
Advanced Certificate in Marketing Audit and Control

Market Research Techniques

Primary research refers to data that is collected firsthand for a specific marketing audit or control objective. A company may commission a focus group to explore consumer attitudes toward a new product concept. The advantage of primary research is its relevance to the particular problem under study, but it often requires significant time and budget. A common challenge is ensuring that the data collection instruments are free from leading questions that could bias responses.

Secondary research involves the use of existing data sources such as industry reports, government statistics, or previously conducted surveys. For example, a marketer might analyse publicly available sales figures from a trade association to benchmark market share. While secondary research is cost-effective, its limitations include potential outdatedness and lack of alignment with the specific research objectives of the audit.

Qualitative research seeks to understand the underlying motivations, feelings, and perceptions of consumers. Techniques include depth interviews, focus groups, and ethnographic observation. An ethnography might involve a researcher spending several weeks in a consumer's home to observe product usage in a natural setting. The richness of qualitative insight is valuable for developing hypotheses, yet the findings are typically not generalisable to the broader population because of small, non-representative samples.

Quantitative research generates numerical data that can be statistically analysed. Surveys with closed-ended questions are the most common method. For instance, a questionnaire might ask respondents to rate brand satisfaction on a 1-10 scale. Quantitative data allow for the calculation of metrics such as mean, variance, and confidence intervals, making it suitable for testing hypotheses and supporting decision-making in marketing control.

Focus group is a moderated discussion with a small, purposively selected group of participants, usually 6-10 individuals, that explores attitudes toward a product or advertising message. The moderator uses a discussion guide to ensure coverage of key topics while encouraging interaction among participants. A practical application is testing concepts for a new beverage; participants can taste prototypes and discuss flavour preferences. Challenges include groupthink, where dominant participants influence others, and the difficulty of translating qualitative insights into quantifiable recommendations.

Depth interview (also called in-depth interview) is a one-on-one conversation that probes a respondent's personal experiences, motivations, and decision-making processes. This technique is particularly useful for uncovering complex consumer journeys, such as the factors influencing high-involvement purchases like automobiles. The interviewer must be skilled at building rapport and probing without leading the interviewee. A common pitfall is interviewer bias, which can shape the respondent's answers.

Observation research involves watching consumers in real or simulated environments without direct interaction. For example, a retailer may install cameras to track shopper movement through aisles, identifying high-traffic zones. Observation provides objective data on actual behaviour, but it may raise privacy concerns and requires careful ethical handling, especially under regulations such as GDPR.

Ethnography extends observation by immersing the researcher in the consumer's cultural context over an extended period. This approach can reveal tacit usage patterns, such as how families integrate a new kitchen appliance into daily routines. While ethnography yields deep insights, it is resource-intensive and its findings are often difficult to scale across larger markets.

Mystery shopping is a covert evaluation technique where trained shoppers assess service quality, compliance with brand standards, or product availability. A fast-food chain may use mystery shoppers to verify that promotional items are correctly displayed and that staff follow the service script. The main challenge is ensuring that mystery shoppers are consistent in their evaluations and that their reports accurately reflect the typical customer experience.

Survey is a structured questionnaire administered to a sample of respondents to collect quantitative data. Surveys can be deployed via online panels, telephone, face-to-face interviews, or mail. A well-designed survey includes clear, unbiased questions and an appropriate response format (e.g., Likert scales, multiple choice). Poorly worded questions can introduce measurement error, reducing the reliability of the data.

Questionnaire design encompasses the process of developing the set of questions, ordering, and formatting. Effective design follows principles such as using simple language, avoiding double-barreled items, and randomising answer options to minimise order effects. Pre-testing the questionnaire through a pilot study helps identify ambiguous items and improve overall validity.

Sampling is the process of selecting a subset of the population to represent the whole. The quality of the sample determines the credibility of the research findings. Sampling techniques fall into two broad categories: Probability and non-probability methods.

Probability sampling ensures that each unit in the population has a known, non-zero chance of selection. This category includes simple random sampling, stratified sampling, cluster sampling, and systematic sampling. For instance, a researcher conducting a national consumer satisfaction study might use stratified sampling to guarantee proportional representation of age groups, genders, and regions.

