SESSION 2B

MAXDIFF + MNL

SPRING 2025

UCSD MSBA – MGTA 495

PLAN FOR TODAY

I. Recap on Fisher's Approach

- An ideal inference machine for statistics and econometrics?

II. MaxDiff

- A substantial improvement over ranking or rating exercises on surveys

III. Multi-Nomial Logit (MNL) Model

- Familiar to students who took MGT 100
- Basis of a large set of Economists' demand models
- Useful for MaxDiff (this week) and Conjoint (coming up)

MLE RECAP

THE LIKELIHOOD AND MLE

The statistical step

You have to assume

- **a probability model** for your data $f(Y_i|\theta)$
- how the data are related (or were collected): in this class, i.i.d.

This leads to the **likelihood function**

$$L_n(\theta) = f(Y_1, \dots, Y_n|\theta) = \prod_{i=1}^{n} f(Y_i|\theta)$$

Almost purely computational

The estimator $\hat{\theta}$ maximizes the (log) Likelihood, and provide a point estimate, the **MLE**:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L_n(\theta) = \arg\max_{\theta} \ell_n(\theta)$$

CONNECTION TO REGRESSION MODELS

If you also observe x_i for each y_i

- Denote the probability model more-completely as $f(y_i|\theta_i, x_i)$
- And possibly model $\theta_i = h(x_i, \beta)$. GLMs use $\theta_i = g^{-1}(x_i'\beta)$

Common (generalized linear) regression models:

- **Linear regression** is $f\left(y_i\middle|\mu_i, \sigma_i^2, x_i\right) = N(\mu_i, \sigma^2)$ with $\mu_i = x_i'\beta$ (notice g is the identity function)
- **Logit/probit regression** is $f(y_i|\pi_i, x_i) = Ber(\pi_i)$ with $\pi_i = \frac{1}{1+\exp(-X_i'\beta)}$ or $\pi_i = \Phi(X_i'\beta)$
- **Poisson regression** is $f(y_i|\lambda_i, x_i) = Pois(\lambda_i)$ with $\lambda_i = e^{x_i'\beta}$

FREQUENTIST INFERENCE FOR MLEs

If you hypothesize $H_0: \theta = \theta_0$,
How likely is θ_0 given your data (and thus estimate $\hat{\theta}_{MLE}$)?

Intuition

We look at the (log) **likelihood ratio** as our test statistic:

$$T^2(\theta_0) = \log \frac{L(\theta_0)}{L(\hat{\theta}_{MLE})} \approx -\frac{1}{2} H^{-1}_{\hat{\theta}_{MLE}} \left(\hat{\theta}_{MLE} - \theta_0\right)^2$$

$$T(\theta_0) = \frac{\hat{\theta}_{MLE} - \theta_0}{s(\hat{\theta}_{MLE})} \xrightarrow{d} N(0,1)$$

Taylor Series Approximation

As usual, we can use this to do inference:

CLT Asymptotics

- **Confidence Intervals:** $\hat{\theta}_{MLE} \pm c_{\alpha/2}\, s(\hat{\theta}_{MLE})$

Slay! Lit! Fire!
Asymptotically G.O.A.T.

- **Hypothesis Tests:** reject $H_0: \theta = \theta_0$ if $|T(\theta_0)| > c_{\alpha/2}$

where e.g. the critical value $c_{\alpha/2} = 1.96$ for a 95% C.I.

BOOTSTRAP STANDARD ERRORS

Suppose $\mathcal{H}_{\hat{\theta}_{MLE}}$ is hard to calculate, but you can find $\hat{\theta}_{MLE}$ reasonably quickly

One general-purpose method to get standard errors is to **bootstrap**

For b in $1, \ldots, B$:

- **Sample (*with replacement*) n obs from the dataset, call this bootstrap dataset b**
- **From it, calculate the quantity of interest, here $\hat{\theta}_{MLE}^{(b)}$**

To get your confidence interval:

- Use the **sample standard deviation** of your B values of $\hat{\theta}_{MLE}^{(b)}$ **as your estimated standard error**, or
- Simply take the **empirical quantiles** of the B values of $\hat{\theta}_{MLE}^{(b)}$ directly **as your confidence interval**

IN-CLASS EXERCISE

Use the Conversion Data to fit a logit regression of Purchase on Index.

1. Fit GLM using built-in command (in R, use `glm`)
   - Calculate confidence interval (CI) for slope coef.
2. Code the likelihood and maximize it
   - Check estimates vs (1)
   - Calculate standard errors from the hessian; check against (1)
   - Get CI for slope coef
3. Bootstrap
   - Get CI for slope coef using sample standard deviation approach
   - Get CI for slope coef using direct quantile approach

BEST-WORST SCALING (AKA MAXDIFF)

INTRO TO RATING

Suppose you have a **list of “items”**: either objects, people, brands, attributes, slogans, messages…

The objective is to **measure** each item on one or more underlying, **latent subjective scales** (ie, you want respondents/consumers to **“rate” the set of items**)

There are **no** first-principles of human behavior that suggest we know what is meant by someone else’s ratings:

- On a 1 to 10 scale, is an 8 for you the same thing as an 8 for me?
- What about someone with dyscalculia or who is innumerate?

But we do know how to analyze **choices**

- In economics, consumer behavior stems from humans choosing most-preferred outcomes
- In statistics, there is a rich sub-discipline of Discrete Choice Modeling

INTRO TO BEST-WORST SCALING (AKA MAXDIFF)

MaxDiff is an analytic package of 3 components:

1. a **survey methodology** that is
2. implemented via an **experimental design**, and
3. analyzed with a **statistical model**

This method is often used in place of ratings-scale questions or a ranking exercise

Marketing Examples:

- Measuring the relative importance of product features
- Testing relative effectiveness of product claims or advertising messages
- Comparing package designs

SURVEY METHODOLOGY

Survey respondents are presented with a set of 3-6 items (per screen) from the overall list of items, and respondents simply **select** their **most-preferred** and **least-preferred** among the subset currently presented. This process is repeated over 6-20 screens for each respondent.

Traditional

MAXDIFF TASK

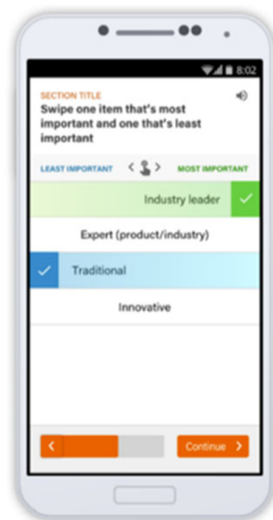
MAX1.

Please select the item you **most** prefer and the one is **least** prefer.

If you see two items with equal appeal, please pick the one that first comes to mind.

Most Likely (choose one)		**Least Likely** (choose one)
	Item 10	
X	Item 7	
	Item 1	
	Item 5	X
	Item 4	

Mobile-first



SURVEY METHODOLOGY

To counter any potential biases in data collection, the items and screens are shown in what **looks like** a **random** order to respondents.

In fact, the subsets are **systematically determined**, and the order is carefully chosen so that, for each respondent, each item appears an equal number of times (**one-way level balance**) and each item appears equally often with each other item (**two-way level balance**).

Each respondent sees a different combination of subsets and orders (called **versions** or **blocks**)

BIBD for 6 objects.

Set	Object codes			Object names		
1[a]	1	2	5	Central Coast beach house	Katoomba upmarket hotel	South Coast, heritage village
2	2	3	6	Katoomba upmarket hotel	Barrington Tops, an isolated setting	Sydney, upmarket hotel
3	3	4	2	Barrington Tops, an isolated setting	Bowral, Southern Highlands	Katoomba upmarket hotel
4	4	1	3	Bowral, Southern Highlands	Central Coast beach house	Barrington Tops, an isolated setting
5	2	5	4	Katoomba upmarket hotel	South Coast, heritage village	Bowral, Southern Highlands
6	3	5	6	Barrington Tops, an isolated setting	South Coast, heritage village	Sydney, upmarket hotel
7	4	6	5	Bowral, Southern Highlands	Sydney, upmarket hotel	South Coast, heritage village
8	1	2	6	Central Coast beach house	Katoomba upmarket hotel	Sydney, upmarket hotel
9	5	1	3	South Coast, heritage village	Central Coast beach house	Barrington Tops, an isolated setting
10	6	4	1	Sydney, upmarket hotel	Bowral, Southern Highlands	Central Coast beach house

PROS

✓ **Cognitively simple**

- Humans are good at trade-offs and regularly assess what is best/worst or most/least preferred

✓ **Built on solid theoretical foundation**

- A highly regarded economic-psychological theory of choices underlies MaxDiff; there is no theory from first principles underlying Likert-scale ratings

✓ **Forces trade-offs and differentiation; can include an "anchor"**

- Respondents can't say "everything is important" which they often do with rating scales

✓ **Scales up to many items**

- Techniques are available to handle large (30+) and huge (100+) item lists (eg, Sparce, Express, and Bandit MaxDiff)
- MaxDiff > Ranking. Ranking long lists is too cognitively demanding on respondents to produce reliable results

✓ **Avoids scale-use bias**

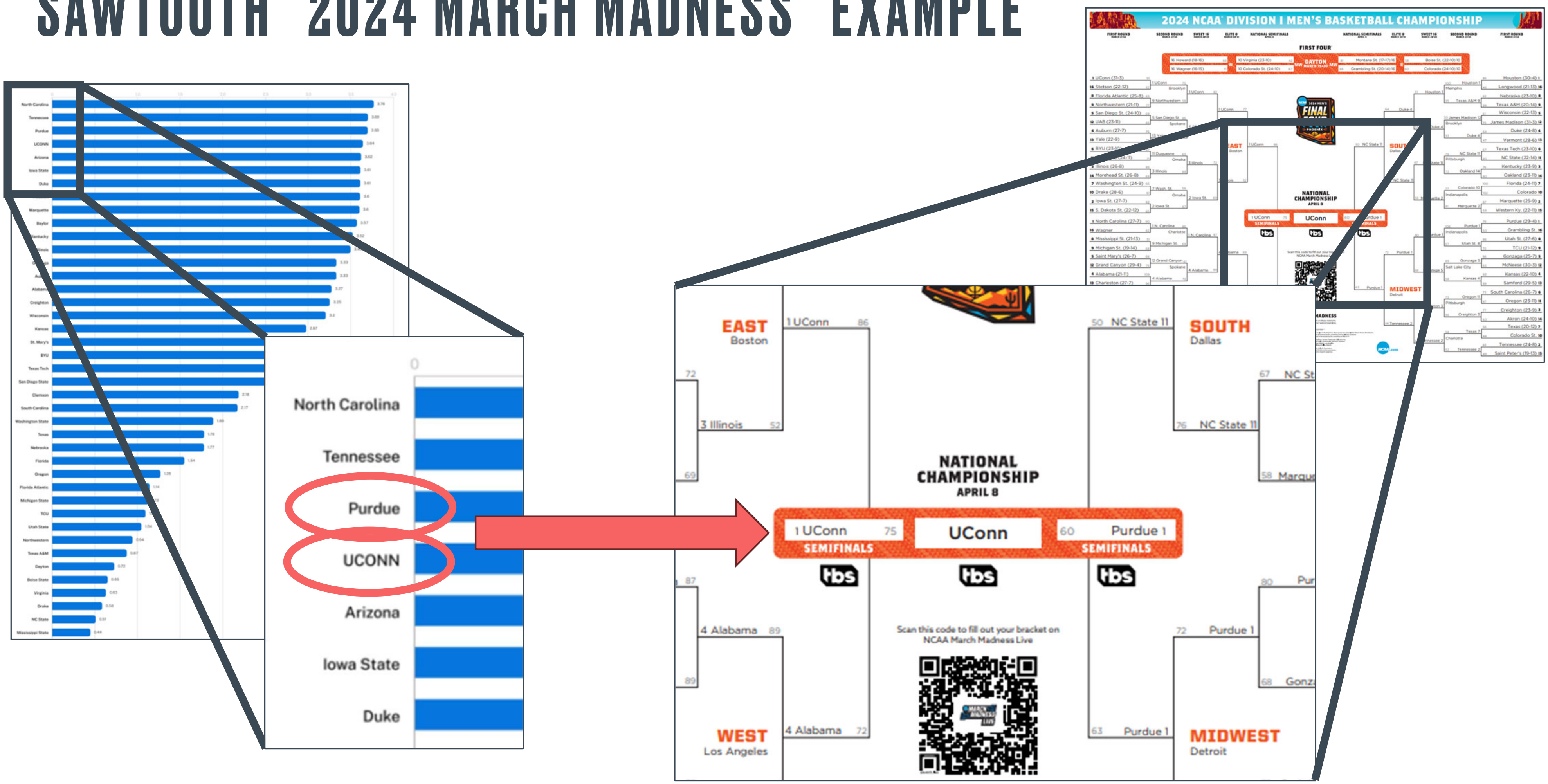
- Respondents from different countries are known to use scales differently; since MaxDiff doesn't use a scale, there is no scale-use bias

CON(S)

✕ **More time-intensive**

- Survey respondents answer more questions than with Likert ratings questions
- Survey development is more complex; you need to create and implement the design
- Data analysis can be more complex; for individual-specific results, you need to fit a statistical model

SAWTOOTH "2024 MARCH MADNESS" EXAMPLE



BASIC ANALYSIS

A statistic $T(\boldsymbol{y})$ is called a **sufficient statistic** for an unknown parameter θ if the conditional distribution of all the data $\boldsymbol{y}$ given $T(\boldsymbol{y})$ does not depend on θ: $f(\boldsymbol{y}|\theta) = h(\boldsymbol{y})g(\theta, T(\boldsymbol{y}))$

For the "aggregate" MNL model discussed next, there is a simple summary statistic for the rating of each item:

- The **percentage of times** it is **selected best** (out of the number of times shown), **minus**
- The **percentage of times** it is **selected worst** (out of the number of times shown)

To measure the precision of these estimates without a statistical model, you can bootstrap.

Source: Marley & Louviere (2005) "Some probabilistic models of best, worst, and best-worst choices." *Journal of Mathematical Psychology*.

IN-CLASS EXERCISE

Calculate scores for MSBA courses from MaxDiff survey data

1	403 AI-Assisted Math/ Programming [Nijs, August]
2	415 Analyzing Unstructured Data [Shang]
3	444 Business Analytics Consulting [Peterson]
4	451 Analytics in Marketing, Finance and Ops [Wilbur, Buti, Shin]
5	452 Collecting and Analyzing Large Data [Hansen]
6	453 Business Analytics [August]
7	454 Business Analytics Capstone Project [multiple]
8	455 Customer Analytics [Nijs]
9	456 Supply Chain Analytics [Park]
10	457 Business Intelligence Systems [Shibler]
11	458 Experiments for Business Analytics [Johnson]
12	464 Big Data Tech: SQL and ETL [Perols]
13	495 Deep Learning[Nijs]

MNL MODEL

MNL DATA

Suppose you have $i = 1, \ldots n$ consumers who each select exactly one item j from a set of J items

The **outcome variable** is the identity of the item chosen $y_i \in \{1, \ldots, J\}$ or equivalently a length-J vector of $J - 1$ zeros and 1 one, where one indicates the selected product

- For example, if the third product was chosen out of four products, the either $y = 3$ or $y = (0,0,1,0)$ depending on how you want to represent it.

Suppose also that you have a **vector of data on each product $\boldsymbol{x_j}$** (e.g., size, price, etc.).

