Application of SAS in Product Testing in a Retail Business

Rick Chambers, Steven X. Yan, Shirley Liu
Customer Analytics, Zale Corporation, Irving, Texas

Abstract:

Testing new products is an important and continuous task for the retail business. A reliable and efficient procedure to evaluate new products is critical to the success of any retailer: Either under- or over- assessment of a product may result in a significant cost and loss to the company. However, in practice many tests are much more difficult to utilize than they first appear because of various sources of uncertainty involved. For example, most tests are carried out only in a small sample of stores due to store space limitations, labor cost, product availability, and other logistic factors. In addition, store sales are affected by internal and external factors such as product inventory, weather, local holidays, etc. In this paper we will introduce a product testing procedure that we have developed using SAS that includes product segmentation, store selection, and product scoring. We will illustrate and evaluate the procedure by a simulation study. Several useful SAS techniques for simulation study are described as well.

Introduction

In the competitive retail business, a company must continuously update its products while maintaining a reliable supply chain of current, high quality products. It is essential for any company's success to establish a reliable and efficient procedure to identify and test new products. Either over- or under- assessment of a product may lead to a significant loss and cost to the company. For example, overstock of a product not only wastes resources to keep and maintain the product but also creates financial restraints to acquire new products. On the other hand, under assessment of a product may cause the company to lose new business opportunities and competitive edge.

Before testing any new products, the company must first identify its existing core products and establish a metric that measures the importance and contribution of each product. This metric system can be developed by product category or by much finer product segmentations, such as by price range, margin ratio, turnover times, and even seasonality. Without such a metric it is hard to measure and interpret how successful a new product is and whether to accept it even after the test is done. In this paper we will introduce a product scoring system that can be easily implemented, readily interpretable and may be used to compare products across different product categories. We will assume that sales of the test products are relatively stable. There are many references on how to forecast sales of new products, see for example, Fisher et al (2004).

After the product segmentation and a scoring metric are established, each test product is compared with the products within an appropriate segment and its score is calculated. These existing products are referred to as base or benchmark products. The test score can then be used as one of measures to decide whether the product should be accepted or rejected. The score can also be used to compare products from different product segments.

Most tests are carried out only in a small sample of stores due to store space limitations, labor cost and other logistic constraints. In addition, the need for fast turnover of test results may also

require the test be conducted in a relatively small scale. There are many ways to select the test stores, see Montgomery (2004) for a general introduction of sampling techniques and analyses. The simplest one would be a random sample with certain control mechanisms to balance the store type, store size, geographical location, and demographics of trading areas. More sophisticated sampling techniques may be used to increase the efficiency or accuracy of a test with a smaller number of test stores, but these techniques usually requires more advanced analytical skills and are based on assumptions that need to be examined more carefully.

When the test score of a product is obtained from a test, a decision must be made as to whether to accept or reject the product. For example, it may be decided from the business perspective that the score of a new product must be 0.20 or greater for it to be accepted. However, like most tests, scores calculated from a test are subject to errors from various sources. For examples, the sale of either base or test products is random by its nature; and selection of test stores creates another source of random errors, or even bias. Therefore, statistical inferences are needed to make a sound and reliable decision and will be illustrated in the simulation study. Some Useful SAS Codes for Simulation are described in the appendix.

Methodology

Let Y_1, Y_2, Y_3, … , and Y_p be store sales of p base products in a specific period. Assume that Y_1, Y_2, Y_3, … , and Y_p are random variables with means $\mu_1 \leq \mu_2 \leq \mu_3 \leq \ldots \leq \mu_p$, and $\mu_i > 0$, for $i=1$, 2, 3, …, p. We define the score of the ith product as

$$S_i = \frac{\sum_{j=1}^{i} \mu_j}{\sum_{j=1}^{p} \mu_j} \ , \quad \text{for } i = 1, 2, 3, \cdots , \ p$$

Clearly, $0 < S_i \leq 1$, with $S_p = 1$. The average score will be used for products with the same mean.

Intuitively, S_i is the cumulative contribution of the i-th product and its inferiors to the total sales of the product segment. Products with lower scores have less impact on the sales than products with higher scores and there is no need to add a new product whose score is below some certain values that can be established in practice. Furthermore, the score of a product is a relative metric that has the same meaning across product segments and thus can be used to compare products from different segments.

