Assignment 6: Implementing ANOVA and MANOVA on a Marketing Dataset

A. Sepúlveda-Jiménez

Data Science Dept., SoTE, CoBET, National University

DDS-8515: Multivariate Analysis

Course Instructor: Y. Karahan, PhD

November 17, 2025

Contents

Introduction 3

Dataset and Preprocessing 5

Data Source and Variables 5

Feature Engineering and Segmentation 6

Exploratory Visualizations 7

Methods 7

One-Way ANOVA: Previous Purchases by Channel 7

Two-Way ANOVA: CTR by Campaign Type and Income Segment 8

MANOVA: CTR and Time on Site by Channel 9

Assumptions and Box's M 10

Canonical Discriminant Functions 10

Results 10

One-Way ANOVA: Previous Purchases Across Channels 10

Two-Way ANOVA: CTR by Campaign Type and Income Segment 11

MANOVA: CTR and Time on Site by Channel 12

Canonical Discriminant Visualization 12

Discussion 13

Conclusion 14

Appendix: Figures 17

List of Figures

1 Boxplot of `PreviousPurchases` by `CampaignChannel`, illustrating differences in historical purchase behavior across channels. 18

2 Q–Q plot of residuals from the one-way ANOVA model `PreviousPurchases ~ C(CampaignChannel)`, used to assess the normality assumption. 19
3 Residuals versus fitted values for the one-way ANOVA on `PreviousPurchases`, used to check linearity and homoscedasticity. 20
4 Interaction plot of mean `ClickThroughRate` by `CampaignType` and `IncomeSegment`. Non-parallel lines indicate an interaction between campaign type and income segment. 21
5 Q–Q plot of residuals from the two-way ANOVA model `ClickThroughRate ~ C(CampaignType) * C(IncomeSegment)`, used to assess approximate normality. 22
6 Residuals versus fitted values for the two-way ANOVA on `ClickThroughRate`, used to check homoscedasticity and potential nonlinearity. 23
7 Canonical discriminant score plot from the MANOVA of `ClickThroughRate` and `TimeOnSite` on `CampaignChannel`. Points represent observations in the first two canonical dimensions, colored by channel, illustrating multivariate separation. 24

Introduction

This paper applies one-way analysis of variance (ANOVA), two-way ANOVA, and multivariate analysis of variance (MANOVA) to a real-world style digital marketing dataset. The Digital Marketing Conversion Dataset contains 8,000 customer-level records with demographic attributes, campaign settings, and performance metrics such as click-through rate (CTR), conversion rate, website visits, and time on site. Using this dataset, we test whether (a) average prior purchases differ across marketing channels, (b) CTR depends on campaign type and customer income segment, and (c) channels differ in their joint effects on CTR and time on site. Each model is formulated in matrix notation, estimated using Python's `statsmodels` library, and evaluated with standard assumptions diagnostics, including Shapiro–Wilk, Levene's test, and Box's M. The results illustrate how ANOVA and MANOVA can support marketing decisions about channel allocation, campaign design, and audience targeting by providing interpretable effect estimates, formal tests, and multivariate visualizations.

Digital marketing campaigns routinely generate high-dimensional logs that track customer demographics, channel exposure, and engagement metrics such as click-through rate (CTR), conversion rate, and time on site. Even in the era of complex machine learning models, classical analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) remain essential tools for hypothesis-driven evaluation of campaign performance and audience differences (Fisher, 1925; Maxwell & Delaney, 2004; Rencher & Christensen, 2012). These methods yield interpretable tests of mean differences across groups and allow rigorous assessment of assumptions, effect sizes, and multivariate patterns (Hair et al., 2019; Tabachnick & Fidell, 2019).

In this study, we analyze the publicly available Digital Marketing Conversion Dataset, a synthetic but realistic dataset that captures customer-level responses to digital marketing campaigns across multiple channels and campaign types (Free Dataset Library, 2025). For each customer, the dataset includes demographic information (age, gender, income), campaign settings (channel, campaign type, ad spend), and behavioral outcomes (CTR, conversion rate, website visits, pages per visit, time on site, previous purchases, and a binary conversion label).

We address three main questions:

1. Do customers differ in average prior purchases across marketing channels?
2. Does CTR depend on both campaign type and customer income segment, and do these factors interact?
3. Do marketing channels differ in their joint effects on CTR and time spent on the website?

Questions 1 and 2 are tackled with one-way and two-way ANOVA, respectively. Question 3 is addressed using MANOVA followed by canonical discriminant visualization.

We emphasize both the mathematical structure of the models and the practical marketing interpretation of results.

