

РОЗДІЛ III

Методологія та методи соціологічних досліджень

УДК 303.68, 303.42

Микола Сидоров

Практична реалізація факторіального дизайну засобами R

У статті розглянуто методичні та практичні особливості реалізації експериментального дизайну в опитуванні. Основну увагу зосереджено на реалізацію методу в середовищі R, поетапне конструювання запитальника, формування масиву віньєток, ефективної вибірки, обробки й презентації результату. Дизайн апробовано на експериментальному дослідженні факторів впливу на вибір студентами першого робочого місця, що представлено у вигляді прикладу реалізації методу в R.

Ключові слова: мова статистичного програмування R, метадані, експериментальний дизайн в опитуванні, віньєтка, побудова запитальника, D-ефективність вибірки.

Постановка наукової проблеми та її значення. Проведення формалізованих анкетних опитувань вже давно стало звичним інструментом збору соціологічної інформації. Використання багатовимірних методів аналізу, таких як регресійний, факторний та інші, є звичною практикою в соціології. Проведення соціологічних експериментів – також звична, хоч і досить складна методика збору соціологічної інформації. Факторіальний, або експериментальний, дизайн дає змогу поєднати експеримент із масовим опитуванням, перевіряти поведінку респондента в експериментальних ситуаціях за допомогою анкетування, а також використовувати комплексну взаємодію параметрів експерименту для статистичного аналізу. Таку практику частково застосовують у психологічних дослідженнях, орієнтованих на порівняно невелику вибірку, тоді як у соціології її використання та інтерпретація результатів ще не увійшли до звичних практик. Відповідно, можливість використання цього методу ще не повністю реалізована в деяких поширених комерційних пакетах, таких як SPSS і Statistica. Зате в пакеті Stata можна як готувати дані для факторіального дизайну, так і аналізувати їх. Проблема полягає в тому, що Stata, як і SPSS та Statistica, є комерційним продуктом, застосування якого вимагає придбання ліцензії. Крім того, обмаль достатньо детальних інструкцій та вказівок для проведення таких досліджень. Альтернативою може бути безкоштовний пакет R, який дає змогу використовувати факторіальний дизайн, про що й ітиметься в цій роботі. Цей підхід достатньо новий у вітчизняній практиці, а комплексна оцінка факторів впливу на вибір робочого місця взагалі ще не здійснювалась.

Мета статті – показати, як практично реалізувати алгоритм факторіального дизайну в середовищі статистичного програмування R від створення анкети з віньєтками до отримання регресійної моделі на прикладі дослідження вибору першого робочого місця магістрами 1–2 років навчання факультету соціології Київського національного університету ім. Тараса Шевченка.

Завдання наукової праці – на основі виділених у [7, 8, 9] та інших дослідженнях базових факторів, що впливають на вибір місця роботи, продемонструвати, (1) як можна виявити взаємозв'язок пріоритетності факторів для різних категорій студентів; (2) визначити їх значущість за регресійним рівнянням, побудованим на цих факторах за допомогою застосування факторіального дизайну.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Методика факторіального дослідження (або експериментального, або «дослідження з використанням віньєток» («vignette survey»)) передбачає наявність у запитальнику одного або низки специфічних запитань, які й

називають «віньєтками» й у яких сформульовано певну ситуацію, щодо якої респонденту потрібно визначитися. У нашій роботі ми використали таке питання.

Уявіть, що Ви отримали пропозицію: *робота за фахом в українській приватній компанії з можливістю професійного та особистісного розвитку на робочому місці, без можливості кар'єрного зростання. Вам запропонували офіційне працевлаштування та часткову зайнятість за фіксованим графіком із можливістю роботи за межами офісу та зарплату від 4000 до 5000 грн.*

Чи можете Ви собі уявити себе працівником цієї компанії? *Позначте свою відповідь на шкалі від -3 до +3, де -3 – «Дуже малоймовірно», а +3 – «Дуже ймовірно».*

Отже, маємо вісім факторів-вимірів (dimensions), у кожному з яких є визначена кількість рівнів (levels) (або альтернатив): 1) *два* рівні виміру ознаки «робота за фахом»; 2) *три* – «характер власності та масштаб діяльності компанії»; 3) *два* – «можливості професійного та особистісного розвитку на робочому місці»; 4) *два* – «кар'єрні перспективи»; 5) *два* – «тип працевлаштування»; 6) *три* – «графік роботи»; 7) *два* – «прив'язка до робочого місця»; 8) *п'ять* – «заробітна платня».

