

Výzkum diskrétních voleb: od experimentu k analýze výsledků

PSY028_E – Statistická analýza dat v psychologii



Michal Šimeček

Centrum dopravního výzkumu, v.v.i.

Líšeňská 33a

Brno

michal.simecek@cdv.cz

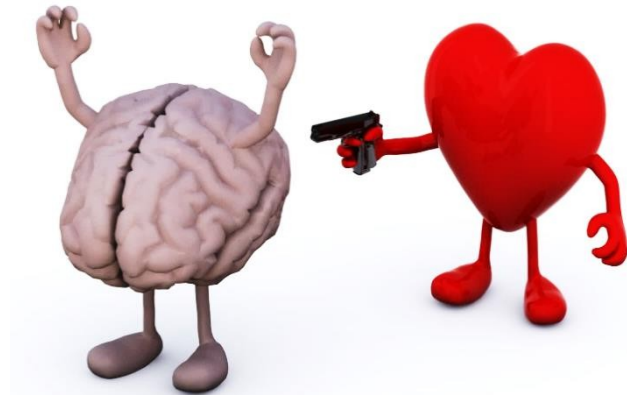
1. Východiska výzkumu diskrétních voleb

- Teorie rozhodování
- Model rozhodování
- Multinomiální logistický regresní model – MNL
- Zdrojová data
- Využití výzkumu rozhodování
- *Společný experiment*

Teorie rozhodování: ekonomická teorie

Vychází z představ neoklasické ekonomie

- Jedinci nebo organizace se rozhodují tak, aby maximalizovali svůj užitek
- Rozhodování je racionální
- K rozhodnutí mají lidé všechny potřebné informace
- Proto lze rozhodovacímu procesu porozumět a předpovídat výsledky



Jak si VY myslíte, že se lidé rozhodují?

Teorie rozhodování: Užitek

Bezrozměrné číslo, které vyjadřuje, jaký užitek by měl člověk z toho, kdyby se rozhodl pro příslušnou alternativu.

Člověk pravděpodobně volí alternativu s nejvyšším užitekem = maximalizace užitku.

John von Neumann a Oskar Morgenstern (1947): teorie maximalizace užitku v teorii her

očekávaný užitek = pravděpodobnost výsledku X užitek výsledku

Je užitek pro všechny stejný?

Vnímají všichni užitek stejně?

Jak měřit užitek?

Teorie rozhodování: měření užitku

Zeptat se na užitek?

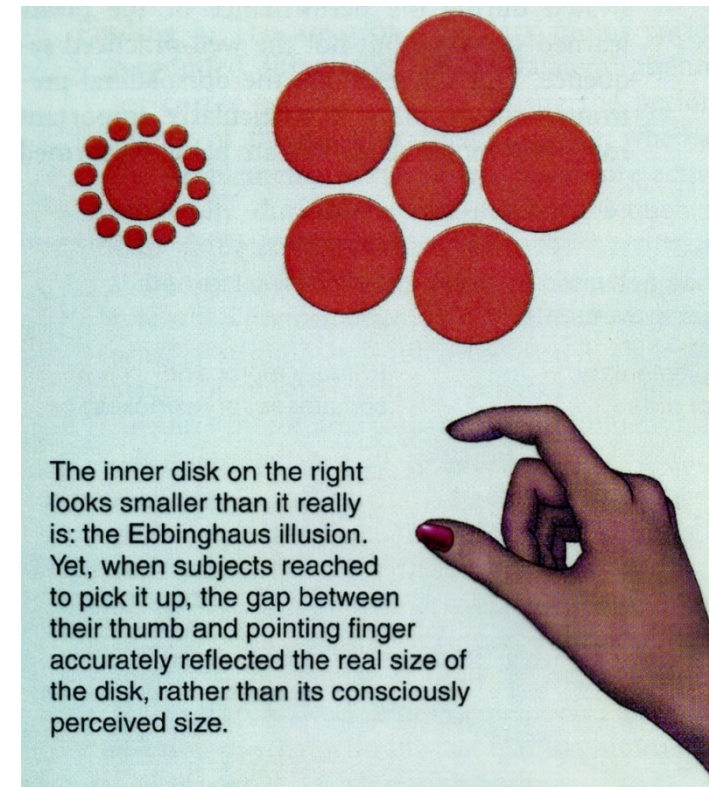
Nechat hodnotit užitek pro různé možnosti?

Nechat řadit možnosti podle užitku?

Pozorovat!

Pozorování rozhodování

Aktér se rozhoduje mezi alternativami. Pro každou alternativu existuje pro aktéra určitý užitek. Z výsledku rozhodnutí a znalosti vlastností alternativ lze odhadovat užitek alternativ, ale hlavně jak se na volbě podílí jednotlivé vlastnosti alternativ.



Model rozhodování: dohoda o tom, co je to rozhodování pro analýzu rozhodování

Rozhodování je proces, v jehož závěru je vybrána (realizována)







- jedna z
- konečného počtu relevantních alternativ,
- vzájemně oddělených (exkluzivních),
- které jsou popsány pozorovatelnými vlastnostmi

Model rozhodování: analýza rozhodování

Analýza rozhodování zkoumá, jak pozorovatelné vlastnosti (atributy) alternativ a vlastnosti aktéra (například jeho sociodemografie) souvisí s jeho volbou.

ATRIBUTY ALTERNATIVY SCÉNÁŘ

Jaký z nabízených způsobů dopravy byste využil/a k cestě z Prahy do Brna nebo z Brna do Prahy?

 Vysokorychlostní železnice	 Obyčejná železnice	 Osobní automobil	 Autobus
Cestovní čas 1h 30min Náklady na cestu 400 Kč	Cestovní čas 2h 28min Náklady na cestu 263 Kč	Cestovní čas 2h 35min Náklady na cestu 326 Kč	Cestovní čas 2h 50min Náklady na cestu 165 Kč

ALTERNATIVA

Model rozhodování: užitková funkce

Užitek z alternativy i pro jedince n je vyjádřen takto:

$$U_{in} = V_{in} + \varepsilon_{in}$$

Diagrammatic description: The equation $U_{in} = V_{in} + \varepsilon_{in}$ is shown with three red arrows pointing to its components. The arrow from 'užitek' points to U_{in} . The arrow from 'deterministická část' points to V_{in} . The arrow from 'náhodná komponenta (chyba)' points to ε_{in} .

Zdroje chyby ε

- Chyba měření atributů (to, co rozhoduje nemusí být přesně to, co umíme měřit)
- Chyba měření rozhodnutí (např. zapomínání)
- Rozdílné marginální užítky (váhy) atributů u různých aktérů
- Nepozorovaný užitek (neznámé atributy)

U chyb se předpokládá normální, všude stejné a nezávislé rozdělení (*independent and identically distributed – IID*).

Model rozhodování: deterministická část užitkové funkce

Deterministická část užítku popisuje vliv nezávislých proměnných na celkový užitek.
Deterministická část užítku z alternativy i pro jedince n může vypadat takto:

$$V_{in} = \alpha_i * 1$$

Alternativně specifická konstanta
Lidé z nějakého důvodu obecně preferují tuto alternativu.

$$+ \beta * x_i$$

Koeficient stejný pro všechny alternativy, proměnná asociovaná s alternativou. Například cena.

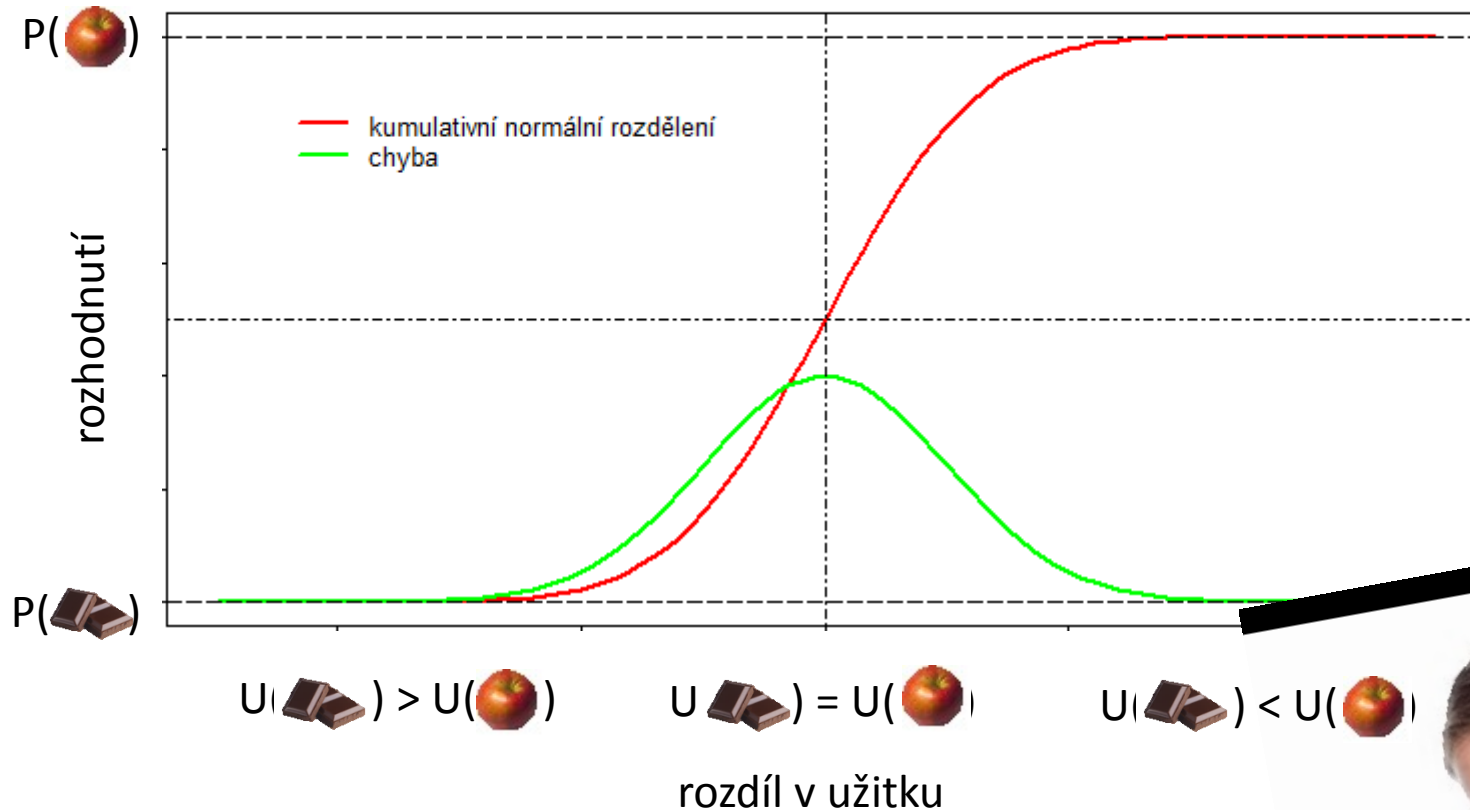
$$+ \gamma_i * y_i$$

Koeficient i proměnná jsou asociované s alternativou.
Například cestovní čas se v různých alternativách liší a zároveň je jinak vnímaný v autě a jinak ve vlaku.

$$+ \delta_i * d_n$$

Koeficient asociovaný s alternativou, proměnná s individuem. Například preference způsobu dopravy pro lidi různého věku nebo sociální skupiny se mohou lišit.