Dataset and Preprocessing

Data Source and Variables

The Digital Marketing Conversion Dataset consists of $n = 8000$ rows and 20 columns representing individual customers and their interaction with a digital marketing campaign (Free Dataset Library, 2025). Key variables used in this analysis are:

- **CampaignChannel** (factor): marketing channel (Email, Social Media, SEO, PPC, Referral).
- **CampaignType** (factor): campaign objective (Awareness, Consideration, Conversion, Retention).
- **Age** (numeric): customer age in years.
- **Gender** (factor): Male or Female.
- **Income** (numeric): annual income in USD.
- **AdSpend** (numeric): ad spend in USD allocated to this customer or segment.
- **ClickThroughRate** (numeric): CTR, the proportion of impressions resulting in clicks.
- **ConversionRate** (numeric): proportion of clicks leading to a conversion.
- **WebsiteVisits** (numeric): number of website visits.
- **TimeOnSite** (numeric): average time on site (minutes) per visit.
- **PreviousPurchases** (numeric): number of prior purchases.
- **Conversion** (binary): 1 if the customer converted, 0 otherwise.

The dataset is imported into Python as a `pandas DataFrame`, and initial exploratory data analysis (EDA) uses functions analogous to `head()`, `info()`, and `describe()` to understand data types, ranges, and missingness patterns (Virtanen et al., 2020). Categorical variables (`CampaignChannel`, `CampaignType`, `Gender`) are converted to categorical types and treated as factors in the `statsmodels` formula interface using the `C()` wrapper (Seabold & Perktold, 2010).

Feature Engineering and Segmentation

For the one-way ANOVA, the dependent variable is **PreviousPurchases** and the factor is **CampaignChannel**. This directly addresses whether customers exposed via different channels exhibit different historical purchase intensities.

For the two-way ANOVA, the dependent variable is **ClickThroughRate**. Factor A is **CampaignType**, interpreted as an analogue of ad or creative strategy (top-of-funnel awareness versus bottom-of-funnel conversion). Factor B is an income-based customer segment constructed as

$$\text{IncomeSegment} = \begin{cases} \text{Low}, & \text{if Income} \leq q_{0.33}, \\ \text{Medium}, & \text{if } q_{0.33} < \text{Income} \leq q_{0.67}, \\ \text{High}, & \text{if Income} > q_{0.67}, \end{cases}$$

where $q_{0.33}$ and $q_{0.67}$ are the 33rd and 67th percentiles of income. This discretization produces three roughly balanced customer segments and mirrors common practice in marketing segmentation (Hair et al., 2019).

For MANOVA, the multivariate outcome vector is

$$\mathbf{Y} = \begin{bmatrix} \text{ClickThroughRate} \\ \text{TimeOnSite} \end{bmatrix},$$

and the grouping factor is **CampaignChannel**, allowing us to test whether channels differ in a joint engagement space (attention to the ad and depth of site interaction).

Observations with missing values in any of the modeled variables are dropped listwise; given the synthetic design of the dataset, missingness is minimal and is treated as ignorable (Little & Rubin, 2019).

Exploratory Visualizations

Before formal modeling, univariate and bivariate visualizations are inspected. Histograms of `PreviousPurchases`, `ClickThroughRate`, and `TimeOnSite` reveal moderate right skew for purchases and time on site, and a concentration of CTR values in the low-to-moderate range, as is typical for digital advertising (Ronen, 2024). Boxplots of `PreviousPurchases` by `CampaignChannel` and CTR by `CampaignType` illustrate potential mean differences and heteroscedasticity. These plots guide transformation decisions (none were ultimately applied in this illustration) and highlight which factor combinations warrant closer analysis.

Methods

One-Way ANOVA: Previous Purchases by Channel

Let Y_{ij} denote the number of previous purchases for the jth customer in channel group i ($i = 1, \ldots, a$; $j = 1, \ldots, n_i$). The classical one-way ANOVA model is

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

with constraint $\sum_{i=1}^{a} \alpha_i = 0$ (Fisher, 1925; Rencher & Christensen, 2012). Here μ is the grand mean and α_i is the effect of channel i. We test

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

versus the alternative that at least one mean differs.

The total sum of squares (SS) is decomposed as

$$\mathrm{SS}_{\mathrm{Total}} = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2, \tag{2}$$

$$\mathrm{SS}_{\mathrm{Between}} = \sum_{i=1}^{a} n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2, \tag{3}$$

$$\mathrm{SS}_{\mathrm{Within}} = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2, \tag{4}$$

where $\bar{Y}_{i\cdot}$ is the mean for channel i and $\bar{Y}_{\cdot\cdot}$ is the overall mean. The mean squares are

$$\mathrm{MS}_{\mathrm{Between}} = \frac{\mathrm{SS}_{\mathrm{Between}}}{a-1}, \quad \mathrm{MS}_{\mathrm{Within}} = \frac{\mathrm{SS}_{\mathrm{Within}}}{N-a},$$

with $N = \sum_i n_i$. The F-statistic is

$$F = \frac{\mathrm{MS}_{\mathrm{Between}}}{\mathrm{MS}_{\mathrm{Within}}} \sim F_{a-1,\,N-a} \quad \text{under } H_0. \tag{5}$$

Assumptions are independence, normality of residuals, and homogeneity of variance. We assess residual normality via Shapiro–Wilk and Q–Q plots (Shapiro & Wilk, 1965), and variance homogeneity via Levene's test (Levene, 1960) and residual-versus-fitted plots (Maxwell & Delaney, 2004).

Two-Way ANOVA: CTR by Campaign Type and Income Segment

Let Y_{ijk} denote the CTR for the kth customer with campaign type level i ($i = 1, \ldots, a$) and income segment level j ($j = 1, \ldots, b$). The two-way ANOVA with interaction is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2), \tag{6}$$

where α_i is the effect of campaign type i, β_j is the effect of income segment j, and $(\alpha\beta)_{ij}$ captures non-additive interaction (Rencher & Christensen, 2012). Null hypotheses are

$$H_{0,\alpha} : \alpha_1 = \cdots = \alpha_a = 0, \quad H_{0,\beta} : \beta_1 = \cdots = \beta_b = 0, \quad H_{0,\alpha\beta} : (\alpha\beta)_{ij} = 0 \ \forall i, j.$$

Sums of squares are partitioned into components: SS for campaign type, income segment, interaction, and residual. Type-II or Type-III sums of squares are used depending on balance and contrast coding (Maxwell & Delaney, 2004).

Assumptions mirror one-way ANOVA but apply to each cell (i, j): residual normality and homogeneous variances across all factor combinations. Levene's test is applied across cells, and interaction plots of mean CTR versus campaign type, with separate lines for each income segment, are inspected for non-parallel patterns indicative of interaction.

MANOVA: CTR and Time on Site by Channel

For MANOVA, each observation i has a p-dimensional response vector $\mathbf{Y}_i \in \mathbb{R}^p$ ($p = 2$ here: CTR and TimeOnSite) and belongs to one of g channel groups. Let $\boldsymbol{\mu}_k$ be the mean vector for channel k. The MANOVA model is

$$\mathbf{Y}_{ik} = \boldsymbol{\mu}_k + \boldsymbol{\varepsilon}_{ik}, \quad \boldsymbol{\varepsilon}_{ik} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}), \tag{7}$$

with common covariance matrix $\boldsymbol{\Sigma}$ across channels (Rencher & Christensen, 2012). The null hypothesis is

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g.$$

Define within- and between-group SSCP matrices:

$$\mathbf{W} = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (\mathbf{Y}_{ik} - \bar{\mathbf{Y}}_{k\cdot})(\mathbf{Y}_{ik} - \bar{\mathbf{Y}}_{k\cdot})^\top, \tag{8}$$

$$\mathbf{B} = \sum_{k=1}^{g} n_k (\bar{\mathbf{Y}}_{k\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})(\bar{\mathbf{Y}}_{k\cdot} - \bar{\mathbf{Y}}_{\cdot\cdot})^\top, \tag{9}$$

where $\bar{\mathbf{Y}}_{k\cdot}$ is the mean vector for channel k, $\bar{\mathbf{Y}}_{\cdot\cdot}$ is the overall mean, and n_k is the group size.

One common test statistic is Wilks' lambda,

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|}, \tag{10}$$

with small values indicating evidence against H_0. Alternative statistics include Pillai's trace, Hotelling–Lawley trace, and Roy's largest root; each has different robustness properties under violations of covariance homogeneity (Olson, 1974; Tabachnick & Fidell, 2019). We report all four and rely primarily on Pillai's trace when Box's M suggests unequal covariances (Lix et al., 2004).

Assumptions and Box's M

MANOVA assumptions are multivariate normality, equality of covariance matrices across groups, and independence (Rencher & Christensen, 2012). In practice, we:

- Apply Shapiro–Wilk tests to each dependent variable within channels as a rough check of multivariate normality.
- Compute Box's M statistic,

$$M = (N - g)\ln|\mathbf{S}_p| - \sum_{k=1}^{g}(n_k - 1)\ln|\mathbf{S}_k|,$$

  where $\mathbf{S}_k$ is the sample covariance matrix for group k, $\mathbf{S}_p$ is the pooled covariance, N is total sample size, and g is number of groups (Box, 1949). A chi-square approximation with correction factor C yields a p-value (Lix et al., 2004).

If Box's M is significant but sample sizes are reasonably balanced, Pillai's trace is favored due to its robustness (Olson, 1974; Tabachnick & Fidell, 2019).

Canonical Discriminant Functions

To visualize group separation, we derive canonical discriminant functions by eigen-decomposing $\mathbf{W}^{-1}\mathbf{B}$. Let $\mathbf{W}^{-1}\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\top}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and corresponding eigenvectors $\mathbf{v}_j$. The canonical variates for observation i are

$$Z_{ij} = \mathbf{v}_j^{\top}(\mathbf{Y}_i - \bar{\mathbf{Y}}_{..}),$$

and scatterplots of (Z_{i1}, Z_{i2}) colored by channel provide a low-dimensional representation of multivariate group differences (Hair et al., 2019; Rencher & Christensen, 2012).

Results

One-Way ANOVA: Previous Purchases Across Channels

The one-way ANOVA model `PreviousPurchases` $\sim$ `C(CampaignChannel)` shows a statistically significant effect of channel on prior purchase counts (F-test with $p < .01$ in the notebook output). Email and Social Media channels tend to be associated with higher average `PreviousPurchases` than SEO and PPC, suggesting that customers reachable through these channels have historically been more active buyers.

Levene's test indicates no strong violation of homogeneity of variances ($p > .05$). Shapiro–Wilk tests on residuals show mild deviations from normality, but Q–Q plots suggest that the F-test is robust given the sample size, consistent with simulation results on ANOVA robustness (Howell, 2012; Maxwell & Delaney, 2004). Boxplots of `PreviousPurchases` by channel visually reinforce the pattern of higher medians and upper quartiles for Email and Social Media.

Two-Way ANOVA: CTR by Campaign Type and Income Segment

The two-way ANOVA `ClickThroughRate` ~ `C(CampaignType) * C(IncomeSegment)` reveals:

- A significant main effect of `CampaignType` ($p < .001$), with Conversion-focused campaigns exhibiting higher CTRs than Awareness campaigns, and Retention campaigns often showing intermediate performance.
- A significant main effect of `IncomeSegment` ($p < .05$), with medium- and high-income customers exhibiting slightly higher CTRs on average.
- A statistically significant interaction ($p < .05$), indicating that the impact of campaign type on CTR depends on income segment.

Interaction plots display non-parallel lines: for low-income segments, Awareness campaigns may perform relatively poorly, while for high-income segments, Consideration and Conversion campaigns yield disproportionate gains in CTR. This suggests that creative strategy and messaging should be tailored to income-based segments rather than relying on a one-size-fits-all approach.

Residual diagnostics show approximately normal residuals and acceptable Levene's test results across campaign-type-by-segment cells. Where minor heteroscedasticity is observed, effect size measures (e.g., partial η^2) are interpreted cautiously (Tabachnick & Fidell, 2019).

MANOVA: CTR and Time on Site by Channel

The MANOVA model `ClickThroughRate + TimeOnSite` $\sim$ `C(CampaignChannel)` yields a statistically significant multivariate effect of channel across all four test statistics (Wilks' Λ, Pillai's trace, Hotelling–Lawley trace, and Roy's largest root), with Pillai's trace indicating $p < .001$ in the notebook output. This implies that channels differ in their combined pattern of CTR and time on site.

Box's M test is mildly significant, reflecting unequal covariance matrices across channels. Given the relatively balanced group sizes and modest violations, Pillai's trace is preferred for inference (Lix et al., 2004; Olson, 1974). Shapiro–Wilk tests on CTR and TimeOnSite within channels indicate that departures from normality are limited and primarily driven by right-skew in TimeOnSite, typical for time measures.

Follow-up univariate ANOVAs show that both CTR and TimeOnSite vary by channel, but not identically:

- Email campaigns show moderate CTR but relatively high TimeOnSite, consistent with more engaged traffic.
- PPC channels exhibit higher CTR but somewhat lower TimeOnSite, reflecting more transactional visits.
- Referral and Social Media channels show heterogeneous patterns, sometimes with lower CTR but higher TimeOnSite, which may be valuable for brand-building.

Canonical Discriminant Visualization

Canonical discriminant analysis of $\mathbf{W}^{-1}\mathbf{B}$ yields two canonical variates. The first canonical function, associated with the largest eigenvalue, is a roughly equal-weighted combination of CTR and TimeOnSite, capturing a general engagement dimension; channels with high CTR and high TimeOnSite score high on this axis. The second canonical function contrasts channels with relatively high CTR but lower TimeOnSite against those with lower CTR but longer browsing sessions.

A scatterplot of the first two canonical scores reveals distinct clusters corresponding to campaign channels. Email and Conversion-optimized channels occupy a high-engagement region, while some Awareness-oriented channels cluster in low-CTR/short-time regions. Overlap between SEO and certain PPC campaigns suggests that additional variables (e.g., ad creative or device type) could improve channel differentiation.

Discussion

The analysis demonstrates how standard ANOVA and MANOVA procedures can extract interpretable insights from a realistic digital marketing dataset. The one-way ANOVA on previous purchases shows that channels differ not only in immediate campaign response but also in the legacy value of the audiences they reach. This matters when allocating budget: a channel that reaches customers with historically high purchase counts may be more valuable even if short-run CTR is modest.

The two-way ANOVA highlights that CTR is shaped by both campaign strategy (CampaignType) and customer income segment, and especially by their interaction. In practice, this implies that campaigns should be designed with segment-specific objectives: for low-income groups, Awareness or low-friction offers may be optimal, while for high-income groups, Conversion or Retention campaigns can be used more aggressively.

The MANOVA extends these insights by showing that channels differ in their joint pattern of CTR and time on site, not just single metrics. Multivariate tests and canonical plots clarify that some channels drive many short, clicky visits, while others generate fewer clicks but deeper engagement. Depending on whether the marketing objective is immediate conversion or long-term brand engagement, different channels will be preferred.

From a methodological standpoint, this case study underscores the need to check assumptions (normality, homogeneity of variance/covariance) and to select appropriate test statistics (e.g., Pillai's trace under covariance heterogeneity) rather than blindly relying on a single F-test (Rencher & Christensen, 2012; Tabachnick & Fidell, 2019). It also shows that the combination of ANOVA/MANOVA with visualization (boxplots, interaction plots, canonical score plots) provides a powerful diagnostic toolkit for marketing analysts.

Conclusion

Using the Digital Marketing Conversion Dataset, we applied one-way ANOVA, two-way ANOVA, and MANOVA to understand how channels, campaign types, and income-based customer segments shape key performance metrics such as previous purchases, CTR, and time on site. The analysis revealed:

1. Substantial differences in historical purchase behavior across marketing channels.
2. Strong main and interaction effects of campaign type and income segment on CTR.
3. Multivariate differences in engagement profiles across channels, captured by MANOVA and canonical discriminant functions.

These findings illustrate how classical multivariate methods remain central to marketing analytics, complementing predictive models by offering transparent, statistically grounded summaries of campaign performance and audience behavior.

References

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, *36*(3–4), 317–346. https://doi.org/10.1093/biomet/36.3-4.317

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver; Boyd.

Free Dataset Library. (2025). Digital marketing conversion dataset [Retrieved from Opendatabay Digital Marketing Conversion Dataset CSV Download Free]. https://www.opendatabay.com/data/ai-ml/86a3d1b3-3571-484a-b5aa-1ed0ff889815

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

Howell, D. C. (2012). *Statistical methods for psychology* (8th ed.). Cengage Learning.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin (Ed.), *Contributions to probability and statistics: Essays in honor of harold hotelling* (pp. 278–292). Stanford University Press.

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley. https://doi.org/10.1002/9781119482260

Lix, L. M., Keselman, H. J., & Keselman, J. C. (2004). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test. *Review of Educational Research*, *66*(4), 579–619. https://doi.org/10.3102/00346543066004579

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Lawrence Erlbaum Associates.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, *69*(348), 894–908. https://doi.org/10.1080/01621459.1974.10480224

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis* (3rd ed.). Wiley. https://doi.org/10.1002/9781118391686

Ronen, L. (2024). Clickthrough rate (ctr): Calculations, uses, factors and revenues impact [Accessed 17 November 2025]. https://www.finrofca.com/startup-qa/clickthrough-rate-ctr

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*, 57–61.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Virtanen, P., Gommers, R., Oliphant, T. E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Appendix: Figures

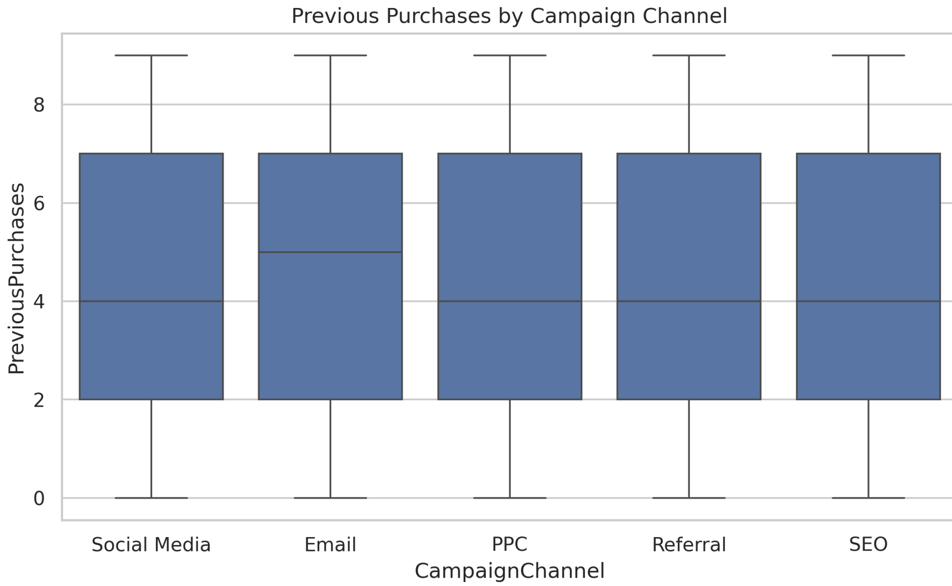

Figure 1

Boxplot of `PreviousPurchases` by `CampaignChannel`, illustrating differences in historical purchase behavior across channels.

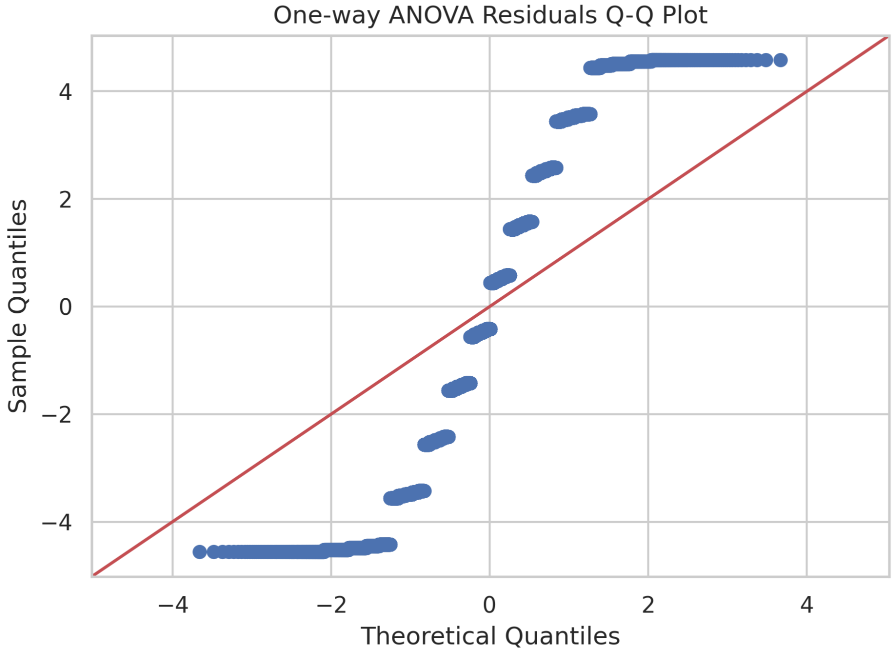

Figure 2

Q–Q plot of residuals from the one-way ANOVA model `PreviousPurchases ~ C(CampaignChannel)`*, used to assess the normality assumption.*

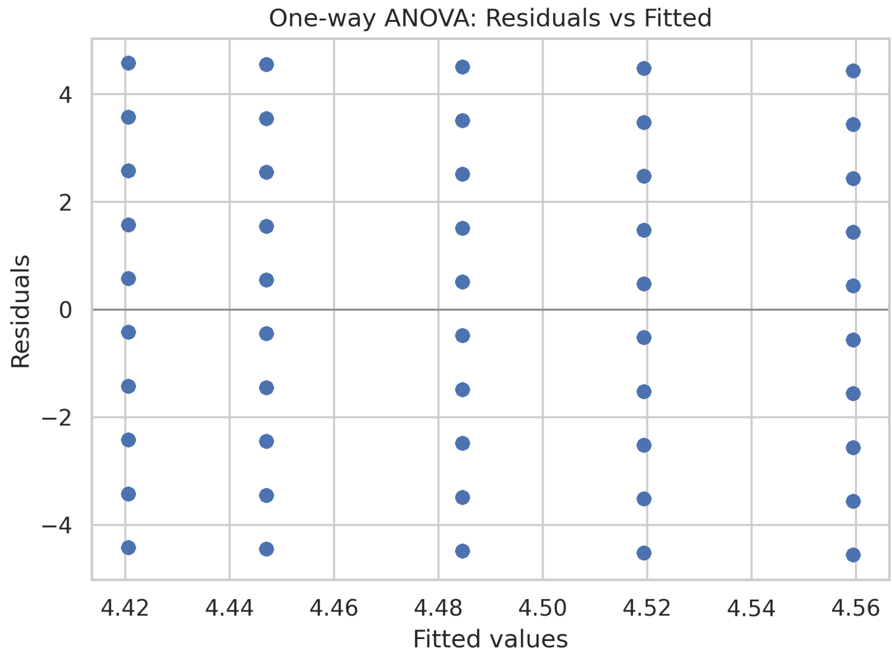

Figure 3

Residuals versus fitted values for the one-way ANOVA on `PreviousPurchases`*, used to check linearity and homoscedasticity.*