Усього можна створити $2 \times 3 \times 2 \times 2 \times 2 \times 3 \times 2 \times 5 = 1440$ різних віньєток із формулюваннями ситуацій, які відрізняють за рівнем хоча б одного виміру. Таку множину називають генеральною сукупністю віньєток, або повним простором віньєток. Якщо в ній наявні такі, які мають безглузді сполучення рівнів, наприклад тип роботи – волонтер, а зарплата – 5000 грн, або професор без вищої освіти, то такі віньєтки примусово видаляються з цього простору.

Щоб повністю й усебічно оцінити взаємозв'язок цих вимірів серед усієї вибірки респондентів, потрібно кожному поставити 1440 запитань, а це досить безглузда справа. Тому існує низка рекомендацій щодо створення запитань такого типу, ці рекомендації стосуються балансу між кількістю вимірів і рівнів у запитанні та кількістю запитань в анкеті.

Перша частина таких рекомендацій стосується організації самої анкети, тобто оптимальної кількості віньєток в анкеті. Тут під оптимальністю маємо на увазі таку максимальну кількість віньєток, за якої буде найменша кількість відмов або випадків автоматичного заповнення. К. Аушпург [2; 1] і Т. Хінц [3; 1] проаналізували як оптимальну кількість віньєток в анкеті, так і парадані, що стосуються часових характеристик проведення опитування, і з'ясували, що оптимальна кількість віньєток в анкеті становить не більше 10. Такою вона була в анкеті нашого дослідження. Блок віньєток у кожній анкеті повинен складатися з однакової кількості хай навіть різних наборів віньєток – сетів. У нашому дослідженні сет складався якраз із 10 віньєток.

Подальші рекомендації пов'язані з тим, як формувати самі сеті, тобто чи повинні повторюватись однакові запитання в різних сетах, як відібрати запитання в сеті, щоб мінімізувати зсуви через віньєтки, що не потрапили до жодного сету, як максимізувати варіативність вимірів у сетах тощо. Для цього існує низка підходів, які розділяють на випадкові й не випадкові (систематичні)¹ [6].

Найбільш оптимальним серед порівняно простих підходів вважають систематичний підхід на основі D-ефективного відбору [1], але він вимагає додаткових затрат часу та наявності відповідного програмного забезпечення. Оскільки для аналізу даних використано оболонку R [4], то в ньому довелося для формування D-ефективної вибірки скористатися бібліотекою AlgDesign, у якій представлено реалізацію алгоритму Федерова у функції optFederov. Ми проводили суцільне опитування, тому планували обсяг респондентів у 67 осіб, але з різних причин (відмови, недоступність) отримали лише 50 повністю заповнених анкет. У кожній із них серед інших запитань ми планували розмістити по 10 віньєток, тому їх вибірка з повного простору становила 670 одиниць. У функції optFederov показником ортогональності виступає параметр D, а нижня границя довірчого інтервалу для D-ефективності є Dea. Якщо вона дорівнює 1, то і сам параметр D-ефективності дорівнює 1, а вибірка є ортогональною для цього простору віньєток й умов їх вибірки. Отже, потрібно з повного простору відібрати 670 віньєток так, щоб показник D-ефективності був якомога ближчим до 1.

Чим більше варіантів розміщень віньєток ми розглядаємо, тим більша ймовірність того, що натрапимо на найбільш ортогональну вибірку заданого обсягу, але загальна кількість таких варіантів

¹ Випадкові: простий випадковий відбір із заміною, простий випадковий відбір без заміни, стратифікований випадковий і кластеризований відбори. Невипадкові: фракційний факторіальний («fractional factorial sampling») та D-ефективний («D-efficiency») відбори, збалансоване перекриття («Balanced Overlap»), теоретична вибірка.

становить $C_{1440}^{670} = \frac{1440!}{(1440 - 670)!670!} \approx e^{990,7972}$, тобто досить велике число. Тож підключаємо пакет

AlgDesign [5] library(AlgDesign).

Спочатку будемо повний простір віньеток та запишемо його в змінну data:

```
data<-gen.factorial (c (2, 3, 2, 2, 2, 3, 2, 5), factors="all", varNames=c ("d1", "d2", "d3", "d4", "d5", "d6", "d7", "d8")),
```

де c (2, 3, 2, 2, 2, 3, 2, 5) – послідовність рівнів вимірів, а varNames – набір міток вимірів. Створюємо послідовність номерів віньеток

```
vnum<-c(1:nrow(data)),
```

та вказуємо першою ознакою номер віньетки, решта – виміри (d1 ... d8)

```
VgnSpace<-data.frame(vnum,data).
```

Згенерований у такий спосіб масив повного простору віньеток складається з дев'яти ознак: 1-ша ознака – номер віньетки та 8-ми вимірів – містить 1440 рядків.

