International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I5P107

Volume 72 Issue 5, 56-64, May 2024
© 2024 Seventh Sense Research Group®

Original Article

Control Group Selection for A/B Testing Through Optimization

Manasa Gudimella

Applied Data Scientist, Cumming, Georgia, USA.

Received: 04 March 2024 Revised: 08 April 2024 Accepted: 24 April 2024 Published: 11 May 2024

Abstract *- A/B testing is critical for big retail giants to adapt to rapidly changing consumer preferences and maintain market leadership. In a situation where the test group is predetermined, effective A/B testing requires selecting control stores that are comparable to test stores, a challenge addressed in this paper. The focus of this study is on the methodology for selecting homogeneous experimental units in physical retail settings evaluating them through a time series of key performance indicators. The methodology demonstrates adaptability to multiple KPIs, enhancing its applicability. Two methods for control store selection are introduced and compared: a statistical sampling technique and an optimization approach, with findings indicating the superiority of the optimization method for achieving more accurate and reliable A/B testing results. This research offers significant insights for retailers aiming to optimize their in-store strategies and improve overall business strategies and performance, making it essential for retail decision-makers seeking to optimize operational efficiency.*

Keywords *- A/B testing, Controlled experiments, Experimental Design, Optimization, Population Sampling, Optimization.*

1. Introduction

Big retail giants navigate through rapidly changing consumer behaviours and preferences to stay ahead of the competition. They employ A/B testing, an experiment where two or more variants of a business scenario are randomly offered to users. This approach allows retailers to experiment with product placements, pricing strategies and promotional activities, and measure the impact of the experiments on business Key Performance Indicators (KPIs). In A/B testing, the experimental results are statistically analysed to determine which variant performs better for a given objective.

The basic requirement for conducting such experiments is to ensure homogeneity among the groups on which different variants are tested. Homogeneity is typically assessed with respect to KPIs, measurable attributes that businesses aim to influence through their promotional activities. Statistically designed experiments deploy randomization to mitigate the adverse effects of experimental units that are supposedly homogeneous [8].

Businesses often face strategic requirements that necessitate selecting a pre-established group of stores for the test group due to the need for a controlled environment, strategic importance, or accurate market representation. In this process, the consideration of experimental units for the control group is often ignored. Identification of control units for such situations post-experimentation is something that has not been addressed in the literature.

This article considers the problem of identifying a control group that is homogeneous with the test group with respect to a multi-dimensional KPI. Two approaches are proposed for this problem. The proposed solution can be extended to similar problems. For example, this methodology can be used in identifying homogeneous customers/users based on multi-dimensional feature vectors involving components such as location, demography, behavioral traits, etc.

2. Problem

The basic requirement for conducting experiments such as A/B testing is to ensure the availability of homogeneous experimental units. Homogeneity can be measured in terms of one or more variables (KPIs), or it could involve a multi-dimensional variable, a KPI vector. For a detailed understanding of vector concepts in this context, see 'Elementary Linear Algebra' [1]. For the problem addressed in this paper, homogeneity is measured through a KPI vector. Given a pre-specified test group in advance, presumably comprising homogeneous experimental units, the objective is to choose units to represent a control group mirroring the characteristics of the test group. In other words, the goal is to select a subset of control units of the size of the test group so that the experimental units, together with test units, are homogeneous with respect to the KPI vector. For the problem addressed, experimental units are stores, KPI is the average weekly sales, and the KPI vector is the time series of the weekly average sales of a store for 51 contiguous weeks of a year prior to A/B testing. For further details on the methodology involving time series analysis, refer to [10]. Two distinct approaches are explored: a statistical method and an optimization method. The optimization methods include linear programming [11] and operations research techniques involving both linear and nonlinear optimization [13]. These methods are evaluated for their effectiveness in selecting an appropriate control group, thus ensuring the reliability of our A/B testing results. Section 4 presents and compares these approaches, shedding light on their respective advantages and limitations in the context of retail A/B testing and thereby contributing to the nuanced understanding of methodological selection in controlled experimental designs.

