交通工程学
Introduction to Traffic Engineering
第 9-10 节　交通中的选择行为[1]

葛乾

西南交通大学　交通工程系
西南交通大学　系统科学与系统工程研究所

[1]仅供教学使用，严禁传播

本节目录

选择问题

- 连续选择
  - 当油价上涨 10% 时，成都市民对汽油的需求会降低多少?
- 离散选择
  - 当油价上涨 10% 时，某一个市民上班选择**开车**、**公交**和**地铁**的概率会怎么变化?

离散选择问题

1. 作为一个工厂的供应商，哪些因素会影响其对配送港口的选择？
2. 作为一个普通的出行者，选择何种出行方式会受到什么因素影响？
3. 作为抖音/B 站/淘宝的推荐算法工程师，如何根据网站/App 的访客特征，为其推荐感兴趣的视频内容或者最可能购买的商品？

为什么不使用连续模型？

线性回归

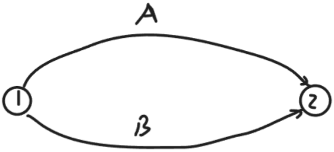
- 考虑一个标准形式 $y_i = \sum_k \beta_k x_{ik} + \epsilon_i$，其中 i 指代某一个体，k 表示某一属性，β_k 表示属性对应的参数，ϵ_i 表示未知变量的分布
- 我们从一组观察数据中，寻找最优的线性曲线，以拟合属性和响应值之间的关系

线性回归的缺陷

- 对于二元选择（仅有两个选择支）的问题，线性回归可以使用。然而，选择概率 y_i 通常取值并不在 $(0,1)$ 这个范围内；
- OLS 用于选择时，不符合经济学对选择行为的假设；
- 无法用于多元选择问题；