Model rozhodování: užitek \leftrightarrow rozhodnutí



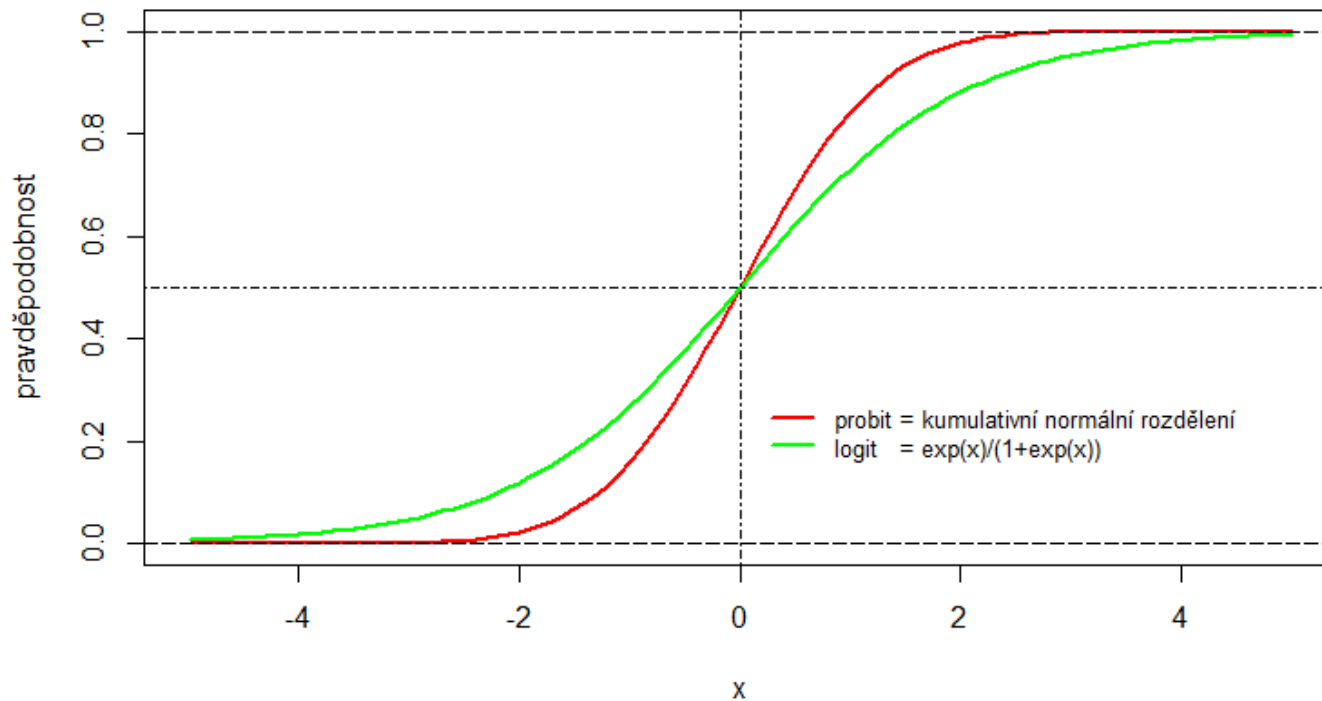
Mezi pravděpodobnostmi určitého rozhodnutí a užitek z tohoto rozhodnutí se používá převodní funkce. Nabízí se funkce kumulativního normálního rozdělení (probitová funkce).



Model Rozhodování: probitová vs. logitová funkce





Probitová funkce je teoreticky správná, ale:

- Na stranách má malé klesání, což může komplikovat odhad
- Je výpočetně složitá



Multinomiální logistický regresní model - MNL

Jaký z nabízených způsobů dopravy byste využil/a k cestě z Prahy do Brna nebo z Brna do Prahy?

 Vysokorychlostní železnice Cestovní čas 1h 30min Náklady na cestu 400 Kč	 Obyčejná železnice Cestovní čas 2h 28min Náklady na cestu 263 Kč	 Osobní automobil Cestovní čas 2h 35min Náklady na cestu 326 Kč	 Autobus Cestovní čas 2h 50min Náklady na cestu 165 Kč
---	---	--	--

$$U_{VR} = ASC_{VR} + \beta_{cena} * cena_{VR} + \beta_{čas} * čas_{VR}$$

$$U_{\check{z}} = ASC_{\check{z}} + \beta_{cena} * cena_{\check{z}} + \beta_{čas} * čas_{\check{z}}$$

$$U_{IAD} = ASC_{IAD} + \beta_{cena} * cena_{IAD} + \beta_{čas} * čas_{IAD} \quad ASC_{IAD} = 0$$

$$U_{Bus} = ASC_{Bus} + \beta_{cena} * cena_{Bus} + \beta_{čas} * čas_{Bus}$$

Obecně. Pravděpodobnost výběru alternativy \underline{m} z možných alternativ \underline{M} se rovná:





$$P_m = \frac{e^{U_m}}{\sum_{i=1}^M e^{U_i}}$$

Zdrojová data

Experiment (stated preference – SP):

- Kontrolujeme scénář, alternativy i jejich vlastnosti.
- Respondent si (většinou) vybírá jen jako
- Je možné zařadit i hypotetické (v realitě neexistující) alternativy.

Jaký z nabízených způsobů dopravy byste využil/a k cestě z Prahy do Brna nebo z Brna do Prahy?

 Vysokorychlostní železnice Cestovní čas 1h 30min Náklady na cestu 400 Kč	 Obyčejná železnice Cestovní čas 2h 28min Náklady na cestu 263 Kč	 Osobní automobil Cestovní čas 2h 35min Náklady na cestu 326 Kč	 Autobus Cestovní čas 2h 50min Náklady na cestu 165 Kč
---	---	---	--

Data o skutečném chování (revealed preference – RP):

- Obvykle máme údaje pouze o zvolené alternativě (pokud vůbec). Ostatní alternativy musíme domodelovat (chyba „měření“).
- Není to jenom jako.
- Data: průzkumy dopravního chování, data o migraci (Vorel, 2014)...

Využití výzkumu rozhodování

Co ovlivňuje agenta (nemusí to být nutně člověk) při určitém rozhodování?

Marketing: Co rozhoduje při výběru určitého produktu (služby)?

Politika: Co rozhoduje při výběru prezidenta?

Zaměstnanost: Co rozhoduje při výběru zaměstnání?

Psychologie: Co rozhoduje při výběru životního partnera?

Podnikání: Co rozhoduje o investicích do start-up firem (fondů a podobně)?

Doprava: Jaká je hodnota času v dopravě?

Jaký podíl v dopravě by mohlo získat VR spojení Praha – Brno?

Jak se změní podíl využívání železnice při zvýšení rychlosti po rekonstrukci tratě?

Pojďme si formulovat vlastní výzkumné otázky (nemusí být nutně z dopravy)...

Společný experiment: volba dopravního módu na trase Brno - Praha

Nyní si představte, že máte podniknout stejnou cestu, o které jste referoval/a. K dispozici máte čtyři možnosti dopravy zobrazené na kartách níže. Zvolte tu možnost dopravy z nabízených, kterou byste si vybral/a.

