

Statistics for Engineering, 4C3/6C3

Assignment 7

Kevin Dunn, kevin.dunn@mcmaster.ca

Due date: 09 April 2013

Note: Assignment objectives

- Performing back-of-the-envelope calculations from designed experiments; and using these models.
- Using and interpreting MLR models from designed experiments.
- Have practice with questions at the level of the final exam. All questions here come from the 2012 final exam.

Question 1 [11 = 2 + 4 + 5]

From the final exam, 2012

Using a 2^3 factorial design in 3 variables (**A** = temperature, **B** = pH and **C** = agitation rate), the profit, y , from a chemical reaction was recorded, in standard order.

Experiment	A	B	C	y
1	–	–	–	72
2	+	–	–	73
3	–	+	–	66
4	+	+	–	87
5	–	–	+	70
6	+	–	+	73
7	–	+	+	67
8	+	+	+	87

- $A = \frac{\text{temperature} - 150^\circ\text{C}}{10^\circ\text{C}}$
- $B = \frac{\text{pH} - 7.5}{0.5}$
- $C = \frac{\text{agitation rate} - 50\text{rpm}}{5\text{rpm}}$

1. Show a cube plot for the recorded data.
2. Estimate the main effects and three 2 factor interactions by hand. Use the fastest way possible, without just writing the final answer.
3. Interpret all the significant factors you identify in part 2 of this question. Clearly explain what any of the significant 2 factor interactions imply and how it can be used to your advantage to improve the process profitability.

Question 2 [16 = 1 + 2 + 2 + 3 + 4 + 4]

From the final exam, 2012

Experiments were conducted by varying the temperature, **T**, and catalyst level, **C**, in order to find conditions that lead to improved conversion, y .

- **T** = 350 K at the low level and 360 K at the high level.
- **C** = 3% catalyst at the low level and 7% catalyst at the high level.

The following experimental data were collected, using coded form:

Experiment	T	C	y
1	0	0	53
2	-1	-1	36
3	+1	-1	45
4	-1	+1	41
5	+1	+1	60
6	0	0	49
7	1.41	0	52
8	0	1.41	49
9	-1.41	0	41
10	0	-1.41	38
11	0	0	51

The following output was obtained when building a model in R from all 11 data points:

Call:

```
lm(formula = y ~ T + C + T * C + I(T^2) + I(C^2))
```

Residuals:

```

      1      2      3      4      5      6      7      8      9
-1.8297  1.2604 -0.7336  2.3564 -2.4565 -1.0423  1.9266  0.5124  3.0021
     10     11
-0.9979 -1.9979
```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  49.9979      1.5518   32.219 5.41e-07 ***
T              5.4550      0.9517    5.732 0.00226 **
C              4.4520      0.9517    4.678 0.00544 **
I(T^2)       -1.6261      1.1356   -1.432 0.21159
I(C^2)       -3.1351      1.1356   -2.761 0.03980 *
T:C           2.5000      1.3439    1.860 0.12193
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

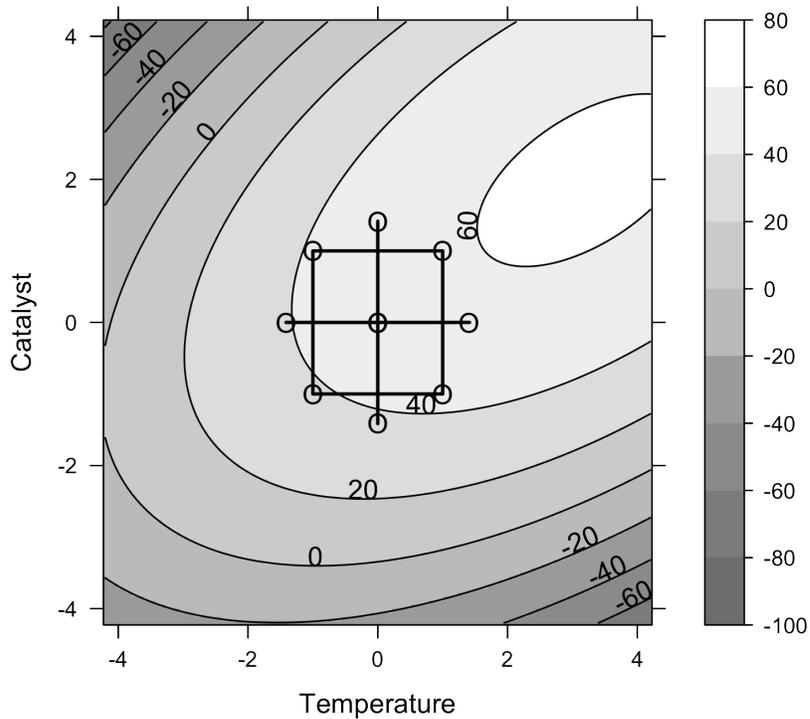
Residual standard error: 2.688 on ___ degrees of freedom