本节目录

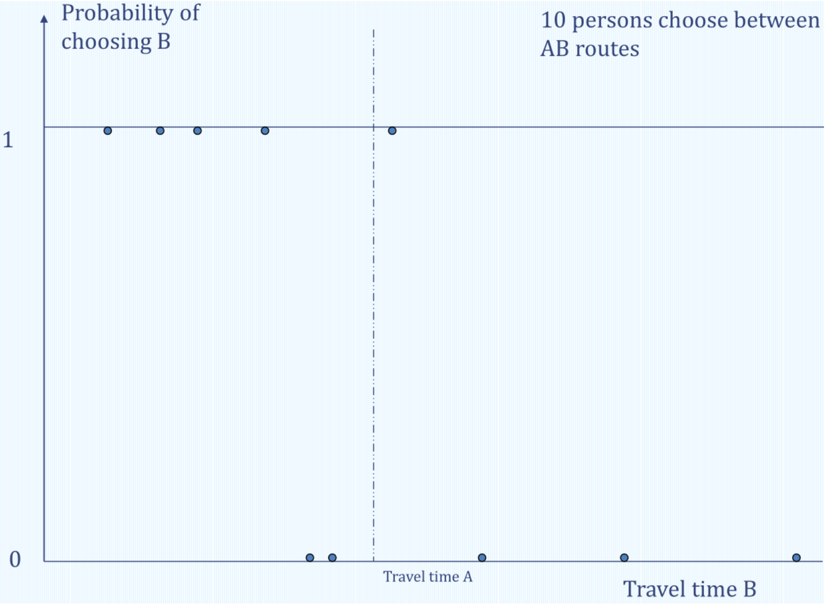
一个例子

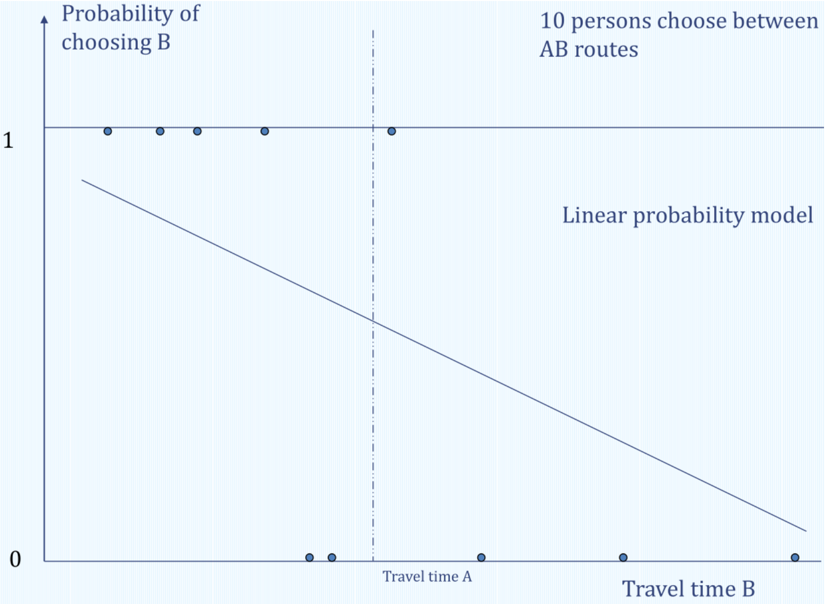


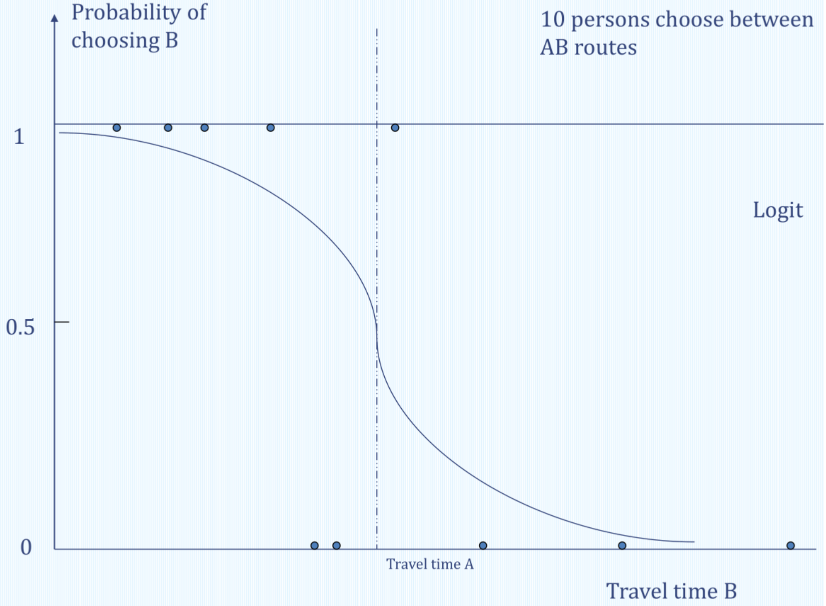
- Consider we have 10 individuals that choose between two routes A and B
- Travel time of A is 9 and of B is 10 minutes
  - some people take route B, because they are familiar with it or they misjudge the travel time
- Let's do a regression of whether or not you have chosen B on the difference between the travel time (TT) of A and B

$$y_B = f(\alpha + \beta(TT_B - TT_A)) + \epsilon$$

where $y_B = 1$ if a person chooses route B and 0 otherwise


Probability of choosing B
10 persons choose between AB routes
1
0
Travel time A
Travel time B


Probability of choosing B
10 persons choose between AB routes
1
Linear probability model
0
Travel time A
Travel time B


Probability of choosing B
10 persons choose between AB routes
1
0.5
0
Logit
Travel time A
Travel time B

随机效用

- As the analyst we realise that we cannot define the utility function precisely-but we can disclose (some) factors which influence individuals' decisions.
- $y = f(x, \epsilon)$. If we know all explanatory components in a utility function it is in essence deterministic
- But ϵ is in random utility theory defined as the unobserved component; in realization that all behavior cannot be explained and/or we cannot identify all explanatory variables.

随机效用与选择概率

- Indirect utility may be given by

$$U_{iA} = V_A(TT_A) + \epsilon_{iA}$$
$$U_{iB} = V_B(TT_B) + \epsilon_{iB}$$

  where V_A, V_B are deterministic utility
- Note that the levels of U_{iA} and U_{iB} are not directly observed due to the random terms!
- If the respondent prefers alternative A to alternative B, we assume: $U_{iA} > U_{iB}$

$$\begin{aligned} P(Y = A) &= P(U_{iA} > U_{iB}) \\ &= P(V_A + \epsilon_{iA} > V_B + \epsilon_{iB}) \\ &= P(V_A - V_B > \epsilon_{iB} - \epsilon_{iA}) \end{aligned}$$

- We observe-the probability of choice, and on the basis of this observation we estimate the parameters that specify V

两个问题

- Which distribution for ϵ's?
  - ϵ 无法观察，因此需要从某个分布中得出
  - Gumbel distribution could generates simple closed-form solutions!→ Logit model
  - Gaussian distribution → Probit model

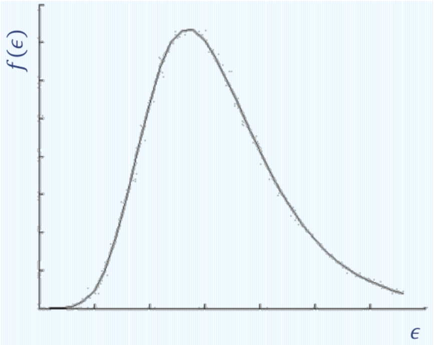


Figure: PDF of Gumbel distribution (used by Daniel McFadden (1964))

选择概率的计算

- Recall $P(Y=A) = P(\epsilon_{iB} - \epsilon_{iA} < V_A - V_B)$
- The next problem is to specify the PDF of $\epsilon_{iB} - \epsilon_{iA}$
- Luckily, if both terms follow Gumbel distribution, their difference ϵ^*_{iAB} follows the logistic distribution

$$F(\epsilon^*_{iAB}) = \frac{e^{\epsilon^*_{iAB}}}{1 + e^{\epsilon^*_{iAB}}}$$

- Easily

$$P(Y=A) = \frac{e^{V_A}}{e^{V_A} + e^{V_B}}$$

直接效用的计算

- Can be any function
- Linear function is often assumed
- Can be extended with multiple variables

$$U_{jA} = \beta p_{jA} + \kappa t_{jA} + \epsilon_{jA}$$
$$U_{jB} = \beta p_{jB} + \kappa t_{jB} + \epsilon_{jB}$$

where p_{jA} is the price of a trip and t_{jA} is travel time of alternative j

- $\beta < 0$, $\kappa < 0$
- Now we have

$$P(Y = A) = \frac{1}{1 + e^{\beta(p_{jB} - p_{jA}) + \kappa(t_{jB} - t_{jA})}}$$

Value of Time

- Value of Time (VOT): How much are you willing to pay to reduce your travel time with one hour, holding utility constant?
- Let's take the deterministic utility function $U_{jA} = \beta p_{jA} + \kappa t_{jA} + \epsilon_{jA}$
- When t_{jA} is measured in hours, the VOT can be written as κ/β
- VOT depends on trip purpose: **Business** €26.25/h, **Commuting** €9.25/h, **Social purpose** €7.50/h
- VOT depends on income: About 50% of net income

多个选项条件下选择概率计算

- Recall the choice probability for two alternatives:

$$P(Y = A) = \frac{e^{x_A}}{e^{x_A} + e^{x_B}}$$

- Usually there are more alternatives in the choice set
  - 地铁，公交，打车
  - 腾讯，阿里，谷歌
  - 从教室到宿舍的路径
- Simply extend the logit formula:

$$P(Y = A) = \frac{e^{x_A}}{e^{x_A} + e^{x_B} + e^{x_c}}$$

Consumer surplus（消费者剩余）

- For policy analysis, the researcher is often interested in measuring the change in consumer surplus that is associated with a particular policy.
  - For example, if a new alternative is being considered, such as building a light rail system in a city, then it is important to measure the benefits of the project to see if they warrant the costs.
- A person's consumer surplus is the utility, in dollar terms, that the person receives in the choice situation.
- **The decision maker chooses the alternative that provides the greatest utility**. $CS = \frac{1}{v}\max_S(U_S)$, where v is the marginal utility of income: $\frac{dU_j}{dY_j}$, with Y_j the income of person j. U_j is not observable, we have V_{jS} and the distribution of unobserved utility instead. Thus

$$\mathrm{E(CS)} = \frac{1}{v}E[\max_S(V_{jS} + \epsilon_{jS})]$$

Logsum

- The aggregate utility derived from the choice set is summarised by the **logsum**

$$E(CS) = \frac{1}{v} \ln(e^{\beta x_A} + e^{\beta x_B} + e^{\beta x_C})$$

  - v is the marginal utility of income
  - Can be used in welfare estimates

- 尽管 logsum 形式与选择概率的分母项相同，然而二者之间仅是数学上的巧合，并不代表选择概率的分母项就是消费者剩余

An example

Assume $\beta x_A = \beta x_B = 10$. Now alternative C is added and $\beta x_C = 1$. The average utility per alternative decreases from 10 to 7 but E[CS] increases → 'Love of variety' effect

