Analytics Problem Set Phase 2: Statistical Inference & Regression

Scenario:

Now that you've developed skills in interval estimation, hypothesis testing, and regression analysis, you've been asked to use the customer and transaction data from Green River Outdoor to study the differences in performance across GRO's five stores in the Boston metro area. The project manager has reached out to describe what they're looking for in the next phase of the analysis.

...

New Message

To: Green River Outdoor Team

Subject: Store performance analysis

Good morning team,

Green River Outdoor has been impressed with our work so far, and we've finalized the scope for the next phase of our analysis. Here are the two key challenges their leadership team is facing:

- **Business performance, as measured by gross margin percentage (GM%) and gross margin dollars (GM$), has been inconsistent.**

  Note: *GM$ reflects both GM% and volume; a category might have high GM% but low GM$ or vice versa. GM% is GRO's primary performance metric, but it's important to keep in mind the size (in GM$) of each store and category when making recommendations.*

  - When GM% is low relative to the company's average, it indicates that items were discounted more than usual (likely because they were slow to sell).
  - There appears to be variation in performance across stores and across product categories, making it hard to determine the root causes of low GM%.

  - **Objective 1:** GRO's team wants to become more data-driven in their decision-making, so they'd like to use hypothesis testing to understand these differences. For now, formulating and conducting **one hypothesis test** (either one-sample or two-sample; related to either population mean or population proportion) will demonstrate the type of analysis we can do and help Husky Analytics land a follow-up project with GRO.

Management believes that too many items are selling at clearance prices, and they want to identify patterns in clearance/discounting behavior.

  - **Objective 2:** GRO needs a better understand of which items go on clearance. It is certain categories? More expensive vs. less expensive products? Does which store an item is in matter? Use a logistic regression model to predict the binary 'clearance' variable, using any relevant predictor variables. They'd like to see a summary of the regression results (keep it simple, including just the information needed to interpret the estimated regression equation) and a brief explanation of how reliable the model is.

I'll be meeting with the GRO team next week, so after you've completed the analysis, please send me your 'elevator pitch' for the work you've done—a ~5 minute video presentation bringing me up to speed so I can pass your key insights on to GRO.

That's it for now—please let me know if you have any questions!

Thank you,
Laura

Send