Далі, використовуючи функцію optFederov, спробуємо відібрати найкращу за критерієм D-efficiency вибірку віньеток. Значення D-efficiency дорівнює 1 не обов'язково на повному просторі віньеток – для нього часто досить меншої кількості. Запишемо таку вибірку до змінної OptSample, але спочатку зафіксуємо генератор випадкового вибору початкової віньетки:

```
set.seed(654354).
```

Цю команду використовуємо для того, щоб «зафіксувати» генератор, адже відбір віньеток функцією optFederov для вибірки з повного простору починається з певної віньетки, обраної випадковим чином. При фіксованому початковому значенні seed алгоритм завжди починатиме відбір з тієї самої віньетки. Це робиться для повторюваності експерименту:

```
OptSample<- optFederov(frml=~., data=VgnSpace[,2:9], nTrials= vgnum, criterion="D", approximate=TRUE),
```

де frml – формула регресійного рівняння, яке братиме участь в аналізі (у нашому випадку ми застосували лінійну множинну регресію); VgnSpace[,2:9] – повний простір віньеток, тобто масив усіх можливих і допустимих комбінацій усіх рівнів усіх вимірів (стовпчики з 2 по 9 – 8 вимірів, у кожній клітинці – значення рівня відповідного виміру, кожен рядок – набір значень рівнів відповідних вимірів однієї віньетки, усього 1440 рядків-віньеток); nTrials – обсяг вибірки віньеток (у нашому випадку маємо вибірку з vgnum=670 віньеток); criterion="D" означає, що розраховується саме D-ефективна вибірка; approximate=TRUE – обчислюється оптимальна з погляду алгоритму Федерова чисельність вибірки, що не перевищує nTrials і дає максимальне з можливих значень D.

У нашому випадку оптимальний обсяг вибірки склав 417 віньеток (хоча планувалося 670), тому розглянемо генерування оптимальної вибірки фіксованого обсягу, а параметр OptSample\$D цього оптимального розміщення буде використаний для підрахунку значення D-efficiency.

Щоб отримати вибірку точно з 670 віньеток, потрібно виключити з рядка запиту approximate=TRUE, тобто Dsample<-optFederov(data= VgnSpace [,2:9],nTrials=vgnum,criterion="D").

У цьому запиті можна не вказувати кількість ітерацій, поки відбуватиметься формування вибірки, але тоді автоматично їх кількість становитиме 100. Для збільшення кількості ітерацій використовують параметри maxIteration, а щоб «прогнати» цю процедуру кілька разів – nRepeats. Ці параметри підбирають уручну для забезпечення найкращої ортогональності, але судячи з [1, с. 34], починаючи з певного моменту, збільшення кількості nRepeats та maxIteration не приводить до покращення вибірки. У нашому випадку застосовано

```
set.seed(654354)
```

```
Dsample<-optFederov(data= VgnSpace[,2:9], nTrials=vgnum,criterion="D", maxIteration=1000, nRepeats=100)
```

з виводом результатів через команду

```
data.frame(Dsample[1:4])
```

D	A	Ge	Dea
10.2142177	6.998639	0.994	0.994.

Отже, нижньою границею довірчого інтервалу показника D-ефективності для отриманої вибірки віньєток є 0.994, що вже саме по собі добре, адже це означає, що істинне значення показника перевищує 0.994². Для обчислення D-ефективності застосовують формулу

$$D - efficiency = \frac{D_{sample} \$ D}{OptSample \$ D},$$

тобто оцінюємо ефективність стосовно найоптимальнішого випадку з можливих для цього набору простору віньєток. Рівень ортогональності в нашому випадку сягнув $D-eficiency=0,9999759$, що є надзвичайно високим показником.

У конструкті Dsample серед інших ознак є й сама вибіркова модель – вона розміщена в змінній Dsample\$design, а множина всіх рядків із номерами віньєток – у змінній Dsample\$rows.

