How do we model mode choice?

Let's say we have a person, and **that person wants to go from Point A to Point B.**

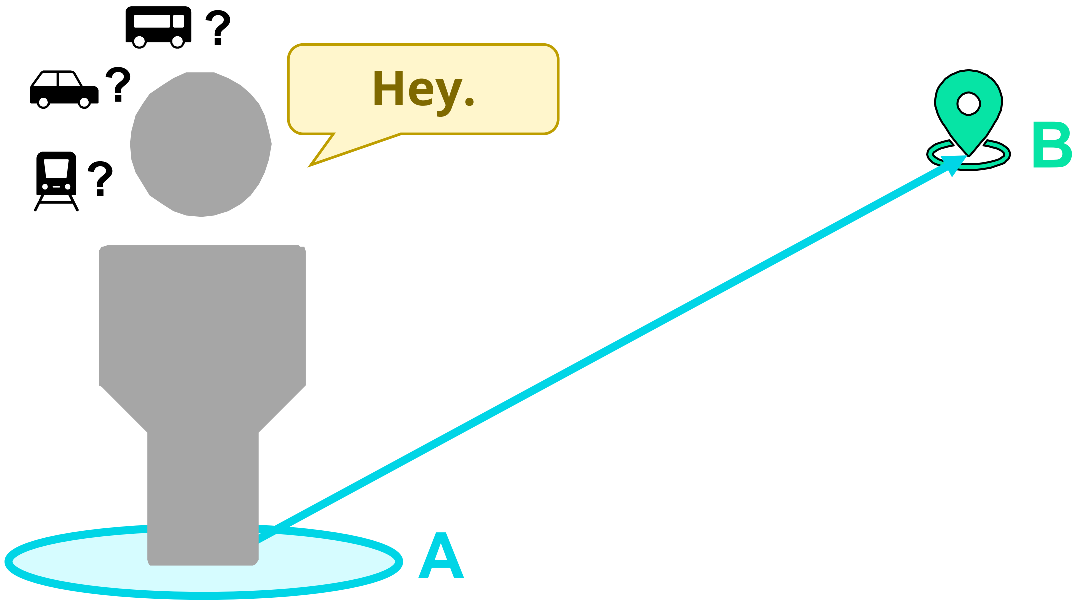
Our guy has a set of options for *how* he gets from A to B.

For any mode i we can express the utility of the mode for this trip as:

$$U_i = V_i + \varepsilon$$

Where V_i can be broken down into:

$$\begin{aligned}\text{Utility}_{\text{Transit}} = & \; a * \text{in-vehicle time} \\ & + b * \text{fare} \\ & + c * (\text{access time} + \text{egress time}) \\ & + d * \text{wait time} \\ & + \text{mode-specific constant}\end{aligned}$$

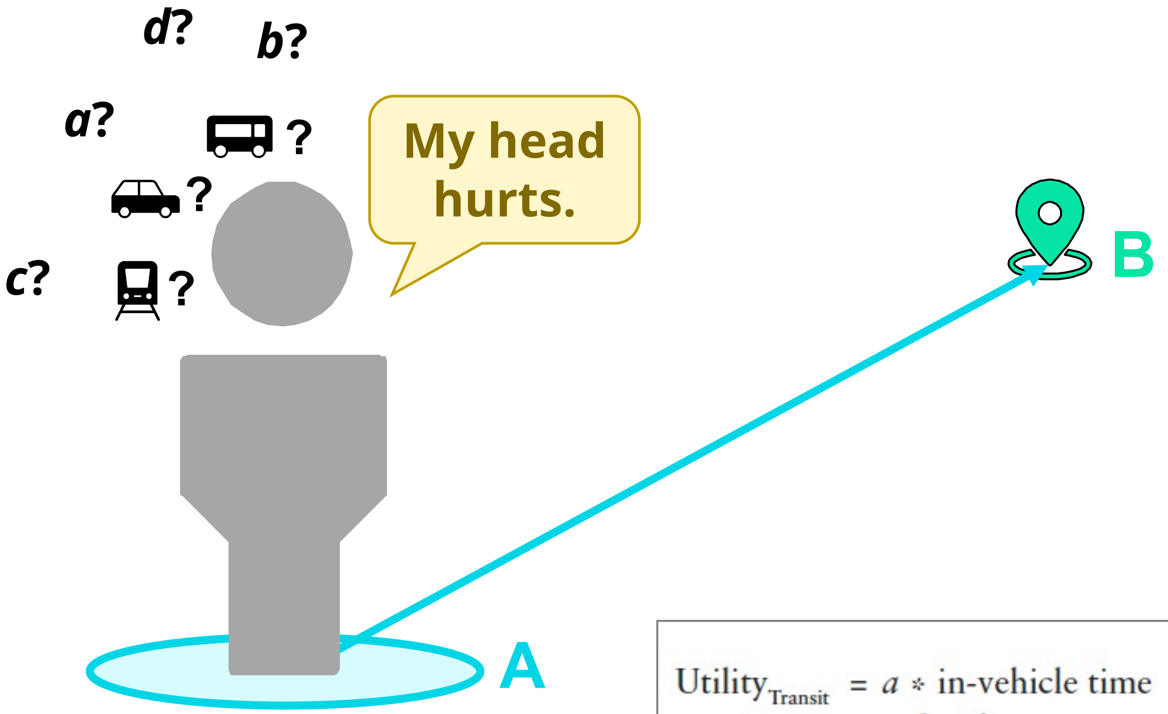


How do we model mode choice?

For each term, the second component is just a characteristic of the trip specific to the mode. **But the first component, the coefficients a through d, need to be estimated.**

Often these are estimated using survey data collected from people in the region willing to keep a travel diary or otherwise submit information on their travel choices.

The result:
By multiplying trip characteristics by these coefficients, we can convert the "value" or "utility" of the trip into utils and compare across diverse modes.



$$
\begin{aligned}
\text{Utility}_{\text{Transit}} = \; & a * \text{in-vehicle time} \\
& + b * \text{fare} \\
& + c * (\text{access time} + \text{egress time}) \\
& + d * \text{wait time} \\
& + \text{mode-specific constant}
\end{aligned}
$$

How do we model mode choice?

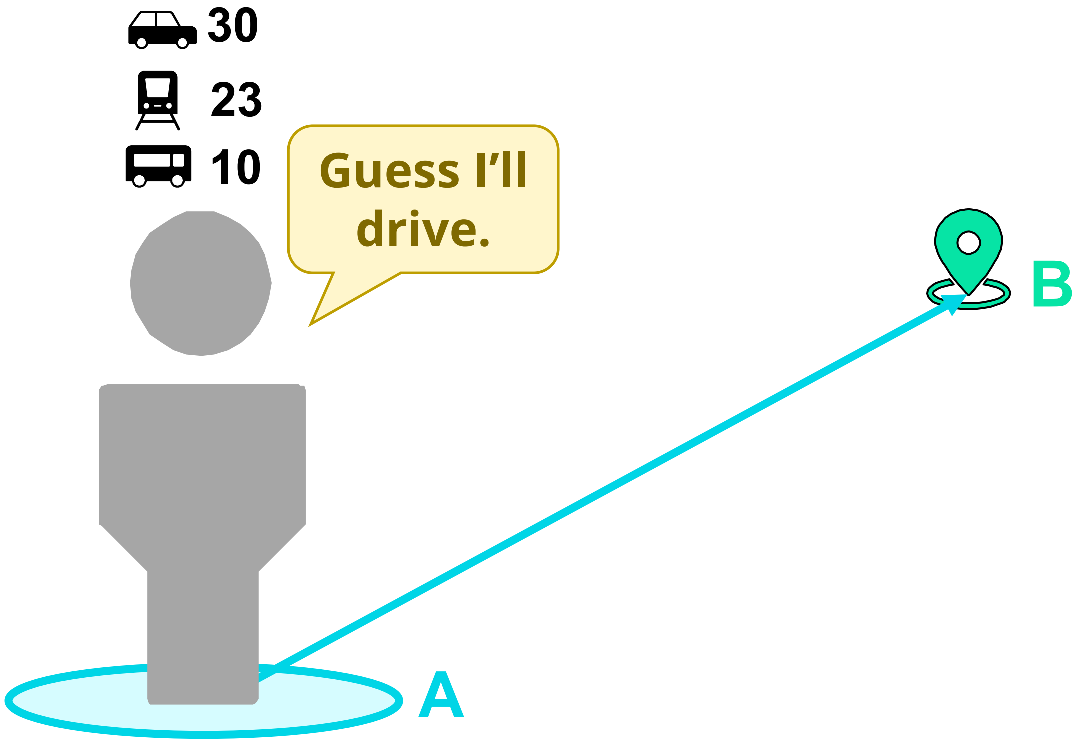
Now, the probability of our person choosing any mode can be expressed as:

$$P(i : C) = Prob\left(U_i \geq U_j, \forall_j \subset C\right)$$

Or, the probability of choosing a mode *i* from all possible modes in the set *C* is equal to the probability that the utility of *i* being greater than the utility of any other mode *j.*

Which is basically just saying “we’re going to pick the mode with the highest utility”.

We can use a **multinomial logit model** to determine mode choices. Unlike a standard logistic regression, which requires a binary response variable, an MLM can accommodate multiple categories.



What about those nested models?

Of course, like most things in life, it's not that simple. The decision to travel to a specific place, at a specific time, via a specific mode does not consist of three independent choices. They are interrelated so we've got to deal with some conditional probabilities.

The critical thing here is the understanding that hierarchical decision making is common and is addressed in ABMs through the calculation of composite utilities.

Also, you may have noticed that the structure of these processes look very tree-like.

FOR REFERENCE ONLY:

Probability of mode *i* is equal to the probability of mode *i* given some previous decision (nest) *n* times the probability of *n*

$$\underbrace{P(i) = P(i \mid n) * P(n)}$$

$$P(i) = \frac{\exp\left(V_{i|n}/\theta_n\right)}{\sum_{j\in n}\exp\left(V_{j|n}/\theta_n\right)} * \frac{\exp\left[V_n + \theta_n \ln\left(\sum_{j\in n}\exp\left(V_{j|n}/\theta_n\right)\right)\right]}{\sum_{\forall m}\left[\exp\left[V_m + \theta_m \ln\left(\sum_{j\in m}\exp\left(V_{j|m}/\theta_m\right)\right)\right]\right]}$$

What mode will I choose?

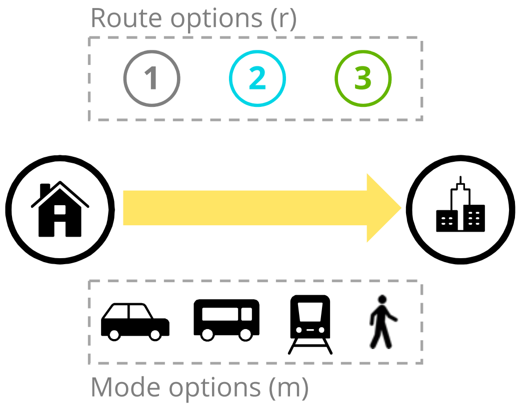
What upstream decision (e.g., destination) did I already make?

The complete process

Person X wants to go from Home to Work. What route and mode do they choose?

Male
50
$$

Value of time (VOT): **15**



ROUTE DETAILS

$$\big(a(time) + b(distance) + c(cost) + d(value\ of\ time)\big) * mode\ specific\ term$$

PERSONA-BASED COEFFICIENTS

$$\big(VOT_X(time_{r,m}) + VOT_X(distance_{r,m}) + PrSen_X(cost_{r,m}) + Pref_X(route_{r,m})\big) * mode\ term$$

Cost of time	Cost of distance (effort)	Cost of trip (parking, tolls, etc.)	Cost related to socioeconomic factors	Constant to represent intangibles

Monte Carlo to make final choice

$$\{ U_{1,car},\ U_{1,bus},\ U_{1,train} \ \dots\ U_{3,train},\ U_{3,walk} \}$$

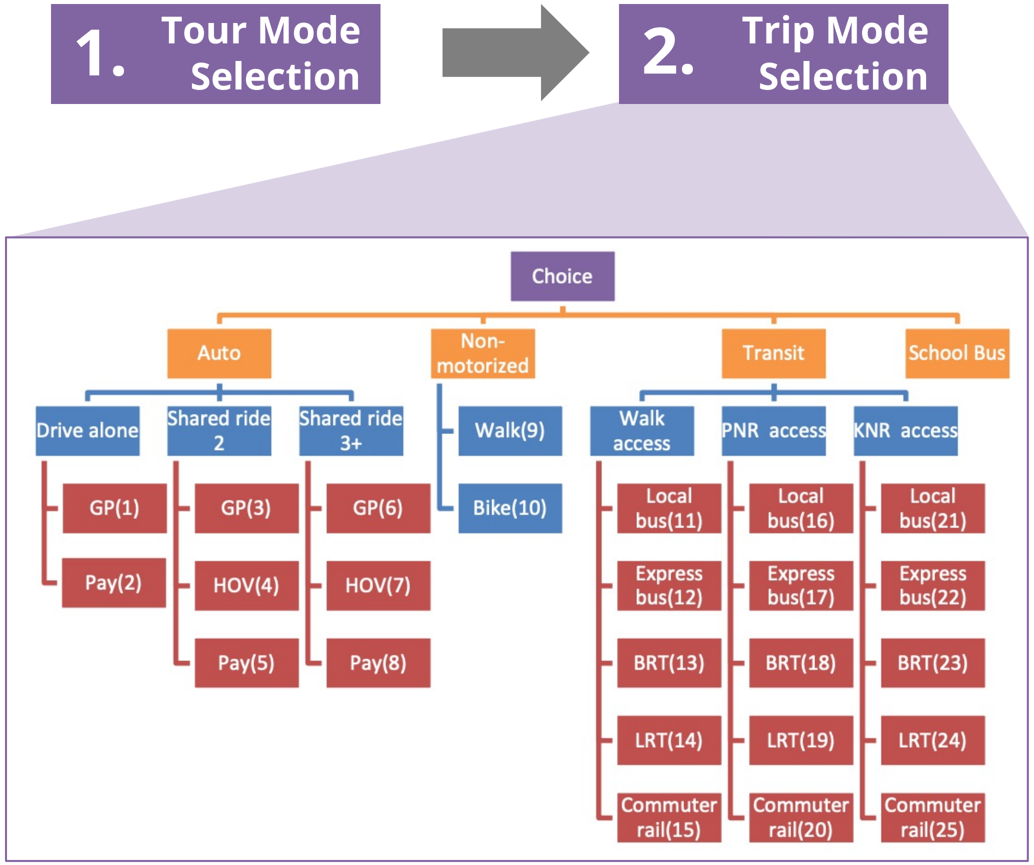
Final choice of mode for trip

A real-life example from SANDAG

Activity-Based Model (ABM)

The SANDAG ABM currently uses a two-step (“switching”) mode choice model: an initial run to determine tour mode and a subsequent run for trip modes which are now constrained by the tour mode choice. For each trip, there are up to 26 possible modes, dependent on the selected tour mode.

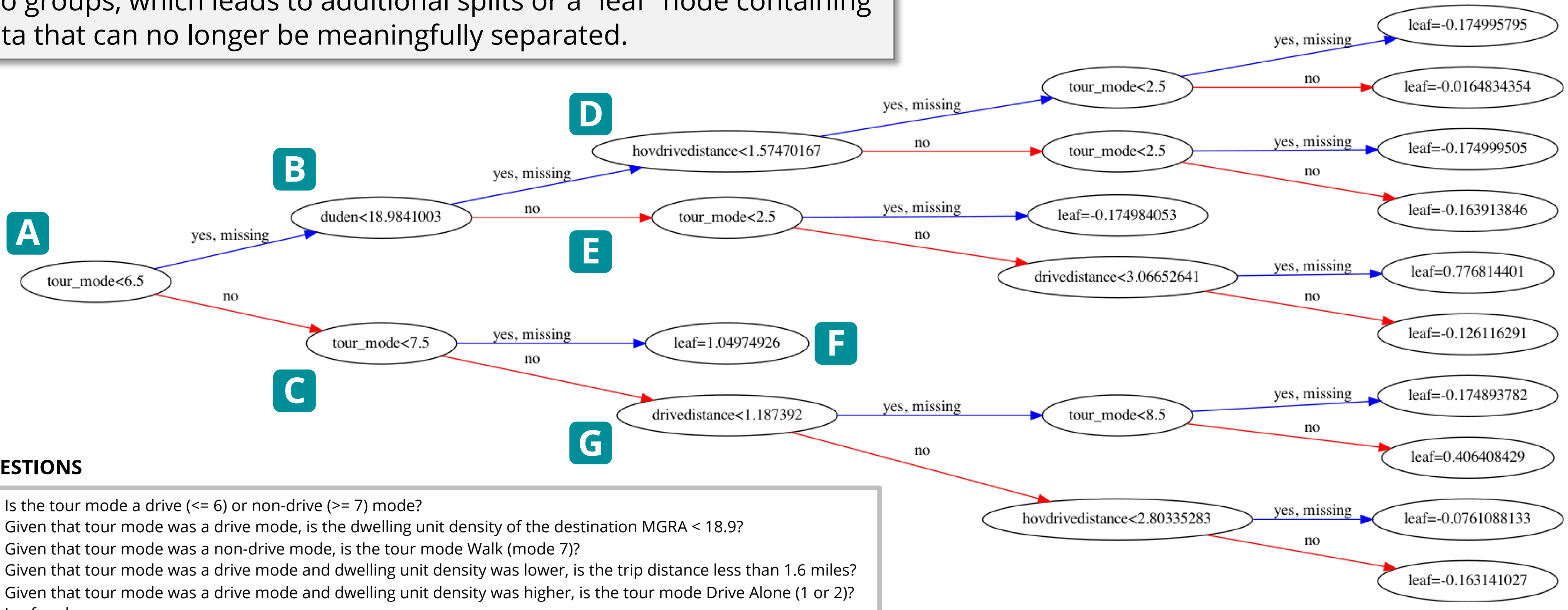
To make these choices, the utility model itself takes into account a large variable set, all of which have associated coefficients that are estimated for each tour purpose.



ACTIVITY-BASED TRAVEL MODEL CALIBRATION AND VALIDATION FOR BASE YEAR 2012, SANDAG, p.22

Trees: A Game of 20 Questions

The graphic below shows a portion of a trained decision tree model. At each node the model asks a question to split the incoming data into two groups, which leads to additional splits or a "leaf" node containing data that can no longer be meaningfully separated.



QUESTIONS

A. Is the tour mode a drive (<= 6) or non-drive (>= 7) mode?
B. Given that tour mode was a drive mode, is the dwelling unit density of the destination MGRA < 18.9?
C. Given that tour mode was a non-drive mode, is the tour mode Walk (mode 7)?
D. Given that tour mode was a drive mode and dwelling unit density was lower, is the trip distance less than 1.6 miles?
E. Given that tour mode was a drive mode and dwelling unit density was higher, is the tour mode Drive Alone (1 or 2)?
F. Leaf node
G. Given that tour mode was a non-drive mode and tour mode was not Walk, was the trip distance less than 1.2 miles?