

Statistics for Engineers, 4C3/6C3

Take-home final exam and DOE project

Kevin Dunn, dunnkg@mcmaster.ca

Due date: 26 March 2012, 16:00

The purpose of this final exam is no different to other exams. The only difference is that this exam will also assess your ability to answer realistic problems that require more time and thought, will require you to read up on other, similar statistical tools learned in the course, and require using computer software. Much in the same way you will solve day-to-day problems as an engineer when you graduate soon.

The questions in this exam use actual data sets, which is one of the aims of this course. You are expected to use any appropriate tools to solve the problems in this exam, particularly the tools learned in this course. You may, however, use any software packages and tools to help solve the problems, as long as they are appropriate.

Important notes

- Complete this exam with 1 other person. Groups greater than 2 will not be accepted.
- You don't need to do the exam in a group, please work on your own, if you prefer.
- 600-level students must complete this exam on their own.
- Identify all your group members *and sources of reference* in your answer submission. **Note: one paper/printed submission per group.**
- The intention of the group work is that you discuss the questions and collaborate with your group in the same way you have done with the assignments.
- **Do not** share or discuss anything with other groups. Take special care with electronic files (e.g. Word documents, source code, Excel files) to safeguard your work.
- Like any other exam, neither the TA nor myself are able to answer direct questions about the exam. Similarly you should not look for help about a specific question from other resources (e.g. asking for help with the question on a website, friends, etc).
- You may use the course notes and any other textbooks though.
- There is no make-up, nor extended time granted for this exam (e.g. you cannot use your late day credits).
- Your answers should preferably be typed up.
- Hand out date: 19 March 2012; hand in date: on, or before, **16:00 on 26 March 2012** in printed form, at the Chemical Engineering office, JHE, 374.

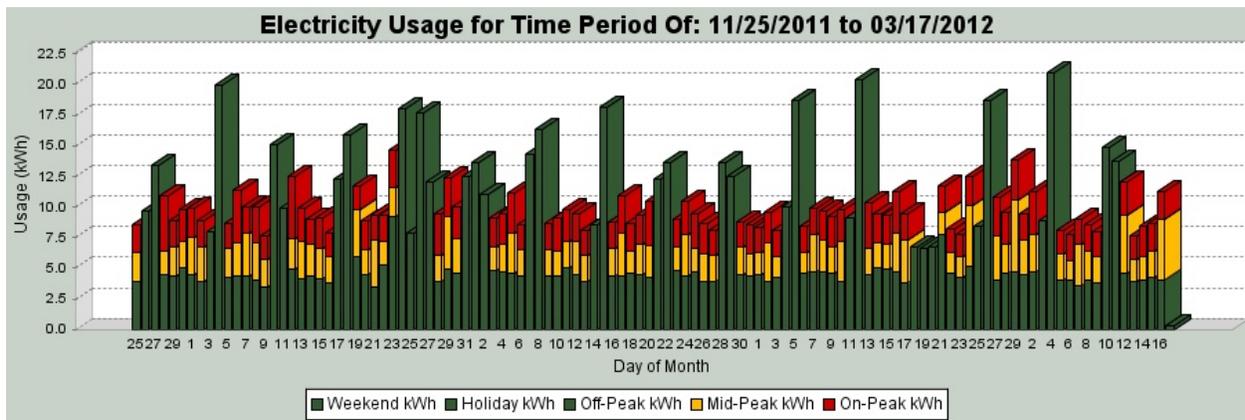
Grading

The 400-level students will be graded out of 100 points; 600-level students will be graded out of 105 points, and 600-level students are expected to show insight and technical accuracy at the graduate student level. 400-level students will get credit for answer 600-level questions.

This take-home final exam and DOE project jointly count 25% of the overall course grade.

Question 1 [8]

Horizon Utilities, the local electricity supplier in Hamilton, makes data from your smart meter available to you via their website. Here's the electricity usage for the past few months from my house:



1. Describe the short comings of Horizon Utilities' visualization.
2. The raw data that Horizon Utilities used to create this plot is given on the [dataset website](#). Use this data to construct a useful data visualization, not necessarily the same as theirs.

Question 2 [8]

A supplier has been providing an important powder raw material to you for many years. Four important characteristics of the material are provided on their certificate of analysis, and have been made available on the [dataset website](#).

