STA 2004F Final Homework. Please work alone.

Due: Dec 19 at 5 p.m. (or sooner).

Submit all the computer code you used as an appendix, but do not submit raw code as your solutions to the problems. You may cut and paste the relevant bits, or retype or rewrite them.

1. An experiment was conducted to study the effects of drying time in hours and temperature in degrees Centigrade on the content of undesirable compounds in a resin. The response is measured in parts per million. The data are given below.

Table 1: Resin impurity content y (ppm); adapted slightly from DV, p.590. Coded values of time and temperature are $x_1 = (time - 7)/4$, $x_2 = (temp - 190)/30$.

Run No.	Time	Temp	x_1	x_2	y
1	7.0	232.4	0	1.41	18.5
2	3.0	220.0	-1	1	22.5
3	11.0	220.0	1	1	17.2
4	1.3	190.0	-1.425	0	42.2
5	7.0	190.0	0	0	28.6
6	7.0	190.0	0	0	19.8
7	7.0	190.0	0	0	23.6
8	7.0	190.0	0	0	24.1
9	7.0	190.0	0	0	24.2
10	12.7	190.0	1.425	0	19.1
11	3.0	160.0	-1	-1	54.1
12	11.0	160.0	1	-1	33.8
13	7.0	147.6	0	-1.41	55.4

The design is a 2^2 factorial, augmented with points at $(0, 0)$, called *center points* and points at $(0, \pm 1.41)$ and $(\pm 1.425, 0)$, called *axial points*. There are 5 center points and 4 axial points.

(a) Ignoring for this part the coded values, fit a linear regression model $y_i = \beta_0 + \beta_1 time_i + \beta_2 temp_i + \epsilon_i$ under the usual second moment assumptions, and report estimates of β_1, β_2 and their standard error. Construct an analysis of variance table showing the sum of squares due to each of time and temperature and residual.

(b) Expand the model fit in part (a) to include the three quadratic terms in $time^2$, $temp^2$, and $time * temp$, and show the corresponding analysis of variance table.[1]

[1] In R use the operator I to introduce squared and product terms: $lm(y \sim time + temp + I(time^2) + I(time * temp) + I(temp^2))$.

Next separate from the residual sum of squares the variation among the center points: i.e. $\sum_{i=5}^{9}(y_i - (1/5)\sum_{i=5}^{9} y_i)^2$. This is sometimes called 'pure error', as it is the observed variability between responses at the same treatment level. Your analysis of variance table should now have 5 lines: *time*, $time^2$, *temp*, $temp^2$, $time \times$ *temp*, *pure error* and *residual*.

(c) Now construct a different analysis of variance table, using the coded values as factor variables. Fit a model with main effects and interactions. How does this differ from the models fit above?

2. (DV 15.12): Suppose you are asked to design an experiment for 6 treatment factors each having two levels. Only 64 observations can be taken in total, and these should be divided into 8 blocks of size 8. Suppose that you decide to confound the interaction contrasts ABD, DEF and $ACDF$.

   (a) Can all the other interaction contrasts be estimated?

   (b) What does the statement "ABD is confounded" mean?

   (c) How would you obtain the 8 blocks? Write out two blocks as an example.

   (d) Suppose the budget is cut before the experiment can take place, and only 8 observations can be taken in total. How would you decide which 8 observations to take? What can be estimated?

3. In the *Applied Statistics* article handed out in the last class, "Modifying a central composite design..." by Kowalski et al, the motivating example is discussed in Sections 2 and 5. Referring to the experimental design laid out in Table 1 on p.624, consider just the first four whole plots.

   (a) What is the treatment structure applied to the whole plots?

   (b) What is the treatment structure applied to the sub plots?

   (c) How do the sixteen runs given in this part of the table differ from a completely randomized 2^4 factorial experiment?

   (d) Why was this particular design chosen for this experiment?

4. A client has come to the consulting service to discuss designing an experiment on peoples' ability to respond to visual stimuli in the presence of various distractions. She wants to know what the difference is between a randomized block design, a Latin square design, and a balanced incomplete block design. She also wants to know which design is best for her experiment. Describe in one or two paragraphs how you would go about helping her design this experiment.

5. Bonus question: Analyse the data given in Table 1 of Kowalski et al referred to in Question 3, and provide an executive summary of your main conclusions. (Ignore the discussion in the paper of the "10-step algorithm".)