3. Related Work

The literature on Controlled Experimentation (CE) highlights several key studies. Hokka [4] addresses the challenges in the retail industry, emphasizing the necessity for CE to understand the causal effects of business decisions. The study outlines a framework for implementing CE effectively in retail contexts. In a similar vein, Kalyanam [5] explores the impact of search engine advertising on brick-and-mortar retail sales through a meta-analysis of 15 field experiments. The findings illustrate a positive influence of search engine advertising and underscore the interconnection between online and offline markets, suggesting that offline effects should be considered in search advertising campaigns. Kohavi [6] details how companies like Bing leverage CE to guide product development and enhance innovation. The paper discusses the challenges involved in scaling experiments, including aspects related to culture, organization, engineering, and trustworthiness. It describes a scalable system to manage multiple concurrent CEs and massive data sets. Koning [7] focuses on the adoption of A/B testing technology by start-ups, illustrating how it facilitates organizational learning and data-driven decision-making, which can lead to quicker scaling or determination of business viability. Additionally, Xu [15] discusses LinkedIn's experimentation platform, XLNT, which oversees A/B testing from design to analysis, addressing large-scale A/B testing challenges in social networks, such as conducting offline experiments and managing network effects. The study emphasizes the integration of A/B testing into the decision-making process and its alignment with business reporting for impactful results. Quin [12] conducted an extensive systematic literature review analysing 141 primary studies on A/B testing, focusing on design, execution, stakeholder roles, and key challenges. The review identified algorithms, visual elements, and workflow/processes as the main targets of A/B testing. It highlighted three principal roles in A/B test design: concept designer, experiment architect, and setup technician. Additionally, it suggested enhancing statistical methods, improving A/B testing processes, and automating A/B testing as areas for future research. Despite the comprehensive coverage, none of the studies mentioned above address the critical issue of identifying an appropriate control group given a test group within the context of CE.

The literature on population sampling primarily focuses on probabilistic and non-probabilistic methods, highlighting the advantages of the former due to its representativeness and ability to minimize bias. Banerjee [2] emphasizes the importance of selecting a representative sample from a target population, outlining various strategies such as random, systematic, stratified, and cluster sampling, each with its applications and limitations. Etikan [3] contrasts probability random sampling with non-probability sampling methods, such as quota and accidental sampling, stressing that probability sampling is preferable for generating unbiased data across the entire population. In contrast, non-probability methods can introduce assumptions and risks. Lohr [8] serves as a comprehensive guide to designing and analysing survey statistics, which is valuable for those interested in survey sampling methods. Taherdoost [14] discusses the necessity of sampling in research and outlines both probability and non-probability sampling techniques, providing formulas and considerations for determining the appropriate sample size. The paper details the steps involved in conducting sampling, from defining the target population to assessing the response rate. However, in the specific context of retail A/B testing for evaluating promotional strategy effectiveness, a more targeted selection process is required for control stores given test stores to ensure a comparable testing environment, which deviates from traditional probability sampling methods.

The existing literature comprehensively explores various facets of controlled experimentation and population sampling. However, a significant gap remains in the identification of appropriate control groups given test groups in the context of controlled experiments, particularly in retail settings. This paper aims to bridge this gap by proposing a novel methodology for selecting control groups that ensure the reliability and validity of experimental outcomes. By addressing this gap, the paper contributes significantly to the body of knowledge, providing a robust framework that can enhance the precision and effectiveness of controlled experiments in the retail industry and beyond.

3.1. Contributions

This paper introduces a novel methodology that selects a homogeneous control group matching the test group based on a multi-dimensional KPI.

4. Data Collection and Preprocessing

In the context of A/B testing, data collection serves as the foundation for comparing the performance of different variables to drive data-informed business decisions.

In this methodology, data must be meticulously gathered not only from the test stores but also from the entire network of eligible stores to identify potential control group candidates, ensuring a valid and comprehensive comparison. For this paper, simulated sales data are used to illustrate the process.

KPIs, typically defined by business objectives such as increasing sales, margins, or volume, are pivotal in aligning the test and control stores. These KPIs form the basis for evaluating store homogeneity and are crucial for the experimental design. Given the inherent variability in sales data, the granularity of data aggregation is essential; weekly aggregation is chosen here to strike a balance between reducing variability and maintaining meaningful insights, a decision guided by exploratory data analysis and the nature of the business's data inflow.

The data structure for this analysis includes four critical columns: store ID, test/control group label, date, and KPI (average weekly sales). The test/control label is particularly important for differentiating between the groups during the analytical phase. To quantify the similarity between stores, l_1 and l_2 distances between their time-series sales data are used, aiming for minimal distances to establish homogeneity. Addressing potential biases and anomalies is a preliminary step in the data collection process. Scrutinizing the data for irregularities, such as zero or negative sales figures, ensures that only representative and reliable data are used for further analysis. This careful approach helps in creating a robust methodological framework for selecting homogenous control stores in retail A/B testing scenarios.

Table 1. Sample data format for the problem

Store ID	Test_Control_Flag	Date	Sales
3496586	Pot. CTRL	9/17/2017	$ 120.30
3496587	Pot. CTRL	7/1/2018	$ 1,431.48
3496589	TEST	7/8/2018	$ 15,312

5. Methodology

Considering the broader applicability of the methodology to the selection of user groups, the following terminology will be used henceforth: 'test group,' 'control group,' and 'potential control'. This terminology is applicable across various settings, including retail environments, online platforms, or any experimental context. A unit of a 'group' can be a store, a user, or any other subject. A 'test group' is where specific changes or interventions are applied to the units of the group to assess the impact of the changes on a KPI. A 'control group' refers to a group of units that are similar to the test group and where changes are not implemented. It is important to note that business typically should dictate the timeframe prior to the start of the A/B test and in the predetermined timeframe, the performance of test and control groups should be similar with respect to the KPI. If changes in the KPI occur after the A/B testing period, attributing these changes to the promotional activities in the test group with statistical significance can be considered. The problem of selecting an appropriate control group arises in the absence of a pre-specified control group. A 'potential control' is any unit that is eligible to be considered as a unit of the control group. Potential control units are those on which the changes under the A/B testing are not implemented.