Далі за цією вибіркою ми сформуваємо масив запитань-віньєток та записали його в змінну Vsample

```
Vsample<-data.frame(Dsample$rows, Dsample$design).
```

Дані віньєтки потрібно також розподілити за сетами й указати їхній порядковий номер у сеті. Оскільки для нашого дослідження не важливо, які з віньєток входять до відповідного сету, то головне, щоб кожна з них увійшла до якогось сету, то послідовно призначимо кожній віньєтці сет і послідовний номер у ньому. Для цього скористаємося командами:

```
vset<-rep(1,67)
for (i in 2:10) vset<-c(vset,rep(i,67))
Vsample<-data.frame(c(1:67),vset, Vsample).
```

Розмірність цього масиву є 670x11, тобто 670 рядків (віньєток), кожен із яких складається з номера сету, послідовного номера віньєтки в сеті, номера віньєтки в повному просторі та 8-ми вимірів.

Тепер потрібно сформувати самі запитання-віньєтки, які фігуруватимуть в анкеті. Для цього створимо допоміжний масив survey, що на цьому етапі має такий самий уміст, що й Vsample:

```
Survey<-Vsample.
```

Записуємо компоненти тексту віньєтки та повні формулювання альтернатив

```
q0_0<-"Робота"
q1<-c ("за фахом", "не за фахом")
q1_0<-"в"
q2<-c ("українській державній", "українській приватній", "міжнародній")
q2_0<-"компанії,"
q3<-c ("з можливістю", "без можливості")
q3_0<-"професійного та особистісного розвитку на робочому місці,"
q4<-c ("без можливості", "з можливістю")
q4_0<-"кар'єрного зростання. Вам запропонували"
q5<-c ("офіційне", "неофіційне")
q5_0<-"працевлаштування та"
q6<-c ("повний робочий день", "часткову зайнятість за фіксованим графіком", "вільний графік")
q6_0<-"
q7<-c ("роботи в офісі", "з можливістю роботи за межами офісу")
q7_0<-"та зарплату"
q8<-c ("до 2000 грн", "від 2000 до 3000 грн", "від 3000 до 4000 грн", "від 4000 до 5000 грн", "понад 5000 грн")
q8_0<-"
Перезаписуємо скорочені альтернативи повними:
```

```
vgnlevels<-list(q1,q2,q3,q4,q5,q6,q7,q8)
for (i in 1:8) Survey[,11+i]<-vgnlevels[[i]][as.numeric(Survey[,i+3])].
```

Формуємо повні формулювання віньєток в анкеті та записуємо їх в останню ознаку масиву (автоматично ознаці призначається ім'я v20):

```
for (i in 1:670)
```

² Тут вважаємо, що D-ефективності набуває значення від 0 до 1, хоча дехто [1] це значення множить на 100.

```
Survey[i,20]<-
paste(q0_0,q1[Survey[i,4]],q1_0,q2[Survey[i,5]],q2_0,q3[Survey[i,6]],q3_0,q4[Survey[i,7]],q4_0,q5[Survey[i,8]
],q5_0,q6[Survey[i,9]],q6_0,q7[Survey[i,10]],q7_0,q8[Survey[i,11]],q8_0,sep=" ")
```

Після цього потрібно «підчистити» в повних фразах можливі повторення пропусків:

```
for (i in 1:670){
  Survey[i,20]<-gsub(" ", " ",Survey[i,20])
  Survey[i,20]<-gsub(" ", " ",Survey[i,20])
  Survey[i,20]<-gsub(" ", " ",Survey[i,20])
}
```

Остаточно ми записуємо цей масив у файл Survey.csv:

```
write.table(Survey, file = "Survey.csv", append = FALSE, quote = TRUE, sep = ";",eol = "\n", na = "NA",
dec = ".", row.names = F,col.names=TRUE, qmethod = c("escape", "double"), fileEncoding = ""),
```

щоб уже з нього сформувати самі анкети, що міститимуть по 1 сету віньєток.

Як зазначалося, опитано 50 респондентів. Результатом збору даних мав стати масив, кожен рядок якого відповідав би окремому респондентові й, крім звичайних 12 запитань та номера сету, складався б із 10 віньєток, кожна з яких була би записана у вигляді набору вимірів із запропонованими у відповідній віньєтці рівнями й відповіді на неї. Оскільки вимірів було по 8 у кожній із віньєток, кожен такий блок віньєтки мав би складатись із 8 значень вимірів конкретної віньєтки та значення відповіді респондента на цю віньєтку. Тобто всього масив складався з 50 рядків і 103 стовпців (12+1+10*(8+1)).

Проте масив даних, записаний у такій формі, не зручний як для введення інформації (потрібно або вручну вводити не лише відповіді на базовий блок запитань і віньєтки, а й усі відповідні значення рівнів вимірів, або ж частину масиву спершу слід згенерувати програмно, прописавши значення рівнів відповідних вимірів, а потім заповнювати), так і для аналізу та побудови регресійних рівнянь для факторіальних досліджень, адже змінні, що відповідають віньєткам, будуть «розкидані» по 10 блоках ознак.

