JSM 2026 — ASA Section on Statistics in Marketing

Boston, MA · August 3–7, 2026 · Thomas M. Menino Convention & Exhibition Center

Dear colleagues, we are pleased to share the program of sessions sponsored by the ASA Section on Statistics in Marketing at JSM 2026 in Boston. We hope to see you there!

Session 1 · Invited Paper Session

A/B Testing Beyond Tech: Novel Methods for a New Era of Experimentation

Monday, August 3 · 10:30 AM – 12:20 PM · Room CC-210C

Organizers: Nicholas Beyler (Stitch Fix), Frank Yoon (Google)

Main Sponsor: Section on Statistics in Marketing Co-sponsors: Business & Economic Statistics Section; Section on Statistical Learning and Data Science

Launch Decision Performance in Underpowered A/B Tests with AI-Assisted Summarization

Nicholas Beyler · Stitch Fix

Underpowered A/B tests, a common challenge given rapid business cycles, can lead to ambiguous results, increasing the risk of Type I and Type II errors in product or feature launch decisions. While traditional statistical approaches, such as Bayesian methods or sequential testing, are employed to improve statistical power and decision reliability, the adoption of Large Language Models (LLMs) to aid in decision-making is a largely unexamined area of research. We investigate the novel utility of LLMs as aids in the probabilistic assessment required for launch decisions. We conduct a simulation study to contrast decisions made using raw statistical metrics versus LLM-generated decision summaries informed by varying degrees of contextual or prior information. Our findings aim to quantify the conditions under which AI assistance can mitigate sub-optimal launch decisions.

Evaluating A/B Testing Methodologies via Sample Splitting: Theory and Practice

Ryan Kessler · Amazon

We develop a theoretical framework for sample splitting in A/B testing environments, where data for each test are partitioned into two splits to measure methodological performance when the true impacts of tests are unobserved. We show that sample-split estimators are generally biased for full-sample performance but consistently estimate sample-split analogues of it. We derive their asymptotic distributions, construct valid confidence intervals, and characterize the bias-variance trade-offs underlying sample-split design choices. We validate our theoretical results through simulations and provide implementation guidance for A/B testing products seeking to evaluate new estimators and decision rules.

When Design Shapes the Policy: Balancing Business Needs with Experimentation Feasibility

Frank Yoon · Google (co-authors: Natalia Ordaz Reynoso, Google; Nitin Jain)

Business decisions are nuanced, complex, and oftentimes inevitably political. In data-driven decision-making, as scientists, we espouse statistical rigor in how impact evidence is generated from evaluation studies; however, rigor can conflict with feasibility. For example, to evaluate the relative impact of one sales program over another, all eligible customers should be randomized with equipoise, but sales executives may be unwilling to reassign customers to the other program due to customer relationships, business needs, and so on. We illustrate how Google adopted a flexible randomization that aligns with business constraints by addressing lack of equipoise in customer assignments. The flexible randomization modifies the business-as-usual assignment through holdbacks, as customers organically move between programs. This approach respects business processes (with minimal disruption) and still provides high-quality evidence to answer the business policy question at hand, including a detailed view of treatment effect heterogeneity.

Session 2 · Invited Paper Session

Statistical Frontiers in Marketing

Tuesday, August 4 · 10:30 AM – 12:20 PM · Thomas M. Menino Convention & Exhibition Center

This session highlights recent advances in statistical and machine learning methods that improve decision-making in marketing. The talks share a common focus on developing and applying methods to estimate policy-relevant causal effects while accounting for uncertainty, heterogeneity, and long-term customer outcomes.

Learning Robust Surrogate Indices from Many Experiments

Madhav Kumar (co-author: Dean Eckles)

Social scientists and decision-makers are often interested in the long-term effects of interventions, but want to make decisions before these outcomes are realized. This has motivated substantial interest in methods for relying instead on proxies observable in the short term. These methods include using observational and/or experimental data to predict long-term outcomes (or effects) from the short-term proxies (or effect estimates). Under particular assumptions, these predictions are termed surrogate indices and enable identification of treatment effects and decision-making. Here we use a large collection of randomized experiments from the internet industry to learn surrogate indices. Of particular interest is the degree to which an index learned on one subset of experiments is useful for improving decisions more generally. We evaluate some methods for attempting to learn a surrogate index that is more robust to changes to the types of interventions.

Online Sequential Decision-Making with Reinforcement Learning: From Robotics to Marketing

Ran Chen · Washington University in St. Louis

Many decision-making tasks in marketing—such as dynamic assortment and pricing or product recommendation—can be naturally formulated as online sequential decision-making problems. Reinforcement learning (RL), originally developed for engineering applications like robotics and games, provides a powerful framework for tackling these challenges. However, marketing problems are distinct in their human-centric nature, introducing new challenges including continuity, heterogeneity, high-dimensionality, interpretability, personalization, and the need to operate in online rather than offline environments. This talk presents work on developing interpretable and personalized RL algorithms for recommendation systems, demonstrating how RL can support real-time, customer-tailored marketing decisions, as well as interpretable methods for joint assortment and pricing addressing high-dimensionality in bandit-type problems.

Customer Scores

Blakeley McShane · Northwestern University

Customer scores are numbers that value customers along some dimension. They are widely used in practice, for example, in customer loyalty programs and targeting. In this paper, we examine three research questions related to customer scores: (i) how well relatively simple scores perform, (ii) how well scores reflect a long-term future outcome, and (iii) how to score across multiple cohorts of customers that vary in tenure. We examine these questions using datasets from three retailers featuring different product categories, different geographies, and the population of customers. We show that three simple scores perform very similarly to one another and to the predictions of a customer lifetime value model considered a gold standard. Differences are de minimis across scoring periods from one to five years, evaluation periods from one to ten years, and hundreds of cohorts varying in tenure by decades.