The criteria for selecting units for the control group should ensure the homogeneity between the test and control groups. Statistically, homogeneity is defined as the condition where all units in both test and control groups have the same distribution for the KPI. Given that this is too rigid a condition for practical applications, it is often common practice to relax the criteria to the following. Both groups should have the same expectations and dispersion. Since the research deals with a multi-dimensional KPI, the expectation becomes a mean vector, and the dispersion becomes a covariance matrix. To reduce complexity, the homogeneity requirement is further relaxed to ensure that the test and control vectors share the same mean vectors.

5.1. Mathematical Formulation

The problem is framed as follows: given a pre-specified test group of size p over m contiguous weeks, let T_{ij} be the KPI value of the i^{th} test unit for the j^{th} week, $i = 1,2, \ldots, p$ and $j = 1,2, \ldots., m$. $T^i = (T_{i1}, T_{i2}, \ldots., T_{im})$ is an m-dimensional random vector. Let n be the number of potential units eligible for possible selection into a control group. Let g be a control group of size p selected from the n potential units. Let S_{ij}^g be the KPI value of the i^{th} unit of g for the j^{th} week, $i = 1,2, \ldots, p$ and , $j = 1,2, \ldots, m$. Then, $S^{ig} = \left(S_{i1}^g, S_{i2}^g, \ldots, S_{im}^g\right)$ is an m-dimensional random vector. Statistically, control and test units are homogenous, provided the random vectors. T^i and S^{ig} Have the same distribution for all i. Let $\mu_{T^i} = E(T^i)$ and $\Sigma_{T^i} = cov(T^i)$ denote the expectation and dispersion matrix of T^i Respectively. Assuming that test units are homogeneous, $E(T^i) = \mu_T$ and $cov(T^i) = \Sigma_T$ For all i within the test group. Similarly, assuming control units within g are homogenous among themselves, $E(S^{ig}) = \mu_S$ and $cov(S^{ig}) = \Sigma_{S^g}$ For all i within g. As a *first-order relaxation of homogeneity*, consider the test and control group g are homogeneous, provided $\mu_T = \mu_{S_g}$ and $\Sigma_T = \Sigma_{S_g}$.

As a second-order relaxation of homogeneity, consider the test and control group g are homogeneous, provided $\mu_T = \mu_{S_g}$. The problem considered in this article is to pick a g that is second-order homogeneous. Note that $\hat{y}_T = \frac{1}{p}\sum_{i=1}^{p} T^i$ and $\hat{y}_{S^g} = \frac{1}{p}\sum_{i \epsilon g} S^{ig}$ are unbiased estimators of μ_T and μ_{S^g} respectively. Therefore, the goal is to pick a g so that. $\hat{y}_{S^g}$ is close to $\hat{y}_T$. This alignment is quantified using vector distance measures, such as the Euclidean distance (l_2-norm) or the absolute distance (l_1-norm) (for distance measures and norms, see https://en.wikipedia.org/wiki/Norm_(mathematics)). The selection involves n potential units, of which p are to be picked.

Considering the selection process, the number of possible combinations is $N = \binom{n}{p}$, where N represents the total number of groups that can be formed. Each group can be uniquely represented by a binary vector $a = (g_1, g_2, \dots, g_n)$ where $g_j = 1$ indicates the inclusion of unit j in the group, and $g_j = 0$ otherwise. For each potential group g, the distance $d(g)$ can be calculated as the distance between $\hat{y}_{S^g}$ and $\hat{y}_T$. The objective is to pick a group g for which $d(g)$ is the minimum over all possible N control groups. The challenge lies in finding the vector g that minimizes d(g). This optimization problem is tackled using two different methodologies: a statistical approach and an optimization approach, both of which are explored and compared in this study.

5.2. Statistical Approach

The statistical approach treats the problem as follows: select a group g from the population of N possible groups with equal probability. Let D be the random variable representing the distance of a randomly selected group g. Since N is large, D may be treated to be a continuous random variable. The distribution of D is assessed through a sample of size k. Suppose F is the cumulative distribution function of D. If $d_0 = d(g_0)$ is the minimum distance observed in the sample, then d_0 represents the α-percentile of D, where $\alpha = F(d_0) = P(D \le d_0)$. Consequently, g_0 is a $100(1-\alpha)\%$ optimal solution to the problem, indicating that approximately $100\alpha\%$ of g selections are better than g_0. This method is particularly effective if F exhibits a negatively skewed distribution.