Один із підходів до організації набору даних для цього дослідження – використання двох масивів одночасно: в одному прописано номери сетів, номери віньєток у сетах та згенеровано рівні вимірів вибірки віньєток (у нашому дослідженні це масив Dsample, який складається з 670 рядків й 11 стовпців: номер сету, номер віньєтки у сеті, номер запитання у повному просторі віньєток і значення рівнів восьми вимірів віньєтки, тобто цей масив є масивом метаданих). Інший масив складається з 50 рядків (усього взяло участь 50 респондентів, хоча «заготовка» робилася на 67) та 23 стовпців: 12 запитань базової частини, номер сету, 10 відповідей на віньєтки сету. Це власне і є масив даних, який досить легко заповнювати, уводячи відповіді з анкет. Таким масивом у нашому дослідженні є vigndata. Цей масив можна заповнити у будь-якій оболонці чи програмі (Excel, SPSS, Notepad) і далі завантажити в R³.

Отже, маємо два масиви: Vsample розміром 670x11 (вибірку віньєток із зазначенням номера сету, номера віньєтки в сеті, номера віньєтки в повному просторі та 8-ми рівнів) і vigndata розміром 50x23 (масив відповідей).

Далі для аналізу потрібно перетворити масив vigndata до вигляду, щоб у кожному рядку була загальна частина анкети й лише одна віньєтка, тобто якщо в сеті 10 віньєток, то кожен респондент з одного рядка (23 ознаки) трансформується в 10 рядків, які міститимуть однаковий запис відповідей на базову частину анкети та одну віньєтку. Таким чином ми отримаємо не 50 респондентів по 10 віньєток, а наче 500 по одній віньєтці. Саме такий формат і використовуватимемо в подальшому аналізі.

Отже, потрібно перетворити vigndata таким чином, щоб кожен рядок складався з відповідей на запитання базової частини, відповіді на одну віньєтку, її номера в сеті та номера сету, тобто розмірність має стати 500x15. Таке перетворення можна зробити за допомогою функції reshape бібліотеки reshape за допомогою запиту:

```
library(reshape)
```

³ Якщо масив даних створювався в SPSS, то прочитати його в R можна за допомогою команд:

```
library(foreign)
vigndata<-read.spss("vignette.sav",to.data.frame=T).
```

Команда vigndata<- vigndata[(!is.na(vigndata\$setnum)),] відсіює анкети, у яких не вказано номер сету, оскільки неможливо ідентифікувати склад віньєток і, відповідно, зіставити запитання з відповідями. Приклад «заготовки» для масиву даних SPSS наведено в додатку 1.

```
longvignette <- reshape(vigndata,
  varying = c("b1", "b2", "b3", "b4", "b5", "b6", "b7", "b8", "b9", "b10"),
  v.names = "vignette", timevar = "qnum", times = c(1:10),
  new.row.names = 1:500, direction = "long"),
```

де b1,...,b10 – відповіді на віньєтки, v.names – ім'я змінної, де зберігається відповідь на віньєтку, timevar – змінна, у якій зберігається порядковий номер віньєтки у сеті, times – нумерація повторів, тобто в нашому випадку номер віньєтки в сеті, що зберігається в змінній timevar, new.row.names – нова нумерація рядків у масиві (оскільки в нашому масиві було 500 віньєток, то нумерацію встановили від 1 до 500), direction = "long" означає, що здійснюватиметься розбиття рядків.

Після перетворення масив даних longvignette складається з 500 спостережень і 23 ознак: номер сету (setnum), 19 запитань постійної частини анкети, номер запитання у віньєтці (qnum), відповідь на віньєтку (vignette) та порядковий номер респондента. Дані масиву longvignette можна розглядати не як результати опитування 50 осіб, котрі дали відповідь на 10 віньєток, а результати опитування 500 осіб, кожна із яких дала відповідь на одну віньєтку.

