Question 1 (22 marks)

A commercial pool management company is seeking to reduce the costs of maintaining swimming pools. The maintenance cost is mostly driven by the amount of chlorine that is needed to keep the water clear and relatively free of germs. The company hired a new graduate from York University ("YorkU") to analyze how much chlorine is needed. It is believed that the speed at which chlorine is depleted is dependent on the temperature (degrees Fahrenheit) of the water (higher temperature uses chlorine faster); which is a measure of the acidity of the water (pH ranges from 0 to 14, where 0 is very acidic and 14 is very alkaline; levels around 7.5 use the least chlorine); and weather (sunshine uses up chlorine). The YorkU graduate recorded the percentage of chlorine depletion during 8-hour days, water temperature, pH level and weather condition (mainly cloudy, sunny and partly sunny). Using Excel's Regression Tool, the YorkU graduate obtained the following Excel output.

Regression Statistics	
Multiple R	**A**
R Square	**B**
Adjusted R Square	**C**
Standard Error	**D**
Observations	210
ANOVA	

	df	*SS*	*MS*	*F*	*Significance F*
Regression	E	6596	**I**	77	0.0000
Residual	F	**H**	**J**		
Total	G	10091			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1003	55.12	18.09	0.0000
Temperature	0.194	0.029	**K**	0.0000
PH Level	-265.6	14.75	-18.01	0.0000
PH Level-Sq	17.76	0.983	18.07	0.0000
Mainly Cloudy	-1.07	0.7	-1.53	0.1282
Sunny	1.16	0.7	1.65	0.0997

1. Clearly define the independent variables being considered.

2. Specify the multiple regression model using all independent variables being considered.

3. Fill in the following table with the correct values of E, F and G.

	ANSWERS
E – Regression	
F – Residual	
G-Total	

4. Fill in the blanks of A, C, D, H, I and K of the ANOVA in the following box. Show all your calculations for each of them. ***Marks will be given ONL* Y *IF appropriate computational work is shown completely in the space below.*** It is recommended in this case that you show computational work to **four digits after the decimal point**.

	FINAL ANSWERS TO BE INSERTED HERE
A	
C	
D	
H	
I	
K	

5. Using a hypothesis test, determine whether the model is significant overall. Use significance level α = 5%. Specify completely the null and alternative hypotheses for testing the overall validity of the multiple regression model. Explain why the model is valid at α = 5%-or why not, as the case may be.

6. Interpret the coefficient of each of the indicator variables in the sample regression equation.

7. Predict the amount of chlorine depletion when the temperature is 75 degrees Fahrenheit and pH level is 6.6 on a partly sunny day.

Question 2 (18 marks)

Othello Bakeries is open for business five days a week. The number of loaves of specialty bread sold daily for the last four weeks is given in the following table.

Week	Business Day	Sales			
1	Monday	310			
	Tuesday	380			
	Wednesday	370			
	Thursday	515			
	Friday	630			
2	Monday	310			
	Tuesday	270			
	Wednesday	375			
	Thursday	405			
	Friday	530			
3	Monday	195			
	Tuesday	245			
	Wednesday	260			
	Thursday	340			
	Friday	400			
4	Monday	280			
	Tuesday	305			
	Wednesday	375			
	Thursday	520			
	Friday	490			

Required:

(a) Is there a seasonal effect over the five-day week? Support your answer with the seasonal indexes for the five business days.

(b) Using the deseasonalized data, the following values were derived.

$\sum t = 210$ $\sum t^2 = 2{,}870$ $\sum Y_t = 7{,}549.75$ $\sum tY_t = 76{,}214$

Develop the equation for the linear trend component for the time series.

(c) Use the equation developed in part (b) to prepare the trend projections for Week 5

(d) Use the seasonal indexes computed in part (a) to adjust the trend projections prepared in part (c) to account for seasonal effects.