AUTOMOBIL	
Cestovní čas	3 h 0 m
Cena za dopravu	600 Kč



AUTOBUS	
Cestovní čas	2 h 0 m
Cena za dopravu	150 Kč



VLAK	
Cestovní čas	2 h 0 m
Cena za dopravu	200 Kč



VYSOKORYCHLOSTNÍ ŽELEZNICE	
Cestovní čas	1 h 30 m
Cena za dopravu	600 Kč



Tato alternativa se vypíná pro ty, kdo nemají auto na cestu k dispozici

Nabízena je jedna z těchto dvou alternativ

HYPERLOOP	
Cestovní čas	0 h 18 m
Cena za dopravu	400 Kč

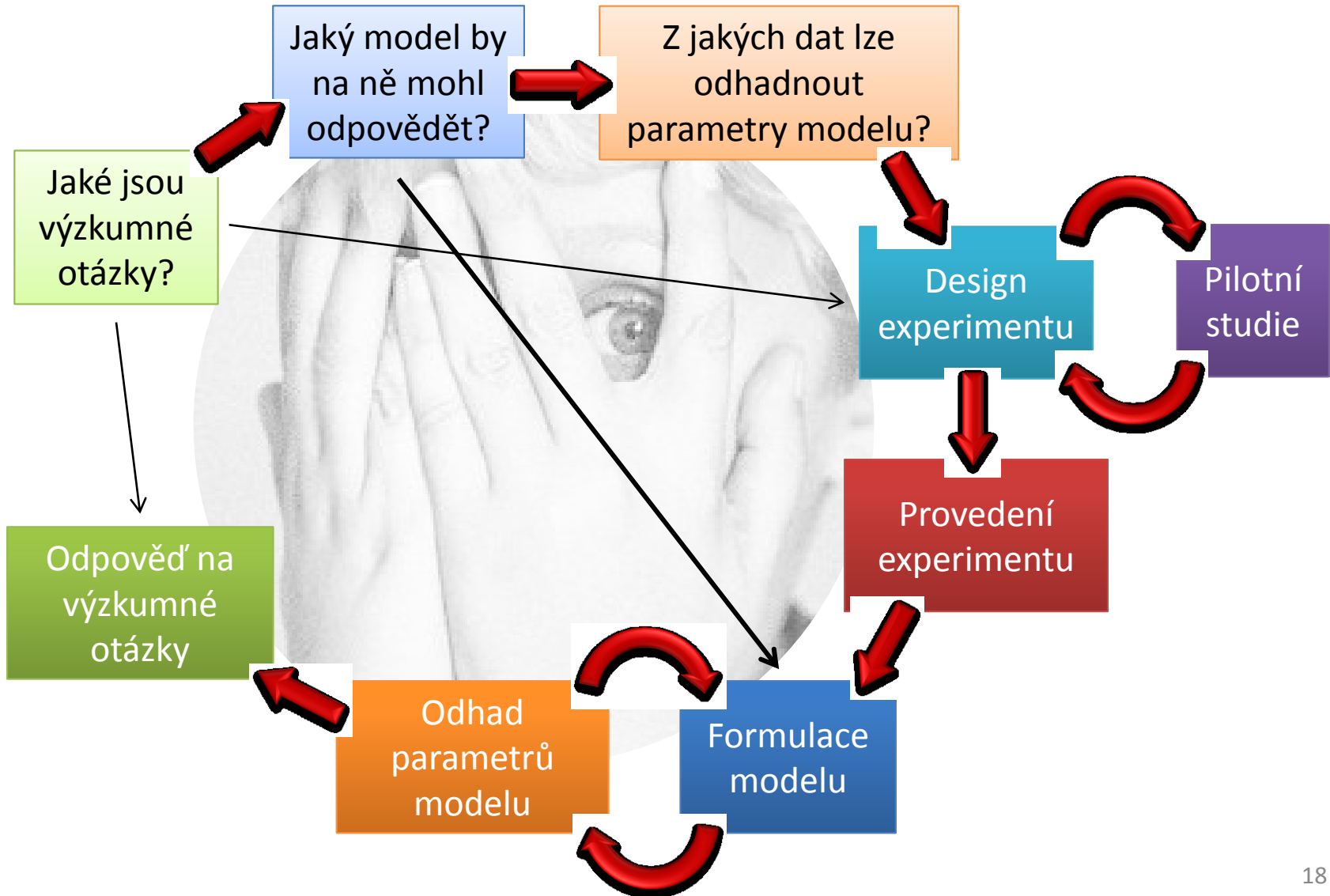


2. Design experimentu

- Fáze výzkumu
- Příprava scénáře
- *Společný experiment: sociodemografie*
- Příprava konkrétní struktury scénáře
- Příklady
- *Společný experiment: alternativy a atributy*

(Henscher, Rose, & Greene, 2007, kap. 5)

Fáze výzkumu



Příprava scénáře

Jací respondenti nás zajímají?

- např. sociodemografie respondenta

Jaké scénáře nás zajímají?

- např. účel cesty, odkud kam se má cestovat

Jaké alternativy přichází do úvahy?

- např. dopravní módy

Jsou předpokládané vlivy na rozhodování známé? Pokud ne, je třeba je nejprve zjistit! → např. focus groups

Jsou předpokládané vlivy na rozhodování pozorovatelné (měřitelné)? Pokud ne, dají se měřit nepřímo?

Společný experiment: sociodemografie

Tento krátký dotazník slouží k tomu, aby prozkoumal preference lidí na cestě mezi Prahou a Brnem a jejich volbu dopravního prostředku. Ptá se také na ochotu využít plánované vysokorychlostní železniční spojení nebo vizionářský dopravní prostředek hyperloop.

Při výběru vepište libovolný znak (třeba "x") do příslušného zeleného políčka nebo se držte instrukcí (vepište rok narození nebo číslo). Nevybraná políčka nechte volná.

Jste muž?
žena?

Vlastníte řidičský průkaz?

Ano
Ne

Ve kterém roce jste se narodil/a?
napište rok:

Jaké je Vaše nejvyšší dosažené vzdělání?

základní
maturita
vysokoškolské

Slyšel/a jste o plánované výstavbě vysokorychlostní železnice?

Ano, mám dobrou představu, o co jde
Ano, mám o tom částečnou představu
Něco o tom se ke mně dostalo
Nevím, co to je

Vysokorychlostní železnice je tratí pro moderní rychlovlaky, které by mezi Prahou a Brnem měly jezdit rychlostí minimálně 250 km/h a mimo stanice kolem 300 km/h.

Jaký je Váš ekonomický status?

Student
Zaměstnanec
Podnikatel
V domácnosti
Důchodce

Slyšel/a jste o tom, co je hyperloop?

Ano, mám dobrou představu, o co jde
Ano, mám o tom částečnou představu
Něco o tom se ke mně dostalo
Nevím, co to je

Hyperloop je futuristický způsob dopravy, založený na pohybu přetlakových kapslí podtlakovými tunely. Plánovaná rychlost přepravy bude mimo stanice kolem 1000 km/h.

Kdy jste naposledy jel/a z Brna do Prahy nebo z Prahy do Brna?

Před několika dny. Odhadněte před kolika:
nebo před několika měsíci. Kolika?

Jaký dopravní prostředek jste na tuto cestu použil/a?

Osobní automobil
Autobus
Vlak

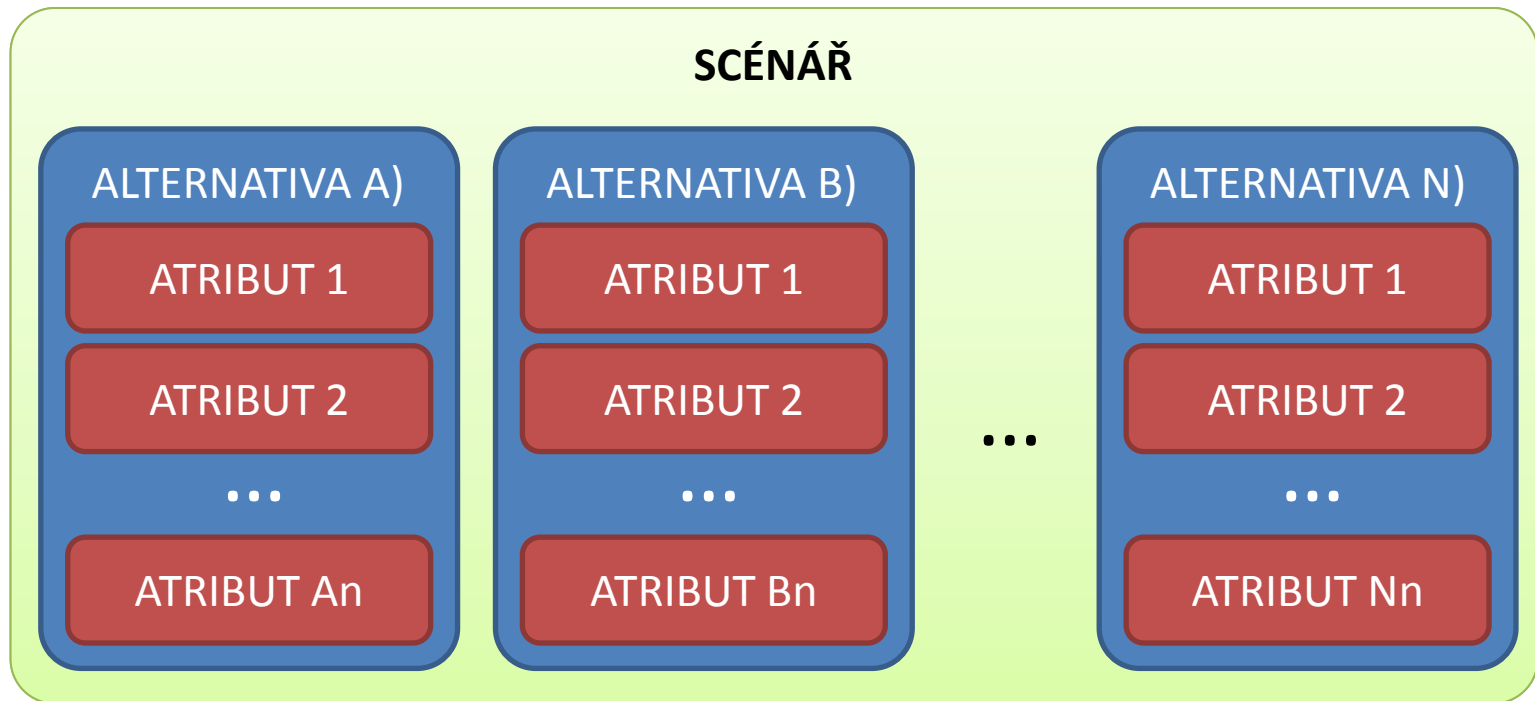
Za jakým účelem jste tuto cestu podnikl/podnikla?

Cesta do/ze zaměstnání
Služební cesta
Jiný účel

Mohl/a jste pro tuto cestu použít automobil?

Ano
Ne

Příprava konkrétní struktury scénáře



Formát odpovědi






- výběr jedné alternativy
- řazení
- Likertovské škály

Kolik může být alternativ?

Kolik může být atributů?







Musí být atributy stejné?

Můžu dát respondentovi víc scénářů?

Brand		SONY	 LG	Nominální proměnné
Price	\$500	\$300	\$400	Metrické proměnné
Resolution	480x720	720x480	1280x800	Kdovíjaké proměnné
Size				Pořadové proměnné

Rozhodování v
enviromentalistice

Fixování na
současný stav

Choice set 1 (Block 2)	STATUS QUO	Future Alternative 1	Future Alternative 2
Windbreaks	20% windbreaks 	100% windbreaks 	50% windbreaks 
Crop residue management	Fire allowed 	Fire banned 	Fire allowed 
Land registration fee <small>Relative to what you pay today</small>	87 Lari/ha 0 Lari/ha	110 Lari/ha +22 Lari/ha	95 Lari/ha +7 Lari/ha
Your choice			



\$25.99

\$11.99

\$17.99

16.99

\$14.49

Considering only the 5 wines above, which one would you be most likely to choose if you had to make a choice, and which one would you be least likely to choose?

Most	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Least	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

SUBMIT

Zvláštní formát
odpovědi

Considering the trip you made Yesterday

Trip	From	To
2	Shop	Your Home

Purpose	Main Mode	Start Hour
Go Home	MRT	6:00 PM

If you had the following options:

Option A	
Total Trip Duration (mins):	35
One-Way Fare:	\$1.90

Option B	
Total Trip Duration (mins):	40
One-Way Fare:	\$1.20

Which option would you prefer?

Strong Preference for Option A	Moderate Preference for Option A	Slight Preference for Option A	No Preference for either Option	Slight Preference for Option B	Moderate Preference for Option B	Strong Preference for Option B
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

Next Game...





Likertovský formát odpovědi získává informaci o „síle“ preference

FIGURE 1 A Typical Question from the Adaptive Stated Preference Survey.

Which one of the following products would you buy?

Aside from the features / characteristics listed, please assume the products to be alike in all other ways.

(1/4)

	Product A	Product B	Product C	Product D
Brand		 HONDA		 TOYOTA
Feature	Remote keyless entry	Electronic stability	Anti-lock brakes (ABS)	Roadside Assistance
Characteristic	Look and Feel	Price	Safety	Brand Recognition
Your Selection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Back Continue

If these are the concept offered to you, which concept would you choose?