Далі потрібно поставити у відповідність номеру віньєтки з масиву longvignette сам їх уміст, тобто певним чином об'єднати масиви longvignette і Vsample. Але спочатку слід «розшифрувати» самі значення рівнів вимірів віньєток у масиві Vsample. Для цього сформуємо перелік вимірів з альтернативами та запишемо його в vgnlevels:

```
dim1<-c ("за фахом", "не за фахом")
dim2<-c ("укр. держ", "укр. приват.", "міжнар.")
dim3<-c ("з профес. зрост.", "без профес. зрост")
dim4<-c ("без можливості ", "з можливістю")
dim4<-c ("без кар'єрн. зрост.", "з кар'єрн. зрост.")
dim5<-c ("офіційне", "неофіційне")
dim6<-c ("повний роб. день", "частк. зайн. за фікс. графік.", "вільний графік")
dim7<-c ("робота в офісі", "можливо за межами офісу")
dim8<-c ("до 2000 грн", "від 2000 до 3000 грн", "від 3000 до 4000 грн", "від 4000 до 5000 грн",
"більше 5000 грн")
vgnlevels<-list(dim1,dim2,dim3,dim4,dim5,dim6,dim7,dim8).
```

Зазначимо, що альтернативи тут дещо відрізняються від альтернатив у масиві survey через те, що для survey потрібні повні їх формулювання, а для Vsample краще використовувати скорочені, бо вони займатимуть менше місця при побудові діаграм.

Наступним кроком ми допишемо до масиву Vsample стовпці з «розшифрованими» значеннями вимірів:

```
for (i in 1:8) {
  Vsample<-cbind(Vsample,vgnlevels[[i]][as.numeric(Vsample[,i+3])])
  names(Vsample)[i+11]<-paste("I",i)}.
```

Остаточо отримаємо вибірку віньєток у масиві Vsample з 19 змінними: номер сету, номер віньєтки в сеті, номер віньєтки в повному просторі віньєток, восьми рівнів (числові значення) та восьми рівнів (самі альтернативи).

Наступний крок – доповнення масиву даних longvignette метаданими з Vsample:

```
numr<-dim(longvignette) [1]
numc<-dim(longvignette) [2].
```

Ця частина коду не є оптимальною, але її найлегше пояснити. Numr – кількість рядків, numc – кількість стовпців масиву longvignette. Цикл:

```
for (i in (1:numr))
  for (j in (1:8))
    longvignette[i,numc+j]<-Vsample[(Vsample$set==longvignette$setnum[i])&
(Vsample$vgset==longvignette$qnum[i]),11+j]
```

формує в масиві longvignette вісім додаткових стовпців, що відповідають значенням рівнів відповідних вимірів віньєток, беручи ці значення з масиву метаданих Vsample так, щоб номер сету й порядковий номер віньєтки в сеті збігався в обох масивах.

Для зручності перейменуємо змінні в longvignette:

```
names(longvignette)<-c("setnum", "Робота", "Досвід", "Спец", "ЗП", "ПланСпец", "Вік", "Стать",  
"ДжДоходу", "Матер", "ТипНасП", "Реґіон", "ДжФіанас", "qnum", "vignette", "id", "Фах.", "Тип.",  
"ПрофЗрост.", "КарЗрост.", "Працевл.", "Час", "Місце", "Зарпл.")
```

У результаті отримуємо масив `longvignette` розмірністю в 500 спостережень за 24 ознаками. За цим масивом уже можна будувати моделі множинної лінійної регресії. Для кращої візуалізації результатів регресійного аналізу можна скористатися пакетом `coefplot2`, який, хоча й не входить до базового переліку доступних пакетів, але є досить зручним:

```
install.packages("coefplot2",repos="http://www.math.mcmaster.ca/bolker/R", type="source")  
library(coefplot2).
```

Далі будується заявлена у формуванні дизайну D-ефективності регресійна модель (тут мається на увазі, що сама модель може бути не простою лінійною множинною, а містити незалежні змінні як у різних степенях, так і у вигляді добутків. Ми ж «замовляли» саме просту лінійну множинну):

```
m1<-  
lm(data=longvignette,vignette~Фах.+Тип.+ПрофЗрост.+КарЗрост.+Працевл.+Час.+Місце.+Зарпл.).
```

Замовляємо виведення результатів моделі: `summary(m1)`.

Висновки й перспективи подальших досліджень. Очевидно, що сучасному соціологу-аналітику потрібно бути на вістрі сучасних методів збору та аналізу даних. У зв'язку з тим, що нові методи й методики часто спрямовані на використання прихованих можливостей та нових шляхів збільшення якості результатів дослідження, їх коректне застосування дає можливість проводити точніший аналіз даних.

Таким у нашій роботі виступає метод факторіального дизайну, за допомогою якого з'являється можливість досліджувати фактори впливу в комплексі, а не поодиноці, як це було раніше у звичайних формалізованих опитувальниках. Це, зі свого боку, уможливило з'ясування важливості кожного з факторів, порівняно з іншими.

Можливість застосування цього підходу повністю реалізована засобами R. У статті показано, як поетапно здійснювати практичну реалізацію методу факторіального дизайну в статистичному середовищі R на прикладі дослідження чинників, що впливають на вибір робочого місця магістрами факультету соціології. Аналіз отриманих результатів є перспективою наступної наукової розвідки.

Джерела та література

1. Auspurg K. Factorial Survey Experiments / K. Auspurg, T. Hinz // Sage, Series: Qualitative Applications in the Social Sciences. – 2015. – № 175. – 143 p.
2. Auspurg K. First Equals Most Important? Order Effects in Vignette-Based Measurement. ISER Working Paper 2012-01 / K. Auspurg, A. Jäckle. – University of Essex : Institute for Social and Economic Research, 2012.
3. Hinz T. Complexity, Learning Effects, and Plausibility of Vignettes in Factorial Surveys [Electronic resource] / T. Hinz, K. Auspurg, S. Liebig // Paper accepted for the ASA-Conference 2009. – Mode of access : <http://kops.uni-konstanz.de/bitstream/handle/123456789/15067/Hinz%20etal.pdf?sequence=2>
4. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Electronic resource]. – 2015. – Vienna, Austria. – Mode of access : <http://www.R-project.org/>
5. Wheeler R. E. Comments on algorithmic design [Electronic resource] / R. E. Wheeler // CRAN documentation, 2009. – P. 59. – Mode of access : <https://cran.r-project.org/web/packages/AlgDesign/vignettes/AlgDesign.pdf>
6. Марченко А. М. Методичні особливості реалізації експериментального дизайну в опитуванні (на прикладі пілотного дослідження ролі ідеологічних чинників у дружніх стосунках) / А. М. Марченко, М. В.-С. Сидоров // Методологія, теорія та практика соціологічного аналізу сучасного суспільства. – 2014. – Вип. 20. – С. 116–122.
7. Працевлаштування студентів факультету соціології. Діяльність після закінчення навчання [Електронний ресурс]. – Режим доступу : <http://www.soc.univ.kiev.ua/uk/nauka/diyalnist-pislya-zakinchennya-navchannya>
8. Працевлаштування студентів факультету соціології. Пошук першої роботи [Електронний ресурс]. – Режим доступу : <http://www.soc.univ.kiev.ua/uk/nauka/poshuk-pershoyi-roboti>
9. Працевлаштування студентів факультету соціології. Фактори вибору робочого місця [Електронний ресурс]. – Режим доступу : <http://www.soc.univ.kiev.ua/uk/nauka/faktori-viboru-roboty>

References

1. Auspurg, K., Hinz, T. (2015), Factorial Survey Experiments, Sage, Series: Qualitative Applications in the Social Sciences, No. 175, 143 p.
2. Auspurg, K., Jäckle, A. (2012), First Equals Most Important? Order Effects in Vignette-Based Measurement. ISER Working Paper 2012-01, University of Essex: Institute for Social and Economic Research.

3. Hinz, T. Auspurg, K., Liebig, S. (2009), *Complexity, Learning Effects, and Plausibility of Vignettes in Factorial Surveys*, Paper accepted for the ASA-Conference 2009, <http://kops.uni-konstanz.de/bitstream/handle/123456789/15067/Hinz%20etal.pdf?sequence=2>
4. *R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing* (2015), Vienna, Austria, <http://www.R-project.org/>
5. Wheeler, R. E. (2009), *Comments on algorithmic design*, CRAN documentation, p. 59, <https://cran.r-project.org/web/packages/AlgDesign/vignettes/AlgDesign.pdf>
6. Marchenko, A., Sydorov, M. (2014), Methodical specificity of factorial surveys (on the example of pilot study of the role of ideological issues in friendships), *Methodology, theory and practice of sociological analysis of contemporary society*, Issue 20, Pp. 116–122.
7. "The job placement of students of the Faculty of sociology. Activities after studies", <http://www.soc.univ.kiev.ua/uk/nauka/diyalnist-pislya-zakinchennya-navchannya>
8. "The job placement of students of the Faculty of sociology. Search the first job", <http://www.soc.univ.kiev.ua/uk/nauka/poshuk-pershoyi-roboti>
9. "The job placement of students of the Faculty of sociology. Factors of choice a workplace", <http://www.soc.univ.kiev.ua/uk/nauka/faktori-viboru-robochogo-miscya>

Сидоров Николай. Практическая реализация факториального дизайна средствами R. В статье рассматриваются методические и практические особенности реализации экспериментального дизайна в опросе. Основное внимание сосредоточено на реализации метода в среде R, поэтапное конструирование вопросника, формирование массива виньеток, эффективной выборки, обработки и презентации результата. Дизайн апробирован на экспериментальном исследовании факторов влияния на выбор студентами первого рабочего места, что представлено в виде примера реализации метода в R.