The supplier guarantees the four characteristics to be within these limits:

- Impurity: 0 and 1 percent
- Particle size X40: 1.80 and 2.40 microns
- Particle size X80: 4.00 and 5.70 microns
- Compressibility: 45.00 and 59.00 percent

Is this supplier capable at the 6-sigma level? You must show your calculations and clearly explain any assumptions you make, including why they might be reasonable assumptions or not.

Question 3 [24 + 5 (for 600-level students)]

A group of students in the 4C3/6C3 class (2011) studied some factors that might affect the consistency of home-made yogurt. Milk, the starting raw material, has almost the same consistency as water. Some starter yoghurt, essentially bacterial culture, is added to the milk to start fermentation. The viscosity increases once the fermentation process curdles the milk and forms yoghurt.

Viscosity (or yoghurt consistency) was measured as the time taken for a marble to fall from the top to the bottom of a narrow glass cylinder filled with the yoghurt. The time duration was measured in triplicate, but only the first two drop values are reported, because shear-thinning caused the third drop to be much shorter than the first two.

The following factors were manipulated:

- **A:** The container material used during fermentation: metal or plastic, because the students had found conflicting recipes for making yoghurt that insisted the container could make a difference.
- **B:** Amount of starting bacterial culture added to the milk: 1 or 3 tablespoons of commercially available yogurt.
- **C:** Milk fat content: 1% or 3.5% milk.
- **D:** Temperature to which the milk is heated just before adding the starter bacterial culture: 25°C or 40°C.

The experimental results are given below in a table.

1. Which other **variables** would you have considered varying as factors? Describe clearly how you would vary each of your additional factors.
2. Which covariates would you have identified and measured in these experiments?
3. Which variables affecting the response might have required blocking?
4. Which variables would you need to control to a constant level?
5. The experiments were obviously not run in randomized order. Why not? Clearly explain the risk that this decision incurred.
6. Draw two cube plots (**A** = plastic and **A** = metal) using the **B**, **C** and **D** factors on the cube axes. What do you observe regarding each of the 4 main effects?
7. Fit a least squares model to the data and determine which factors and their interactions affect yoghurt consistency based on a Pareto plot. Does this match the cube plots?
8. Refit the least squares model after removing insignificant effects, and calculate confidence intervals for the remaining effects. Does it match the initial selection of effects from the Pareto plot?
9. Explain clearly why the model coefficients did not change when you refit the model.
10. A yoghurt consistency corresponding to 10 seconds for the marble to drop is desirable from an organoleptic and mouth-feel point. If the cost of starting bacterial culture is negligible, which settings would you recommend to use in the future to achieve this value? Pay careful attention to, and explain, the **BD** interaction term in the context of this question.
11. This set of experiments required a significant amount of work to complete in a full factorial manner. Generate the half fraction of experiments using the **A = BCD** generator. Then select the 8 experiments, which would have been used in this half fraction, from the full set of 16 below. Build a new least squares model on the half-fraction data, and compare it with the model used in parts 7 and 8 of this question.

Explain how and why (or how not and why not) you would have learned the same information from the half fraction experiments?
12. **600-level students:** Instead of using the average of the marble drop durations for each experiment, fit a least-squares model using the raw $n = 32$ observations, still fitting $k = 16$ parameters. What is the advantage of this approach?

Experiment	Order	A	B	C	D	Marble drop duration (s)
1	4	-	-	-	-	0.9 and 0.8
2	2	+	-	-	-	0.9 and 0.8
3	3	-	+	-	-	3.5 and 3.0
4	1	+	+	-	-	2.4 and 1.9
5	8	-	-	+	-	0.7 and 0.8
6	5	+	-	+	-	0.9 and 0.6
7	7	-	+	+	-	7.0 and 7.7
8	6	+	+	+	-	12.8 and 10.4
9	14	-	-	-	+	4.5 and 4.5
10	16	+	-	-	+	3.8 and 4.8
11	13	-	+	-	+	5.0 and 4.6
12	15	+	+	-	+	4.2 and 4.2
13	9	-	-	+	+	10.9 and 9.8
14	11	+	-	+	+	10.0 and 10.0
15	12	-	+	+	+	10.1 and 9.0
16	10	+	+	+	+	8.4 and 8.3

Question 4 [20]

This question must be handed in at class, or electronically, on Thursday, 29 March, at or before 18:30. Late hand-ins will not be accepted as the solution will be given in class on Thursday.