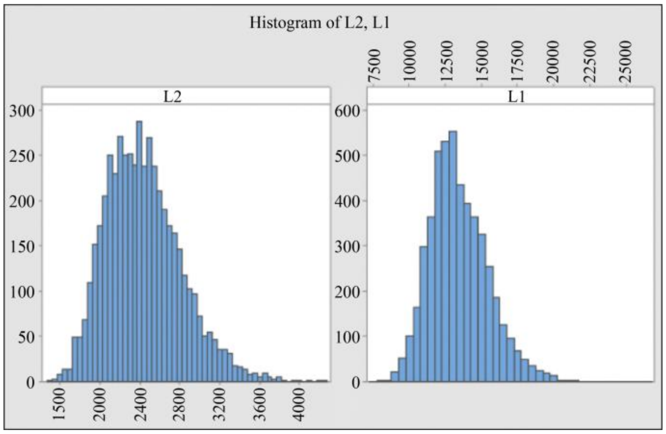


Fig. 1 Sample distributions of l_2 and l_1 distances based on 5000 samples

Consider an example where $m = 51, n = 592$, and $p = 100$, using a sample size $k = 5000$. Two distance measures are evaluated - the l_2-norm and the l_1-norm for k groups (gs) randomly selected. For each group g so selected, $\hat{y}_{S^g}$ and $d(g)$ are computed. For a simulated data set, the histograms of D under these two measures reveal distinct contributions in Figure 1. Their basic statistics are also provided for analysis. In the case of the l_1-norm, a probability plot suggests that a 3-parameter log-normal distribution closely represents the sample data, with parameters indicating location, scale, and threshold (the probability plot is shown in Figure 2). However, the minimum l_1 distance observed in the 5000 samples is 7441, which is significantly higher than the estimated threshold of 3508 obtained from the l_1 distance's probability plot. This discrepancy indicates that a much larger sample size is needed to identify a group that closely matches the test group.

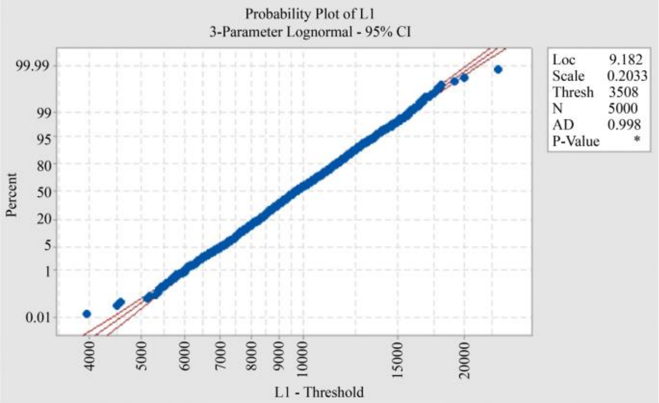


Fig. 2 3-parameter log-normal for l_1 distance

5.3. Optimization Approach

In the optimization approach, the problem is formulated mathematically, aiming to find a binary vector g with exactly p ones such that $d(g)$ is minimized. Two formulations are presented, one for the l_2-norm and the other for the l_1-norm. For brevity of the presentation, the following notation is introduced. Let y be the transpose of $\hat{y}_T$. Let X be the $m \times n$ matrix whose. i^{th} column is the transpose of the KPI vector of the i^{th} potential control unit multiplied by the constant $\frac{1}{p}$. Then $y - Xg = \hat{y}_T - \hat{y}_{S^g}$.

Using l_1-norm: The problem is cast as a binary integer linear programming problem. Introducing two new nonnegative variable vectors q and r of order $m \times 1$ facilitates this formulation. For any g, the difference $y - Xg$ may have positive, negative, or zero elements. Thus, $y - Xg = q - r$, where $q_i = \max(y_i - u_i, 0)$ and $r_i = \max(u_i - y_i, 0)$, with $u = Xg$. The l_1-norm of y – u is equal to $\sum_{i=1}^{m}(q_i + r_i)$. Consequently, the optimization problem is formulated as follows:

$$Minimize: \sum_{i=1}^{m}(q_i + r_i)$$

$$subject\ to: q - r + Xg = y,$$

$$\sum_{i=1}^{m} g_i = p,$$

$$q \ge 0, r \ \ge 0,$$

$$g_i \in \{0,1\}\ for\ all\ i.$$

Using l_2-norm: Here, the objective function is $(y - u)\,(y - u)^T$, which forms a quadratic function in g, given $u = Xg$. This results in a mixed integer non-linear programming problem.

6. Experimental Results

Using l_1-norm: This problem was solved using a commercial Operations Research (OR) solver, yielding a feasible solution g with $d(g)$ = 833 in 40 seconds, with no significant reduction in the objective value after 7 minutes of processing. The final solution had a distance of 827, a substantial improvement compared to the minimum distance of 7441 obtained through the sampling approach.

Using l_2-norm: The mixed integer non-linear programming problem proved computationally intensive, with no feasible solution found within 15 minutes of processing. However, for the l_1-norm solution, the corresponding l_2 the distance can be computed. For example, for g with an l_1 distance of 833, the l_2 distance is found to be 160.