Question 3 (20 marks)
A construction company is considering bidding on one of *two* projects for the upcoming season. The first project could potentially lead to a profit of $100,000 if the market return is low or a profit of $300,000 if it is high. Project 2 on the other hand has the potential to return profits of $400,000 if the market return is low or $200,000 if it is high. The probability of the market return being high is 40%.

a) Should the construction company select project 1 or project 2?

b) What is the Expected value of Perfect Information?

c) The report classifies the market condition as either Favorable (F) or Non-favorable (N). There is a 56% chance that the market is classified as Favorable and a 44% chance that it is classified as Non-favorable. Given a favorable report, the market condition will be high with probability 57%. Given the report is Non-favorable the probability of a low market condition is 82%.Construct a decision tree for the whole problem. Include all probabilities and payoffs.

d) What is the optimal decision strategy?

e) How much is the team willing to pay for Pre-Construction Survey Report? (Calculate EVSI)

Question 1 (22 marks)

A commercial pool management company is seeking to reduce the costs of maintaining swimming pools. The maintenance cost is mostly driven by the amount of chlorine that is needed to keep the water clear and relatively free of germs. The company hired a new graduate from York University ("YorkU") to analyze how much chlorine is needed. It is believed that the speed at which chlorine is depleted is dependent on the temperature (degrees Fahrenheit) of the water (higher temperature uses chlorine faster); which is a measure of the acidity of the water (pH ranges from 0 to 14, where 0 is very acidic and 14 is very alkaline; levels around 7.5 use the least chlorine); and weather (sunshine uses up chlorine). The YorkU graduate recorded the percentage of chlorine depletion during 8-hour days, water temperature, pH level and weather condition (mainly cloudy, sunny and partly sunny). Using Excel's Regression Tool, the YorkU graduate obtained the following Excel output.

Regression Statistics	
Multiple R	**A**
R Square	**B**
Adjusted R Square	**C**
Standard Error	**D**
Observations	210
ANOVA	

	df	*SS*	*MS*	*F*	*Significance F*
Regression	E	6596	**I**	77	0.0000
Residual	F	**H**	**J**		
Total	G	10091			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	1003	55.12	18.09	0.0000
Temperature	0.194	0.029	**K**	0.0000
PH Level	-265.6	14.75	-18.01	0.0000
PH Level-Sq	17.76	0.983	18.07	0.0000
Mainly Cloudy	-1.07	0.7	-1.53	0.1282
Sunny	1.16	0.7	1.65	0.0997

1. Clearly define the independent variables being considered.

$$x_1 = water\ temperature$$

$$x_2 = PH\text{-}Level$$

$$x_2^2 = (PH\text{-}Level)^2$$

$$I_1 = \begin{cases} 1\ if\ weather\ is\ mainly\ cloudy \\ 0\ otherwise \end{cases}$$

$$I_2 = \begin{cases} 1\ if\ weather\ is\ sunny \\ 0\ otherwise \end{cases}$$

2. Specify the multiple regression model using all independent variables being considered.

$$\textbf{Population Model}: \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + \beta_4 I_2 + \beta_5 I_2 + \varepsilon$$

$$\textbf{Estimate}: \quad \hat{y} = 1003 + 0.194 x_1 - 265.6 x_2 + 17.76 x_2^2 - 1.07 I_2 + 1.16 I_2$$

3. Fill in the following table with the correct values of E, F and G.

	ANSWERS
E – Regression	5
F – Residual	204
G-Total	209

[E] = df (Regression) = k = 5 (Number of independent variables)

[F]= df (Residual or Error) = n – k – 1 = 210 – 5 - 1 = 204

[G]= df (Total) = n – 1 = 209

4. Fill in the blanks of A, C, D, H, I and K of the ANOVA in the following box. Show all your calculations for each of them. ***Marks will be given ONL* Y *IF appropriate computational work is shown completely in the space below.*** It is recommended in this case that you show computational work to **four digits after the decimal point**.