Let Y be the store sale of a test product with mean μ. Assume $\mu_{k-1} \leq \mu < \mu_k$, where k is a fixed but unknown integer, $1 \leq k < p$. The score of the new product is defined to be

$$S = \omega \ S_{k-1} + (1-\omega) \ S_k,$$

where $\omega = (\mu_k - \mu) / (\mu_k - \mu_{k-1})$.

The purpose of the test is to estimate S. Scores calculated from a test are always subject to at least three sources of random variations: sales of the base product, sales of the test product, and test store selection. If a threshold for the score has been decided to reject or accept a new product, then a bound for the test score can be found through simulation by which a product is rejected or accepted.

The test score can be calculated in the following two steps.

1. Estimation of base scores. Average store sales of base products can be calculated from test stores, non-test stores, or both, depending on how the test stores are selected and how the test product affects the base product. For example, if test stores are selected randomly and the test product has little impact on the base product, then the sales of base products should be calculated from all stores. On the other hand, if the test product has a relatively large impact on the base product, then the product sales should be estimated from the test stores. But even under the latter case the sales of non-test stores may help to smooth the sales of test stores. Store sales can also be smoothed by other statistical methods, for example, see Robertson et al (1988).

2. Interpolation of test scores. Interpolation can be performed on the single or several adjacent base products on each side of the test item, See Figure 1. The bandwidth is defined to be the number of base products on either side. A wider bandwidth may provide more robust estimate when sales are clustered or have large variations as shown in the simulation, (Table 1 and 2). The bandwith will not affect the estimation of base scores.

Figure 1. Estimation Method - Bandwith
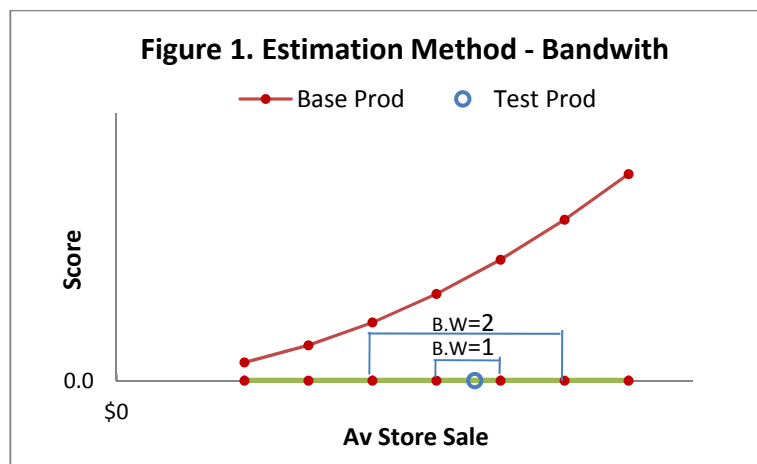
Simulation Result

Suppose that a retail company has 500 stores, of which 100 stores are selected randomly for the test. Suppose that there are 15 base products in a specific product category and 3 new products to be tested, see Figure 2. We assume that store sales of these products are normally distributed with mean μ specified in Tables 1 and 2. The difference between the two tables is the standard deviation σ that is larger in Table 2.

For each simulation, we calculate scores from the test stores and from the base stores, each with two product bandwidths. The performance of the estimation method is measured by the mean square error (MSE), which combines both the bias and variance of the method:

$$MSE = (Est. Bias)^2 + (Est\ Variance)$$

A method with a smaller MSE provides more accurate and reliable estimates than a method with a larger MSE. The square root of MSE (RMSE) is calculated in Tables 1 and 2.
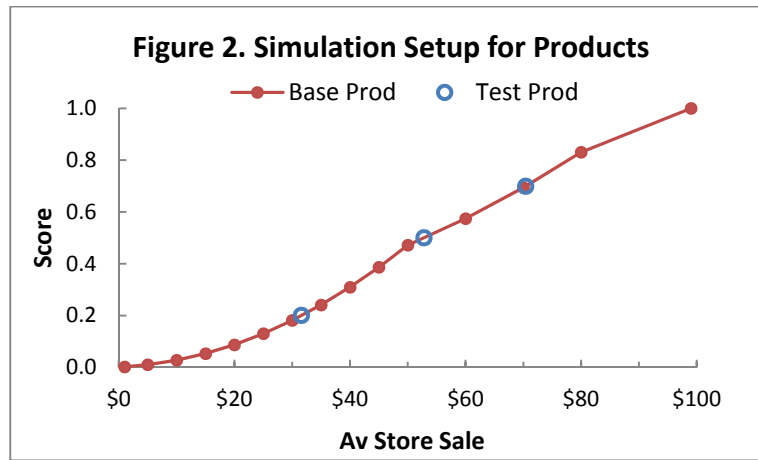