- With Conjoint, we will have many features in x_j, but for MaxDiff x_j is just a length-J binary vector indicating which item is item j

MNL PROBABILITIES

The **multinomial logit (MNL) model** posits consumer i get utility u from product j:

$$U_{ij} = x_j'\beta + \varepsilon_{ij} \qquad \Rightarrow \qquad P_i(j) = \frac{e^{x_j'\beta}}{\sum_{k=1}^{J} e^{x_k'\beta}}$$

For example, if there are 4 products, the probability that consumer i chooses product 3 is:

$$P_i(3) = \frac{e^{x_3'\beta}}{e^{x_1'\beta} + e^{x_2'\beta} + e^{x_3'\beta} + e^{x_4'\beta}}$$

Notice: this simplifies to the binary logit model if there are only two options ($j \in \{0,1\}$) with the x's normalized such that $x_0 = 0$ and $x_1 = x$:

$$P_i(1 \text{ from } \{0,1\}) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} = \frac{1}{1 + e^{-x'\beta}}$$

NO CUSTOMER ATTRIBUTES (YET)

You can't simply add customer attributes x_i to the MNL model
They cancel out, leaving choice probabilities unaffected and their coefficients are not identified.

Suppose: $U_{ij} = x_j'\beta + x_i'\delta + \varepsilon_{ij}$

Then: $$P_{ij} = \frac{e^{x_j'\beta + x_i'\delta}}{\sum_{k=1}^{J} e^{x_k'\beta + x_i'\delta}}$$

$$= \frac{e^{x_i'\delta} \times e^{x_j'\beta}}{e^{x_i'\delta} \times \left(\sum_{k=1}^{J} e^{x_k'\beta}\right)}$$ because $e^{a+b} = e^a \times e^b$

$$= 1 \times \frac{e^{x_j'\beta}}{\sum_{k=1}^{J} e^{x_k'\beta}}$$

No x_i's in here!

24

MNL INDIVIDUAL LIKELIHOOD

A clever way to write the individual likelihood function for consumer i is the product of the J probabilities, each raised to the power of an indicator variable (δ_{ij}) that indicates the item chosen

$$L_i(\beta) = P_i(1)^{\delta_{i1}} \times \cdots \times P_i(J)^{\delta_{iJ}} = \prod_{j=1}^{J} P_i(j)^{\delta_{ij}}$$

e.g., if consumer i selected item $j = 3$, then $\delta_{i3} = 1$ while $\delta_{i1} = \delta_{i2} = \delta_{i4} = 0$, and the likelihood is

$$L_i(\beta) = P_i(1)^0 \times P_i(2)^0 \times P_i(3)^1 \times P_i(4)^0 = P_i(3) = \frac{e^{\beta_3}}{\sum_{k=1}^{J} e^{\beta_k}}$$

MNL JOINT LIKELIHOOD

The joint likelihood (across all consumers) is the product of the n individual likelihoods

$$L_n(\beta) = \prod_{i=1}^{n} L_i(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} P_i(j)^{\delta_{ij}}$$

And the joint log-likelihood function is:

$$\ell_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J} \delta_{ij} \log\big(P_i(j)\big)$$

Coding this up yourself will be an exercise for HW3.

The “hard” part of the MNL is keeping track of the 3 dims (n consumers, J items, K X variables)

LOOKING FORWARD

So far, all of this statistical machinery will only get you as far as those simple summary statistics

However, because we have multiple choice tasks per respondent, we can estimate **respondent-specific** β vectors (i.e., β_i)

This “explodes” the number of parameters from K to $\mathrm{n}K$, so some form of “regularization” will be required.

This is a topic called hierarchical models, discussed next week, and estimated with Bayesian methods

ONE-SLIDE RECAP

Maximum Likelihood

- Enables you to estimate a rich set of statistical and regression models
- One way to do inference is via the bootstrap

MaxDiff is a method to gather respondent preferences toward a list of items

- You can analyze the data with the MNL model

COMING UP

Homework 2

- Due Wednesday, May 7

Next Time:

- Conjoint Analysis
- Extending the MNL to be Hierarchical; estimating with Bayes

TODO:

- HW2
- Read CASI chapter 3 and DCMS chapters 5 & 12
- Read Sawtooth blog posts on conjoint