Hw 14: Retail Space Testing

Intro

The problem is determining whether the amount of shelf space is correlated and even causal to the amount of sales for a product. With real life applicability, the amount of sales essentially refers to profit. Two of the main challenges in this problem is obtaining useful data and avoiding confounding factors.

A key aspect of this is in experimental design. While raw data is easily gathered, not all of the data will be useful in making our decision. While we can model all we want, we also need to think of optimizing sales during the experimentation period. This is especially true if the variability in sales per day of a product is very high, since this will invoke the need for longer experiments or more trials.

The process will be as follows:

1. Experimental Design & Data Gathering
2. Modeling & Testing Data
3. Analyzing Results

The **Given** for all of the following include:

- Multiple stores exist
- Previous study results (Supplementary sales, adjacency of product)

1. Experimental Design

Two Phase
A two step process to gathering data can include choosing items within the store to increase the shelf space for. Since there are many products in place, we will have many variables to try this on. Furthermore, increasing the shelf space for a product means decreasing the space for another, which allows us to see if the sales decreased for the latter.

Blocking
Blocking is defined as a split in data on a factor that is similar but not of primary importance to the experiemnt's purpose.
After a specified period, which can be determined by the potential seasonality of products (e.g. would not want to increase or decrease candy space during Halloween season, as the lecture mentioned) we can reverse the increase or decrease in adjacently placed products. Note that the adjacency here does not specifically refer to complementary products. By doing so, we now have three sets of data - the original amount of sales, sales after increasing, and sales after decreasing. These trials are "blocked" into batches depending on the order of the trials since the timing can have variability, but should be accounted for heuristically depending on the product at hand. (Explained in Notes below).

Finally, we need a control group that does not change shelf space. The products chosen for all of these divisions should be spread across possible types of items (food, hardware, entertainment) and have similar adjacent complimentary products to not influence the sales.

Factorial

The experiments should be split by median income or variability of income geographically, if they are significantly different unless the relative number of sales for each type of product at our store is constant throughout. For example, higher class neighborhoods might have a significantly higher affinity towards organic vegetables than frozen, and might change how shelf space affects sales.

Optimization & Design

An important part of this is looking at the cost of changing shelf space. We can build an optimization model to see which product spaces we can alter with regards to constraints on seasonality, storage stocks, contracts with recurring imports, and the cost of increasing spaces for a product. The last one is important because we need to know if increasing shelf space costs much more money. Another decision point here is in determining what shelf space really means. Should we be increasing the space relative to the product, space on the shelf, or the number of products dispalyed?

Also, we only want to test changes on products that will give increased revenue if it is successful. For example, increasing the amount of space for toothbrushes should not need testing since it does not generate revenue. This would be in relation to products next to it as well, since we want a simple change that does not reorganize the entire store for the sake of one product.

Lastly, space limitation is important to this as well since produce requires specific accomodations like refridgeration to be sold without losing value. Should we test produce separately from non-perishable goods?

Notes

Two important notes.
First, an initial increase in sales prefacing a decreases may have a separate affect than reversing the order of change. For half of the store, we should first decrease then increase. This can help determine whether or not the more important factor is a change in space or the amount of space available.
Second, after an initial change, the decrease or increase should put the shelf space to the either lower or higher than the original spacing, respectively. Otherwise, we would only have data on the original shelf space after a change occured.

2. Modeling & Testing Data

Significance Testing

Given Change in sales, amount of sales, marginal profit, and relative profit from the experiment
Use ANOVA testing, Wilcoxon Median (if assumptions are met based on the samples' distribution)
To Determine whether the change in sales or amount of relative profit is significant

Notes
It is important to specify whether we are looking at the relative sales or than the flat sales. First, should be examining the sales of each grocery store in the same batch and even perform a

significance test on the sales to see if they are directly comparable, or if we need to scale the data. Then, we can apply our significance testing.

Also, the results of our testing might not be direct. If we see a significant difference for increasing in one batch but not in the other, what should we do? Remember, a batch is differentiated by increasing then decreasing versus decreasing then increasing the shelf space.

The testing should be applied separately to each split in factors.

Classification

Given Number of sales, profit, price of sales, stock availability, shelf size, change in shelf size, time period

location, affluence of neighborhood, seasonal or time of year

Use SVM, knn, logistic regression, random forest classification

To Predict whether or not a product's shelf space has been changed (increased or decreased) or not

Notes

We have already seen whether or not the sales changes are significant. However, significance testing has its limitations. For example, many non-parametric test results are influenced by the number of instances and become obsolete with too many observations.

Random Forest should be used as a means of analyzing the data, similar to linear regression analysis. Which factors are more significant? Do we see other splits in data, such as the affluence in neighborhoods causing a significant, new branch?

SVM and KNN can be used to actually try classifying whether a change in shelf space has occured. If we can accurately determine the changes, we can support the hypothesis that there is a relationship between sales.

Logistic Regression can than be applied for a deeper understanding of the probability of a change happening. With SVM and KNN, we may be able to make stronger predictions since they are better tuned for multinomial classification, but if we want to look specifically at a difference between factors all else held constant, a binary logistic regression classifer can help us see how powerful the relationship is based on the probabilities predicted.

3. Analyzing Results

Scenario 1: Cannot reject any null hypothesis, there is no relationship with shelf space vs sales. Also, we cannot successfully classify them using basic machine learning models.

If this is the case, we must decide whether to conclude the study or look at the experiment in detail. Are there confounding factors? Did we choose the wrong product spaces to alter? Were the adjacent products for each factor held constant?

Scenario 2: Find relationship with ANOVA/non-parametric testing and/or modeling.

Predictability in shelf space changes might not mean it has the desired relationship with the sales price.
For example, what if we can accurately classify what changes happened to the shelf space, but our random forest classifier decides that the sales price had very little predictive power in the model? But at the same time, our significance testing tells us that the sales price results are significant? It might mean our experiment itself has a problem since the two are intuitively contradictory, and we have confounding factors that we were not aware of.

An optimal result would be with seeing significant differences in our statistical tests and seeing powerful results from the models.

Conclusion

In prototyping and experimentation, the question might be about whether or not increasing shelf space increases sales. But in reality, we needed to define what about the sales is important - total profit.
Ultimately, the challenge is in optimizing the experimental design to gather significant data for our analytic job. Many of the problems above are also limited by the number of stores potentially available. For example, are there enough stores and employees to make this a full factorial experiment with blocking? Otherwise, we might have to go with a more efficient but less comprehensive design structure.