Figure 2. Simulation Setup for Products

Table 1. Moderate Variation: σ = μ$^{1/2}$ (MSE is multiplied by 1,000)

Product Set up				Estimation by Test Stores (100)				Estimation by Other Stores (400)			
				Bandwidth=1		Bandwidth=2		Bandwidth=1		Bandwidth=2	
Type	Num	μ	Score	Av	RMSE	Av	RMSE	Av	RMSE	Av	RMSE
Base	1	$1.00	0.002	0.002	0.17	0.002	0.17	0.002	0.08	0.002	0.08
Base	2	$5.00	0.010	0.010	0.42	0.010	0.42	0.010	0.21	0.010	0.21
Base	3	$10.00	0.027	0.027	0.67	0.027	0.67	0.027	0.34	0.027	0.34
Base	4	$15.00	0.053	0.053	0.92	0.053	0.92	0.053	0.47	0.053	0.47
Base	5	$20.00	0.087	0.087	1.17	0.087	1.17	0.087	0.58	0.087	0.58
Base	6	$25.00	0.130	0.130	1.40	0.130	1.40	0.130	0.69	0.130	0.69
Base	7	$30.00	0.181	0.181	1.60	0.181	1.60	0.181	0.80	0.181	0.80
Base	8	$35.00	0.241	0.241	1.78	0.241	1.78	0.241	0.88	0.241	0.88
Base	9	$40.00	0.309	0.309	1.91	0.309	1.91	0.309	0.95	0.309	0.95
Base	10	$45.00	0.386	0.386	2.03	0.386	2.03	0.386	1.02	0.386	1.02
Base	11	$50.00	0.472	0.472	2.06	0.472	2.06	0.472	1.04	0.472	1.04
Base	12	$60.00	0.574	0.574	2.03	0.574	2.03	0.574	1.03	0.574	1.03
Base	13	$70.00	0.694	0.694	1.90	0.694	1.90	0.694	0.95	0.694	0.95
Base	14	$80.00	0.831	0.831	1.56	0.831	1.56	0.831	0.77	0.831	0.77
Base	15	$99.00	1.000	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00
Test	16	$31.60	0.200	0.200	8.22	0.205	8.41	0.200	7.12	0.205	8.04
Test	17	$52.80	0.501	0.500	9.46	0.491	13.50	0.500	8.05	0.491	12.85
Test	18	$70.40	0.699	0.700	14.38	0.700	8.81	0.700	11.90	0.700	7.34

Table 2. Large Variation: σ = μ (MSE is multiplied by 1,000)

Product Set up				Estimation by Test Stores (100)				Estimation by Other Stores (400)			
				Bandwidth=1		Bandwidth=2		Bandwidth=1		Bandwidth=2	
Type	Num	μ	Score	Av	MSE	Av	MSE	Av	MSE	Av	MSE
Base	1	$1.00	0.002	0.002	0.18	0.002	0.18	0.002	0.09	0.002	0.09
Base	2	$5.00	0.010	0.010	0.93	0.010	0.93	0.010	0.47	0.010	0.47
Base	3	$10.00	0.027	0.027	2.39	0.027	2.39	0.027	1.04	0.027	1.04
Base	4	$15.00	0.053	0.054	6.00	0.054	6.00	0.053	1.75	0.053	1.75
Base	5	$20.00	0.087	0.089	12.49	0.089	12.49	0.087	2.80	0.087	2.80
Base	6	$25.00	0.130	0.133	20.95	0.133	20.95	0.130	4.97	0.130	4.97
Base	7	$30.00	0.181	0.185	31.65	0.185	31.65	0.182	9.01	0.182	9.01
Base	8	$35.00	0.241	0.246	43.19	0.246	43.19	0.242	14.65	0.242	14.65
Base	9	$40.00	0.309	0.315	55.13	0.315	55.13	0.311	21.18	0.311	21.18
Base	10	$45.00	0.386	0.388	61.99	0.388	61.99	0.388	28.23	0.388	28.23
Base	11	$50.00	0.472	0.458	63.64	0.458	63.64	0.467	23.93	0.467	23.93
Base	12	$60.00	0.574	0.580	65.56	0.580	65.56	0.575	17.51	0.575	17.51
Base	13	$70.00	0.694	0.699	77.44	0.699	77.43	0.696	27.76	0.696	27.76
Base	14	$80.00	0.831	0.815	76.79	0.815	76.79	0.827	24.33	0.827	24.33
Base	15	$99.00	1.000	0.988	44.66	0.988	44.66	1.000	5.63	1.000	5.63
Test	16	$31.60	0.200	0.203	44.49	0.208	41.66	0.202	39.50	0.206	38.36
Test	17	$52.80	0.501	0.492	82.41	0.490	78.47	0.494	70.96	0.490	71.48
Test	18	$70.40	0.699	0.703	102.88	0.700	93.23	0.703	93.05	0.701	83.75