本节目录

公式的性质

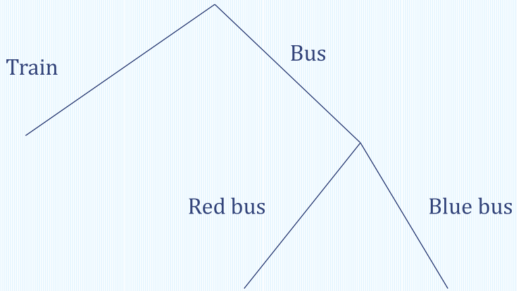
- The ratios of choice probabilities for A and B do not depend on whether or not C is in the choice set
- Independence of irrelevant alternatives (IIA)

$$\frac{P(Y=A)}{P(Y=B)} = \frac{\frac{e^{x_A}}{e^{x_A}+e^{x_B}+e^{x_c}}}{\frac{e^{x_B}}{e^{x_A}+e^{x_B}+e^{x_c}}} = \frac{e^{x_A}}{e^{x_B}}$$

The 'Red Bus-Blue Bus' problem

- Choice set 1: Train, red bus. Assume *V*s are 2.54, 1; then the market shares are 0.823, 0.177
- Choice set 2: Train, red bus, **blue bus**. Assume *V*s are 2.54, 1, 1; then the market shares are 0.70, 0.15 and 0.15
- The problem: Higher probability to take the bus in choice set 2 → not very realistic as red buses and blue buses are identical

- When some alternatives are more similar than other alternatives, the use of multinomial choice model may be misleading
- Use nested logit instead.



Figure

- Nested logit takes into account correlation between alternatives → But the structure of each nest is defined by the analyst!

NL 的效用函数

- Let us define utility as follows:

$$U_{jg} = V_j + W_g + \epsilon_{jg}$$

  V_j only differs within nests between alternatives j; W_g only differs between nests g

- We may write the probability to choose an alternative:

$$P(d_j = 1) = P(g) \cdot P(j|g)$$

  where $P(j|g) = \frac{e^{V_j/\lambda_g}}{\sum_{k\in g} e^{V_k/\lambda_g}}$, $P(g) = \frac{W_g+\lambda_g I_g}{\sum_h W_h+\lambda_h I_h}$ and I_g is the logsum for nest g, $I_g = \ln(\sum_{j\in g} e^{V_j/\lambda_g})$

- λ_g measures the correlation of alternatives across each nest.
  - $\lambda_g = 1$ no correlation (multinomial logit)
  - $\lambda_g \to 0$ perfect correlation (red bus/blue bus)
- If choices j and k belong to the same nest, we have

$$\frac{P(d_j = 1)}{P(d_k = 1)} = \frac{e^{W_g+V_j}/\lambda_g}{e^{W_g+V_k}/\lambda_g} = \frac{e^{W_g+V_j}}{e^{W_g+V_k}} = \frac{e^{V_j}}{e^{V_k}}$$

NL in summary

- Nested Logit probability depends on
  - Probability to choose a nest
  - Probability to choose an alternative within the nest
- But Nested Logit does not imply a sequential choice

How to deal with large choice sets

- Number of observations in your regressions is $\#of\, alternatives \times respondents$

Methods

- Model aggregate choices
- Random selection of alternatives
- Estimate count data models (Poisson)

Model aggregate choices

- Modelling location choice
  - Focus on aggregate areas (zones)
- Choice of cars
  - Only distinguish between brands
- However, lack of detail makes results less credible

Random selection of alternatives

- McFadden (1978)
  - Choose a random subset of J alternatives for each choice set, including the chosen option
  - This should not affect the consistency of the estimated parameters
  - Small-sample properties are yet unclear
- How large should J be?
- Applied in many good papers, e.g. Bayer et al. (2007, JPE)

Estimate count data models

- Estimate Conditional Logit by means of a Poisson model
- A Poisson regression is a count data model
  - Dependent variable is integer
  - $\cdots$and should be Poisson distributed
  - $C_j = e^{\beta' x_j} + \epsilon$ where where C_j is the # of decision makers that have chosen a certain alternative
- Convenient interpretation of β: When x_j increases with one, C_j increases with $\beta \times 100$ percent
- A Poisson model should give identical parameters to the Conditional Logit
  - Maximum likelihood functions are identical up to a constant(Guimarães et al. 2003)
- Hence, group observations based on their chosen alternatives
  - $\cdots$the number of firms choosing a certain location
  - $\cdots$the number of people buying a certain car

Estimate count data models (cont.)

