

Statistics for Engineering, 4C3/6C3

Assignment 5

Kevin Dunn, kevin.dunn@mcmaster.ca

Due date: 12 March 2014

Note: Once again, I strongly recommend you submit this [assignment electronically](#) (see instructions on the course website), so that you can practice using the electronic system for the course project.

Note: For this assignment you will benefit from studying the [R tutorial on vectors and matrices](#).

Question 1 [12]

A company has 3 reactors that are identical. Typical production schedules split the raw material equally between the 3 reactors. Data [on the website](#) contain the brittleness values of the product produced from the three reactors for the past few days.

1. Compare the brittleness values between reactors TK104 and TK107, using a regular test for differences we learned about earlier. Feel free to use the `t.test(...)` function, but make sure you can get the same results by hand.
2. What is the interpretation of your confidence interval?
3. Next, build a least squares model where the brittleness values are predicted using a single integer variable, d , which is coded as 0 for TK104, and coded as 1 for TK107. *Hint* use the `c(...)` function in R to combine vectors, and use the `numeric(...)` function to create vectors

Report the R^2 and standard error values for the model.

4. Calculate the slope coefficient for variable d and report a confidence interval for it.

What is your interpretation of the confidence interval?

Question 2 [12]

A factorial experiment is used to investigate settings to *minimize* the production of an unwanted side product. Two factors being investigated are called **A** and **B** for simplicity, but are:

- **A** = reaction temperature: low level was 440 K, and high level was 450 K
- **B** = amount of surfactant: low level was 8 kg, high level was 12 kg

A full factorial experiment was run, randomly, on the same batch of raw materials, in the same reactor. The recorded amount, in grams, of the side product was:

Experiment	Run order	A	B	Side product formed
1	2	440 K	8 kg	89 g
2	1	450 K	8 kg	268 g
3	3	440 K	12 kg	179 g
4	4	450 K	12 kg	448 g

1. Write out a least squares model that will predict the amount of side product formed given the settings for **A**, **B** and the **AB** interaction.
2. Write out the **X** matrix and **y** vector that can be used to estimate the model coefficients using the equation $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$.

- Solve for the coefficients of your linear model, by using $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ directly.

Show your calculations that you've done by hand.

Feel free though to compare your answer to R, Minitab, Excel, or other software.

- Give a clear interpretation of the slope coefficient of **A** and the slope coefficient for **B**.
- What happens when you try to calculate confidence intervals? Explain clearly.

Question 3 [12]

We considered data from a lab-scale bioreactor, y , earlier in the course. In class, we looked at an example where the reactor temperature, batch duration, impeller speed and reactor type (one with with baffles and one without) were used to judge the effect on yield, y .

Here are the data once again, and [on the website](#):

Temp = T [°C]	Duration = d [minutes]	Speed = s [RPM]	Baffles = b [Yes/No]	Yield = y [g]
82	260	4300	No	51
90	260	3700	Yes	30
88	260	4200	Yes	40
86	260	3300	Yes	28
80	260	4300	No	49
78	260	4300	Yes	49
82	260	3900	Yes	44
83	260	4300	No	59
64	260	4300	No	60
73	260	4400	No	59
60	260	4400	No	57
60	260	4400	No	62
101	260	4400	No	42
92	260	4900	Yes	38

- Demonstrate that you get the same regression slope when building the following two models:
 - a model using only temperature to predict yield;
 - as in part (a) above, but first mean center the temperature vector;
 - as in part (b) above, but also mean center the yield vector.
- Next, build a linear model to predict the yield from all remaining variables. See the [R tutorial](#) for help to build and interpret linear models containing integer variables.
 Show your model, and interpret each variable in the model. If you are using R, then the `confint(...)` function will be helpful as well.
- What is the predicted yield for a new batch, operating at 95°C for 260 minutes, at a speed of 4000 rpm in a tank with no baffles?