Ключевые слова: язык статистического программирования R, метаданные, экспериментальный дизайн в опросе, виньетка, построение вопросника, D-эффективность выборки.

Sydorov Mykola. Practical Implementation of the Factorial Design with R. In the article considered the methodological and practical features of the implementation of experimental design (factorial design) in the survey. This type of survey design is come from social psychology and looks very similar to conjoint analysis but gives possibility to use this method in mass surveys. We focused on the implementation of this method in R statistical program. This program was selected as one of the most popular (10th place in common rating of programming languages) and modern free programming language which is developing mostly for data analysis. Therefore we describe all steps of realization of this method in R: construction of the questionnaire, forming an array of vignettes, modeling of orthogonally efficient sample, processing and presentation of results. Design tested in pilot study of factors influencing the selection of students (for this pilot survey was selected 50 students of magistracy of Sociology Faculty of Taras Shevchenko National University of Kyiv) first job, which was presented as an example implementation of the method in R. We tried to write whole code with maximum explanation. Finally some results was presented as linear regression model.

Key words: statistical programming language R, metadata experimental design in the survey, vignette, construction of questionnaire, D-effectiveness sampling.

Додаток 1

DATA LIST FREE / setnum a0 a1 a2 a3 a4 b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 c1 c2 c3 c4 c5 c6 c7.

VARIABLE LABEL

setnum "Номер сету"

a0 "Чи працюєте Ви на даний момент?"

a1 "Вкажіть, будь ласка, Ваш досвід роботи"

a2 "Чи працюєте Ви за спеціальністю?"

a3 "Вкажіть, будь ласка, якою була Ваша середньомісячна заробітна платня за останній рік?"

a4 "Чи плануєте Ви працювати за спеціальністю?"

b1 "Відповідь на віньєтку 1"

b2 "Відповідь на віньєтку 2"

...

b10 "Відповідь на віньєтку 10"

c1 "Вкажіть, будь ласка, Ваш вік"

c2 "Зазначте Вашу стать"

c3 "Назвіть, будь ласка, основні джерела вашого доходу під час навчання в університеті"

c4 "Вкажіть, будь ласка, ступінь матеріального забезпечення Вашої батьківської родини на момент Вашого вступу на бакалаврат"

c5 "Вкажіть тип населеного пункту, у якому пройшла переважна частина Вашого життя до моменту вступу в КНУ"

c6 "Вкажіть область (міста Київ та Севастополь, за межами України), у якій Ви проживали до вступу в КНУ"

c7 "Вкажіть, будь ласка, джерело фінансування вашого навчання на бакалавраті (магістратурі)".

VALUE LABELS a0

1 Так

2 Ні.

VALUE LABELS a1

1 До 1 року

2 Від 1-го до 2-х років

3 Більше за 2 роки.

VALUE LABELS a2

1 Так

2 Ні.

VALUE LABELS a3

1 До 2000 грн

2 Від 2000 до 4000 грн

3 Від 4000 до 6000 грн

4 Від 6000 до 8000 грн

5 Більше за 8000 грн.

VALUE LABELS a4

1 Так

2 Ні

3 Важко відповісти

VALUE LABELS c2

1 Чоловіча

2 Жіноча

VALUE LABELS c3

1 стипендія

2 матеріальна допомога від держави (пенсія за інвалідністю, допомога сиротам)

3 заробітна плата

4 допомога від родичів

5 дохід від підприємницької діяльності

6 інше.

VALUE LABELS c4

1 Коштів вистачає лише на продукти харчування

2 Коштів вистачає на продукти харчування і одяг, однак купівля речей тривалого використання проблематична

3 Коштів вистачає на все необхідне, але нам не до заощаджень

4 Коштів вистачає на все необхідне, робимо заощадження

5 Живемо в повному достатку

6 Важко сказати.

VALUE LABELS c5

1 Київ

2 Севастополь

3 обласний центр України

4 районний центр України

5 невеличке місто на території України

6 смт на території України

7 село на території України

8 місто за межами України

9 село за межами України.

VALUE LABELS c7

1 держзамовлення

2 контрактна форма навчання.

VARIABLE LEVEL a0 to a2 a4 c2 to c3 c5 to c7 (NOMINAL).

VARIABLE LEVEL a3 b1 to b10 c4 (ORDINAL)

Статтю отримано 21.11.2015 р.