Predictive Analytics: Direct Mail Simulation

One of the illusions magicians try to create is the ability to accurately forecast complex future events. The goal of predictive analytics is similar and rather magical, but it is not an illusion. As the name implies, predictive analytics is a broad set of statistical, machine learning, and data mining techniques used to forecast complex future events with a significant degree of accuracy.

In the appendix of the Hawkes Learning courseware, there is a role playing game called Direct Mail. The purpose of the game is for the player to develop a predictive strategy for direct mail marketing.

The direct mail game is a simulation game in which you will play the role of a junior marketing manager in charge of direct mail marketing for a product you have been given to help manage. You are a new hire and are anxious to impress the marketing manager with some statistical predictive analytics.

In your company the marketing department will develop a brochure to send out to members of various mailing lists. The only thing you will know about the mailing lists are that the marketing manager has given you a set of lists which she believes may be potentially profitable using the developed brochure. You have been asked to develop a strategy to evaluate the risks and opportunities associated with each list for any potential profitability. If the marketing manager believes your initial assessments are good, you will be allowed to do the remainder of the lists on your own. Is this just a guessing game, or is there some statistical science that can be applied to aid you? As you likely guessed, the latter is the correct answer.

In direct mail marketing one of the big problems is getting the recipient to open your mail. How personal the mail looks will influence the probability it will be opened. In addition to choosing whether to mail a list or not, you can define your own mailing tactic that may impact whether your mail is opened. There are two mailing choices. First, you must decide whether to use first class or bulk mail. First class costs $0.68 per piece and bulk mail costs $0.49 for each piece mailed. First class looks more personal, but bulk is cheaper; which will be best for your particular list? Another personalizing feature you can invest in is whether you want to use mailing labels or have the envelopes typed with the name and address.

The marketing manager is anxiously awaiting your predictive analytic model and mailing strategy.

10.4 Exercises

Basic Concepts

1. Describe, in layman's terms, how a confidence interval is constructed for a population proportion.
2. It seems that estimating proportions produces estimates which are much more precise than those for means. Explain why this is the case or is not the case.
3. When determining the sample size required to estimate a population proportion within some specified error level with a specific level of confidence, what is the guideline to follow when there is no estimate available for the population proportion? Why is this done?

Exercises

4. Acid rain accumulations in lakes and streams in the northeastern part of the United States are a major environmental concern. A researcher wants to know what fraction of lakes contain hazardous pollution levels. From 200 randomly selected lakes, it is determined that 45 of them have an unsafe concentration of acid rain pollution.

a. Calculate the best point estimate of the population proportion of lakes that have unsafe concentrations of acid rain pollution.

b. Determine a 95% confidence interval for the population proportion.

c. If a local politician states that only 20% of the lakes are contaminated, does the study provide overwhelming evidence at the 95% level to contradict his views?

5. *The Richland Gazette*, a local newspaper, conducted a poll of 1000 randomly selected readers to determine their views concerning the city's handling of snow removal. The paper found that 650 people in the sample felt the city did a good job.

a. Compute the best point estimate for the percentage of readers who believe the city is doing a good job of snow removal.

b. Construct a 90% confidence interval for this percentage.

6. The clinical testing of drugs involves many factors. For example, patients that have been given placebos, which are harmless compounds that have no effect on the patient, often will still report that they feel better. Assume that in a study of 500 random subjects conducted by the Poppins Sucre Drug Company, the percentage of patients reporting improvement when given a placebo was 37%.

a. What would be a 95% confidence interval for the true proportion of patients who exhibit the placebo effect? Interpret this interval in terms of the problem.

b. What would the 99% confidence interval be?

c. To gain the additional 4% of confidence how much wider did the interval become?

7. The Peacock Electric Company thinks that 40% of their customers would be interested in bundling solar power purchasing with their electric bill. A random sample of 400 households reveals that 110 of the households are interested in this.

a. Construct a 99% confidence interval for the true proportion of households interested in bundling solar with their utilities.

b. Do you feel the company is accurate in its belief about the proportion of customers who have interest in bundling solar power with their utilities? Justify your answer.

8. Running continues to be a very popular sport in America. At a major race, there may be over 10,000 people entered to run. The race promoters for a road race in the Pacific Northwest took a random sample of 750 runners out of the 5000 runners entered to estimate the number of runners who will need hotel accommodations. Five hundred runners indicated they would need hotel accommodations.

a. Construct a 90% confidence interval for the true proportion of runners who will need hotel accommodations.

b. Is the confidence interval obtained sufficiently narrow to be of help in planning the number of hotel rooms which will be necessary to accommodate the runners? Justify your answer.

9. Over the past few years the national rate for homeownership has been stable near 64% likely due to low mortgage interest rates and an active real estate market. For various reasons, which could include short supply of affordable homes and demographical factors, the homeownership rate in West Palm Beach, Florida is suspected to be lower than the national rate. A sample of 80 households was randomly selected from the population of Palm Beach County in Florida. Suppose that 47 of the households sampled were owned by the residents of the homes.

a. Construct a 95% confidence interval for the proportion of households in the area sampled that are owned by the residents of the homes.

b. Is there evidence at the 95% level that the proportion of the households in the area sampled that are owned by residents is less than the national rate?

10. In the *Gallup Poll Monthly*, it was reported that 31% of the people surveyed in a recent poll claimed that vegetables were their least favorite food. Surprisingly, only 14% responded with liver, and 10% of those surveyed did not submit a response because they claimed that they liked everything. The poll was based upon a sample of 1001 people. Assume that a random sample of Americans was chosen, and construct a 90% confidence interval for the percentage of all Americans who say that vegetables are their least favorite food.

11. The Big Green Poster Company wants to estimate the percentage of poster sites controlled by their competition, Bird's Billboard Service. What sample size would be necessary to estimate this percentage to within 3% with 95% confidence? (They think Bird's controls about 33% of the boards.)

12. Researchers working in a remote area of Africa feel that 40% of families in the area are without adequate drinking water either through contamination or unavailability. What sample size will be necessary to estimate the percentage without adequate water to within 5% with 99% confidence?

13. Companies that provide environmental cleanup for hazardous waste and toxic chemicals are growing rapidly. W.R. Gross is thinking about entering this field with a subsidiary called Saf-t-Soil. They wish to estimate the true proportion of U.S. corporations that produce hazardous waste as a by-product of their manufacturing process to within 10% with 80% confidence. What sample size will be needed?

14. The public relations manager for a political candidate would like to determine if the registered voters in the candidate's district agree with the politician's view on a particular issue. Find the sample size necessary for the public relations manager to estimate the true proportion to within 5% with 85% confidence.