Notes:

g GE Power Systems

Lurking Variables

- *Definition*:*
  - *A variable that has an important effect and yet is not included among the factors under consideration because:*
    - *Its existence is unknown*
    - *Its influence is thought to be negligible*
    - *Data on it are unavailable*
- *Safeguard:*
  - *Randomize the order of the experimental trials to protect against the effect of lurking variables*
- *Action:*
  - *If the lurking variable creates a trend it can be compensated for in the numerical analysis*
  - *Conclusions can then be drawn from the original factors that are not affected by such lurking variables*

*Source: Joiner, Brian L. "Lurking Variables: Some Examples," *The American Statistician*, November 1981, Volume 35, No. 4, pp. 227-233.

Notes:

An Example of a Lurking Variable

The Agricultural Sciences Department at North Carolina State University developed a new and improved chicken feed that would supposedly promote plumper and meatier chickens. The school contracted with a local poultry provider (Holly Farms) and conducted a series of studies testing the new product. The NC State Mathematics Department was asked to develop a DOE to support the above tests.

Preliminary calculations were made and two populations of chickens were identified, tagged (this becomes very important later), and segregated. One population was fed the standard feed and the other fed the new and improved feed. After feeding the two populations of chickens, statistically significant samples from each population were slaughtered and weighed (what we in Power Systems would refer to as 'Destructive Testing'). The outcome of the experiment was obviously trying to prove that chickens on the new feed weighed more than those on the old feed. In this case the Y = Weight and X = Type of feed.

After reviewing the data, the scientists were surprised to learn that there was no statistical difference between the two populations. The average weight (from the samples) was actually slightly higher (although not statistically higher - p-value > 0.05) for those chickens fed the standard feed. Obviously, this baffled the scientists involved in the experiment.

After a few weeks of evaluating the experiment and the data, one of the grad students asked Holly Farms for a map of their facility.

Please take a few minutes to read this example.

Notes:

g GE Power Systems

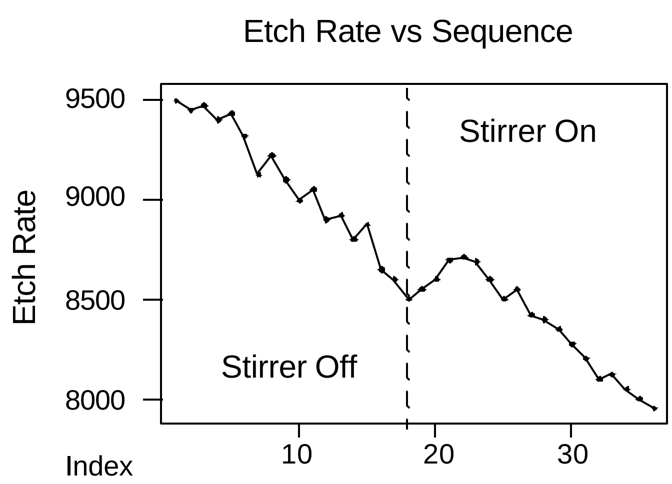
Lurking Variable Example (cont.)

After reviewing the map, the student noticed that some of the chicken houses were located immediately next to the slaughter house. This raised a question in the student's mind and he decided to drive out to the farm for some first-hand observations. He was escorted to the slaughter house area and immediately noticed that the chickens located in the houses next to the slaughter house demonstrated significantly higher levels of activity - i.e. clucking, pecking, and running around like... well... like chickens. After another review of the experimental data (by tag number), it was discovered that all of the chickens on the new feed had been located in the house immediately adjacent to the slaughter house - a lurking variable had been identified. (NOTE: without the tag numbers being recorded, this lurking variable may have never been discovered once again illustrating the importance of proper, thorough data collection). After reviewing his findings with the team, it was decided to introduce a new variable into the experiment - chicken house location.

The experiment was re-run with the new variable included (i.e. - chicken locations were randomized) and the results analyzed. On the second attempt, the results validated the scientists original hypothesis - the new feed produced plumper, meatier chickens. Evidently, those chickens located next to the slaughter house experienced higher stress levels and subsequent weight loss. As a result, Holly Farms opted to use the new feed AND also relocated all chicken houses AWAY from their slaughter houses.

Moral of the Story: Don't keep your chickens too close to the slaughter house.

Why Randomize?



Bath degradation obscures factor effects in a design run in standard order.

This shows an example of how etch rate appears to decrease over time.

If you ran your experiment in standard order, all of the tests on the last factor would be at the end of the run.

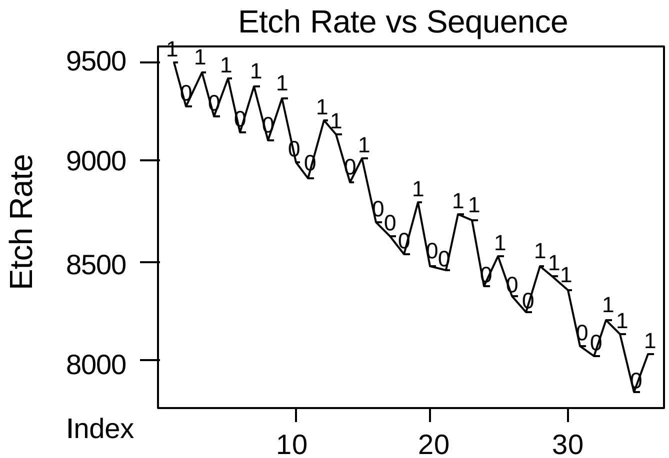
In this example the last factor was stirrer (on or off).

This graph would lead us to conclude that the stirrer being on has no effect or even a negative effect on etch rate. The lurking variable of bath degradation has caused us to incorrectly assume that having the stirrer on results in a shallower etch depth.

Note a T-test for means with H_a = mean of stirrer off > mean of stirrer on gave a P value of 0.

Notes:

Example: Why Randomize? (cont.)



0=stirrer off, 1=stirrer on

In the randomized experiment the stirrer effect is visible despite the effect of bath degradation.

In this case the experimental conditions were randomized.

Now the true effect of the stirrer on etch rate is made evident as the randomization has successfully averaged out the effect of the lurking variable - bath degradation.

This graph leads us to conclude that stirrer on yields a greater average etch rate than stirrer off.