This demonstrates the efficiency of the optimization approach compared to the statistical method. Upon identifying a group similar to the target group, to find another comparable group, the process involves excluding the units identified in the first solution and selecting a new group from the remaining units.

6.1. Limitations of this Methodology

6.1.1. Generalization of Results

One possible concern is whether the findings from the selected control and test scores can be generalized to other stores or the entire retail chain.

6.1.2. Bias in Selection

The method for selecting control stores based on matching test stores' sales trends could introduce biases, as it may overlook other factors influencing sales that are not related to promotional activities.

6.1.3. Impact Isolation

The methodology must robustly isolate the impact of promotional activities from other variables that could affect sales, ensuring that the observed differences are truly attributable to the promotions.

7. Open Research Problems

Providing experiment owners with guided insights, moving beyond just the "what" to understanding the "why" of experiment outcomes.

When A/B testing is not feasible, can quasi-experimental designs like propensity score matching be used to approximate control and treatment groups?

To address the potential loopholes, it would be important to discuss how the methodology accounts for these factors and ensure that the results are robust, reliable, and applicable to broader retail operations.

A methodology for selecting a control group under the first-order relaxation of homogeneity appears to be more complex. Developing a methodology according to this criterion may be explored.

8. Conclusion

The fundamental aim of this study was to underscore the importance of identifying homogenous units for control group selection to enable A/B testing. Two methodologies are proposed to address this problem, one using a statistical approach and the other using an optimization approach. The statistical methodology proposed highlighted the necessity of a substantial sample size to find a control group with low l_1 and l_2 distances to the test group. However, the optimization approach demonstrated superior efficiency by quickly identifying a control group with minimal l_1 distance, suggesting its effectiveness over traditional statistical sampling methods. Specifically, the optimization method yielded a control group with an l_1 distance of 833, positioning it at the 0.05-percentile point of the l_1 distribution, indicative of a highly homogenous group. Furthermore, this study illustrated the applicability of the methodology using a single KPI across a time frame, represented as a column vector. However, the approach's versatility can be extended to multi-dimensional KPIs, such as sales, volume, and margin, which can be analysed as a time-series size over a 3-dimensional vector to assess l_1 and l_2 distances. In conclusion, the operations research approach not only proved to be superior in identifying an optimally homogenous control group but also represents a novel contribution to the field of experimental design in retail analytics. This innovation lays the groundwork for more accurate and reliable experimental designs, thereby enhancing the capacity for data-driven decision-making in the retail industry.

References

[1] Howard Anton, *Elementary Linear Algebra*, 9th Edition, John Wiley & Sons, 1973. [Google Scholar] [Publisher Link]

[2] Amitav Banerjee, and Suprakash Chaudhury, "Statistics without Tears: Populations and Samples," *Industrial Psychiatry Journal*, vol. 19, no. 1, pp. 60-65, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ilker Etikan, and Kabiru Bala, "Sampling and Sampling Methods," *Biometrics & Biostatistics International Journal*, vol. 5, no. 6, pp. 215-217, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[4] Julius Hokka, "*Controlled Experiments for Data-driven Retail Optimization*," Master's Thesis, pp. 1-82, 2023. [Google Scholar] [Publisher Link]

[5] Kirthi Kalyanam et al., "Cross Channel Effects of Search Engine Advertising on Brick & Mortar Retail Sales: Meta Analysis of Large Scale Field Experiments on Google.com," *Quantitative Marketing and Economics*, vol. 16, pp. 1-42, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Ron Kohavi et al., "Online Controlled Experiments at Large Scale," *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1168-1176, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[7] Rembrand Koning, Sharique Hasan, and Aaron Chatterji, "Experimentation and Start-up Performance: Evidence from A/B Testing," *Management Science*, vol. 68, no. 9, pp. 6355-7064, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sharon L. Lohr, *Sampling: Design and Analysis*, Duxbury Press, pp. 1-494, 1999. [Google Scholar] [Publisher Link]

[9] Douglas C. Montgomery, *Design and Analysis of Experiments*, John Wiley & Sons, pp. 1-734, 2017. [Google Scholar] [Publisher Link]

[10] Douglas C. Montgomery, Cheryl L. Jennings, and Murat Kulahci, *Introduction to Time Series Analysis and Forecasting, John Wiley & Sons*, 2015. [Google Scholar] [Publisher Link]

[11] Katta G. Murty, *Linear Programming*, Springer, pp. 1-482, 1983. [Publisher Link]

[12] Federico Quin et al., "A/B Testing: A Systematic Literature Review," *Journal of Systems and Software*, vol. 211, pp. 1-28, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Hamdy A. Taha, *Operations Research: An Introduction*, Prentice Hall Upper Saddle River, NJ, USA, 2013. [Google Scholar] [Publisher Link]