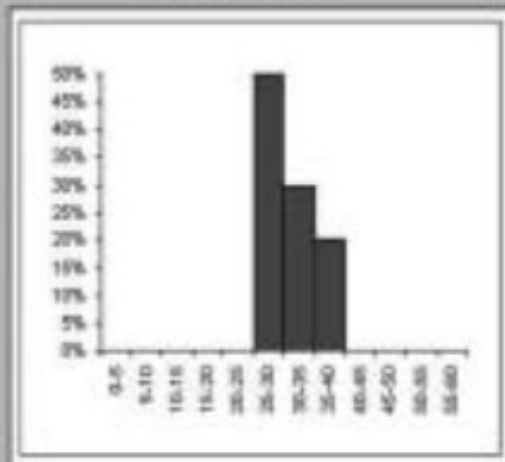
Pack Size	6 beer package size	12 beer pack size	24 beer pack size	None of these
Beverage				
Package	Cans	Pints: 330ml	Bottles: 650 ml	
Occasion	After work	On weekends	With dinner	
Choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Jak se do výsledků promítnou relevantnější atributy?

Zanáší současný „stav“ respondenta do jeho rozhodování. Zvyšuje validitu experimentu.

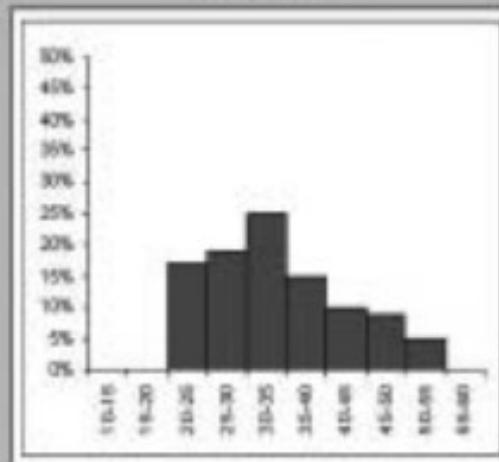
Think about your morning commute to work. Suppose, you have exactly two routes to choose from for your commute. If the distribution of your travel time and the toll on each route were as shown below, which route will you choose?

Route 1



Toll: \$ 1.75

Route 2



Toll: \$ 0.00

Your Choice

1

2

Next

Společný experiment: alternativy a atributy

AUTOMOBIL Cestovní čas 3 h 0 m Cena za dopravu 600 Kč	AUTOBUS Cestovní čas 2 h 0 m Cena za dopravu 150 Kč	VLAK Cestovní čas 2 h 0 m Cena za dopravu 200 Kč	VYSOKORYCHLOSTNÍ ŽELEZNICE Cestovní čas 1 h 30 m Cena za dopravu 600 Kč	HYPERLOOP Cestovní čas 0 h 18 m Cena za dopravu 400 Kč
■	■	■	■	■

Variance atributů u jednotlivých alternativ

Je tohle taky atribut?

osobní automobil		autobus		vlak		vysokorychlostní železnice		hyperloop	
tt [h]	cost [Kč]	tt [h]	cost [Kč]	tt [h]	cost [Kč]	tt [h]	cost [Kč]	tt [h]	cost [Kč]
2	300	2	100	2	100	0.5	300	0.3	0
2.5	400	2.5	150	2.5	200	1	400	0.5	200
3	600	3	200	3	300	1.5	600	0.75	400
3.5	800	3.5	250	3.5	400	2	800	1	800

VRT:

http://www.vysokorychlostni-zeleznice.cz/wp-content/uploads/2012/10/odhad_cestujicich_VRT_Brno_Praha_Kordis_2013.jpg

<http://www.vysokorychlostni-zeleznice.cz/vysokorychlostni-trat-praha-brno/>

Hyperloop:

https://ekonomika.idnes.cz/hyperloop-brno-podepsalo-memorandum-s-htt-fja-/eko-doprava.aspx?c=A170118_141802_eko-doprava_rny

https://ekonomika.idnes.cz/hyperloop-a-spojeni-brna-svetem-dnt-/eko-doprava.aspx?c=A160819_094653_eko-doprava_rny

3. Design experimentu: generování scénářů

- Full factorial design
- Úrovně atributů
- Ortogonální kódování
- Optimální design
- *Optimální design v R*
- Panelový design
- Adaptivní metody
- Výběr vzorku respondentů

Full factorial design

Full factorial design obsahuje scénáře, které jsou kombinacemi všech úrovní všech parametrů. Těch je L^{MA} .

M=počet (nazvaných) alternativ; A=počet atributů; L=počet úrovní

Zajišťuje, aby atributy byly skutečně nezávislé (nekorelované).

Alternativa	Cena	Cestovní čas
Auto	4 úrovně	4 úrovně
Vlak	4 úrovně	4 úrovně
Autobus	4 úrovně	4 úrovně
VRT	4 úrovně	4 úrovně
Hyperloop	4 úrovně	4 úrovně

Počet kombinací: $4^{5*2} = 1048576$



**Co s tím?
Redukce úrovní?**

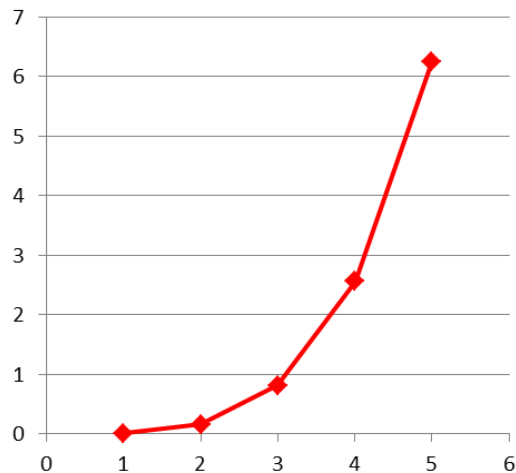
Úrovně atributů

Víc úrovní atributů zvětšuje počet kombinovaných scénářů

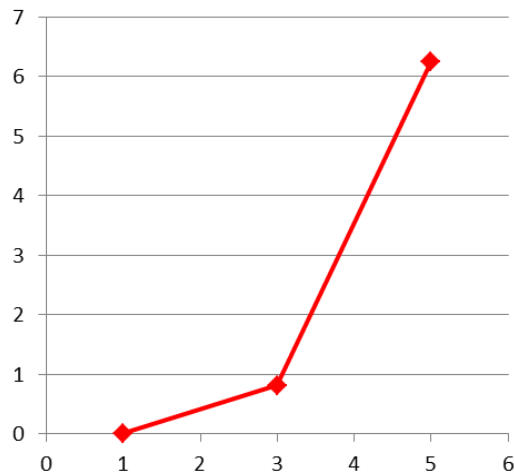
Méně úrovní atributů může opomenout některé důležité nelinearity

Dvě úrovně umožňují odhadovat lineární trend.

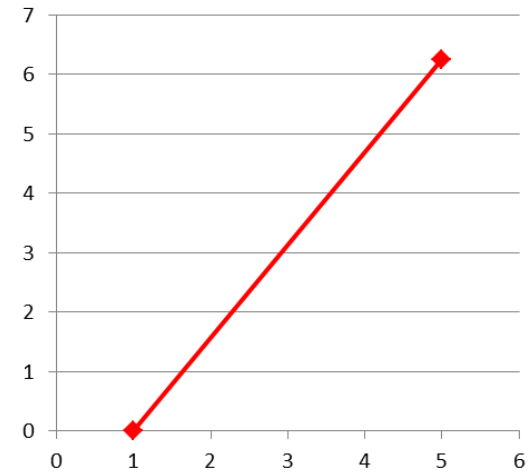
5 úrovní



3 úrovně



2 úrovně



Redukce na dvě úrovně: $2^{5*2} = 1024$

Ortogonalní kódování

Ortogonalní kódování – součet všech hodnot atributu se rovná nule

Úroveň	1.	2.
Ortogonalní	-1	1

Úroveň	1.	2.	3.
Ortogonalní	-1	0	1

Úroveň	1.	2.	3.	4.
Ortogonalní	-2	-1	1	2

Úroveň	1.	2.	3.	4.	5.
Ortogonalní	-2	-1	0	1	2

Optimální design

Cílem je redukovat počet scénářů a současně nekorelovat nezávislé proměnné.

Pokud nás zajímají jen hlavní efekty, neměly by korelovat ortogonálně kódované řádky proměnných.

Pokud nás zajímají interakce, přidáme násobky příslušných proměnných. Ani ty by pak neměly korelovat.

- Náhodný výběr
- Optimalizace – např. knihovna AlgDesign v R.

Optimální design v R

```
library("AlgDesign")
```

4 proměnné se
dvěma úrovněmi

```
ff=gen.factorial(c(2, 2, 2, 2)) →
```

```
rf=optFederov(~., ff, 6)
```

Požadovaný
počet pokusů

Požadovaný model: všechny
proměnné v lineární kombinaci

```
$D [1] 0.940863  
$A [1] 1.111111  
$Ge [1] 0.818  
$Dea [1] 0.801  
$design
```

Indikátory nezávislosti

```
  X1 X2 X3 X4  
2   1 -1 -1 -1  
3  -1  1 -1 -1  
8   1  1  1 -1  
12  1  1 -1  1  
13 -1 -1  1  1
```

Výsledná matice



	X1	X2	X3	X4
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	-1	1	-1	-1
4	1	1	-1	-1
5	-1	-1	1	-1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	-1
9	-1	-1	-1	1
10	1	-1	-1	1
11	-1	1	-1	1
12	1	1	-1	1
13	-1	-1	1	1
14	1	-1	1	1
15	-1	1	1	1
16	1	1	1	1

Wheeler, B., & Braun, M. J. (2015). Package 'AlgDesign'.

<https://mirror.its.sfu.ca/mirror/CRAN/web/packages/AlgDesign/AlgDesign.pdf>

Panelový design

Získat respondenta je náročné.

Udělat SP experiment je snadné.

Proč nenechat jednoho respondenta rozhodovat nad více scénáři?

