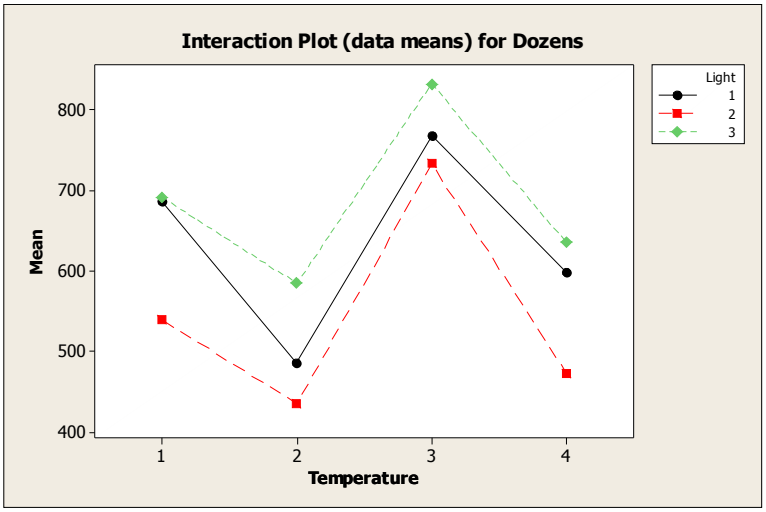
One-way ANOVA: Dozens versus Temperature

Source	DF	SS	MS	F	P
Temperature	3	378041	126014	18.72	0.000
Error	32	215367	6730		
Total	35	593408			

S = 82.04 R-Sq = 63.71% R-Sq(adj) = 60.30%

```
                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level  N    Mean  StDev  ---------+---------+---------+---------+----
1      9  638.22  93.28               (-----*-----)
2      9  501.33  80.60  (-----*-----)
3      9  777.78  59.97                             (-----*-----)
4      9  568.00  90.16        (-----*-----)
                          ---------+---------+---------+---------+----
                             500       600       700       800
```

Pooled StDev = 82.04



Model 3:

Two-way ANOVA: Dozens versus Light, Temperature

Source	DF	SS	MS	F	P
Light	2	121611	60805	_____	_____
Temperature	3	378041	126014	_____	_____
Interaction	6	16231	2705	_____	_____
Error	24	77525	3230		
Total	35	593408			

S = 56.84 R-Sq = 86.94% R-Sq(adj) = 80.95%

Model 4:

Two-way ANOVA: Dozens versus Light, Temperature

Source	DF	SS	MS	F	P
Light	2	121611	60805	_____	_____
Temperature	3	378041	126014	_____	_____
Error	30	93756	3125		
Total	35	593408			

S = 55.90 R-Sq = 84.20% R-Sq(adj) = 81.57%

Residual Plots for Model 4: