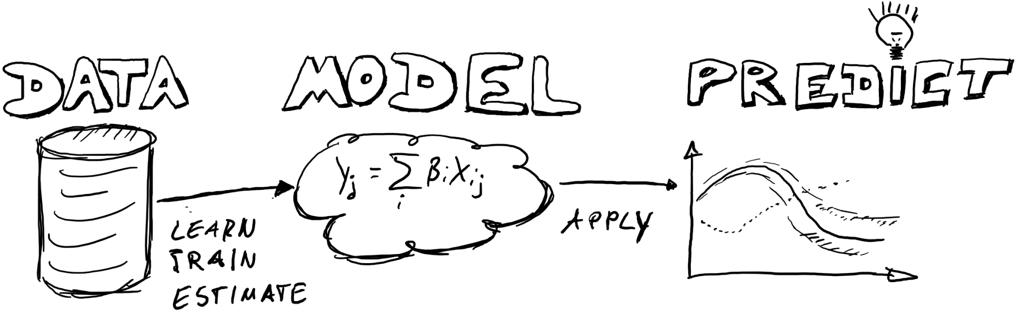
# Predictive Analytics

## Exercise 3: Binary Choice

**Univ.-Prof. Dr. Sven Müller**

Business Administration, esp. Operations Management
Otto-von-Guericke-University Magdeburg

# Our Overall Objective



# To achieve this objective, we ...

1. imply a (theory-driven) **model** structure: select and specify variables $X$, i.e., deterministic utility $V$. For instance
   - $V_{ni} = \beta_1 \mathsf{TT}_i$ or,
   - $V_{ni} = \beta_1 \mathsf{TT}_i + \beta_2 \mathsf{TC}_i$
2. **estimate** model coefficients $\hat{\beta}$:
   - maximum likelihood procedure: select values of $\hat{\beta}$ such that the assumed model structure fits (replicates) the choices $y_{ni}$ in the **data** as good as possible
   - roughly sayin': highest choice probability $P_{ni}$ for the actual chosen alternative $(y_{ni} = 1)$
3. compute deterministic utility values $\hat{\beta} X_{ni}$ and choice probabilities:

$$P_{ni} = \frac{e^{\hat{\beta} X_{ni}}}{\sum_j e^{\hat{\beta} X_{nj}}}$$

4. consider different scenarios (values) of $X$ and **predict** corresponding $P_{ni}$

# Mode Choice Example

- Estimate a series of models along the following slides
- Try to implement the different specifications

- Get data and R script from moodle

# Specifications

- Baseline specification

$$
\begin{aligned}
U_{n,\text{rail}} &= \beta_0 + \beta_1 \text{TC}_{n,\text{rail}} + \epsilon_{n,\text{rail}} \\
U_{n,\text{car}} &= \beta_1 \text{TC}_{n,\text{car}} + \epsilon_{n,\text{car}}
\end{aligned}
$$

- Specification 1

$$
\begin{aligned}
U_{n,\text{rail}} &= \beta_0 + \beta_1 \text{TC}_{n,\text{rail}} + \beta_2 \text{TT}_{n,\text{rail}} + \epsilon_{n,\text{rail}} \\
U_{n,\text{car}} &= \beta_1 \text{TC}_{n,\text{car}} + \beta_2 \text{TT}_{n,\text{car}} + \epsilon_{n,\text{car}}
\end{aligned}
$$

- Specification 2

$$
\begin{aligned}
U_{n,\text{rail}} &= + \beta_1 \text{TC}_{n,\text{rail}} + \beta_2 \text{TT}_{n,\text{rail}} + \epsilon_{n,\text{rail}} \\
U_{n,\text{car}} &= \beta_0 + \beta_1 \text{TC}_{n,\text{car}} + \beta_2 \text{TT}_{n,\text{car}} + \epsilon_{n,\text{car}}
\end{aligned}
$$

# Specifications

- Specification 3

$$
\begin{aligned}
U_{n,\text{rail}} &= \beta_0 + \beta_1 \mathsf{TC}_{n,\text{rail}} + \beta_2 \mathsf{TT}_{n,\text{rail}} + \epsilon_{n,\text{rail}} \\
U_{n,\text{car}} &= \beta_1 \mathsf{TC}_{n,\text{car}} + \beta_3 \mathsf{TT}_{n,\text{car}} + \epsilon_{n,\text{car}}
\end{aligned}
$$

- Specification 4

$$
\begin{aligned}
U_{n,\text{rail}} &= \beta_0 + \beta_4 \text{gender}_n + \beta_5 \text{business}_n + \beta_1 \mathsf{TC}_{n,\text{rail}} + \beta_2 \mathsf{TT}_{n,\text{rail}} + \epsilon_{n,\text{rail}} \\
U_{n,\text{car}} &= \beta_1 \mathsf{TC}_{n,\text{car}} + \beta_3 \mathsf{TT}_{n,\text{car}} + \epsilon_{n,\text{car}}
\end{aligned}
$$

# Challenge Questions

1. Try to implement the following specification 5

$$
\begin{aligned}
U_{n,\text{rail}} &= \beta_0 + \beta_4 \text{gender}_n + \beta_5 \text{business}_n + \beta_6 \text{income}_n + \beta_1 \text{TC}_{n,\text{rail}} + \beta_2 \text{TT}_{n,\text{rail}} + \epsilon_{n,\text{rail}} \\
U_{n,\text{car}} &= \beta_7 \text{income}_n + \beta_1 \text{TC}_{n,\text{car}} + \beta_3 \text{TT}_{n,\text{car}} + \epsilon_{n,\text{car}}
\end{aligned}
$$

2. Specify and estimate a model based on specification 5 that also considers whether wifi is available (see $\text{service}_{\text{rail}}$ in apollo manual)
3. How does your model improve? Interpret the meaning of the coefficient!

OM