Simple random sampling selects respondents purely by chance, often using random number generators. While conceptually straightforward, it may be impractical for large, geographically dispersed populations because of the cost of contacting randomly selected individuals.

Stratified sampling divides the population into homogeneous sub-groups (strata) such as income brackets, then draws random samples from each stratum. This approach improves precision by reducing sampling error within each subgroup, making it ideal for studies where the analyst wants to compare attitudes across

distinct market segments.

Cluster sampling groups the population into clusters (e.G., Neighborhoods) and randomly selects entire clusters for study. It is cost-effective when the sampling frame is unavailable at the individual level, but it may increase sampling error if clusters are internally heterogeneous.

Systematic sampling involves selecting every k th element from an ordered list after a random start. For example, a researcher could select every 10th customer from a retailer's loyalty database. Systematic sampling is easy to implement but can produce bias if the list has a hidden periodic pattern.

Non-probability sampling does not provide each population member a known chance of selection. Techniques include convenience sampling, quota sampling, and purposive sampling. While faster and cheaper, non-probability samples limit the ability to generalise findings and often require caution when interpreting results for marketing control decisions.

Sample size determines the statistical power of the study. Larger samples reduce the margin of error and increase confidence in the estimates, but they also raise costs. Sample size calculators typically require inputs such as desired confidence level (commonly 95%), expected proportion, and acceptable margin of error (often 3-5%). A common challenge is balancing statistical requirements with budgetary constraints.

Confidence interval provides a range within which the true population parameter is expected to fall, given a certain confidence level. For example, a 95% confidence interval of $45\% \pm 3\%$ for brand awareness indicates that the researcher is 95% confident the true awareness lies between 42% and 48%.

Margin of error is the half-width of the confidence interval and reflects sampling variability. A smaller margin of error implies higher precision but typically requires a larger sample. When reporting audit findings, it is essential to disclose the margin of error to stakeholders to contextualise the reliability of the data.

Reliability refers to the consistency of a measurement instrument across time, items, or observers. High reliability indicates that repeated administrations of the same questionnaire would yield similar results. Common reliability metrics include test-retest reliability and internal consistency measured by Cronbach's alpha.

Test-retest reliability assesses stability by administering the same instrument to the same respondents at two different points in time. A high correlation (e.G., >0.80) suggests that the measure is not overly sensitive to transient influences.

Internal consistency evaluates how well items that are intended to measure the same construct correlate with each other. Cronbach's alpha values above 0.70 are generally considered acceptable for marketing research scales such as brand attitude or purchase intention.

Validity examines whether a measurement truly captures the intended construct. Types of validity include

content validity, construct validity, and criterion validity. Content validity is ensured through expert review of questionnaire items, while construct validity may be established via factor analysis. Criterion validity involves correlating the new measure with an established benchmark, such as comparing a new loyalty index with actual repeat purchase data.

Content validity ensures that the questionnaire covers all relevant aspects of the concept under investigation. For a new product satisfaction survey, experts might verify that items address quality, usability, and value for money.

Construct validity confirms that the instrument measures the theoretical construct it purports to measure. Factor analysis is a statistical technique commonly used to assess construct validity by revealing underlying dimensions of a set of items.

Criterion validity is demonstrated when the instrument's scores predict an external criterion, such as actual sales. A high correlation between a brand preference score and subsequent purchase behaviour strengthens the case for using the score in strategic decision-making.

Data collection methods encompass the various ways researchers gather information. Modern marketers increasingly rely on online panels, where pre-recruited respondents complete surveys via the internet. Telephone interviews remain valuable for reaching older demographics less active online, while face-to-face interviews are still the gold standard for in-depth exploration in certain cultural contexts.

Online panel recruitment provides rapid access to large, diverse samples, but panel fatigue can lead to lower data quality. Researchers must monitor response speed and consistency to detect satisficing behavior, where respondents give superficial answers to complete the survey quickly.

Telephone interview offers the advantage of real-time clarification, yet it may suffer from declining response rates due to call screening and the increasing use of mobile phones with unknown numbers.

