1. Cluster Analysis

Definition

Rooted in the core principle that ***"all customers differ,"*** cluster analysis is a segmentation tool that assumes customer heterogeneity and groups individuals with similar behaviours or preferences into clusters that are internally cohesive and externally distinct (Nur & Siregar, 2024). It enables marketers to tailor marketing strategies and offerings for each cluster identified.

Types

- **K-Means:** Pre-specifies the number of clusters (K) and iteratively assigns observations to nearest centroids (efficient for large datasets) (Sharma, 2019).
- **Hierarchical:** Builds a nested cluster structure without pre-setting cluster count (often visualized with dendrograms).

Key Characteristics

- **Sample Size:** No ideal sample size rule (Idlette-Wilson, 2022).
- **Data Type Required:** Continuous, Categorical and Mixed-type data.
- **Similarity Metric:** Manhattan and Euclidean distance (most used) (Yadav, 2025).

$$\text{Euclidean distance formula} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

- **Assumptions:** Requires standardization if variable scales differ (Z-score standardization most common).
- **Limitations:** Sensitive to outliers, cluster interpretation is subjective and no definitive criteria for optimal number of clusters.

Working

Step 1: Clean and standardize data

↓

Step 2: Select appropriate clustering method

↓

Step 3: Select appropriate distance metric

↓

Step 4: Form clusters based on distance proximity

↓

Step 5: Profile clusters

Managerial Implications

- Adjust online pricing in real-time for clusters that are less sensitive to price changes.
- Enables 1:1 marketing by tailoring messaging campaigns for price-sensitive vs. quality-driven clusters.
- Identify at-risk clusters early and deploy loyalty rewards or personalized discounts accordingly.
- Assign high-value clusters to skilled sales reps and low-value clusters to cost-efficient channels like chatbots or email automation.

2. Choice Models

Definition

Choice models are predictive analytics tools that estimate the likelihood of a customer selecting one option over others based on observed attributes (Feng et al., 2021). Grounded in the core principle that ***"all customers change,"*** these models help anticipate future behaviour as customer preferences shift due to price, promotion or product feature variations.

Types

- **Multinomial Logit (MNL):** Estimates the probability of choosing among multiple discrete alternatives (Tapas Mahanta, 2020).

$$\text{MNL Model Formula} = P(i) = \frac{e^{U_i}}{\sum_{j=1}^{J} e^{U_j}}$$

$$\text{Utility Function} = U_i = \beta_{1x1i} + \beta_{2x2i} + \cdots + \beta_{kxki}$$

- **Latent Class Choice Models:** Incorporates unobserved heterogeneity by segmenting users into probabilistic classes.

Key Characteristics

- **Dependent Variable:** Discrete choices i.e., product a or product b.
- **Independent Variables:** Attributes of the choices i.e., price, brand and delivery time.
- **Data Type Required:** Individual-level panel or transaction data.
- **Assumptions:** Rational choice behaviour and logit models assume independence from irrelevant alternatives.
- **Limitations:** Sensitive to model specification.

Working

Step 1: Define the choice set

↓

Step 2: Specify the utility function

↓

Step 3: Estimate model coefficients using regression

↓

Step 4: Calculate choice probabilities

↓

Step 5: Use model outputs to inform marketing decisions

Managerial Implications

- Identify how free delivery or extended warranty messages shift customer choices.
- Model how removing a middle-priced option impacts customer migration to higher or lower tiers.
- Assess whether a limited-time discount increases preference for a lower-performing product.
- Determine how bundling a product with a streaming service or device changes customer choices.

3. Conjoint Analysis

Definition

Rooted in the principle that "***all competitors react,***" conjoint analysis is a quantitative method used to measure customer preferences by evaluating how individuals choose between different product features (Lee et al., 2006). It helps marketers design offerings that stand out in competitive markets.

Types

- **Full-Profile Conjoint:** Respondents evaluate entire product profiles with multiple attributes (Qualtrics, 2018).
- **Choice-Based Conjoint (CBC):** Customers choose from sets of product alternatives, simulating real market scenarios (most widely used) (Gell, 2023).

Key Characteristics

- **Required Sample Size:** Start with 300 respondents (Rule of thumb) (Halversen, 2020)
- **Dependent Variable:** Preferences between product profiles.
- **Independent Variable:** Product features i.e., price, color or delivery time.
- **Data Type Required:** Preference data from customer surveys.
- **Assumptions:** Customers make rational trade-offs in real-world scenarios.
- **Limitations:** Hypothetical bias and limited number of attributes.

Working

Step 1: Select attributes and levels

Step 2: Generate product combinations

Step 3: Collect customer responses

Step 4: Use regression models to estimate the utility of each attribute

Step 5: Predict customers response levels

Managerial Implications

- Quantify customer willingness to pay for product features like battery life or extended warranty.
- Model how a price drop by a competitor affects your product's share of market preference.
- Use preference data to create faster delivery + free return bundles that appeal to target segments.
- Identify low priority features to be excluded to meet cost constraints without minimising value.