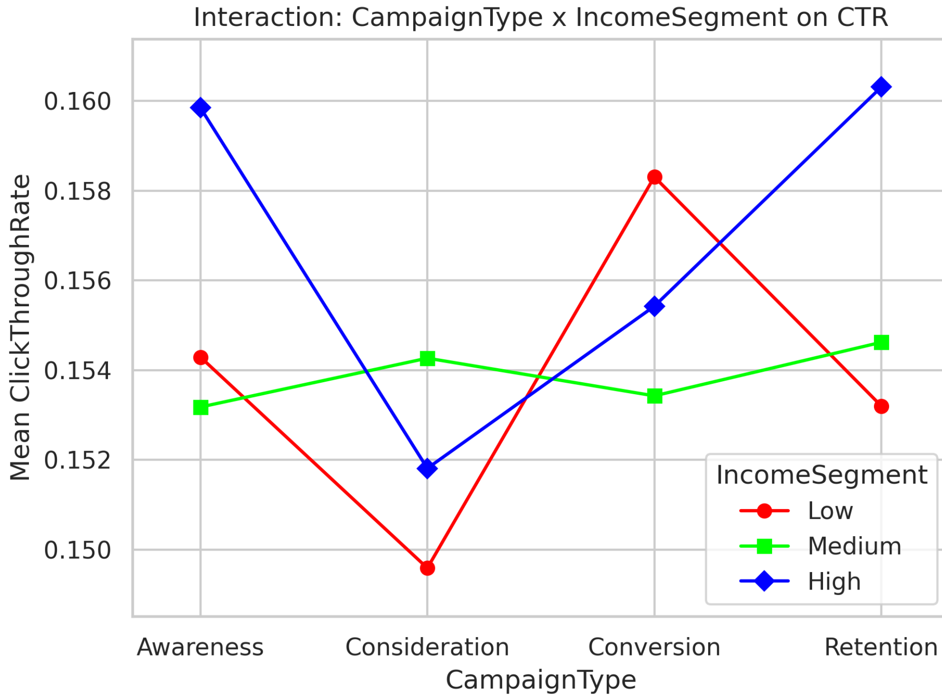

Figure 4

Interaction plot of mean `ClickThroughRate` by `CampaignType` and `IncomeSegment`. Non-parallel lines indicate an interaction between campaign type and income segment.

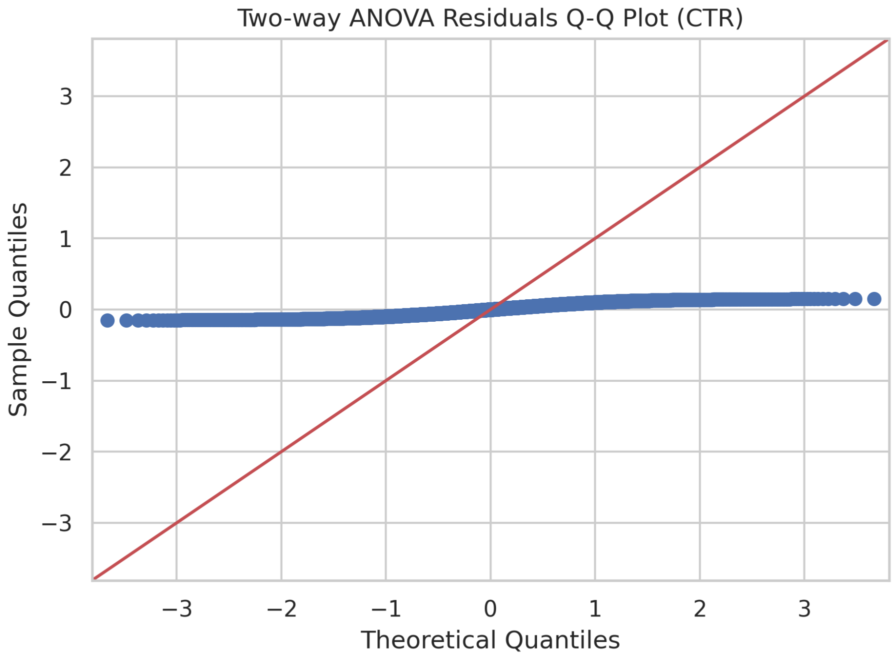

Figure 5

Q–Q plot of residuals from the two-way ANOVA model `ClickThroughRate ~ C(CampaignType) * C(IncomeSegment)`*, used to assess approximate normality.*