Face-to-face interview enables observation of non-verbal cues and can improve response rates, especially in regions with limited internet penetration. However, it is costlier and may introduce interviewer bias if the interviewer's demeanor influences answers.

Mail survey and mobile survey are alternative channels. Mail surveys can reach households without internet but have long turnaround times. Mobile surveys leverage smartphones to capture respondents on the go, often incorporating features like geolocation to enrich data. The challenge with mobile surveys is ensuring that the questionnaire layout is responsive and that the length does not exceed respondents' willingness to engage on a small screen.

Data analysis transforms raw data into actionable insights. It typically begins with descriptive statistics to summarise central tendency, dispersion, and distribution shape. Descriptive results are presented in tables, bar charts, or histograms to provide a clear overview for audit stakeholders.

Descriptive statistics include measures such as mean, median, mode, standard deviation, and frequency counts. For example, a mean satisfaction score of 7.2 (On a 10-point scale) indicates a generally positive perception, while the standard deviation of 1.5 Reveals moderate variability among respondents.

Inferential statistics enable researchers to draw conclusions about the broader population based on sample data. Techniques such as hypothesis testing, confidence intervals, and regression analysis are central to inferential work.

Cross-tabulation (crosstab) examines the relationship between two categorical variables. A marketer might cross-tab brand awareness (aware/unaware) with age groups to identify demographic patterns. Chi-square tests assess whether observed differences are statistically significant.

Regression analysis estimates the impact of one or more independent variables on a dependent variable. Linear regression can predict purchase intention based on factors like price sensitivity, brand trust, and advertising exposure. The coefficients indicate the direction and magnitude of each predictor, guiding resource allocation in marketing control.

Factor analysis reduces a large set of items into underlying dimensions, facilitating scale development. For instance, a set of 12 attitude statements about a service may load onto three factors: Reliability, responsiveness, and empathy. Confirmatory factor analysis can then validate the structure in a separate sample.

Cluster analysis groups respondents based on similarity across multiple variables, often used for market segmentation. A retailer may discover distinct clusters such as "price-sensitive bargain hunters" and "premium experience seekers," informing targeted communication strategies.

Conjoint analysis (also known as choice-based conjoint) quantifies the relative importance of product attributes by presenting respondents with a series of hypothetical product profiles. The output includes part-worth utilities that can be used to simulate market share for new product configurations. A practical challenge is ensuring that the design incorporates realistic attribute levels to avoid unrealistic trade-offs.

Choice modeling extends conjoint by incorporating probabilistic models such as multinomial logit to predict the likelihood of selecting each alternative. Researchers can test the impact of pricing strategies, promotional offers, or feature bundles on choice probabilities. Calibration of the model requires a sufficiently large sample to estimate parameters with precision.

MaxDiff (Maximum Difference Scaling) is an efficient method for measuring preference intensity across a set of items. Respondents repeatedly select the most and least important items from a subset. The resulting scores are on a ratio scale, allowing marketers to rank features or benefits with greater discrimination than traditional Likert scales.

Segmentation is the process of dividing a market into distinct groups of consumers who share similar

characteristics or behaviours. Segmentation enables marketers to tailor the marketing mix, optimise media spend, and improve the relevance of communications.

Demographic segmentation groups consumers by observable traits such as age, gender, income, education, or family size. For example, a luxury car brand may target high-income males aged 35-55.

Psychographic segmentation focuses on lifestyle, values, attitudes, and personality. A sports apparel company might target "active achievers" who prioritize performance and self-expression.

Behavioural segmentation categorises consumers based on actions such as purchase frequency, brand loyalty, usage occasion, or benefit sought. Loyalty program data can reveal "high-value repeat purchasers" versus "occasion-based shoppers."

Geographic segmentation divides markets by region, climate, urbanicity, or cultural zones. A beverage firm may launch a citrus-flavoured product in warm climates where demand for refreshing drinks is higher.

Target market is the specific segment(s) a firm decides to serve. Selecting a target involves evaluating segment size, growth potential, competitive intensity, and alignment with the firm's strengths. A misaligned target can lead to wasted resources and poor ROI.

Positioning defines how a brand or product is perceived relative to competitors in the minds of the target audience. Positioning statements articulate the unique value proposition, such as "the most reliable compact SUV for families." Successful positioning is reinforced through consistent messaging across all touchpoints.

Brand equity represents the added value a brand contributes to a product beyond its functional attributes. It encompasses brand awareness, perceived quality, brand associations, and brand loyalty. Measuring brand equity often involves surveys that assess these dimensions, as well as financial metrics such as price premium.

Brand awareness measures the extent to which consumers can recognise or recall a brand. Two common metrics are aided recall (recognition) and unaided recall (spontaneous recall). For instance, a 70% aided recall rate indicates that most respondents recognise the brand when presented with a list of names.

Brand recall (unaided) reflects the ability of consumers to retrieve a brand from memory without prompts. High unaided recall suggests strong top-of-mind presence, which is crucial in high-involvement purchase categories.

Brand loyalty captures the degree of repeat purchase and resistance to switching. Loyalty can be measured through behavioural data (repeat purchase rate, share of wallet) and attitudinal data (likelihood to recommend, emotional attachment). The Net Promoter Score (NPS) is a widely used loyalty indicator that asks respondents how likely they are to recommend the brand on a 0-10 scale.

Net Promoter Score classifies respondents into promoters (9-10), passives (7-8), and detractors (0-6). The NPS is calculated by subtracting the percentage of detractors from the percentage of promoters. A positive NPS indicates more promoters than detractors, signalling strong brand advocacy.

Customer satisfaction gauges how well a product or service meets or exceeds expectations. The American Customer Satisfaction Index (ACSI) and the Customer Satisfaction Score (CSAT) are common measures. High satisfaction is a prerequisite for loyalty but does not guarantee it; therefore, marketers track both metrics.

Customer experience (CX) encompasses the entire journey a consumer has with a brand, from initial awareness through post-purchase support. Mapping CX helps identify friction points, such as long call-center wait times, that can erode satisfaction and loyalty.

Purchase intent reflects the likelihood that a respondent will buy a product in the near future. Intent is often measured on a Likert scale (e.G., "Definitely will," "probably will," "might," "unlikely," "definitely not"). While intent does not always translate into actual sales, it serves as a leading indicator for demand forecasting.

Purchase behaviour analyses the actual actions taken by consumers, such as frequency, basket size, and channel choice. Transaction data from point-of-sale (POS) systems or e-commerce platforms provide objective evidence of behaviour, complementing self-reported intent.

Decision journey maps the stages a consumer moves through from problem recognition to post-purchase evaluation. The classic funnel (awareness → consideration → purchase → loyalty) is now often depicted as a circular loop, acknowledging the role of advocacy and repurchase. Understanding the journey enables marketers to allocate resources to the most influential touchpoints.

Funnel analysis tracks conversion rates at each stage of the decision journey, highlighting drop-off points. For example, an e-commerce site may see a 30% conversion from cart addition to checkout, indicating a need to optimise checkout flow.

A/B testing (split testing) compares two variants of a marketing element (e.G., Headline, button colour) to determine which performs better on a predefined metric such as click-through rate (CTR). The test requires random assignment of users to each version and sufficient sample size to achieve statistical significance.

Multivariate testing extends A/B testing by evaluating multiple variables simultaneously, allowing researchers to identify interaction effects. For instance, testing combinations of image, copy, and call-to-action can reveal the most effective overall design.

Eye tracking records where viewers look on a screen or page, providing insights into visual attention and information hierarchy. Heat maps generated from eye-tracking data show zones of high focus, informing layout decisions for print ads or website design.

Heat map visualises data intensity using colour gradients. In web analytics, heat maps illustrate click density, scroll depth, and mouse movement, helping marketers optimise page elements for higher engagement.

Sentiment analysis applies natural language processing (NLP) to classify text (e.G., Social media posts, reviews) as positive, negative, or neutral. Automated sentiment scores enable real-time monitoring of brand perception, though accuracy can be affected by sarcasm, slang, and language nuances.

