Exercises 5 · More uncertainty

Due Monday, February 29, 2016

(1) Does 1+1=2?

The data from "semicon.csv" are (almost) from a real patent lawsuit brought by one semiconductor firm against another. I say "almost" because I am not allowed to distribute the real data. Therefore I have simulated data that looks like the real data in every important respect. Everything else, including the following story, is real.

Company A develops a nano-scale manufacturing process for etching semiconductor wafers via photolithography. (Photolithography entails passing a light beam over the wafer and inscribing a pattern into it.) The company applies for, and receives, a patent for this novel manufacturing process.[1]

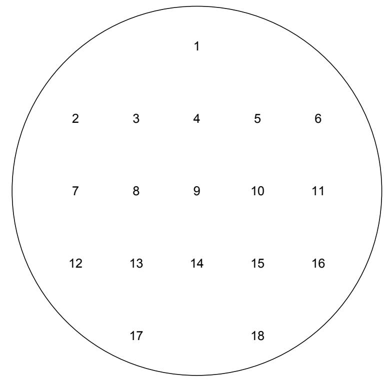
[1] http://en.wikipedia.org/wiki/Photolithography

Subsequently, Company B begins etching their semiconductor wafers using a manufacturing process that, according to Company A, is identical to the process that A has under patent. A sues B for $1.6 billion in actual and punitive damages.

Company B claims that its nanofabrication process is different from that of A in one crucial respect. Let's say we can measure the "dosage" of the light beam applied to the wafer as some number, D, in arbitrary units. Company B applies a light beam of "dosage" wD, where w is between 0 and 1. Then they wait a bit and apply a second light beam of "dosage" $(1 - w)D$. At the end of the process, the semiconductor wafer has received the same total dosage as it would have under Company A's process, but in two passes, separated in time. Company B claims that this is a novel process resulting in a higher average development rate of the photoresist, which is the thin layer of material used to transfer the circuit pattern to the semiconductor substrate. Don't worry too much about what the "development rate" is, so long as you understand that company B claims it is higher under its own process than under A's.

Thus the lawsuit turns on the question of, essentially, whether 1+1=2. A and B both etch their wafers using the same total dosage; B just applies that dosage in two passes. Do the two processes result in the same rates of development of the photoresist? If not, B can claim its manufacturing process as novel, successfully defend itself against A's lawsuit, and file a patent of its own.

An independent laboratory is retained to assess the strength of the factual claims made by company B. The lab conducts an experiment, etching 18 different physical sites on each of 27 different semiconductor wafers: 3 wafers for each of 9 different wafer materials, each material from a different manufacturer. The 18 sites on each wafer are arranged just like those on the right. At each of the 18 sites, a specific dose sequence is applied. For example, a dose sequence of 60–40 means that 60% of the dose was applied in the first pass; 20–80 means that 20% was applied in the first pass; and so on. Remember, of course, that there are three wafers for each material. For one of these wafers, Company A's process was used to etch all 18 sites. Thus the dose sequence here was 100–0. Then for the other two wafers, a specific pattern of dose sequences was applied to each of the 18 different sites using Company B's process.



The data from this lab's experiment—or rather, data essentially the same as the real data—is in "semicon.csv." Each row is a measurement of the photoresist development rate at a single site on a single wafer. The variable codes are:

rate : average development rate of the photoresist, measured at that site, in nanometers per second.

pass1: the fraction of the total dosage applied in the first pass

pass2: the fraction of the total disage applied in the second pass

split: the minimum of pass1 and pass2.

material: a categorical variable, coded 1 through 9, indicating the raw material of the wafer. These have been anonymized, so that 1 is the first material, 2 is the second material, and so forth. Each material is manufactured by a different company.

wafer: a categorical variable, coded 1, 2, or 3. On Wafer 1 of each material, Company A's process was used, resulting in a "dose sequence" of 100–0. On Wafers 2 and 3, Company B's process was used, with varying levels of the dose sequence from 100–0 to 0–100. For each material, Wafers 2 and 3 are replicates of the same set of dose sequences at each site.

site: the etching site on the wafer, corresponding to the numbers in the diagram on the previous page.

Imagine that the judge in the case has asked you to analyze the data and give your opinion. Does B's process—wherein the same total dosage is used as in A's one-pass process, but applied in two different passes—result in a higher average development rate of the photoresist? Write a report for the judge (who you may optimistically assume is statistically literate) that summarizes your analysis and conclusions.

(2) More cheese

This question considers data on sales volume, price, and advertisting display activity for packages of sliced cheese, available as "cheese.csv" on the course website. For each of 88 stores (store) in different US cities, we have repeated observations of the weekly sales volume (vol, in terms of packages sold), unit price (price), and whether the product was advertised with an in-store display during that week (disp = 1 for display). Altogether there are 5,555 observations in the data set.

Address the following questions thoroughly but concisely. Make sure to include the appropriate plots, statistical summaries, and measures of uncertainty to illustrate and support your conclusions.

(A) Ignoring price, do the in-store displays appear to have an effect on sales volume? Use an appropriate transformation for modeling consumer demand. In light of your analysis, complete the following two sentences. "I estimate that in-store displays increase/decrease sales by —%. I am 95% confident that this quantity is between —% and —%."

(B) Is there reason to suspect that your result in (A) is confounded by pricing strategies? Show evidence either way. If the answer is yes, propose a model that allows you to adjust for price in assessing the marginal effect of in-store displays on sales volume. Remember back to our milk sales-versus-price data that a typical model for price elasticity of demand is of the form $\hat{y}_i = K x_i^{\beta}$, where $\hat{y}$ is expected sales, x is price, K is a constant, and β is the elasticity—that is, the marginal effect of price on sales volume. You should recall how to use linear least squares to fit such a model; now modify it to account for the effect of in-store displays.

As above, in light of your analysis, complete the following two sentences. "I estimate that in-store displays increase/decrease sales by —%, once the effect of price is accounted for. I am 95% confident that this quantity is between —% and —%." Again, make sure you properly account for differences in overall sales volume among stores.

(C) Does price elasticity for cheese appear to be changed by the presence of in-store advertisement? (Hint: remember about interaction terms in models with numerical and categorical predictors.) As above, quote an appropriate confidence interval that addresses this question. Can you think of a possible economic explanation for your result here?

(D) What should Kroger's in Dallas/Ft. Worth charge for cheese in no-display weeks? What should they charge in display weeks? Assume that the wholesale cost of cheese is \$1.50 per unit.