# SP experiment: Výběr studentského podnájmu

V Brně je velmi rozšířený trh s podnájmy studentského bydlení, protože je zde velké množství studentů a kapacity kolejí jsou naprosto nedostatečné. Charakter nabízených podnájmů je různý a také způsob jejich popisu v inzerátech se liší. S frekventanty kurzu analýzy diskrétních voleb na FSS MU jsme se rozhodli prozkoumat preference studentů, kteří si hledají podnájem. Kromě ceny se měl experiment zaměřit také na další aspekty bydlení, které jsou pro studenty důležité.

## Příprava experimentu

Před započetím přípravy experimentu proběhl rozhovor mezi samotnými výzkumníky, kterými byli studenti psychologie na FSS MU. Tito studenti by mohli patřit mezi osoby, které studentský podnájem skutečně hledají. Cílem rozhovoru bylo nalézt další relevantní aspekty volby podnájmu studentského bydlení, které by stálo zato do experimentu zahrnout.

Skupina vybrala tyto atributy a k nim příslušící úrovně:

* Měsíční cena podnájmu na osobu – 2200-2700 KČ, 4800-5300 KČ a 7400-7200 KČ
* Počet osob na pokoji – 1, 2 a 3 osoby na pokoji
* Zařízení pokoje – nezařízení, zařízený, nově zařízený
* Vzdálenost od školy – do 10 minut, 30 minut

V rámci každé úrovně ceny podnájmu hodnota náhodně variuje v rozsahu 500 KČ, aby se v jednom scénáři nevyskytovaly zcela stejné ceny, což by vypadalo divně.

Skupina dále rozhodla, že se jeden scénář bude skládat ze čtyř alternativ a každému respondentovi se budou administrovat tři různé scénáře. Jako nástroj pro webovou administraci dotazníku byl zvolen software Qualtrics.



Z počtu zamýšlených atributů a jejich úrovní vyplývá, že všech kombinací je 3x3x3x2 = 54. Protože těchto kombinací je málo a protože nebylo dopředu známo, kolik respondentů se podaří získat, rozhodli se výzkumníci, že použijí plně faktoriální design.

## Výběrový soubor

Celkem bylo získáno 234 pozorování od 78 respondentů. Respondenti se rekrutovali nahodilým výběrem z facebookových skupin v období duben až květen 2018. Základní sociodemografické údaje výběrového souboru jsou uvedeny v tabulce \*\*\*.

|  |  |  |
| --- | --- | --- |
|  | n | % |
| Celkem | 78 | 100 |
| Pohlaví | | |
| muži | 26 | 33 |
| ženy | 52 | 67 |
| Dosažené vzdělání | | |
| základní | 0 | 0 |
| střední odborné | 0 | 0 |
| střední s maturitou | 33 | 42 |
| bakalářské | 35 | 45 |
| magisterské/inženýrské | 5 | 6 |
| doktorské | 5 | 6 |
| Zaměstnání (alespoň 0.2 úvazku) | | |
| ano | 72 | 92 |
| ne | 6 | 8 |

Ambicí výzkumníků bylo zjišťovat také věk respondentů, to se ale z technických důvodů nepodařilo.

## Příprava dat

Tabulka dat je ve formátu, vhodném po analýzu v programu *Biogeme*. Na každém řádku je jedno pozorování. Je zde číslo respondenta (*id*), údaje o respondentově pohlaví, vzdělání a zaměstnání (*gen*, *edu*, *zam*), číslo pokusu (*trial*). Dále následují proměnné jednotlivých alternativ, což je počet osob na pokoji (*posob*), zařízení pokoje (*zariz*), vzdálenost privátu od školy (*vzdal*) a výše měsíčního nájmu na osobu (*najem*). Proměnné pro jednotlivé alternativy jsou označeny číslem alternativy za podtržítkem. Po té následuje sloupec s číslem vybrané alternativy (*Choice*). Nakonec je proměnná (*ok*), která označuje řádky, dostatečně kompletní pro další analýzu.