[14] Hamed Taherdoost, "Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research," *SSRN*, vol. 5, no. 2, pp. 18-27, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[15] Ya Xu et al., "From Infrastructure to Culture: A/B Testing Challenges in Large Scale Social Networks," *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2227-2236, 2015. [CrossRef] [Google Scholar] [Publisher Link]

Appendix

Simulation of KPI Data

In the model formulated for the problem (Section 4.1), the problem background is defined as follows. There are N units. Each unit has KPI values for m weeks. Consequently, each of the N units is represented by an m-dimensional KPI vector. The simulation process is composed of two stages: (i) simulating the N m-dimensional vectors and (ii) selecting p test stores from the N stores. The N m-dimensional vectors will be simulated from the distribution of an m-dimensional random vector x. For simulation, x is assumed to follow a multivariate normal distribution, with mean vector $\mu = (\mu_1, \mu_2, \dots, \mu_m)$ and covariance matrix $\Sigma = (\sigma_{ij})$ serving as the parameters of x.

For clarity, units can be considered as stores belonging to a major business across the United States, where the KPI is the weekly sales of a particular product. For simulation purposes, three types of stores are identified: Good (type 1), Moderate (type 2) and Dull (type 3). Each type is characterized mathematically by distinct sets of parameters μ and Σ for the sales vector x. Specifically, (μ^1, Σ^1) represents Good stores, (μ^2, Σ^2) represents Moderate stores and (μ^3, Σ^3) Represents Dull stores. The mean vector μ^k is defined as $\mu_j^k = a_k + jb_k, j = 1,2,\dots,m, \ k = 1,2,3,$ where a_k represents average sales in the first week and b_k indicates the growth rate.

Store classification is based on the value of b_k. For the covariance matrix Σ, two scenarios are considered: (i) different Σ^k for each store type, or (ii) a common Σ for all types of stores.

Thus, there are three populations, one for each type of store, denoted as $normal(\mu^k, \Sigma^k), k = 1,2,3$. Additionally, it is assumed that $100q_k$ Percent of the stores in the population belongs to type k, for $k = 1,2,3$. The initial task involves simulating one m-dimensional sales vector for each of the N stores, resulting in a data matrix D of order $N \times m$ whose. i^{th} a row represents the sales vector for store i.

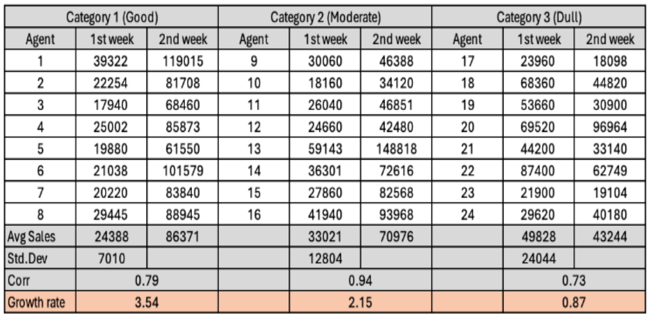
The second task is to randomly select p test stores through simple random sampling of p integers from the set $\{1, 2, \dots, N\}$ without replacement.

Steps for simulating D:

- Select the type k: This selection process involves choosing a type based on the probabilities q_k.
- Generate a random number u from $U(0, 1)$.
- Determine k:
  - If $u \le q_1$, assign k=1.
  - If $q_1 < u \le q_1 + q_2$, assign k=2.
  - If $q_1 + q_2 < u$, assign k=3.
- Sample from the Distribution: Draw a sample $\boldsymbol{x}$ from the distribution $normal(\boldsymbol{\mu^k}, \boldsymbol{\Sigma^k})$ Corresponding to the selected type and append it to $\boldsymbol{D}$.
- Repeat the above steps N times.

Simulation Instance

To derive the parameters for simulation, consider the data presented in the figure below. These values represent actual sales amounts (in dollars) for a specific medicine during two weeks in January. Due to a confidentiality agreement, the data source is withheld. Sales are promoted by 24 agents, as shown in the Figure. For this analysis, weekly sales will be simulated over 52 weeks.



Category 1 (Good)			Category 2 (Moderate)			Category 3 (Dull)		
Agent	1st week	2nd week	Agent	1st week	2nd week	Agent	1st week	2nd week
1	39322	119015	9	30060	46388	17	23960	18098
2	22254	81708	10	18160	34120	18	68360	44820
3	17940	68460	11	26040	46851	19	53660	30900
4	25002	85873	12	24660	42480	20	69520	96964
5	19880	61550	13	59143	148818	21	44200	33140
6	21038	101579	14	36301	72616	22	87400	62749
7	20220	83840	15	27860	82568	23	21900	19104
8	29445	88945	16	41940	93968	24	29620	40180
Avg Sales	24388	86371		33021	70976		49828	43244
Std.Dev	7010			12804			24044	
Corr	0.79			0.94			0.73	
Growth rate	3.54			2.15			0.87	