	FINAL ANSWERS TO BE INSERTED HERE
A	0.8085
C	0.6619
D	4.1391
H	3495
I	1319.2
K	6.6897

[H] = SSE = SSTotal – SSR = 10091 – 6596 = 3495

[I] = MSR = SSR/k = 6596/5 = 1319.2

[J] = MSE = SSE/(n-k-1) = 3495/204 = 17.1324

[D] = $s_{\varepsilon} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{3495}{204}} = \sqrt{17.13235} = 4.1391$

[B] = R^2 = SSR/SSTotal = 6596/10091 = 0.6537

[A] = R = $\sqrt{R^2} = \sqrt{0.6537} = 0.8085$

[C] = Adj. $R^2 = 1-\left[\frac{n-1}{n-k-1}\right]\left(1-R^2\right) = 1-\left[\frac{209}{204}\right](1-0.6537) = 1-\left[\frac{209}{204}\right](1-0.6537) = 0.6619$

[K] = $t = \frac{b_i}{s_{b_i}} = \frac{0.194}{0.029} = 6.6897$

5. Using a hypothesis test, determine whether the model is significant overall. Use significance level α = 5%. Specify completely the null and alternative hypotheses for testing the overall validity of the multiple regression model. Explain why the model is valid at α = 5%-or why not, as the case may be.

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: At least one of the β_is is not 0.

$\alpha = 0.05$

Rejection region is $F > F_{\alpha,k,n-k-1} = F_{0.05,\,5,204} \approx 2.258$

Since F-stat = 77 > 2.258, we reject the null hypothesis. There is enough evidence to infer that the model is valid.

6. Interpret the coefficient of each of the indicator variables in the sample regression equation.

$$I_1 = \begin{cases} 1 \; if \; weather \; is \; mainly \; cloudy \\ 0 \; otherwise \end{cases}$$

b_4 = -1.07; percentage of chlorine depletion is 1.16 lower when the weather is mainly cloudy than when it is partly sunny, provided other variables are held constant.

$$I_2 = \begin{cases} 1 \; if \; weather \; is \; sunny \\ 0 \; otherwise \end{cases}$$

b_5 = 1.16; percentage of chlorine depletion is 1.16 higher when the weather is sunny than when it is partly sunny, provided other variables are held constant.

7. Predict the amount of chlorine depletion when the temperature is 75 degrees Fahrenheit and pH level is 6.6 on a partly sunny day.

The estimated model: $\hat{y} = 1003 + 0.194x_1 - 265.6x_2 + 17.76x_2^2 - 1.07I_1 + 1.16I_2$

Predicted amount of chlorine depletion

$\hat{y} = 1003 + 0.194(75) - 265.6(6.6) + 17.76(6.6)^2 - 1.07(0) + 1.16(0) = 38.2156$

Question 2 (18 marks)

Othello Bakeries is open for business five days a week. The number of loaves of specialty bread sold daily for the last four weeks is given in the following table.

Week	Business Day	Sales			
1	Monday	310			
	Tuesday	380			
	Wednesday	370			
	Thursday	515			
	Friday	630			
2	Monday	310			
	Tuesday	270			
	Wednesday	375			
	Thursday	405			
	Friday	530			
3	Monday	195			
	Tuesday	245			
	Wednesday	260			
	Thursday	340			
	Friday	400			
4	Monday	280			
	Tuesday	305			
	Wednesday	375			
	Thursday	520			
	Friday	490			

Required:

(e) Is there a seasonal effect over the five-day week? Support your answer with the seasonal indexes for the five business days.

Yes. There is a seasonal effect over the five-day week. The seasonal indexes are above and below 1.

Week	Business Day	Sales	CMA	Seasonal Irregular	Seasonal Index
1	Monday	310			0.7244
	Tuesday	380			0.7620
	Wednesday	370	441	0.8390	0.9279
	Thursday	515	441	1.1678	1.1492
	Friday	630	419	1.5036	1.4366
2	Monday	310	420	0.7381	
	Tuesday	270	398	0.6784	
	Wednesday	375	378	0.9921	
	Thursday	405	355	1.1408	
	Friday	530	350	1.5143	
3	Monday	195	327	0.5963	
	Tuesday	245	314	0.7803	
	Wednesday	260	288	0.9028	
	Thursday	340	305	1.1148	
	Friday	400	317	1.2618	
4	Monday	280	340	0.8235	
	Tuesday	305	376	0.8112	
	Wednesday	375	394	0.9518	
	Thursday	520			
	Friday	490			

Seasonal Irregular Values

Week	Monday	Tuesday	Wednesday	Thursday	Friday	
1			0.8390	1.1678	1.5036	
2	0.7381	0.6784	0.9921	1.1408	1.5143	
3	0.5963	0.7803	0.9028	1.1148	1.2618	
4	0.8235	0.8112	0.9518			Total
Average	0.7193	0.7566	0.9214	1.1411	1.4266	4.965
Adjusted	0.7244	0.7620	0.9279	1.1492	1.4366	5.000

(f) Using the deseasonalized data, the following values were derived.

$$\sum t = 210 \qquad \sum t^2 = 2{,}870 \qquad \sum Y_t = 7{,}549.75 \qquad \sum tY_t = 76{,}214$$

Develop the equation for the linear trend component for the time series.