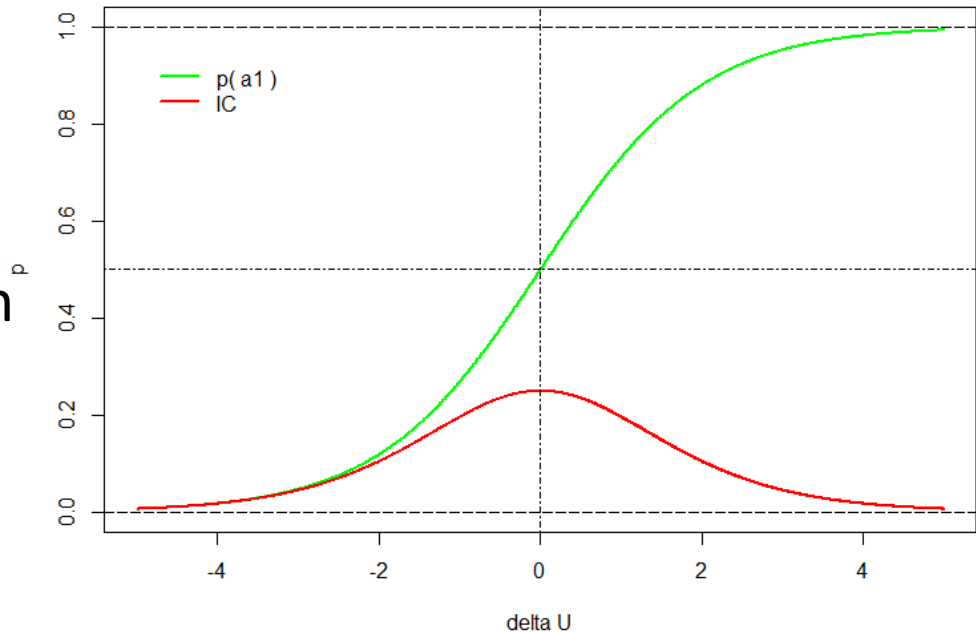
Má to svá úskalí:

- Jednotlivá rozhodnutí se mohou ovlivňovat
- Kvalita rozhodnutí může postupně degradovat
- Model musí počítat s efekty jednotlivých respondentů. Tyto efekty totiž nejsou nezávislé.

Adaptivní metody

Nastavení úrovní atributů na základě průběžných výsledků tak, aby se co nejvíc zvýšil informační přínos následujících pokusů.

$$IC = p(a1) p(a2) = p(a1) (1 - p(a1))$$



Individuální zjištění parametrů krátkým adaptivním experimentem.

Fowkes, A.S.; Shinghal, N. (2002). The Leeds Adaptive Stated Preference Methodology. *Bilingualism*, 110, 115–122.

Fowkes, T. (2007). The design and interpretation of freight stated preference experiments seeking to elicit behavioural valuations of journey attributes. *Transportation Research Part B: Methodological*, 41(9), 966-980.

Výběr vzorku respondentů

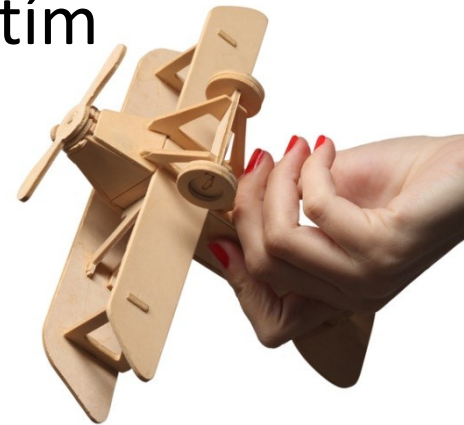
- **Náhodné výběry** umožňují zobecnění na populaci, ale jsou velice drahé. Navíc se zvětšuje non-response.
 - Prostý náhodný
 - **Stratifikovaný** – umožňuje počítat s tím, že jen některých se rozhodování týká
- **CBS (*Choice-Based Sample*)** – průzkum přímo mezi těmi, kteří se rozhodují
 - RP data
 - Průzkum v dopravě - umožňuje určit variability atributů respondentů
- **Over / undersampling** - relativní navýšení vzorku málo zastoupené a zmenšení vzorku více zastoupené skupiny.

4. Pilotní průzkum

- K čemu je pilotní průzkum
- *Program Biogeme a EasyChoice*
- *Formulace modelu v Biogeme*
- *Formulace modelu v R*

K čemu je pilotní průzkum

- Ověření srozumitelnosti dotazování
 - Realističnost – dává to respondentům smysl?
 - Kognitivní přístupnost – vypořádají se s tím respondenti?
- Ověření procedury dotazování
- Nastavení úrovní parametrů – jsou úrovně parametrů nastaveny tak, aby variovaly vybírané alternativy?



Oslovit co nejširší pole respondentů – desítky dotazování

Program Biogeme

Autor: [Michel Bierlaire](#),
[Ecole Polytechnique Fédérale de Lausanne](#),
Switzerland.



Open source multiplatformní program, který je stále vyvíjen.
Nyní jsou dvě verze *Pythonbiogeme* a *Bisonbiogeme*.

<http://biogeme.epfl.ch/>

Program má vlastní GUI, já ale používám svoje.

Rozbalte si někde balík EasyChoice.

Balík obsahuje Biogeme, GUI EasyChoice a příklady, se kterými budeme pracovat.

EasyChoice: GUI pro Biogeme

NAČTE DATA A MODEL

UDĚLÁ ODHAD PARAMETRŮ MODELU

ULOŽÍ DATA A MODEL POD JMÉNEM

OKNO DAT

PARAMETRY ODHADU

REPORT PRŮBĚHU ODHADU

PARAMETRY MODELU

VÝSLEDKY ODHADU MODELU

OKNO MODELU

OKNO VÝSLEDKŮ

The screenshot displays the EasyChoice v. 0.4 - GUI for Biogeme interface. It features a menu bar with 'Load All', 'Save Project (Model + Report) As...', 'ESTIMATE', 'Options', and 'About'. Below the menu is a data table with columns for 'id', 'zs', 'ss', 'vs', 'stud', 'zam', 'podnik', 'doma', 'pens', 'dlic', 'vrt3', 'vrt2', 'vrt1', 'vrt0', 'hl3', 'hl2', 'hl1', 'hl0', 'tripd', and 'tripv'. The 'ESTIMATE' button is highlighted. Below the data table are two panes: 'Model' and 'Report'. The 'Model' pane shows a list of parameters with their values, lower and upper bounds, and status (0=variable, 1=fixed). The 'Report' pane shows the results of the estimation, including correlation coefficients and a table of coefficients.

id	zs	ss	vs	stud	zam	podnik	doma	pens	dlic	vrt3	vrt2	vrt1	vrt0	hl3	hl2	hl1	hl0	tripd	tripv
1	0	0	1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	1	0
2	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0
4	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0
5	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	6	0

```
Model Parameters
Load Save As model1.mod Undo | 24:13
*****
MODEL
*****
Michal Šimeček,
Easy choice Brn
*****

// Name value LowerBound UpperBound status (0=variable, 1=fixed)
asc_a 0 -10000 10000 1
asc_b 0 -10000 10000 0
asc_h 0 -10000 10000 0
asc_r 0 -10000 10000 0
asc_v 0 -10000 10000 0
bcost 0 -10000 10000 0
btt 0 -10000 10000 0

//[Mu]
// In general, the value of mu must be fixed to 1. For testing purposes, you
// may change its value or let it be estimated.
// Value LowerBound UpperBound Status
//1 -1 1 0

//in-parameter expression (beta1*x1 + beta2*x2 + ... )
asc_b * one + btt * a_tt + bcost * a_cost
asc_b * one + btt * b_tt + bcost * b_cost
3 vvlak one asc_r * one + btt * r_tt + bcost * r_cost
4 vrt v_avl asc_v * one + btt * v_tt + bcost * v_cost
5 hyperloop h_avl asc_h * one + btt * h_tt + bcost * h_cost

[Expressions]
```

```
Report LogFile
Save As model1.rep
-1.44 0.806 -1.78
*****
functions
*****
to a_avl asc_a * one + btt * a_tt + bcost * a_cost
robust one asc_b * one + btt * b_tt + bcost * b_cost
rloop h_avl asc_h * one + btt * h_tt + bcost * h_cost
vloop v_avl asc_v * one + btt * v_tt + bcost * v_cost
r_cost
v_cost

Correlation of coefficients
*****
Coeff1 Coeff2 Covariance Correlation t-test Rob. covar. Rob. correl. Rob. t-test
-----
asc_h asc_v 1.74 0.660 0.05 * 1.21 0.501 0.04
asc_b bcost 0.00114 0.370 -0.15 * 0.00185 0.516 -0.13
asc_h bcost -0.00150 -0.270 0.27 * -0.000771 -0.140 0.28
asc_b asc_h -0.101 -0.0442 -0.30 * -0.255 -0.105 -0.29
asc_v bcost -0.000949 -0.265 0.36 * 0.000688 0.194 0.38
asc_b asc_v 0.125 0.0850 -0.39 * 0.235 0.150 -0.39
asc_r bcost 0.00140 0.435 -0.44 * 0.00167 0.494 -0.43
asc_h asc_r -0.111 -0.0467 0.44 * -0.345
asc_b asc_r 1.07 0.811 0.49 * 1.34
asc_r asc_v 0.0801 0.0523 -0.58 * 0.75
asc_r btt -0.222 -0.234 0.58 * -0.271
asc_b btt -0.266 -0.293 0.81 * -0.315 -0.331 0.75
asc_h btt 0.857 0.526 1.13 * 0.799 0.546 1.20
asc_v btt 0.505 0.480 1.64 * 0.608 0.646 2.00
bcost btt 0.000646 0.292 1.78 * 0.000614 0.285 1.89

Smallest singular value of the hessian: 0.191222
```


Formulace modelu v Biogeme

[Choice]

PROMĚNNÁ, KTERÁ OBSAHUJE VOLBU

CHOICE

Jméno proměnné, která nabývá číselných hodnot vybraných alternativ (viz níže)

[Beta]

SEZNAM ODHADOVANÝCH PARAMETŮ

```
// Name Value LowerBound UpperBound status (0=variable, 1=fixed)
asc_a 0 -10000 10000 1
asc_b 0 -10000 10000 0
asc_h 0 -10000 10000 0
asc_r 0 -10000 10000 0
asc_v 0 -10000 10000 0
bcost 0 -10000 10000 0
btt 0 -10000 10000 0
```

Jméno
parametru

Počáteční
hodnota

Horní a
dolní mez

Má být velikost
parametru
odhadována?
0 = ano

Formulace modelu v Biogeme

ROVNICE UŽITKOVÝCH FUNKCÍ

```
[Utilities]
// Id  Name      Avail  linear-in-parameter expression (beta1*x1 + beta2*x2 + ... )
  1  auto      a_avl  asc_a * one + btt * a_tt + bcost * a_cost
  2  autobus  one    asc_b * one + btt * b_tt + bcost * b_cost
  3  vlak      one    asc_r * one + btt * r_tt + bcost * r_cost
  4  vrt      v_avl  asc_v * one + btt * v_tt + bcost * v_cost
  5  hyperloop h_avl  asc_h * one + btt * v_tt + bcost * v_cost
```

Jméno
alternativy

Má být

alternativa
povolena?