Text mining extracts patterns and themes from unstructured textual data. Techniques include keyword extraction, topic modelling (e.G., Latent Dirichlet Allocation), and phrase clustering. Text mining can uncover emerging consumer concerns that are not captured in structured surveys.

Big data refers to extremely large and complex datasets that exceed the processing capabilities of traditional tools. Sources include transaction logs, clickstreams, sensor data, and social media feeds. Harnessing big data requires advanced analytics platforms, but it offers granular insights into consumer behaviour at scale.

Data mining involves discovering hidden patterns, correlations, and anomalies within large datasets. Algorithms such as association rule mining (e.G., Market-basket analysis) reveal product combinations frequently purchased together, informing cross-selling strategies.

Predictive analytics uses historical data and statistical models to forecast future outcomes. In marketing, predictive models can estimate churn probability, lifetime value (LTV), or response likelihood to a campaign.

Forecasting projects future sales, market size, or demand based on past trends and external variables. Time-series techniques such as ARIMA, exponential smoothing, and seasonal decomposition are common methods.

Trend analysis identifies long-term movements in market indicators, such as rising demand for plant-based foods. Recognising trends early allows firms to adjust product portfolios and positioning before competitors.

Market share measures a firm's sales relative to the total market sales. Tracking changes in market share over time provides a clear indicator of competitive performance and the effectiveness of marketing control initiatives.

Competitive analysis examines rivals' strengths, weaknesses, strategies, and market positions. Tools such as SWOT (Strengths, Weaknesses, Opportunities, Threats) and Porter's Five Forces help structure the analysis.

SWOT analysis summarises internal capabilities (strengths, weaknesses) and external conditions (opportunities, threats). For a new product launch, a SWOT may reveal strong R&D capabilities (strength) but limited distribution channels (weakness), alongside rising health-conscious consumer trends (opportunity) and intense price competition (threat).

PESTLE (Political, Economic, Social, Technological, Legal, Environmental) analysis evaluates the macro-environmental forces shaping the market. Understanding PESTLE factors is essential for strategic planning and risk mitigation in marketing audits.

Porter's Five Forces assesses industry attractiveness by analysing competitive rivalry, threat of new entrants, bargaining power of suppliers, bargaining power of buyers, and threat of substitutes. A high threat of substitutes, such as streaming services for traditional TV, may compel a broadcaster to innovate its content offering.

Benchmark provides a point of reference against which performance is measured. Benchmarks can be internal (historical performance) or external (industry averages). For example, a conversion rate benchmark of 2% for e-commerce sites helps marketers gauge the effectiveness of their own site.

KPI (Key Performance Indicator) quantifies progress toward strategic objectives. Common marketing KPIs include CTR, cost per acquisition (CPA), return on investment (ROI), and customer lifetime value (CLV). Selecting relevant KPIs ensures that audit findings align with business goals.

ROI calculates the financial return generated by a marketing investment relative to its cost. $ROI = (\text{Revenue} - \text{Cost}) / \text{Cost}$. Accurate ROI measurement requires reliable attribution of sales to specific marketing activities, which can be complex in multi-channel environments.

Cost per acquisition (CPA) measures the average cost incurred to acquire a new customer. For a digital campaign, $CPA = \text{total spend} / \text{number of conversions}$. High CPA may indicate inefficiencies in targeting or creative, prompting optimisation.

Lifetime value (LTV) estimates the total net profit a customer is expected to generate over the entire relationship with the brand. LTV calculations incorporate average purchase value, purchase frequency, gross margin, and churn rate. Comparing LTV to CPA helps assess the profitability of acquisition strategies.

Churn rate quantifies the proportion of customers who discontinue a relationship within a given period. In subscription services, a monthly churn rate of 5% signals that the business must replace 5% of its subscriber base each month to maintain revenue.

Attrition is another term for churn, often used in B2B contexts. High attrition may result from inadequate product performance, poor service, or more attractive competitor offers.

Loyalty program incentivises repeat purchases through rewards, points, or tiered benefits. Measuring the program's effectiveness involves tracking enrollment rates, redemption frequency, and impact on average order value.

Loyalty index aggregates multiple loyalty dimensions—repeat purchase, advocacy, and emotional attachment—into a single score. A rising loyalty index suggests strengthening customer relationships, which can be a leading indicator of future revenue stability.

