Exercises 7 · Cause and effect, probability

Problem 1: green buildings

Pick up where you left off with our in-class case study on green buildings, and imagine the following scenario.

An Austin real-estate developer is interested in the possible economic impact of "going green" in her latest project: a new mixed-use building on East Cesar Chavez, just across I-35 from downtown. The developer has access to a data set of some green-certified commercial properties, together with some of their neighboring non-green buildings. She had someone on her staff, who has been described to her as a "total Excel guru from his MBA statistics course," run some numbers on this data set and make a preliminary recommendation. Here's how this person described his process.

> I looked at the green buildings and non-green buildings separately. The median market rent in the non-green buildings was $25 per square foot per year, while the median market rent in the green buildings was $27.60 per square foot per year: about $2.60 more per square foot. (I used the median rather than the mean, because there were still some outliers in the data, and the median is a lot more robust to outliers.) Because our building would be 250,000 square feet, this would translate into an additional $250000 \times 2.6 = \$650{,}000$ of extra revenue per year if we build the green building.
>
> If our expected baseline construction costs are $50 million, with a 7% expected premium for green certification, that means we should expect to spend an extra $3.5 million on the green building. Based on the extra revenue we would make, we would recuperate these costs in $3{,}500{,}000/650{,}000 \approx 5.4$ years. Even if our occupancy rate were only 90%, we would still recuperate the costs in about 6 years. Thus from year 7 onwards, we would be making an extra $650,000 per year in profit, which is 18.5% per year of our extra $3.5 million in going green. Since the building will be earning rents for 30 years or more, it seems like a good financial move to build the green building.

The developer listened to this recommendation and understood the analysis, but felt unconvinced. Something feels off about the analysis run by the "guru," but she can't put her finger on it. She has therefore asked you to revisit the question, so that she can get a second opinion.

You can accept some of this person's assumptions, like that the building will cost $50 million to build, and that the extra costs associated with "going green" are about 7% of baseline construction costs. But do you agree with his estimate that the green building is likely to earn an extra $650,000 of revenue per year?

EXERCISES 7 · STA 371H   2

Analyze the data you've been given to come up with an estimate of the causal effect of green certification on rental revenue. Write a memo to the real-estate developer (let's say her name is Ms. Jane Alexander). This memo should not exceed two pages in length, and it should concisely address three issues:

1. What, if anything, is wrong with the original analysis that leads to the estimate of $2.60 extra revenue per square foot?

2. How would you approach the problem instead, and why is this a better way of doing it?

3. What do you think is a better number here? If you think the estimate of $2.60 extra revenue per square foot is wrong, propose a specific alternative estimate, with error bars/confidence interval.

Your memo should include at least one figure in support of any of these three items—although it can have more than one figure if you choose. Moreover, it should be written in terms that a non-technical reader can understand.

In addition to the 2-page memo, you should include a technical appendix with any other figures, technical descriptions, model output, etc. that you feel are important to have in there, in case someone else more statistically savvy were to try to interpret what you've done in detail, and to reproduce your results. This part should be clear, but it need not be especially polished.

Note: this is a pretty typical way to structure a report in the real world: a short memo for a decision-maker that summarizes the main issue and the bottom-line numbers, together with a detailed appendix that can support any necessary follow-up inquiry by technical staff, consultants, etc.

Problem 2: probability

Here's a question a friend of mine was asked when he interviewed at Google. Visitors to your website are asked to answer a survey question before they get access to the content on the page. Among all users, there are two categories: Random Clicker (RC), and Truthful Clicker (TC). There are two possible answers to the survey: yes and no. Random clickers click either one with equal probability. You are also giving the information that the expected fraction of random clickers is 0.3.

After a trial period, you get the following survey results: 65% said Yes and 35% said No. What fraction of people who are truthful clickers answered yes?