Fig. Actual sales (in $) under 3 categories of agents

The data reveals that agents are grouped into 3 categories based on their performance. The growth rate for each category is calculated as the ratio of the average sales in the second week to the first week. Agents in Category 1 are classified as Good, with a growth rate of $b_1 = 3.54$. Similarly, Category 2 agents are considered Moderate, and Category 3 agents are Dull, with growth rates of $b_2 = 2.15$ and $b_3 = 0.87$ respectively. The mean vectors, μ^k are established using the first week's average sales for each category: $a_1 = 24388$ for Category 1, $a_2 = 33021$ for Category 2, and $a_3 = 49828$ for Category 3. How should Σ^k be chosen? For simplicity, assume constant variance across all weeks, meaning that $Var\left(x_j^k\right) = \sigma_{jj}^k = \sigma_{11}^k = \sigma_k$ for $j = 1,2, \ldots, m$ (where $m = 51$ for this paper). The covariance between x_g^k and x_h^k, is given as $\sigma_{11}^k \rho_k^{|g-h|}$, where $g, h \in \{1, .., m\}$ and $\rho_k = corr(x_1^k, x_2^k)$, which simplifies the assumptions further. Using these assumptions, the structure of Σ^k is

$$\Sigma^k = \sigma_k \begin{pmatrix} 1 & \rho_k & \rho_k^2 & \cdots & \rho_k^{11} \\ \rho_k & 1 & \rho_k & \cdots & \rho_k^{10} \\ \rho_k^2 & \rho_k & 1 & \cdots & \rho_k^9 \\ \vdots & \vdots & \ddots & \vdots & \\ \rho_k^{11} & \rho_k^{10} & \rho_k^9 & \cdots & 1 \end{pmatrix}$$

In this instance, ρ_k is taken as the correlation between first and second week sales within the respective category. Based on the data, the variances are:

$$\sigma_1 = 1753^2, \ \sigma_2 = 3201^2, \sigma_3 = 6011^2,$$

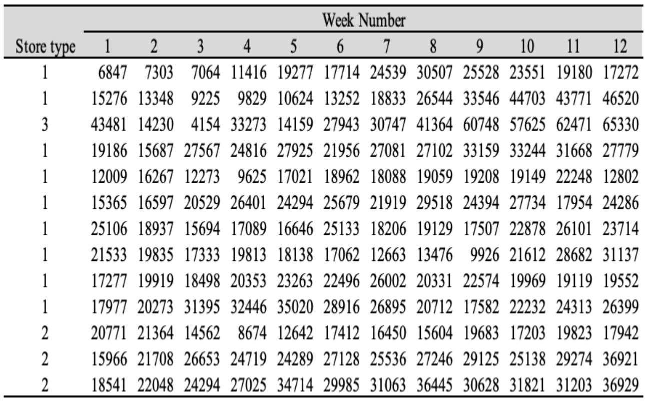
The correlations are:

$$\rho_1 = 0.79, \quad \rho_2 = 0.94 \text{ and } \rho_3 = 0.73.$$

Summarizing, the (g, h) element of Σ^k is given by:

$$\sigma_{gh}^k = \sigma_{11}^k \rho_k^{|g-h|}, \ g, h \in \{1,2, \ldots, m\}.$$

A Python was developed to implement this procedure. Sales data for 100 stores were simulated using probabilities of $q_1 = 0.3, q_2 = 0.5$ and $q_3 = 0.2$. For clarity and due to space constraints, only the first quarter's (12 weeks) sales data are presented in the figure below. The first column identifies the store type. Once the sales data are simulated for the N units, p units can be selected as the test group, and the remaining units can be considered as the potential control group.



	Week Number											
Store type	1	2	3	4	5	6	7	8	9	10	11	12
1	6847	7303	7064	11416	19277	17714	24539	30507	25528	23551	19180	17272
1	15276	13348	9225	9829	10624	13252	18833	26544	33546	44703	43771	46520
3	43481	14230	4154	33273	14159	27943	30747	41364	60748	57625	62471	65330
1	19186	15687	27567	24816	27925	21956	27081	27102	33159	33244	31668	27779
1	12009	16267	12273	9625	17021	18962	18088	19059	19208	19149	22248	12802
1	15365	16597	20529	26401	24294	25679	21919	29518	24394	27734	17954	24286
1	25106	18937	15694	17089	16646	25133	18206	19129	17507	22878	26101	23714
1	21533	19835	17333	19813	18138	17062	12663	13476	9926	21612	28682	31137
1	17277	19919	18498	20353	23263	22496	26002	20331	22574	19969	19119	19552
1	17977	20273	31395	32446	35020	28916	26895	20712	17582	22232	24313	26399
2	20771	21364	14562	8674	12642	17412	16450	15604	19683	17203	19823	17942
2	15966	21708	26653	24719	24289	27128	25536	27246	29125	25138	29274	36921
2	18541	22048	24294	27025	34714	29985	31063	36445	30628	31821	31203	36929

Fig. Simulated sales of the first 12 weeks

Simulation with Common Covariance Matrix

In the approach outlined above, three different covariance matrices were used. If a common covariance matrix is required, the shared Σ should be calculated as the average of the three Σ^k Matrices

Listing 1: Sample Python Code for Generating Data

```python
# Required Python infrastructure
# This code uses the 'numpy' library for numerical computations.
# To install it, use: pip install numpy

import numpy as np

# Function to set the parameters for the multivariate normal distribution
# based on the type of store: good, moderate, or dull.
def set_parameters(k_value):
    """
    Set parameters based on the type of store.
    Parameters:
    k_value (int): A numeric value representing the store type.
                   1: Good stores
                   2: Moderate stores
                   3: Dull stores

    Returns:
    tuple: a, b, s (variance), and r (correlation coefficient).
    """

    if k_value == 1:  # Good stores
        a = 24388      # E(x_1), initial sales
        b = 3.54       # Incremental increase per period
        s = 7010**2    # Variance of sales
        r = 0.79       # Correlation coefficient between periods

    elif k_value == 2:  # Moderate stores
        a = 33021      # E(x_1), initial sales
        b = 2.15       # Incremental increase per period
        s = 12804**2   # Variance of sales
        r = 0.94       # Correlation coefficient between periods

    else:  # k_value == 3 (Dull stores)
        a = 49828      # E(x_1), initial sales
        b = 0.87       # Incremental increase per period
        s = 24044**2   # Variance of sales
        r = 0.73       # Correlation coefficient between periods

    return a, b, s, r

# Example usage:
# good_store_params = set_parameters(1)
# moderate_store_params = set_parameters(2)
# dull_store_params = set_parameters(3)



# Required Python infrastructure
# This function uses the 'scipy' library for multivariate normal distribution.
# To install it, use: pip install scipy

from scipy.stats import multivariate_normal
import numpy as np

# Function to generate a single sample from a multivariate normal distribution.
def generate_multivariate_normal_samples(m, a, b, s, r):
    """
    Generate a single sample from a multivariate normal distribution.

    Parameters:
    m (int): The number of time periods, 51 in the paper.
    a (float): The initial sales value.
    b (float): The incremental change per variable.
    s (float): Variance of the variables.
    r (float): Correlation coefficient between variables.

    Returns:
    np.ndarray: A single sample from the multivariate normal distribution.
    """
    # Create the mean vector with an incrementing pattern based on 'a' and 'b'
    mean_vector = np.array([a + (j - 1) * b for j in range(1, m + 1)])

    # Construct the covariance matrix using the given variance 's' and correlation 'r'
    covariance_matrix = np.array([[s * r**abs(i - j) for j in range(1, m + 1)] for i in range(1, m + 1)])

    # Create a multivariate normal distribution object with the calculated mean and covariance
    mvn = multivariate_normal(mean=mean_vector, cov=covariance_matrix)
```

```python
    # Generate a single sample from the multivariate normal distribution
    sample = mvn.rvs(size=1)

    return sample

# Example usage:
# sample_data = generate_multivariate_normal_samples(5, 24388, 3.54, 7010**2, 0.79)


# Required Python infrastructure
# This code requires the 'pandas' library for DataFrame manipulation.
# To install it, use: pip install pandas

import pandas as pd
import numpy as np

# Number of weeks (columns for the sample data)
m = 12

# Initialize an empty DataFrame with columns representing store type and weeks
columns = list(range(m + 1))
df = pd.DataFrame(columns=columns)

# Rename columns for clarity: 'Type' for store type, and 'Week1', 'Week2', etc., for weekly data
new_columns = {0: 'Type'}
for i in range(1, m + 1):
    new_columns[i] = f'Week{i}'
df = df.rename(columns=new_columns)

# Example usage:
# Define store types and their respective probabilities for random selection
k_values = [1, 2, 3]  # 1: Good, 2: Moderate, 3: Dull
probabilities = [0.3, 0.5, 0.2]  # Probability distribution among the store types

# Number of stores to generate data for
num_stores = 100

# Generate store sales data using the `set_parameters` and `generate_multivariate_normal_samples` functions
for i in range(num_stores):
    # Randomly select a store type based on the defined probabilities
    k = np.random.choice(k_values, p=probabilities)

    # Retrieve parameters based on the selected store type
    a, b, s, r = set_parameters(k)

    # Generate a single sample of weekly sales data
    y = generate_multivariate_normal_samples(m, a, b, s, r)

    # Insert the store type as the first column
    y = np.insert(y, 0, k)

    # Convert the generated data to a DataFrame and rename columns accordingly
    y_df = pd.DataFrame(y.reshape(1, -1), columns=new_columns)

    # Append the new row of store data to the main DataFrame
    df = pd.concat([df, y_df], ignore_index=True)

# Round the values to the nearest integer for clarity
df = df.round(0)

# The DataFrame 'df' now contains the generated data for the specified number of stores
```