Solution:

$$\bar{t} = \frac{\sum t}{n} = \frac{210}{20} = 10.5 \qquad \bar{Y}_t = \frac{\sum Y_t}{n} = \frac{7549.75}{20} = 377.4875$$

$$b_1 = \frac{\sum tY_t - (\sum t \sum Y_t)/n}{\sum t^2 - (\sum t)^2/n} = \frac{76214 - (210 * 7549.75)/20}{2870 - (210)^2/20} = \frac{-3058.375}{665} = -4.5991$$

$$b_0 = \bar{Y} - b_1 \bar{t} = 377.4857 - (-4.5991)(10.5) = 425.7776$$

Equation for the linear trend: $\hat{y} = 425.7776 - 4.5991x$

(g) Use the equation developed in part (b) to prepare the trend projections for Week 5

Solution:

Trend projections for Week 5

	t	**trend projection**
Monday	21	$\hat{y} = 425.7776 - 4.5991(21) = 329.1965$
Tuesday	22	$\hat{y} = 425.7776 - 4.5991(22) = 324.5974$
Wednesday	23	$\hat{y} = 425.7776 - 4.5991(23) = 319.9983$
Thursday	24	$\hat{y} = 425.7776 - 4.5991(24) = 315.3992$
Friday	25	$\hat{y} = 425.7776 - 4.5991(25) = 310.8001$

(h) Use the seasonal indexes computed in part (a) to adjust the trend projections prepared in part (c) to account for seasonal effects.

Solution:

	t	**trend projection**	**× Seasonal Index**	**= Seasonal Forecast**
Monday	21	329.1965	× 0.7244 =	**238.4699**
Tuesday	22	324.5974	× 0.762 =	**247.3432**
Wednesday	23	319.9983	× 0.9279 =	**296.9264**
Thursday	24	315.3992	× 1.1492 =	**362.4568**
Friday	25	310.8001	× 1.4366 =	**446.4954**

Question 3 (20 marks)
A construction company is considering bidding on one of *two* projects for the upcoming season. The first project could potentially lead to a profit of $100,000 if the market return is low or a profit of $300,000 if it is high. Project 2 on the other hand has the potential to return profits of $400,000 if the market return is low or $200,000 if it is high. The probability of the market return being high is 40%.

c) Should the construction company select project 1 or project 2?

Solution:

	Low	High	EV
Project 1	$100,000	$300,000	$100,000*.6 + $300,000*.4 = $180,000
Project 2	$400,000	$200,000	$400,000*.6 + $200,000*.4 = $320,000
Probability	0.6	0.4	

The construction company should select project 2 since EV(Project 2) is higher.

d) What is the Expected value of Perfect Information?