Multiple R-squared: 0.9298, Adjusted R-squared: 0.8596

F-statistic: 13.25 on 5 and 5 DF, p-value: 0.006562

1. How many degrees of freedom are available to estimate confidence intervals?
2. Calculate the confidence interval for factor **T** at the 95% level using the above software output.
3. Why might the experimenters have included runs 1, 6 and 11?
4. After finishing experiments 1 to 6, **what** would the experimenters have seen in their data from experiments 1 to 6 to make them decide to spend extra time and money to add the next 4 experiments? Provide any necessary calculations to justify your answer (do not perform any calculations from the data for experiments 7, 8, 9 and 10).
5. Using the above R model, the contour plot shown here was generated and the contours extrapolated. The 11 experiment locations are superimposed.

Contour plot for factors T and C (coded units)



At which conditions of **T** and **C**, in real-world units, would you run the next experiment to see if you can achieve maximum conversion? Explain how you choose your values of **T** and **C**.

- Calculate the predicted conversion at your chosen conditions of **T** and **C**; also give an approximate prediction interval for your prediction.

Question 3 [10 = 3 + 1 + 2 + 4]

From the final exam, 2012

- Give the generators *and* defining relationship, in terms of factors **A**, **B**, **C**, **D**, **E**, and **F**, for a set of fractional factorial experiments using 6 factors, in the fewest number of runs. However, we require that main effects not be confounded with two factor interactions.
- How many experiments would be required?
- What is the resolution and projectivity of these experiments?
- In general, list some advantages of fractional factorial designs and describe how these designs should be used in practice.

Question 4 [600-level student question: 10 = 7 + 3 (though I strongly recommend all students attempt this question)]

From the final exam, 2012

Biological drugs are rapidly growing in importance in the treatment of certain diseases, such as cancers and arthritis, since they are designed to target very specific sites in the human body. This can result in treating diseases with minimal side effects. Such drugs differ from traditional drugs in the way they are manufactured – they are produced during the

complex reactions that take place in live cell culture. The cells are grown in lab-scale bioreactors, harvested, purified and packaged.

These processes are plagued by low yields which makes these treatments very costly. Your group has run some experiments to learn more about the system and find better operating conditions to boost the yield. The following factors were adjusted in the usual factorial manner:

- **G** = glucose substrate choice: a binary factor, either **G**– at the low level code or **G**+ at the high level.
- **A** = agitation level: low level = 15 rpm and high level = 25 rpm, but can only be set at integer values.
- **T** = growth temperature: 30°C at the low level, or 35°C at the high level, and can only be set at integer values in the future.
- **C** = starting culture concentration: low level = 1100 and high level = 1500, and can only be adjusted in multiples of 50 units.

A fractional factorial in 8 runs, created by aliasing **C = GAT**, has given the following 8 model coefficients, when **C**, **G**, **A** and **T** are centered and scaled (coded) in the usual factorial way:

- **I + GATC** = 24
- **G + ATC** = +3.5
- **A + GTC** = –1.5
- **T + GAC** = +4.0
- **C + GAT** = +3.5
- **GA + TC** = –0.18
- **GT + AC** = –0.09
- **GC + AT** = +0.20

The aim is to find the next experiment that will improve the yield, measured in milligrams, the most. Since your manager has seen that temperature has a strong effect, she has requested the next experiment be run at 40°C, which also happens to be the highest level you can adjust the bioreactor to.

1. Give the experimental conditions for all 4 factors for the next experiment. The conditions are to be reported in both real-world units, as well as in the usual coded units of the experiment, presented in a table.
2. What is the expected yield at your proposed experimental conditions?

END