Consumer journey mapping visualises every interaction a consumer has with a brand across channels and touchpoints. Mapping helps identify moments of truth where the experience can significantly influence perception and behaviour.

Touchpoint is any point of contact between the consumer and the brand, such as advertising, website visits, in-store interactions, or customer service calls. Optimising each touchpoint contributes to a cohesive omnichannel experience.

Omni-channel strategy integrates multiple channels to provide a seamless experience, allowing customers to switch between online, mobile, and physical stores without friction. Measuring omni-channel performance requires linking data across systems, a challenge often addressed through customer data platforms (CDPs).

Digital analytics encompasses the collection and interpretation of data generated by digital interactions. Web analytics tools track metrics such as page views, bounce rate, and session duration, informing optimisation of the online customer experience.

Web analytics provides insight into how users navigate a website. Key metrics include unique visitors, average pages per visit, and conversion rate. Segmentation by traffic source (organic, paid, referral) helps allocate marketing spend efficiently.

Click-through rate (CTR) measures the proportion of users who click on a link or ad relative to the number of impressions. $CTR = \text{clicks} / \text{impressions} \times 100\%$. Low CTR may indicate mismatched messaging or poor creative relevance.

Conversion rate quantifies the percentage of visitors who complete a desired action (e.g., Purchase, sign-up). Conversion rate optimisation (CRO) involves iterative testing and refinement of landing pages, forms, and calls-to-action.

Bounce rate represents the share of single-page visits where the user leaves without interacting further. A high bounce rate can signal irrelevant traffic, poor page load speed, or unengaging content.

Dwell time measures how long a user remains on a page before navigating away. Longer dwell time often correlates with higher engagement, though context matters (e.g., A long dwell time on a support article may indicate difficulty finding the answer).

Session duration aggregates the total time a user spends on a site during a single visit. Tracking session duration helps assess the overall attractiveness of the site's content and navigation.

Funnel drop-off identifies where users abandon the conversion funnel. For an e-commerce checkout, a significant drop-off at the payment stage may point to concerns about security or lack of preferred payment options.

Attribution modeling assigns credit for conversions to the various marketing touchpoints that contributed to the sale. Common models include first-touch, last-touch, linear, time-decay, and data-driven (algorithmic) attribution.

First-touch attribution gives 100% credit to the initial interaction that introduced the consumer to the brand. This model is useful for evaluating brand-awareness campaigns but overstates their impact on final conversion.

Last-touch attribution attributes all credit to the final interaction before conversion. While simple, it neglects the influence of earlier touchpoints that may have nurtured the consumer.

Multi-touch attribution distributes credit across multiple interactions, providing a more balanced view of the customer journey. Data-driven attribution uses machine learning to weigh each touchpoint based on its actual contribution to conversion.

Data visualization translates complex data sets into graphical formats such as charts, maps, and dashboards, facilitating rapid insight extraction. Effective visualisation follows principles of clarity, relevance, and appropriate use of colour.

Dashboard aggregates key metrics into a single interface, allowing managers to monitor performance in real time. A marketing control dashboard might display KPI trends for spend, CPA, ROI, and NPS side by side.

Reporting communicates research findings to stakeholders through structured documents, presentations, or interactive platforms. Clear reporting includes methodology, key insights, actionable recommendations, and limitations.

Ethics in market research ensures that data collection respects participant rights and complies with legal standards. Core principles include informed consent, confidentiality, and avoidance of deception.

Informed consent requires that participants understand the purpose of the research, what data will be collected, and how it will be used, and that they voluntarily agree to participate.

Confidentiality protects respondents' identities and personal information from unauthorized disclosure. Researchers must store data securely and anonymise records where possible.

Data protection regulations such as GDPR (General Data Protection Regulation) impose strict requirements on personal data handling, including lawful basis for processing, data minimisation, and the right to be forgotten.

Bias refers to systematic errors that distort research findings. Common sources include sampling bias, response bias, and measurement bias.