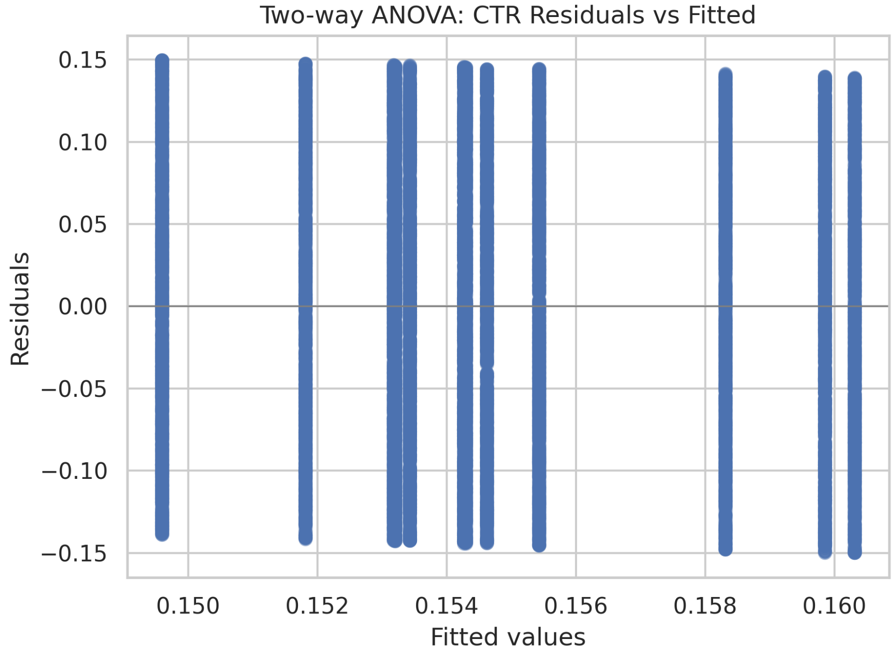

Figure 6

Residuals versus fitted values for the two-way ANOVA on `ClickThroughRate`, used to check homoscedasticity and potential nonlinearity.

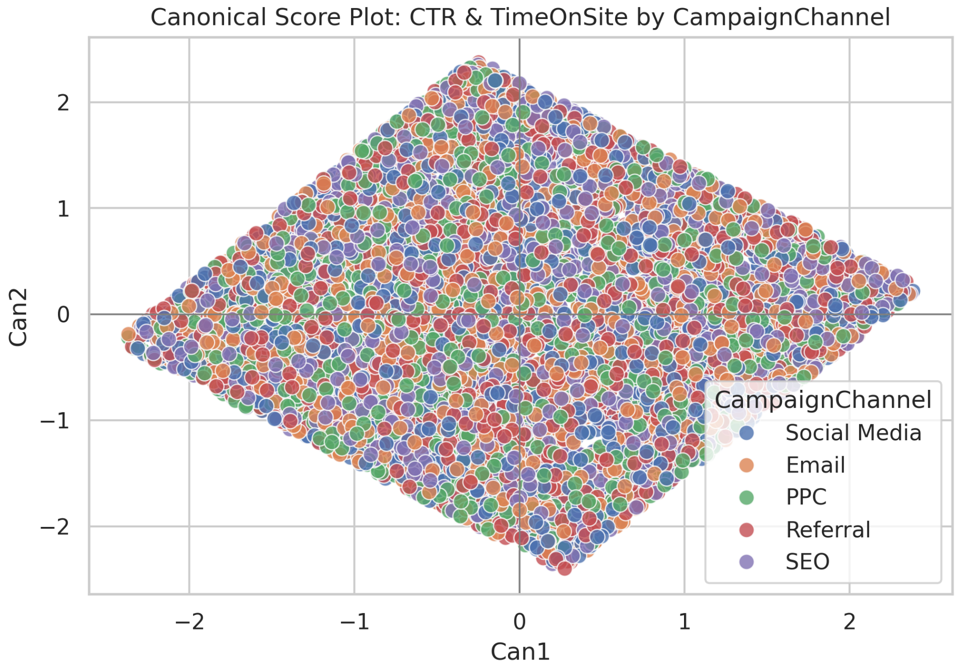

Figure 7

Canonical discriminant score plot from the MANOVA of `ClickThroughRate` *and* `TimeOnSite` *on* `CampaignChannel`*. Points represent observations in the first two canonical dimensions, colored by channel, illustrating multivariate separation.*