1 = povolená
0 = zakázaná

Číslo alternativy
(těchto hodnot
nabývá proměnná
v sekci Choice)

Regresní rovnice
(proměnné a operátory musí být
odděleny mezerou)

Formulace modelu v Biogeme

DALŠÍ VÝPOČTY

```
[Expressions]
// Define here arithmetic expressions for name that are not directly
// available from the data
one = 1
```

TYP MODELU

```
[Model]
// Currently, the following models are available
// Uncomment exactly one of them
//$BP // Binary Probit Model
//$OL // Ordinal logit
$MNL // Multinomial Logit Model
//$NL // Nested Logit Model
//$CNL // Cross-Nested Logit Model
//$NGEV // Network GEV Model
```

Typy modelů, se kterými umí Biogeme pracovat (viz dokumentace k Biogeme).

KTERÉ ŘÁDKY DAT MAJÍ BÝT VYNECHÁNY

```
[Exclude]
( err == 1 )
```

Řádek s err=1 obsahuje chybu

Formulace modelu v R

Alternativně
specifická
konstanta

Koeficient asociovaný s
alternativou, proměnná
s individuem.

$$V_{mn} = \alpha_m * 1 + \beta * cost_m + \delta_m * purp_n + \gamma_m * tt_m$$

Koeficient stejný pro všechny
alternativy, proměnná
asociovaná s alternativou.

Koeficient i proměnná
jsou asociované s
alternativou.

 `m = nlogit(cost | purp | tt)`

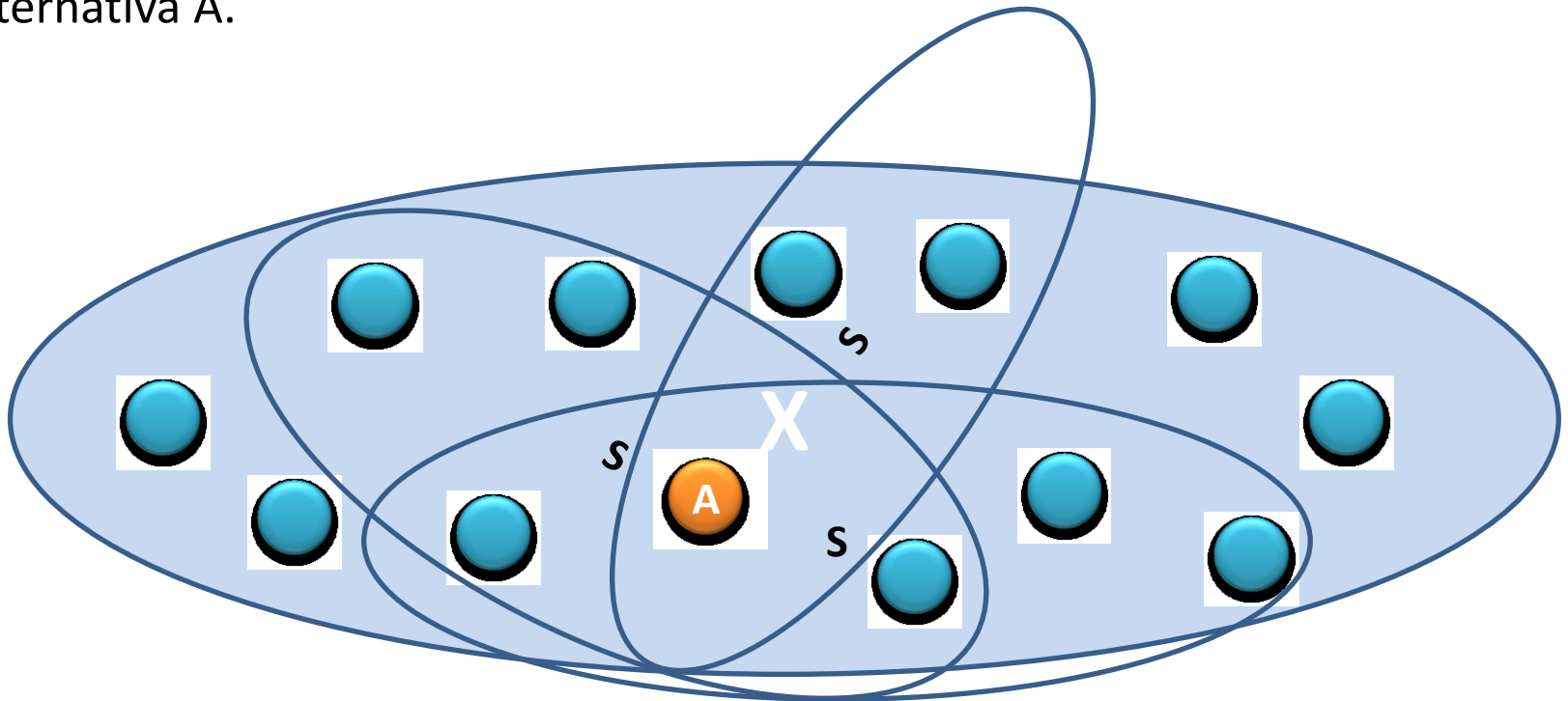
Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit Packages. *R package version 0.2-2*. URL: <http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf>

5. Analýza diskrétních voleb

- IIA – independence of irrelevant alternatives
- IIA a nested logit
- *Nested logit v Biogeme*
- Vícečetné alternativy
- Mixed logit pro panelová data
- *Mixed logit v Biogeme*

IIA – independence of irrelevant alternatives

Pokud je z množiny alternativ X vybrána alternativa A , pak z jakékoli množiny alternativ S , která je podmnožinou X a obsahuje alternativu A , bude vybrána alternativa A .



To platí pro individuální rozhodnutí. Jak je to ale pro soubor rozhodnutí?

IIA – independence of irrelevant alternatives

CHOICE SET 1



$p = 0,5$
odds = 1:1



$p = 0,5$
odds = 1:1

Odds = šance

CHOICE SET 2



$p = 0,333$
odds = 1:1



$p = 0,333$
odds = 1:1



$p = 0,333$
odds = 1:1



IIA a nested logit

CHOICE SET 1



$p = 0,5$
odds = 1:1



$p = 0,5$
odds = 1:1

CHOICE SET 2



$p \approx 0,5$
odds \approx 1:1



Hnízdo (*nest*)
 $p \approx 0,5$
odds \approx 1:1
Volba mezi autobusy
následuje po volbě
hnízda „autobus“.

Do modelu se zavádí škálovací parametr μ pro alternativy z hnízda „bus“.

$$V_{bus} = \frac{1}{\mu_{bus}} \ln(e^{\mu_{bus}V_{bus1}} + e^{\mu_{bus}V_{bus1}} + \dots)$$

Nested logit v Biogeme

TYP MODELU

```
[Model]
```

```
NL // Nested Logit Model
```

— Vybraný je nested logit model

SPECIFIKACE NESTED MODELU

```
[NLNests]
```

```
// Name paramvalue LowerBound UpperBound status list of alt
```

```
Individual
```

```
1
```

```
0
```

```
1000
```

```
1
```

```
1
```

```
2
```

```
3
```

```
4
```

```
5
```

1 = auto

```
Public
```

```
21.9
```

```
1
```

```
1000
```

```
0
```

```
2
```

```
3
```

```
4
```

```
5
```

2 = autobus

3 = vlak

4 = VRT

5 = hyperloop

Počáteční hodnota
škálovacího
parametru μ

Horní a dolní mez
škálovacího
parametru μ

Alternativy
patřící do
hnízda

Název
hnízda

Pro které hnízdo se má
odhadovat škálovací parametr?

0 = ano

1 = ne

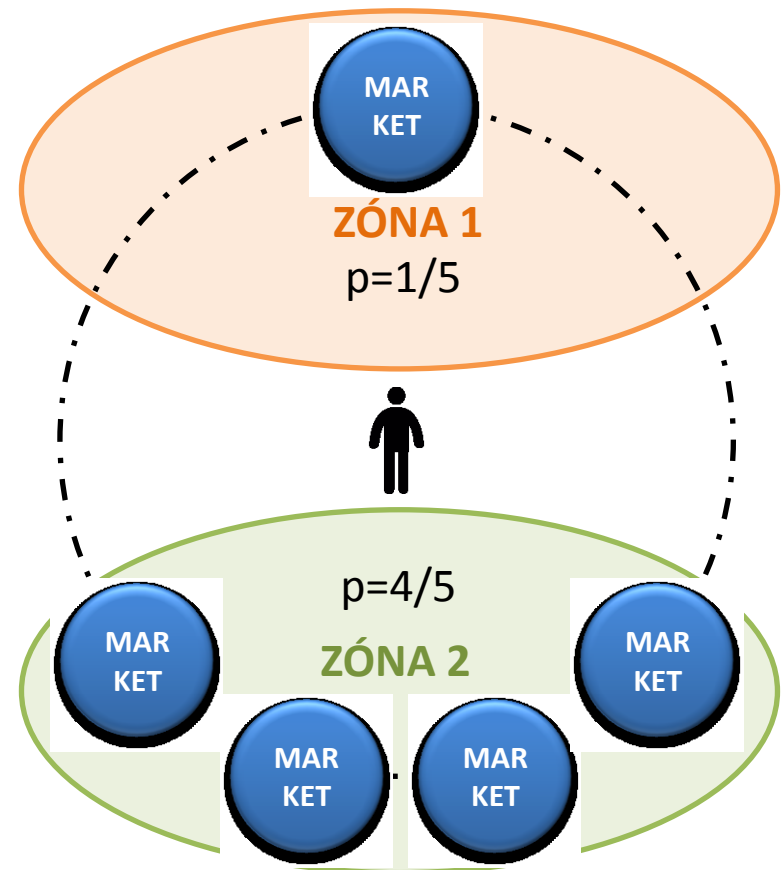
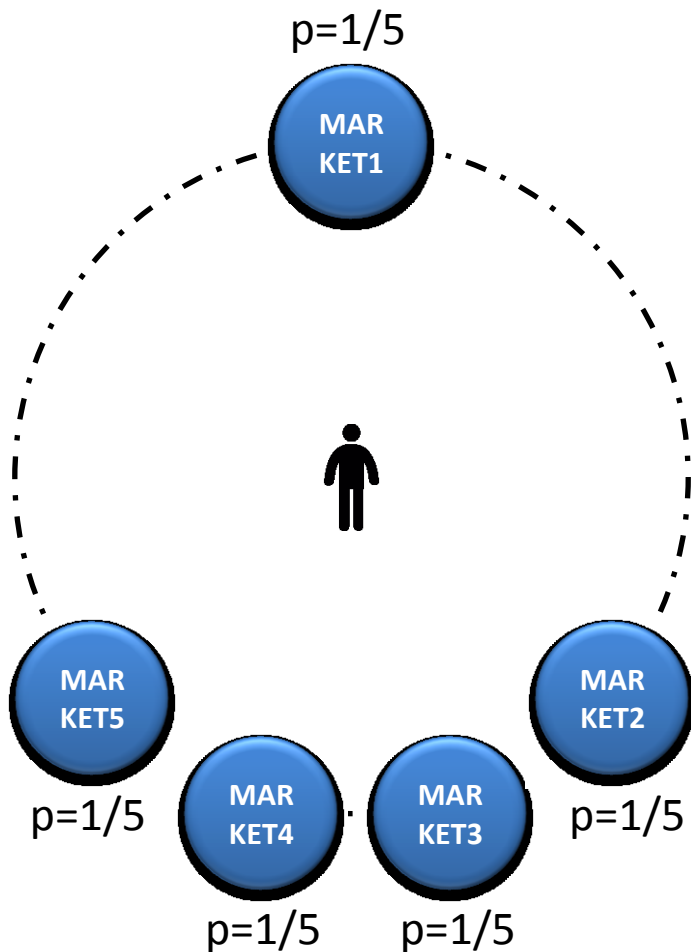
Vícečetné alternativy

$$U_{market1} = U_{market2} = U_{market n}$$

$$U_{zóna1} \neq U_{zóna2}$$

$$U_{zóna1} = 1 \quad U_{zóna2} \neq 4 = \ln(4)$$

4 jako 4 obchody
v zóně 2



Mixed logit pro panelová data

Proč nenechat jednoho respondenta rozhodovat nad více scénáři?

Model musí počítat s efekty jednotlivých respondentů r na každou alternativu a :

$$U_{ar} = \beta x_{ar} + [\eta_{ar} + \varepsilon_{ar}]$$

Přičemž η_{ar} je náhodná veličina, která není IID, ale může být závislá na respondentovi a alternativě.

Hensher, D., & Greene, W. (2001). The Mixed Logit Model: The State of Practice and Warnings for the Unwary, (October 2011), 39. <https://doi.org/10.1.1.195.1039>

Mixed logit v Biogeme

```
[Beta]
asc_a    0    -10000    10000    1
asc_b    0    -10000    10000    0
asc_h    0    -10000    10000    0
asc_r    0    -10000    10000    0
asc_v    0    -10000    10000    0
bcost    0    -10000    10000    0
btt      0    -10000    10000    0
s        0     0        10000    0
```

Směrodatná odchylka „s“
náhodné proměnné je normálně
odhadovaná veličina, která nemůže
nabývat záporných hodnot.

```
[Utilities]
1 auto      a_avl  asc_a * one + btt [ s ] * a_tt + bcost * a_cost
2 autobus  one    asc_b * one + btt [ s ] * b_tt + bcost * b_cost
3 vlak     one    asc_r * one + btt [ s ] * r_tt + bcost * r_cost
4 vrt      v_avl  asc_v * one + btt [ s ] * v_tt + bcost * v_cost
5 hyperloop h_avl  asc_h * one + btt [ s ] * v_tt + bcost * v_cost
```

```
[Draws]
1000
```

POČET SIMULACÍ

```
[PanelData]
```

SPECIFIKACE PANELU

id Proměnná „id“ identifikuje respondenty
btt_s Podle ní se vytváří náhodná proměnná

Náhodná proměnná má rovnoměrné
rozložení (hranaté závorky) a
směrodatnou odchylku „s“

Model zohledňuje, jak se respondenti systematicky liší v hodnotě času.

6. Interpretace výsledků

- Interpretace kvality modelu
- Kvantifikace koeficientů
- Hodnota času a ochota platit za služby
- *Hodnota času na trase Brno - Praha*
- Elasticita
- *Inkrementální model volby dopravního módu na trase Brno - Praha*

Interpretace kvality modelu

DC modely jsou odhadovány metodou maximální věrohodnosti. Ukazatelem kvality je log-likelihood funkce LL a dále například likelihood ratio test a McFaddenovo R-square.

Likelihood ratio test LR vychází z log-likelihood modelu LL_m a log-likelihood nulového modelu (bez nezávislých proměnných) LL_{h0} a rozdělení má χ^2 .

$$LR = 2 * \ln \left(\frac{\exp(LL_m)}{\exp(LL_{h0})} \right)$$

McFaddenovo R-square lze interpretovat podobně jako koeficient determinace: podíl vysvětleného rozptylu.

$$R^2_{McFadden's} = 1 - \frac{LL_m}{LL_{h0}} \qquad R^2 = 1 - \frac{SS_m}{SS_{h0}}$$

Interpretace kvality modelu

```
// This file has automatically been generated.  
// 02/16/18 08:27:23  
// Michel Bierlaire, EPFL
```

```
biogeme 2.2 [Thu Mar 15 14:58:02 WEST 2012]  
Michel Bierlaire, EPFL
```

```
Model: Mixed Multinomial Logit for panel data  
Number of draws: 1000  
Number of estimated parameters: 5  
Number of observations: 19  
Number of individuals: 12  
Null log-likelihood: -24.901  
Init log-likelihood: -24.901  
Final log-likelihood: -17.270  
Likelihood ratio test: 15.263  
Rho-square: 0.306  
Adjusted rho-square: 0.106  
Final gradient norm: +6.736e-005  
Diagnostic: Convergence reached.  
Iterations: 88  
Run time: 00:18  
Variance-covariance: from finite difference hessian  
Sample file: data.dat
```

Typ modelu

Liší se, protože se jedná o panelová data

LL_{h0}

LL_m

LR

$R^2_{McFadden's}$

Jak dopadl odhad

Interpretace koeficientů

Utility parameters

Name	Value	Std err	t-test	p-val	Rob. std err	Rob. t-test	Rob. p-val
asc_a	0.00	--fixed--					
asc_b	0.00	--fixed--					
asc_h	0.00	--fixed--					
asc_r	-0.370	0.688	-0.54	0.59 *	0.592	-0.63	0.53 *
asc_v	0.259	0.972	0.27	0.79 *	0.849	0.30	0.76 *
bcost	-0.00619	0.00243	-2.54	0.01	0.00178	-3.48	0.00
btt	-1.59	0.641	-2.47	0.01	0.409	-3.88	0.00
s	0.0282	0.806	0.03	0.97 *	0.0859	0.33	0.74 *

Hodnoty jsou fixované a nemění se

Odhadnuté
velikosti
parametru

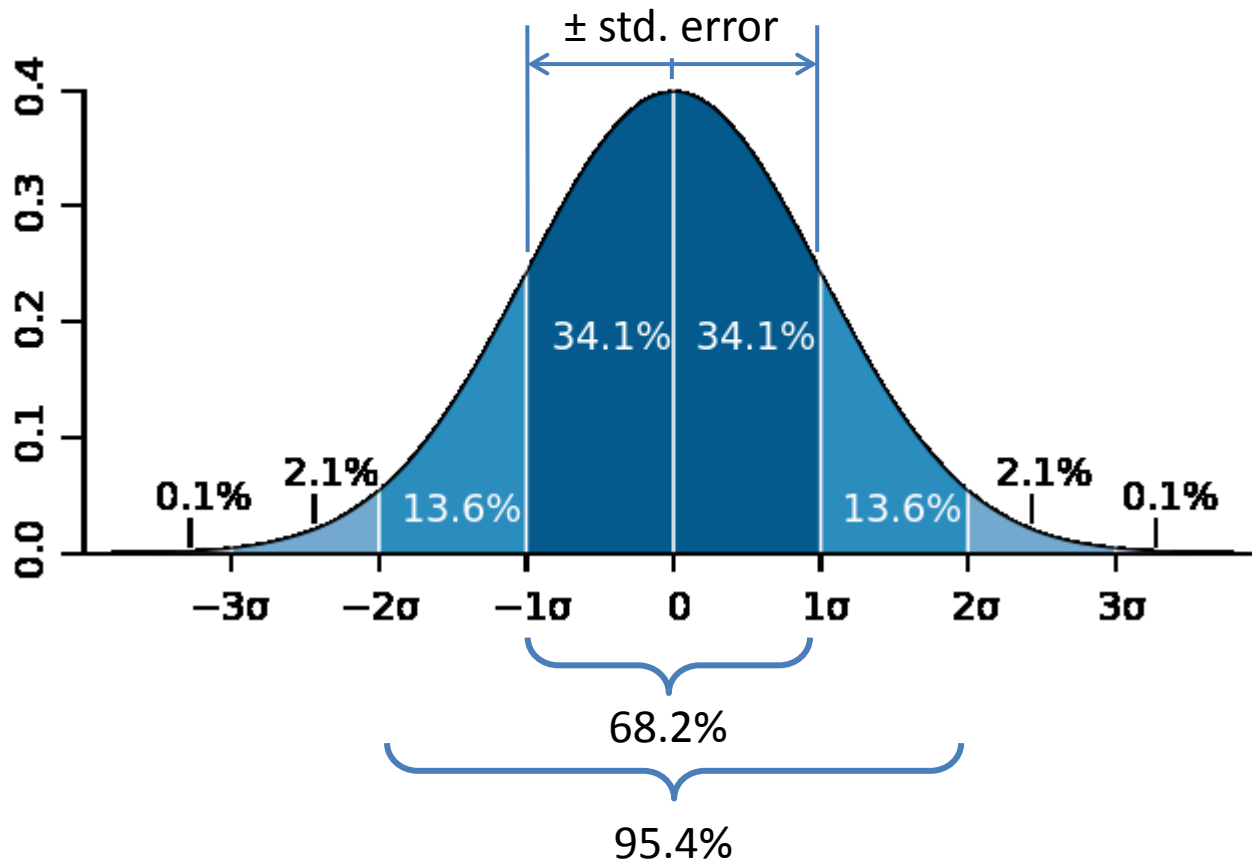
Standardní
chyba odhadu

t-test
významnosti
rozdílu od nuly

významnost
t-testu

Interpretace koeficientů

Standardní chyba odhadu znamená, že se skutečná velikost parametru pohybuje ve vzdálenosti \pm tato chyba od hodnoty odhadu s pravděpodobností 68.2%.



Kvantifikace koeficientů

$$U = \beta_1 x_1 + \beta_2 x_2 \dots$$

Užitek U je bezrozměrná veličina. Nezávislé proměnné $x_1, x_2 \dots$ mají různé jednotky (např. minuty, koruny...). Jaké veličiny mají mít bety ($\beta_1, \beta_2 \dots$), aby rovnice dávala smysl?

Bety mají rozměr převrácené hodnoty svých .

Hodnota času (VoT) neboli hodnota ušetřeného cestovního času (VTTS)

$$U = \beta_{tt} tt + \beta_{cost} cost \dots$$

$$tt[*min*] \quad \rightarrow \quad \beta_{tt} \left[\frac{1}{*min*} \right]$$

$$cost[CZK] \quad \rightarrow \quad \beta_{cost} \left[\frac{1}{CZK} \right]$$

$$VoT (VTTS) = \frac{\beta_{tt}}{\beta_{cost}} \rightarrow \quad VoT \left[\frac{CZK}{*min*} \right]$$

Stejně jako hodnotu času můžete konstruovat jiné veličiny.