Both tables show that the estimation can be improved significantly by using the non-test stores when the test product has little impact on the base product, in which case using all the stores provides even better result. In addition, using a wider bandwidth can improve the result for sales with large variations, (Product 18 in Table 1 and Products 16-18 in Table 2).

As discussed in the introduction, the score calculated from a test is subject to several sources of random errors and it is important to apply statistical inferences with business expertise to make a decision that may have a big impact on the company. For example, suppose that the company decides to accept a new product only if its score is 0.20 or better. Then the test score must be 0.283 or better for it to be accepted for the setup in Table 2, (bandwidth=1, one-sided, α = 0.05).

Summary

Testing new products is an important task for a retail business. The test can be accomplished by establishing a product scoring system. Estimation of base product scores can be improved by using appropriate stores and smoothing techniques. For base products with close sales and large variations the estimate can be improved by using more neighboring base sales. In addition, test scores are subjected to random errors, but statistical inferences can be made based on sales history and simulation.

Author Contact Information

Rick Chambers, VP
Customer Analytics
Zale Corporation, Irving, Texas
Email: rchamber@zalecorp.com

Steven X. Yan, Sr. Manager
Customer Analytics
Zale Corporation, Irving, Texas
Email: syan@zalecorp.com

Shirley Liu, Manager
Customer Analytics
Zale Corporation, Irving, Texas
Email: sliu@zalecorp.com

References

Fisher L.M., and Rajaram K. 2000. Accurate retail testing of fashion merchandise: Methodology and application. Marketing Science Summer 2000 vol. 19 no. 3 266-278

Montgomery D.C. 2004. Design and Analysis of Experiments, Sixth Edition. John Wiley & Sons, Inc.

Robertson T., Wright F.T., and Dykstra R.L.1988. Order Restricted Statistical Inference. John Wiley & Sons, Inc.

Whitlock, H.I 1997. CALL EXECUTE: How and Why. Proceedings of the 22nd Annual SUGI Conference, San Diego, CA. 410-414.

APPENDIX Some Useful SAS Codes for Simulation

Simulation usually creates a large amount of repetitive logs and outputs that may use up memory/storage and are unnecessary to keep. The following SAS code redirects the log and output to the dummy file.

```
filename dummyLog dummy;
filename dummyLst dummy;

proc printto log=dummyLog print=dummyLst;
   run;
```

To return to the default output,

```
proc printto;
```

```
        run;
```

Simulation repeats the same set of code many times. This can be done with the %do loop in a SAS macro. However, a better way to run the macro for simulation is by calling the EXECUTE routine in a SAS data step. See Whitlock (1997) for more details about the function.

```
data _null_;
   do isml=1 to &nsml;
      randomSeed=put(20120918+isml,12.);
      call execute('%SML_prodTest('||randomSeed||')');
      end;
   run;
```

In the above code, the macro %SML_prodTest is the core segment of the simulation, accepting an integer as the random seed to generate pseudo-random numbers for each simulation.

The following code simulates sales of 500 stores, of which the first 100 are test stores. The store sale of each product is normally distributed with the mean and standard deviation given in the SAS table *prodStat*. The STREAMINIT routine is called to create reproducible streams of pseudo-random. The RAND function generates random numbers from various continuous and discrete distributions and is very easy to use.

```
%let randomSeed=1234567;
%let nStore=500;
%let nTestStore=100;
data storeSale;
   set prodStat(keep=prod type meanSale stdSale);
   if _N_=1 then call streaminit(&randomSeed);
   do store=1 to &nstore;
      testStore=(store<=&nTestStore);
      sale=rand('norm',meanSale,sdSale);
      if type='Base' or testStore then output;
      end;
   run;
```

Simulation accumulates the result from each iteration. The APPEND procedure is the ideal way for this purpose. However, make sure that the table to accumulate the result is clean at the start of the simulation.

```
proc datasets lib=work nolist;
   delete simulation_all;
   quit;
   run;

proc append data=simulation_each base=simulation_all;
   run;
```