- Implications
  - You cannot include characteristics of the decision maker (because you sum up all choices)!
  - Homogeneous parameters across the population
- Extensions
-
  - Include fixed effects
  - Negative binomial regression
  - Zero-inflated models
  - See Guimarães et al. (2004) for details

本节目录

Likelihood function

- A sample of N decision makers is obtained for the purpose of estimation. Since the logit probabilities take a closed form, the traditional maximum-likelihood procedures can be applied.
- The probability of person j choosing the alternative that he was actually observed to choose can be expressed as

$$\prod_S (P_{jS})^{y_{jS}}$$

where $y_{jS} = 1$ if person j chose S and zero otherwise. Since $y_{jS} = 0$ for all nonchosen alternatives and Pni raised to the power of zero is 1, this term is simply the probability of the chosen alternative.

- Assuming that each decision maker's choice is independent of that of other decision makers, the probability of each person in the sample choosing the alternative that he was observed actually to choose is

$$L(\beta) = \prod_{j=1}^{N} \prod_S [P_{jS}(\beta)]^{y_{jS}}$$

Estimation

The log-likelihood function can be re-expressed as

$$
\begin{aligned}
LL(\beta) &= \sum_{j=1}^{N} \sum_{S} y_{jS} \ln P_{jS}(\beta) \\
&= \sum_{j=1}^{N} \sum_{S} y_{jS} \ln \frac{e^{\beta' x_{jS}}}{\sum_{S'} e^{\beta' x_{jS}}} \\
&= \sum_{j=1}^{N} \sum_{S} y_{jS} (\beta' x_{jS}) - \sum_{j=1}^{N} \sum_{S} y_{jS} \ln (\sum_{S'} e^{\beta' x_{jS}})
\end{aligned}
$$

Estimation of Logit model

The derivative of the log-likelihood function then becomes

$$\frac{\mathrm{d}LL(\beta)}{\mathrm{d}\beta} = \frac{\sum_{j=1}^{N}\sum_{S} y_{jS}(\beta' x_{jS})}{\mathrm{d}\beta} - \frac{\sum_{j=1}^{N}\sum_{S} y_{jS}\ln(\sum_{S'} e^{\beta' x_{jS}})}{\mathrm{d}\beta}$$

$$= \sum_{j=1}^{N}\sum_{S} y_{jS}x_{jS} - \sum_{j=1}^{N}\sum_{S} y_{jS}\sum_{S'} P_{jS}x_{jS}$$

$$= \sum_{j=1}^{N}\sum_{S}(y_{jS} - P_{jS})x_{jS}$$

Setting this derivative to zero gives the first-order condition, which is a system of nonlinear equations.

Goodness of fit(拟合优度)

- A statistic called the likelihood ratio index is often used with discrete choice models to measure how well the models fit the data.
- It measures how well the model, with its estimated parameters, performs compared with a model in which all the parameters are zero (which is usually equivalent to having no model at all).
- This comparison is made on the basis of the log-likelihood function, evaluated at both the estimated parameters and at zero for all parameters. The likelihood ratio index is defined as

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

  where $LL(\hat{\beta})$ is the value of the log-likelihood function at the estimated parameters and $LL(0)$ is its value when all the parameters are set equal to zero.
- If the estimated parameters do no better, in terms of the likelihood function, than zero parameters (that is, if the estimated model is no better than no model), then $LL(\hat{\beta}) = LL(0)$ and so $\rho = 0$. This is the lowest value that ρ can take (since if $LL(\hat{\beta})$ were less than $LL(0)$, then $\hat{\beta}$ would not be the maximum likelihood estimate).

假设检验

- Standard t-statistics are used to test hypotheses about individual parameters in discrete choice models, such as whether the parameter is zero
- Consider a null hypothesis H that can be expressed as constraints on the values of the parameters. Two of the most common hypotheses are (1) several pa- rameters are zero, and (2) two or more parameters are equal.
- The constrained maximum likelihood estimate of the parameters (labeled $\hat{\beta}^H$) is that value of β that gives the highest value of LL without violating the constraints of the null hypothesis H.
- Define the ratio of likelihoods, $R = \frac{L(\hat{\beta}^H)}{L(\hat{\beta})}$,where $\hat{\beta}^H$ is the (constrained) maximum value of the likelihood function (not logged) under the null hypothesis H, and $\hat{\beta}$ is the unconstrained maximum of the likelihood function
- As in likelihood ratio tests for models other than those of discrete choice, the test statistic defined as $2\ln R$ is distributed chi-squared with degrees of freedom equal to the number of restrictions implied by the null hypothesis. Therefore, the test statistic is $2(LL(\hat{\beta}^H) - LL(\hat{\beta}))$.

两种检验

- Null Hypothesis I: **The Coefficients of Several Explanatory Variables Are Zero**
  - To test this hypothesis, estimate the model twice: once with these explanatory variables included, and a second time without them (since excluding the variables forces their coefficients to be zero). Observe the maximum value of the log-likelihood function for each estimation; two times the difference in these maximum values is the value of the test statistic. Compare the test statistic with the critical value of chi-squared with degrees of freedom equal to the number of explanatory variables excluded from the second estimation.
- Null Hypothesis II: **The Coefficients of the First Two Variables Are the Same**
  - To test this hypothesis, estimate the model twice: once with each then with the first two variables replaced by one variable that is the sum of the two variables (since adding the variables forces their coefficients to be equal). Observe the maximum value of the log-likelihood function for each of the estimations. Multiply the difference in these maximum values by two, and compare this figure with the critical value of chi-squared with one degree of freedom.

本节目录

选项应该满足以下条件

- Mutually exclusive
  - If confusion arises, choice categories - A and B only
- Choice set must be exhaustive
  - If presenting choice scenario - heating in house (gas, electricity, wood) should we include no heating? Or
  - Exclude respondents who do not have heating
- # of alternative must be finite
  - # of levels on attributes must be finite
  - Ex: 0, 1, or 2 more cars in he household

调查的设计

- Which type of alternatives to include?
  - Status Quo
  - Opt-out (alternatives other than present) – should include when relevant
- # of alternatives presented in choice scenario: 2, 3 or 4? (Health economics-2, environment 3)

Number of alternatives

- If more than 2 alternatives: multinomial logit
- If 2 alternatives: binary logit
- Two models have underlying characteristics

两种数据

- Revealed preference (RP) data
  - Observed or reported actual behavior
- Stated preference (SP) data
  - Respondents are confronted with hypothetical choice sets
- Combinations of RP and SP

RP 数据的优点

- Based on actual behavior!!
- Use existing (large) data sources
  - Cheaper
  - No expensive experiments
- Panels of the same individuals over a long time

RP 数据的缺点

- Lack of variability
- Collinearity (e.g. price and travel times)
- Lack of knowledge on the choice set
- Not possible with new choice alternatives (opt-out)
- Actual behavior may not be first choice
  - 大学招生计划
- Perception errors and imperfect information
  - Airline tickets

Example of stated preference question

What alternative will you choose?

Option 1	Option 2
Price: €1,000 Handling time: 3 days % doesn't arrive: 1.0%	Price: €750 Handling time: 1 week % doesn't arrive: 1.3%

SP 数据的优点

- New alternatives
- New attributes
- Large variability is possible
- Problems of collinearity can be solved
  - 'Orthogonal design'
- Choice set is clearly defined

SP 数据的缺点

- Information bias
  - The respondent has incorrect information on the context
  - Make your experiment as realistic as possible
- Starting point bias
  - Respondents are influenced by the set of available responses to the experiment
  - Test your design and choose realistic attribute values
- Hypothetical bias
  - Individuals tend to respond differently to hypothetical scenarios than they do to the same scenarios in the real world.
  - Cognitive incongruity with actual behavior
  - Again: make your experiment as realistic as possible
  - But otherwise hard to mitigate⋯
- Strategic bias
  - Respondent wants a specific outcome
  - (S)he fills in answers that are in line with desired outcomes
- Errors
  - Boredom
  - Respondents do not carefully read instructions
  - Respondents do not understand the questions

本节目录

开源软件

- biogeme
- R+mlogit
- python+choicemodels

课下训练

- Kenneth Train. Chap 1-4

谢谢!