Valuing Winners: When and How to Correct for Selection Bias in Randomized Experiments

Walter Zhang · University of Pennsylvania

Decision-makers often deploy the best-performing treatment from a randomized experiment, creating a winner's curse: selection favors treatments whose observed outcomes are high partly because of statistical noise, so the naive estimate of the winner is upward biased. We distinguish two forms of winner's curse—bias relative to the true best treatment (global) and bias relative to the selected treatment's true mean (selective)—and link them to regret from deploying a suboptimal treatment. This framework defines seven decision-relevant evaluation targets. We show that methods that perform well on one target can perform poorly on others, so corrections should be matched to the manager's objective. Across simulations with varying effect sizes, multiple-arm settings, and data calibrated to an online A/B testing platform, no method dominates uniformly. We also introduce an adaptive empirical likelihood procedure that delivers asymptotically valid confidence intervals across settings without the tuning sensitivity of resampling-based methods.

Session 3 · Topic-Contributed Paper Session

2026 ASA Statistics in Marketing Doctoral Research Award Finalists Presentation

Tuesday, August 4 · 4:00 PM – 5:50 PM · Room CC-156A

Organizers: Shibo Li, Gourab Mukherjee (University of Southern California) Chair: Shibo Li

AI for Customer Journeys: A Transformer Approach

Zipei Lu · University of Maryland

When analyzing a sequence of customer interactions, it is important for firms to understand how these interactions align with key objectives, such as generating qualified customer leads, driving conversion events, or reducing churn. We introduce a transformer-based framework that models customer interactions in a sequence similar to how a sentence is modeled as a sequence of words by Large Language Models. We propose a heterogeneous mixture multi-head self-attention mechanism that captures individual heterogeneity in touchpoint effects. The model identifies self-attention patterns reflecting both population-level trends and unique relationships between touchpoints within each customer journey. This results in more accurate predictions, enabling precise targeting and outperforming existing approaches such as hidden Markov models, point process models, and LSTMs. Extensive simulations further establish the model's superiority over competing approaches across various data generating processes.

Blind Targeting: Personalization under Third-Party Privacy Constraints

Anya Shchetkina · The Wharton School

Major advertising platforms recently increased privacy protections by limiting advertisers' access to individual-level data. Instead of providing granular raw data, platforms only allow a limited number of aggregate queries further protected by differentially private noise. This paper studies whether and how advertisers can design effective targeting policies within these restrictive environments. The proposed method, based on Bayesian optimization, introduces two innovations: (i) integral updating of posteriors to select the best data regions to query, and (ii) a targeting-aware acquisition function that dynamically selects the most informative regions for the targeting task. Applied to the Criteo AI Labs dataset (14M users), a naive benchmark achieves only 33% of non-privacy-preserving targeting potential in some cases, while the proposed method achieves 97–101% and is statistically indistinguishable from Causal Forest, a state-of-the-art method.

Correcting Endogeneity via Nonparametric Copula Control Functions

Xixi Hu · University of British Columbia

We propose a new framework that addresses endogenous regressors using a novel conditional copula endogeneity model to capture the regressor-error dependence unexplained by exogenous regressors. Building on the model, we develop a two-stage nonparametric copula control function approach (2sCOPE-np) for endogeneity correction without relying on instrumental variables. The method relaxes the restrictive Gaussian copula assumption and unifies and generalizes existing copula-based endogeneity correction methods. Notably, 2sCOPE-np can handle discrete endogenous regressors (e.g., binary or count) by leveraging exogenous control variables to smooth discrete conditional CDFs. Simulation studies demonstrate that the proposed method outperforms existing methods, illustrated via an empirical example of store sales estimation.

How Visual Designs Drive Success: Interpretable Generative AI for Data-Driven Design

Lan Luo

Visual designs are widely used in marketing (e.g., packaging, ads, media covers) to achieve outcomes such as improved sales, click-through rates, and brand attitudes. This paper develops a novel methodological framework to automatically discover interpretable features that make visual designs successful. A fine-tuned generative text-to-image model (Stable Diffusion 3.5) enables designs to be described by low-dimensional representations; an adaptation of sparse autoencoders then discovers a taxonomy of interpretable and managerially relevant features predictive of success. Applied to book cover redesigns and sales on Amazon.com (160,000+ books), just 30 discovered features improve variation explained in sales by nearly as much as prices and by more than reviews. Back-of-the-envelope calculations suggest a large publisher could increase annual revenue by over $9.1 million—equivalent to an 8.5% price discount. A lab study provides causal evidence that the framework can redesign covers to significantly improve preferences.

Reference-Dependent Pricing

Ningyin Xu · University of Chicago

Firms rely heavily on promotions to drive short-term sales, but frequent promotions may inadvertently lower consumers' reference prices, reducing the effectiveness of future regular pricing. This paper provides the first field evidence of the causal reference price effect through a large-scale multi-stage experiment with a food vending machine company. A 10% increase in the first-stage reference price improves second-stage net revenue and quantity sales by about 1–3%. A micro-founded structural model of demand based on Kőszegi and Rabin (2006) allows clean separation of price sensitivity from reference dependence and enables counterfactual analysis. Simulations show that reference-aware

firms achieve about 6% higher profits through less frequent but more strategic promotions, as myopic firms misattribute loss aversion to price sensitivity and run excessive promotions that erode long-term profitability.

Section Mixer

Section on Statistics in Marketing Mixer

Tuesday, August 4 · 6:00 PM – 7:30 PM · Westin Boston Seaport District, Room W-Revere

Join fellow section members for networking and conversation. All are welcome!

We look forward to seeing you in Boston!

ASA Section on Statistics in Marketing