4. Market Response Models

Definition

Market Response Models quantify how marketing inputs such as price or promotions affect outputs like sales, market share or customer acquisition (Hanssens et al., 2005). Grounded in the principle that "***all resources are limited***," these models help marketers optimize ROI and make data-backed budget allocation decisions.

Types

- **Linear Regression Models:** Measure the impact of one or more marketing variables on KPIs (Slava Kisilevich, 2023).
- **Log-Linear Models:** Used when effects are multiplicative or for interpreting elasticities (Bensmida, 2023).

$$\text{Ouput} = \beta_0 + (\beta_1 \times \text{Input1}) + (\beta_2 \times \text{Input2}) + \epsilon$$

Key Characteristics

- **Dependent Variable:** Output variables i.e., sales, market share, revenue etc.
- **Independent Variables:** Input variables i.e., price, ad spend, promotions, distribution etc.
- **Data Type Required:** Time-series, cross-sectional or panel data.
- **Assumptions:** Linear relationship for basic models, independent variables not highly correlated.
- **Limitations:** Multicollinearity and unexplained model variance.

Working

Step 1: Collect and prepare historical data

Step 2: Select appropriate model type

Step 3: Estimate model coefficients

Step 4: Simulate future performance under real world scenarios

Managerial Implications

- Allocate budgets toward paid search, retargeting ads or influencer partnerships delivering the highest marketing returns.
- Adjust pricing in segments identified with lower price sensitivity (price-elasticity insights).
- Promote limited-time discounts or flash sales according to periods of highest customer responsiveness.
- Forecast the impact of sales from product launch to optimize marketing campaign scope and investment.

References

Bensmida, K. (2023, July 18). *MassTer*. MASS Analytics. https://doi.org/10952337412/C-o2COKH5OsYEITQveYo

Feng, Q., Shanthikumar, J. G., & Xue, M. (2021). Consumer Choice Models and Estimation: A Review and Extension. *Production and Operations Management*. https://doi.org/10.1111/poms.13499

Gell, T. (2023, February). *Explaining Choice-Based Conjoint Analysis [With Examples]*. Www.driveresearch.com. https://www.driveresearch.com/market-research-company-blog/choice-based-conjoint-analysis/

Halversen, C. (2020, December 29). *Sample Size Rule of Thumb for a Choice-Based Conjoint (CBC) Study - Sawtooth Software*. Sawtoothsoftware.com. https://sawtoothsoftware.com/resources/blog/posts/sample-size-rules-of-thumb

Hanssens, D. M., Leeflang, P. S. H., & Wittink, D. R. (2005). Market response models and marketing practice. *Applied Stochastic Models in Business and Industry*, *21*(4-5), 423–434. https://doi.org/10.1002/asmb.584

Idlette-Wilson, A. (2022, January 17). *Sample Size for Cluster Analysis*. Data Viz for Fun. https://medium.com/business-data-quality-analyst/sample-size-for-cluster-analysis-72260a40e41d

Lee, K. C., Choi, H.-J., Lee, D. H., & Kang, S. (2006). Quantitative Measurement of Quality Attribute Preferences Using Conjoint Analysis. *Lecture Notes in Computer Science*, 213–224. https://doi.org/10.1007/11752707_18

Nur, M. F., & Siregar, A. (2024). Exploring the use of cluster analysis in market segmentation for targeted advertising. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, *5*(2), 158–168. https://doi.org/10.34306/itsdi.v5i2.665

Qualtrics. (2018, November 19). *Conjoint Analysis, Conjoint Types & How to Use Them | Qualtrics*. Qualtrics. https://www.qualtrics.com/en-gb/experience-management/research/conjoint-analysis/

Sharma, P. (2019, August 19). *The Most Comprehensive Guide to K-Means Clustering You'll Ever Need*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/

Slava Kisilevich. (2023, June 14). *Exploring Different Approaches to Generate Response Curves in Marketing Mix Modeling*. Medium; TDS Archive. https://medium.com/data-science/exploring-different-approaches-to-generate-response-curves-in-marketing-mix-modeling-ff6dcc7927f7

Tapas Mahanta. (2020, April 24). *Linear Probability, Logistic and Choice Model - Tapas Mahanta - Medium*. Medium. https://mahanta-tapas.medium.com/linear-probability-logistic-and-choice-model-b29ca3cd4b7a

Yadav, P. (2025, March 2). *Euclidean Distance vs. Manhattan Distance (Machine Learning)*. Medium. https://medium.com/@pawan329/euclidean-distance-vs-manhattan-distance-machine-learning-655ba36439e5