Pro analýzu v prostředí *R* je data potřeba upravit. Úprava původní tabulky probíhá ve dvou krocích. Tabulka je načtena a odstraněny jsou nekompletní řádky (proměnná *ok*). Proměnné *zariz* a *posob* jsou nominální proměnné a proto jsou převedeny na faktory.

d=read.table("data\_privat.dat",header = T)

d = d[d$ok==1,]

d$zariz\_1 = as.factor(d$zariz\_1)

d$zariz\_2 = as.factor(d$zariz\_2)

d$zariz\_3 = as.factor(d$zariz\_3)

d$zariz\_4 = as.factor(d$zariz\_4)

d$posob\_1 = as.factor(d$posob\_1)

d$posob\_2 = as.factor(d$posob\_2)

d$posob\_3 = as.factor(d$posob\_3)

d$posob\_4 = as.factor(d$posob\_4)

Samotná analýza probíhá pomocí knihovny *mlogit* (Croissant, 2011, 2012). Knihovna se načte a data se připraví do tvaru pro modelování.

library("mlogit")

md = mlogit.data(d, "Choice", shape="wide", varying=7:22, sep="\_", id.var="id")

Příkazu *mlogit.data* se určí, s jakou tabulkou dat se bude pracovat, jaké jméno má proměnná, která obsahuje volbu respondenta, jaký typ matice má vzniknout. Dále jaké sloupce (podle jejich indexů v tabulce) představují proměnné jednotlivých alternativ a jak jsou separována označení alternativy v názvech těchto proměnných. Nakonec zadáme proměnnou, která určuje číslo respondenta. Příkaz má samozřejmě ještě celou řadu dalších nastavení. V proměnné *md* jsou nyní data v podobě, vhodné pro analýzu.

## Model 1

Model se skládá z jedné užitkové funkce, která je stejná pro všechny alternativy. Alternativy nejsou pojmenované, a proto nemají žádnou alternativně specifickou složku[[1]](#footnote-1). Vzhledem k malému počtu respondentů nejsou do modelu zahrnuty rovněž sociodemografické údaje, které slouží k popisu výběrového souboru. Pro alternativu by mohla deterministická část vypadat takto:

V prostředí *R* vypadá zadání modelu následovně. Nula v rovnici znamená, že se v modelu nemají použít alternativně specifické konstanty.

m1 = mlogit(Choice ~ najem + posob + vzdal + zariz + 0, md)

summary(m1)

Výpis modelu obsahuje především tabulku odhadů parametrů modelu a jejich statistickou významnost od nuly. Pro kategoriální proměnné (*posob*, *zariz*) se odhadují parametry pro jednotlivé hodnoty proměnných.

Call:

mlogit(formula = Choice ~ najem + posob + vzdal + zariz + 0,

data = md, method = "nr", print.level = 0)

Frequencies of alternatives:

1 2 3 4

0.23077 0.25214 0.27778 0.23932

nr method

6 iterations, 0h:0m:0s

g'(-H)^-1g = 0.000376

successive function values within tolerance limits

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

najem -0.00082395 0.00008106 -10.1647 < 2.2e-16 \*\*\*

posob2 -1.86536549 0.26248559 -7.1065 1.190e-12 \*\*\*

posob3 -3.69928484 0.37332388 -9.9090 < 2.2e-16 \*\*\*

vzdal -0.05080133 0.01118997 -4.5399 5.628e-06 \*\*\*

zariz1 -0.19582217 0.26253155 -0.7459 0.45573

zariz2 0.47133194 0.24434014 1.9290 0.05373 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -153.69

Nespočítá se ovšem McFaddenovo R2, to se vypíše pro modely s alternativně specifickými konstantami. Výpis ale vrátí lig-likelihood modelu, který nám pro výpočet R2 stačí. Musíme spočítat log-likelihood nulového modelu, což je jednoduše logaritmus pravděpodobnosti výběru kterékoli alternativy (jedna čtvrtina), násobený počtem pozorování. Log-likelihood nulového modelu vychází na -324,39.

LL0 = log(1/4)\*nrow(d)

cat("R2(m1) = ",1-as.numeric(m1$logLik)/LL0,"\n")

cat("AIC(m1)= ",2\*length(m1$coefficients)-as.numeric(m1$logLik),"\n")

Log-likelihood modelu vychází na -153,69. R2 pak vychází na 0,526. Zároveň můžeme vypočítat například Akaikovo informační kritérium, které pro tento model vychází na 165,7.

Model předpokládá, že chyby modelu budou mezi pozorováními na sobě nezávislé. My ale víme, že každému respondentovi byly předloženy tři scénáře. Individuální rozdíly mezi respondenty mohou vstupovat do rozdílů v percepci jednotlivých atributů (jejich váhy při rozhodování). Tyto rozdíly v modelu vstupují do chyby modelu. Chyby všech pozorování pro jednoho respondenta tedy mohou být na sobě závislá. Vzniká tak riziko špatného odhadu parametrů modelu, což lze odstranit použitím smíšeného multinomiálního modelu pro panelová data.