Další kvantifikace (monetarizace)

Ochota platit
za služby:

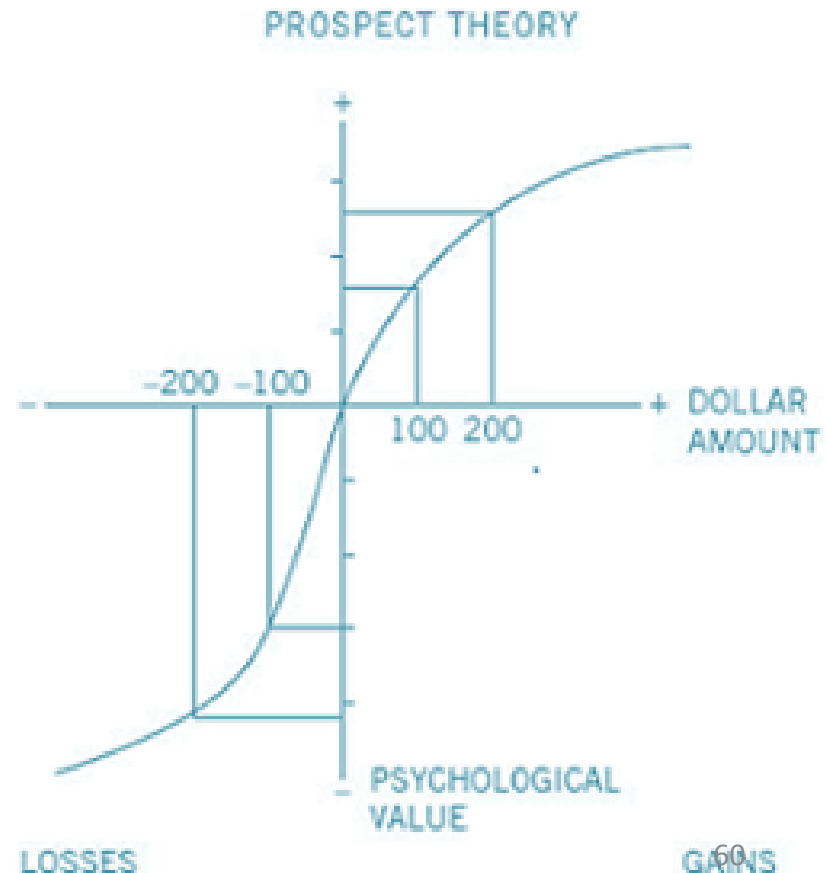
$$U = \beta_{služba}služba + \beta_{cost}cost \dots$$

Služba nabývá hodnot 1, 0 (přítomnost/nepřítomnost služby).

$$WTP = \frac{\beta_{služba}}{\beta_{cost}}[CZK/slужba]$$

Kolik peněz průměrně jsou
respondenti ochotná zaplatit
za službu.

Podobně funguje ochota akceptovat
(WTA). Není to ale zrcadlový opak,
a to kvůli averzi ke ztrátě (teorie
Prospektu Kahnemana a Tverského).



Chyba odhadu kvantifikovaných parametrů

Kvantifikované ukazatele jsou složeny z parametrů modelu. Pro jejich odhady jsou určeny standardní chyby odhadu.

Výpočet chyby kvantifikovaného ukazatele vychází z obecného vzorce výpočtu chyby pro nepřímá měření (Tichá, 2004).

Jestliže nepřímo měřená veličina y je funkcí f přímo měřených veličin x_1 až x_n , pak chyba nepřímo měřené veličiny y , ε_y , se rovná odmocnině součtu mocnin násobků parciálních derivací podle každé veličiny x .

$$\varepsilon_y = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \varepsilon_{x_i} \right)^2}$$

Chyba odhadu VoT (VTTS)

Jakou chybu má kvantifikovaný parametr, například *hodnota času*?

$$VoT = \frac{\beta_{tt}}{\beta_{cost}}$$

Přibližně platí, že jeho standardní chybu odhadu (SE – *standard error*) lze vypočítat ze standardních chyb odhadu parametrů času ($SE_{\beta_{tt}}$) a ceny ($SE_{\beta_{cost}}$) podle vzorce:

$$SE_{VoT} \cong \sqrt{\left(\frac{1}{\beta_{cost}} SE_{\beta_{tt}}\right)^2 + \left(\frac{\beta_{tt}}{\beta_{cost}^2} SE_{\beta_{cost}}\right)^2}$$

Tento vzorec platí pro chyby, které se limitně blíží nule, proto je rovnice ve skutečnosti pouze přibližná.

Různé kvantifikované parametry

Ochota platit za službu

Ochota platit za zmenšení rizika zpoždění

Ochota akceptovat přestup za zkrácení cestovního času

Ochota platit za zlepšení účinnosti léčby

Ochota akceptovat vedlejší příznaky za zlepšení účinnosti léčby

Hodnota času na trase Brno - Praha

$$VoT = \frac{\beta_{tt}}{\beta_{cost}}$$

$$SE_{VoT} \cong \sqrt{\left(\frac{1}{\beta_{cost}} SE_{\beta_{tt}}\right)^2 + \left(\frac{\beta_{tt}}{\beta_{cost}^2} SE_{\beta_{cost}}\right)^2}$$

	Value	SD
β_{cost}	-0.007	0.0017
β_{tt}	-1.75	0.152
VoT	246.48	62.779

Hodnota času cestujícího podle příručky indikátorů OPD je 163 Kč/h.

Elasticita

Bezrozměrná míra, která vyjadřuje vztah mezi procentuální změnou určité proměnné a procentuální změnou míry poptávky (je-li všechno ostatní stejné).

Direct-elasticity – vztah mezi změnou atributu a změnou poptávky po **příslušné** alternativě.

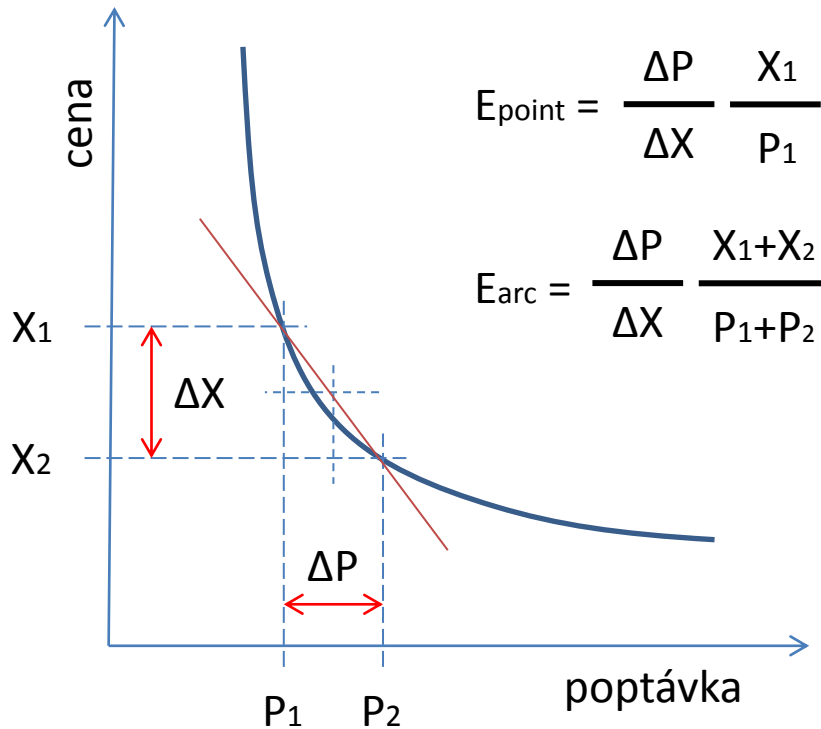
Jak se změní poptávka po MHD, pokud se jízdenka zlevní o 10 korun?

Cross-elasticity – vztah mezi změnou atributu a změnou poptávky po **konkurenční** alternativě.

Jak se změní poptávka po automobilové dopravě ve městě, pokud se jízdenka MHD zlevní o 10 korun?

Elasticita

Bezrozměrná míra, která vyjadřuje vztah mezi procentuální změnou určité proměnné X a procentuální změnou míry poptávky P (je-li všechno ostatní stejné).



Jak to vypočítat pomocí odhadnutého modelu?

Musíme znát podíly alternativ na poptávce (tzv. market share).

Direct-point elasticity:

$$E_a = \beta_{ai} x_{ai} (1 - P_a)$$

Cross-point elasticity:

$$E_{j \neq a} = -\beta_{ai} x_{ai} P_a$$

Elasticita – inkrementální model

$$P'_a = \frac{P_a \times e^{\Delta u_a}}{\sum_{i=1}^k (P_i \times e^{\Delta u_i})}$$

- P_a - současná pravděpodobnost alternativy a ,
- P'_a - revidovaná pravděpodobnost alternativy a ,
- Δu_a - změna utility alternativy a ,
- k - počet alternativ.

Když se v alternativě i nemění nezávislé proměnné, nemění se ani utilita.

Potom $\Delta u_i = 0$ a tím pádem $e^{\Delta u_i} = 1$.

Pokud se v alternativě i proměnné mění, změna utility Δu_i se vypočítá takto:

$$\Delta u_i = \sum_{j=1}^n \beta_j \Delta x_j$$

- Δx_j - změna v j -té proměnné,
- β_j - parametr j -té proměnné.

Inkrementální model nepočítá elasticity, ale přímo dopady na podíly alternativ.

Nevýhodou je práce s absolutními změnami atributů – lze přeformulovat.

Inkrementální model volby dopravního módu na trase Brno - Praha

Chmelík, J. (2015). Assessments of modal split in long-distance passenger transport. *Review of Economic Perspectives*, 15(1), 49-69.

Model stávajících módů

	AUTOMOBIL	AUTOBUS	VLAK	
P	68%	15%	17%	100%
β_{cost}	-0.001	-0.001	-0.001	
β_{tt}	-0.154	-0.154	-0.154	
cost [CZK]	500	150	200	
tt [h]	2.5	2.5	3	
cost [CZK]	500	150	200	
tt [h]	2.5	2.5	1.5	
$\Delta cost$	0	0	0	
Δtt	0	0	-1.5	
ΔU	0	0	0.231	
$exp(\Delta U)$	1.000	1.000	1.260	
$P * exp(\Delta U)$	0.680	0.150	0.214	
P'	65%	14%	21%	100%

VoT podle tohoto modelu je 154 (± 31) Kč/h.
To není úplně málo.