EVPI = EVwPI – EVwoPI
EVwPI = 0.6(400,000) + 0.4(300,000) = $360,000
EVPI = $320,000 - $360,000 = $40,000

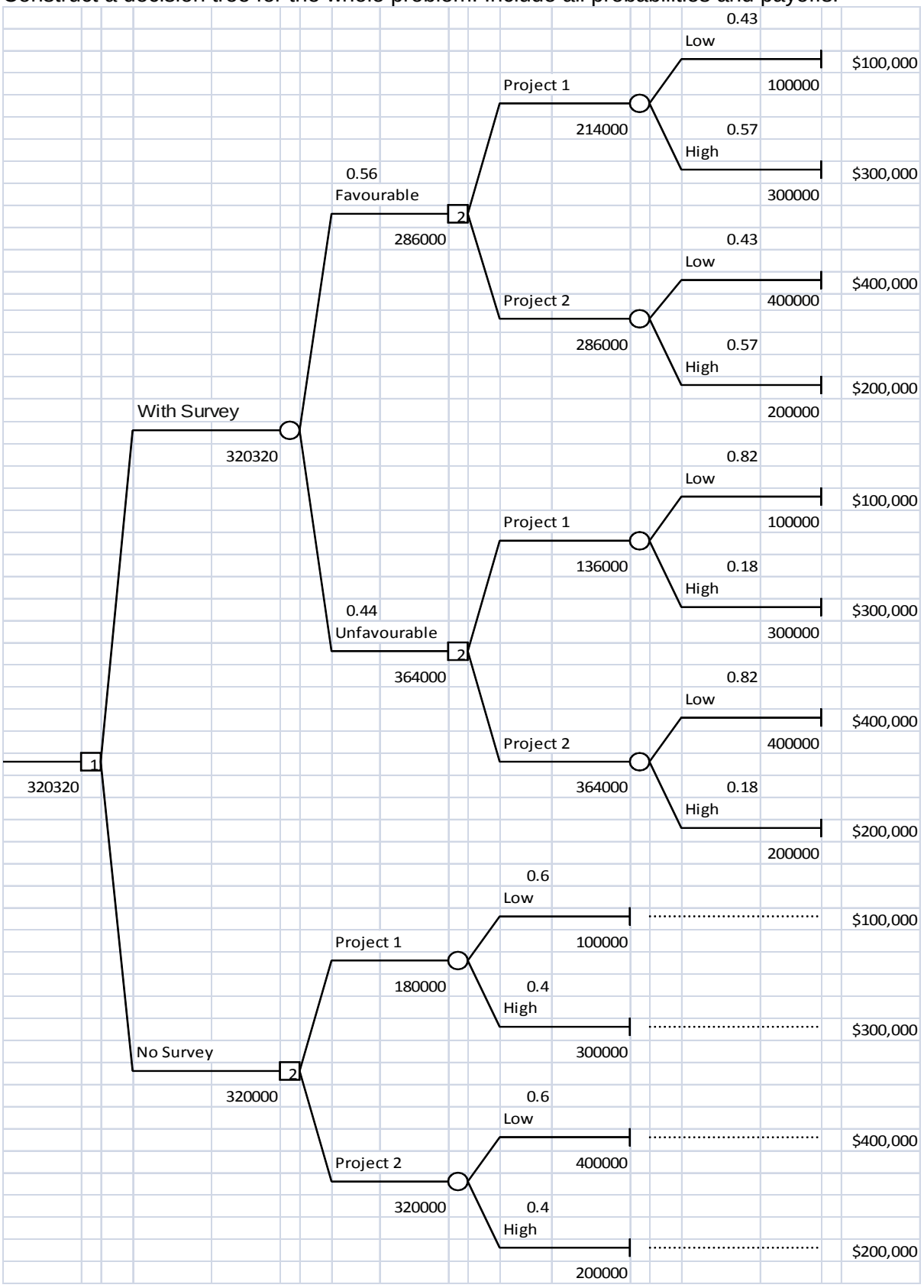
The company can buy a pre-construction survey report which assesses future market conditions.

The report classifies the market condition as either Favorable (F) or Non-favorable (N). There is a 56% chance that the market is classified as Favorable and a 44% chance that it is classified as Non-favorable. Given a favorable report, the market condition will be high with probability 57%. Given the report is Non-favorable the probability of a low market condition is 82%.

Favourable (F)				
State	**Prior**	**Conditional**	**Joint**	**Posterior**
Low	**0.6**			***0.43***
High	**0.4**			**0.57**
	1		**0.56**	**1**

Non-favorable (N)				
State	**Prior**	**Conditional**	**Joint**	**Posterior**
Low	**0.6**			***0.82***
High	**0.4**			**0.18**
	1		**0.44**	**1**

e) Construct a decision tree for the whole problem. Include all probabilities and payoffs.



d) What is the optimal decision strategy?

The optimal decision strategy will be to conduct the survey if its cost is less than EVSI ($320) and select project 2.

e) How much is the team willing to pay for Pre-Construction Survey Report? (Calculate EVSI)

EVSI = EVwSI – EvwoSI = 320320 – 320000 = $320