## Model 2

Smíšený model pro panelová data znamená, že se budou odhadovat některé parametry s ohledem na každého respondenta. Bude se předpokládat, že tyto odhadované parametry modelu budou variovat mezi respondenty v rámci nějakého typu rozdělení a s určitou variancí. Tuto varianci bude model odhadovat. Které parametry budou v modelu vystupovat jako náhodné, je třeba vyzkoušet. V našem případě se ukázalo, že pro model je nejvýhodnější jako náhodné uvažovat parametry pro nájem a počty osob.

m2=mlogit(Choice ~ najem + posob + vzdal + zariz + 0, md, panel=T, rpar=c(najem="n",posob2="n",posob3="n"), R=1000, start=c(c(m1$coefficients),0,0,0))

summary(m2)

Zadání modelu se nyní změnilo. Je určeno, že se má jednat o panelová data a každému odhadovanému parametru je přidělen typ rozdělení, ve kterém má variovat mezi respondenty. Označení „n“ znamená normální rozdělení, ale k dispozici jsou i další, se kterými lze experimentovat. Dále se zadává počet simulací náhodných proměnných (v tomto případě 1000). Čím je počet simulací větší, tím lépe jsou parametry náhodných proměnných odhadovány, tím ale déle trvá odhad modelu. Nakonec je třeba zadat výchozí velikosti odhadovaných parametrů. Bez nich model často není schopen najít řešení. Průměry parametrů jsou vzaty z odhadu předchozího modelu a směrodatné odchylky jsou nastaveny na nulu.

Call:

mlogit(formula = Choice ~ najem + posob + vzdal + zariz + 0,

data = md, start = c(c(m1$coefficients), 0, 0, 0), rpar = c(najem = "n",

posob3 = "n", posob2 = "n"), R = 1000, panel = T)

Frequencies of alternatives:

1 2 3 4

0.23077 0.25214 0.27778 0.23932

bfgs method

28 iterations, 0h:0m:18s

g'(-H)^-1g = 7.39E-07

gradient close to zero

Coefficients :

Estimate Std. Error t-value Pr(>|t|)

najem -0.00194314 0.00049790 -3.9027 9.514e-05 \*\*\*

posob2 -3.63790963 0.97069034 -3.7478 0.0001784 \*\*\*

posob3 -8.53685617 2.07048992 -4.1231 3.738e-05 \*\*\*

vzdal -0.10525769 0.02999990 -3.5086 0.0004505 \*\*\*

zariz1 -0.69942864 0.48634812 -1.4381 0.1503990

zariz2 1.18317496 0.51317117 2.3056 0.0211322 \*

sd.najem 0.00090143 0.00027031 3.3348 0.0008537 \*\*\*

sd.posob2 3.07778107 1.09653424 2.8068 0.0050032 \*\*

sd.posob3 -4.00494770 1.36125473 -2.9421 0.0032599 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -132.27

random coefficients

Min. 1st Qu. Median Mean 3rd Qu. Max.

najem -Inf -0.002551148 -0.001943142 -0.001943142 -0.001335136 Inf

posob3 -Inf -11.238152347 -8.536856171 -8.536856171 -5.835559994 Inf

posob2 -Inf -5.713841417 -3.637909633 -3.637909633 -1.561977848 Inf

Výpis nyní ukazuje navíc odhady směrodatných odchylek pro náhodné parametry a také popis jejich rozdělení. Log-likelihood modelu stoupl na -132,27 a R2 je nyní 0,592. Akaikovo informační kritérium pro tento model vychází na 150,3. Je tedy u tohoto modelu menší ztráta informace oproti modelu verze 1.

Odhady parametrů smíšeného modelu jsou v absolutní hodnotě větší, než u první verze modelu. To je proto, že část chyby modelu je nyní „vysvětlena“ náhodnými parametry. Jestli jsou oba modely skutečně podstatně rozdílné, uvidíme při kvantifikaci parametrů.

## Kvantifikace parametrů

V modelech používáme cenu měsíčního pronájmu za jednu osobu. Vůči parametru této proměnné budeme posuzovat ostatní parametry. Pro výpočet chyby odhadu budeme kromě odhadu parametrů také potřebovat standardní chyby odhadu. Ty jsou uvedeny ve výpisu v téže tabulce. Odhad chyby kvantifikovaných parametrů můžeme dále zpřesnit započtením kovariancí mezi standardními chybami parametrů. To se pro udělá příkazem *vcov*. Z výsledné kovarianční tabulky nás budou zajímat pouze kovariance s proměnnou *najem*. Pro druhý model vypadá příkaz následovně:

vcov(m2,"coefficient",type=c("cov"))

Výsledky pro oba modely s vypočtenými kvantifikacemi parametrů (ochotou platit nebo ochotou akceptovat) jsou shrnuty v tabulce. Na první pohled je vidět, že zatímco odhadnuté parametry se u obou modelů dosti liší, odhad kvantifikace parametrů je v rámci chyby srovnatelný.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | | | Model 2 | | | | |
|  | parametr | std. chyba | kovariance | kvantifikace | | parametr | std. chyba | kovariance | kvantifikace | |
|  | odhad | chyba | odhad | chyba |
| najem | -0.00082 | 8.11E-05 | 6.57E-09 | - | - | -0.00194 | 0.000498 | 2.48E-07 | - | - |
| posob2 | -1.86537 | 0.262486 | 9.21E-06 | 2264 | 299 | -3.63791 | 0.97069 | 4.19E-04 | 1872 | 254 |
| posob3 | -3.69928 | 0.373324 | 1.63E-05 | 4490 | 430 | -8.53686 | 2.07049 | 9.17E-04 | 4393 | 518 |
| vzdal | -0.0508 | 0.01119 | 2.36E-07 | 62 | 13 | -0.10526 | 0.03 | 1.06E-05 | 54 | 11 |
| zariz1 | -0.19582 | 0.262532 | 1.40E-06 | 238 | 318 | -0.69943 | 0.486348 | 8.82E-05 | 360 | 233 |
| zariz2 | 0.471332 | 0.24434 | -2.60E-06 | -572 | 294 | 1.183175 | 0.513171 | -1.27E-04 | -609 | 231 |

Přesto budeme interpretovat smíšený model (model 2), protože je věcně správnější a lépe odpovídá datům.

Největší vliv na rozhodování respondentů má z nefinančních atributů počet osob na pokoji. Respondenti by byli ochotni akceptovat jednoho spolubydlícího za snížení nájmu o 1872 Kč měsíčně. Pro dva spolubydlící je to 4393 Kč měsíčně, což je o něco více, než dvojnásobek. Nové zařízení podnájmu respondenti cení kladně a jsou za ně ochotni zaplatit 609 Kč měsíčně proti nezařízenému pokoji. Naopak, pokud by pokoj byl zařízený starým nábytkem, budou to respondenti vnímat přibližně stejně, jako nezařízený pokoj. Rozdíl je statisticky nevýznamný. Vzdálenost od školy hraje překvapivě významnou roli. Proměnná vzdálenost je dána v minutách a proto vychází, že každé minuty potřebné k cestě do školy si respondenti cení na 54 Kč měsíčně. V experimentu byly použity dvě úrovně tohoto atributu, a to 10 minut a 30 minut. Vysokou hodnotu času, stráveného na cestě do školy lze přičíst skutečnosti, že zatímco při desetiminutové cestě do školy půjde respondent pěšky, pro cestu, která trvá 30 minut, bude patrně nucen využít veřejnou dopravu nebo nějaký individuální dopravní prostředek. Studenti jsou ochotni platit za bydlení vzdálené 10 minut od školy proti bydlení 30 minut od školy navíc 1080 Kč měsíčně. Navíc, kdyby student absolvoval cestu do školy 2x za školní den, a ten byl 20x za měsíc, znamenala by jedna minuta vzdálenosti 40 minut cestovního času měsíčně. To je hodnota cestovního času 1,35 Kč za minutu nebo 81 Kč za hodinu. Pěší VoT u nás. V této perspektivě se zdá částka na minutu vzdálenosti od školy naopak jako nízká.

## Závěr

Experiment na vyjádřené preference při volbě studentského podnájmu proběhl na vzorku 78 respondentů vybraných nahodilým výběrem. Z výsledků vyplývá, že studenti jsou ochotni akceptovat spolubydlícího za slevu 1872 Kč na měsíčním nájmu. Dále jsou ochotni platit za nově zařízený pokoj o 609 Kč měsíčně navíc proti nezařízenému pokoji, ale i pokoji zařízenému starým nábytkem. Dále jsou ochotni platit za bydlení vzdálené 10 minut od školy proti bydlení 30 minut od školy 1080 Kč měsíčně navíc.

1. To, zda pořadí alternativy nemá vliv na pravděpodobnost výběru lze testovat přidáním alternativně specifické konstanty. Tyto konstanty však pro naše data nejsou statisticky významně rozdílné od nuly, a proto nejsou do modelu zahrnuty. [↑](#footnote-ref-1)