KEY QUESTION ANSWERS
(DOE WEEK 1)

KQ1: What does KISS stand for?

This stands for keep it simple statistically. DOE can be very mathematical and statistical in nature, however, by using rules of thumb the statistical complexity can be reduced so that non-statisticians can apply DOE. See appendix M for rules of thumb in DOE.

KQ2: What does confounding mean?

See the glossary. Two factors (inputs) are considered to be confounded when their test profiles contain the same pattern of test settings, thus making it impossible to evaluate the two factors independently.

KQ3: What kind of effect is missed if you do one factor at a time (OFAT) experimentation?

You will miss interaction effects. It is widely accepted that most processes are dominated by main effects (effect of an input) and 2-way interactions (the combined effect of 2 inputs). Some chemical processes often have important 3-way interactions. Thus the ability to estimate interactions is often important, so OFAT is a bad idea.

KQ4: What two quality tools should be used **before** undertaking a designed experiment?

To help minimize extraneous variation and promote experimental discipline, the experimenter should build a process flow diagram and a cause and effect diagram. This is best done in a cross-functional team environment. See additional reading for more information on flow charting and cause and effect diagrams.

KQ5: Based on customer focus groups it has been determined that customers are only willing to wait for up to 10 seconds for a web page to load. If the page does not load quickly they skip to another site. You just designed a homepage that you want customers to be able to quickly access when they click on a link from a search engine. You've tested the page over various machines that your customers typically have. You've found that the average load time is 5 seconds and the standard deviation is 2 seconds. What is the process capability (C_{pk})? How acceptable is this?

From the problem we have: USL = 10 seconds, $\bar{y} = 5$ seconds, $\hat{\sigma} = 2$ seconds. Since we only have a one-sided SPEC limit we do not need to worry about the minimum.

$$C_{pk} = \left(\frac{\text{upper specification} - \text{process center}}{3 \times \text{process variability}}\right) = \left(\frac{\text{USL} - \bar{y}}{3 \times \hat{\sigma}}\right) = \left(\frac{10-5}{3 \times 2}\right) = \left(\frac{5}{6}\right) = 0.83$$

Since the Cpk is less than 1, by conventional standards this is unacceptable. To be "World Class" we would like to see a value of 1.5

KQ6: Where should specification limits come from?

Specification limits should come from customers.

KQ7: To improve process capability and/or reduce loss. What two actions are needed?

Improvement will occur when we center the process (reduce off target distance $|\bar{y} - T|$) and reduce process variability $\hat{\sigma}$. In simple terms, we need to hit the target and hit it often.

KQ8: How much time should be designated for planning the experiment?

Up to 50% of your energy should be spent on the planning phase of the experiment.

KQ9: When brainstorming for a designed experiment is complete, what should you have to show for it?

- A detailed process map / flow chart.
- List of outputs of interest and their customer driven specifications.
  Note: Some texts will call the outputs: y-variable, response variables, outcomes, and dependent variables.
- A cause and effect diagram (fishbone) showing all the possible inputs.
  Note: Some texts will call the inputs: x-variable, factors, predictors, and independent variables.
- Every input should be identified as a C=constant factor, N=noise factor, or X=experimental factor. This is often an iterative process and may take some time to decide which factors should be experimented with.
- For every X input that we want to experiment with we should know the range over which we wish to experiment.
- For every C input we should develop standard operating procedures (SOPs) to ensure those inputs are truly controlled and held constant.

KQ10: Typically, do you need more output data for continuous responses or for binary responses?

You typically need more data for pass/fail, sell/don't sell, defect/non defect (i.e binary data).

KQ11: How many experimental conditions (runs) would you have if you tested 4 inputs (factors) each at 2 levels (lo and hi)?

Runs = #levels$^{\#\text{inputs}}$ = $2^4 = 2 \times 2 \times 2 \times 2 = 16$

KQ12: Why do we code experimental inputs to –1 and +1?

Experimental inputs are on different scales. Coding puts everything on the same scale so that we can compare the relative importance of the inputs. Coding also helps simplify the calculations involved with analyzing experiments. Although in today's modern computer age this is not that big of a deal. We'll learn later that coding also helps us with a mathematical property known as orthogonality.

KQ13: In a design matrix, how do you obtain the coding for the interactions?

You simply multiply the columns. For example if the design matrix is as follows:

Run	Input A	Input B
1	+1	+1
2	-1	-1
3	-1	-1
4	+1	-1

The AB interaction would be:

Run	**Input A**	**Input B**	**AB**
1	+1	+1	(+1)x(+1) = +1
2	-1	-1	(-1)x(-1) = +1
3	-1	-1	(-1)x(-1) = +1
4	+1	-1	(+1)x(-1) = -1

KQ14: Why do we try to randomize in our experiments?

Randomization guards against uncontrolled noise factors. By randomizing the experimental tests, the effects of the noise factors are spread throughout the experiment as opposed to being concentrated in one area. In essence we are averaging out the effect of the uncontrolled noise factors.

KQ15: In simple terms, what do we mean by "orthogonal design?"

The design is vertically balanced.
Each column has an equal number of –1's and +1's. The column values will sum to 0.

and

The design is horizontally balanced.
For each level of a column, there is an equal number of –1's and +1's in the remaining columns.

KQ16: Define the term "aliased" in reference to design of experiments.

From the glossary, when two factors or interaction terms are set at identical levels throughout the entire experiment (i.e., the two columns are 100% correlated).