Soap is commercially manufactured at high volumes; being able to maximize the profit from this process, even by a small amount, can lead to large financial gains. The process involves sodium hydroxide and fats, which are mixed together and heated to between 90 and 110°C. Saponification occurs to create soap, after which the soap is precipitated by adding sodium chloride.

There are some constraints on the system:

- T = saponification temperature, between 80 to 120°C
- D = contact time of the reactants before precipitation, between 30 and 60 minutes
- C = sodium chloride concentration use to precipitate: low or high
- $3T + 5D \leq 600$ is a safety constraint that avoids simultaneous high temperatures and long reaction durations.

The baseline conditions are at $T = 98^\circ\text{C}$, $D = 35$ minutes using $C = \text{high NaCl concentration}$. The 3 factors can be moved anywhere, as long as they obey the constraints. You have a budget of 25 experiments (about \$5,000 per experiment), and your objective is to find a new operating point that gives the highest profit, measured in cents per kilogram of product.

You are expected to use all the tools learned in this course to solve this problem; in particular: using clear visualization plots, such as contour or gradient plots and interaction plots, linear models, design of experiments and particularly response surface methods.

A simulation of the process has been computerized, and is available on the [course website](#). Nominate *one of your group members* and email their name and student number to the course instructor, together with the name of the other group member. This will activate an account for your group. Once you sign into the account you will be able to specify the levels of the 3 factors and the server will return the response (i.e. the server will “run” the experiment for you).

The grading for this question will be marked mostly on the **systematic methodology** used to approach the optimum.

Since the cost of each experiment is so high, you must report your approach clearly, justifying to your manager why you chose *every experiment's conditions*, and what you planned to do with that new result. In particular, you should predict the result of the next experiment before you run it (of course this doesn't apply to the first few experiments). Then use that result in the way you planned, and see if it met your expectations. Please reread this paragraph again.

You might realize after you complete this question that you would have done things differently. If so, report what you would have done.

Your final answer must report:

1. why you decided to stop with the particular number of experiments you actually ran,
2. the optimum operating levels for factors T , D , and C that you will recommend to your manager
3. give the expected profit at this optimum
4. describe why you are convinced you are at or near the optimum, showing a plot of the expected contours at the optimum
5. provide a detailed list of things you learned about this process as you were doing the experiments.

The grade you earn for this question will be further adjusted according by adding (or subtracting) the following amount:

$$5.0 \times \frac{\text{Your optimum} - \text{Baseline}}{\text{True optimum} - \text{Baseline}} - 0.25N + 3.0, \text{ where } N \text{ is your number of experiments.}$$

Please note:

- There is error in the response variable in the order of 0.3 to 1.5 c/kg; please take this into account.
- For the same levels of T , D and C , the simulation will return different results for different groups.

- Please enter your conditions carefully – if you use the wrong settings you will have to work with those results.
- You must wait 1.5 hours between each experimental condition; please plan your time accordingly.
- The server will also keep track of and display all your previous experiments on a results sheet.
- Do not try to retroactively justify your experiments. The order in which the experiments were performed is clear from the time stamps on the results sheet. “Trial and error” is not a systematic methodology to approach the optimum, and is wasteful of your budget.

Once you have completed the question, print out the result sheet and submit that with your answer. (The true optimum and operating point for the optimum will be available after the exam is handed in).

DOE project [40]

As described in more detail in [the project handout](#), the grading will be for:

- Describe your objective for the system under investigation. What is/are the outcome variable/s you are investigating; how are they measured? [4]
- Outline the factors that you expect will influence the outcome variable. How will you measure the factors, over what range will you vary them? State how you expect each factor to affect the response(s); do you expect any interactions? [5]
- Disturbances: which factors are known to affect the response but not being investigated here? How do you control for them? [4]
- Plan an experimental program that will change the system’s factors and control for disturbances. Be specific on how you chose your design and the levels of the factors. [9]
- Execute the experimental program, logging all relevant details (e.g. experiments that are “weird”, unusual events). Take photos/keep a log sheet. Scan and attach this as an appendix. [3]
- Analyze the experimental results using the tools introduced in the course. [11]
- The conclusions, related back to your original objectives. What would be the next set of experiments you run? [4]

The project is handed in separately, either on 26 March or 29 March, at your preference. Please submit a printed hand-in, *but also email an electronic copy of your DOE report* (PDF preferred) to me.

END