Leading question subtly prompts a respondent toward a particular answer, compromising data validity. For example, "Don't you agree that our product is the best on the market?" Is a leading question.

Response bias occurs when respondents answer in a way they think is socially acceptable rather than

truthfully. Social desirability bias can inflate positive ratings for a brand.

Social desirability bias leads participants to over-report favourable behaviours (e.G., Recycling) and under-report unfavourable ones. Techniques such as indirect questioning or anonymity can mitigate this bias.

Non-response bias arises when individuals who do not participate differ systematically from those who do, potentially skewing results. High non-response rates in telephone surveys may under-represent younger demographics.

Sampling bias results from an unrepresentative sample selection, such as using only online panels for a product that appeals to older, offline consumers. Corrective weighting can partially address this issue.

Measurement error reflects inaccuracies in data collection instruments, such as poorly worded questions or faulty sensors. Pre-testing and calibration help reduce measurement error.

Data quality encompasses accuracy, completeness, timeliness, and relevance of data. High-quality data underpin reliable insights and sound marketing control decisions.

Data cleaning involves detecting and correcting errors, inconsistencies, and duplicate records before analysis. Common steps include removing outliers, handling missing values, and standardising variable formats.

Outlier is an observation that deviates markedly from the rest of the data set. Outliers can indicate data entry errors, unusual consumer behaviour, or genuine market shifts; analysts must decide whether to retain or exclude them.

Missing data occurs when respondents skip questions or data collection fails. Strategies for handling missing data include listwise deletion, imputation (mean, regression, multiple), or using modelling techniques robust to missingness.

Imputation fills missing values with estimated ones, preserving sample size. Simple mean imputation can bias variance estimates, whereas multiple imputation provides more accurate uncertainty estimates.

Weighting adjusts sample results to reflect known population characteristics (e.G., Age, gender distribution). Weighting improves representativeness but can increase variance, so it must be applied judiciously.

Panel conditioning refers to changes in respondent behaviour caused by repeated participation in surveys, potentially leading to more socially desirable answers or reduced attention. Rotating panel members and limiting survey frequency mitigate conditioning effects.

Longitudinal study tracks the same respondents over time, allowing analysis of changes in attitudes, behaviours, or outcomes. Panel studies are a common longitudinal design, useful for measuring brand

equity evolution.

Cross-sectional study captures data at a single point in time, providing a snapshot of the market. While faster and cheaper than longitudinal designs, cross-sectional studies cannot infer causality.

Experimental design manipulates independent variables to observe causal effects on dependent variables. Randomised controlled trials (RCTs) are the gold standard, ensuring internal validity through random assignment.

Quasi-experimental designs lack random assignment but still include a treatment and a comparison group. Techniques such as matched pairs or propensity score matching help approximate experimental conditions.

Control group does not receive the experimental treatment, serving as a baseline for comparison. In a pricing experiment, the control group would see the standard price while the treatment group sees a discounted price.

Treatment group receives the experimental manipulation. Measuring differences between treatment and control groups reveals the impact of the intervention.

Blind test hides the treatment condition from participants to prevent expectation effects. In sensory testing of a new beverage, participants may be unaware of whether they are tasting the new formula or the existing one.

Double-blind extends blinding to both participants and researchers administering the test, further reducing bias. Double-blind designs are essential in clinical trials and increasingly relevant in marketing experiments involving product samples.

Placebo effect occurs when participants experience perceived changes due to expectations rather than the actual treatment. In marketing, a "free trial" may generate higher satisfaction simply because users expect benefits.

Hypothesis testing evaluates whether observed data support a predefined claim (hypothesis). The null hypothesis (H_0) posits no effect, while the alternative hypothesis (H_1) suggests a relationship exists.

Null hypothesis serves as the default position that any observed difference is due to random chance. Statistical tests aim to reject the null hypothesis in favour of the alternative.

Alternative hypothesis represents the researcher's conjecture, such as "the new ad increases purchase intent."

P-value quantifies the probability of observing data as extreme as the sample results, assuming the null hypothesis is true. A p-value below the significance level (commonly 0.05) Leads to rejection of the null.

Significance level (α) defines the threshold for accepting statistical evidence against the null hypothesis.