

Асимптотична статистика

4 курс, статистика, Шкляр Ірина Володимирівна

Завдання 2, варіант 7

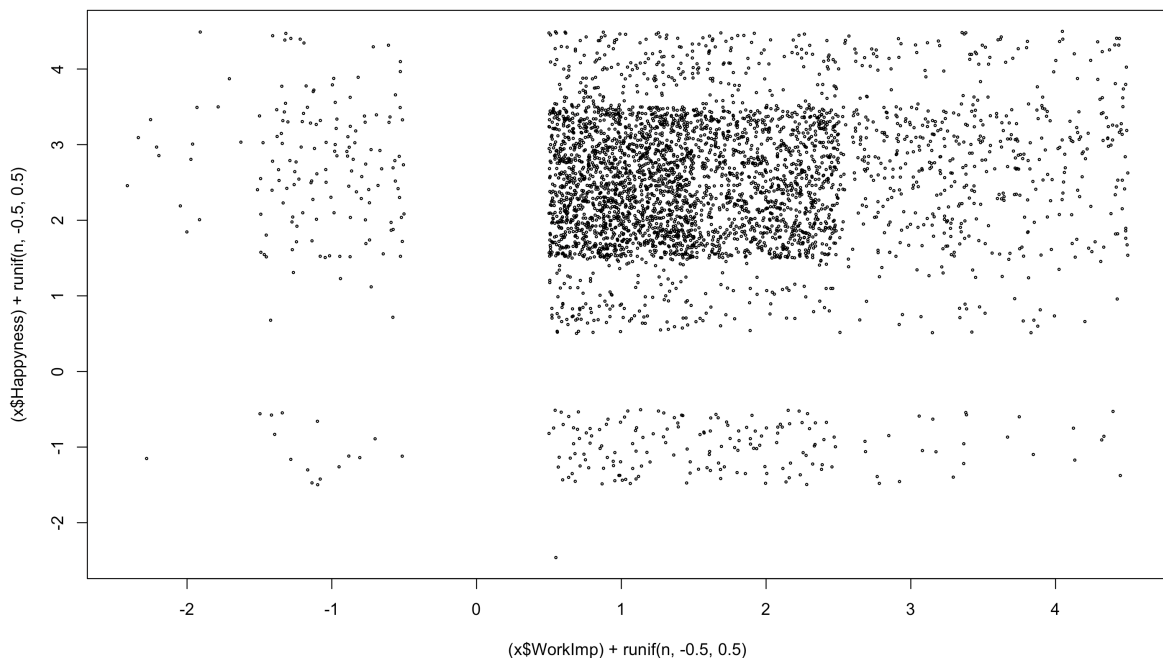
У роботі ми маємо перевірити незалежність між змінними WorkImp і Happiness у м. Києві за допомогою тесту χ^2 для перевірки незалежності.

H0: змінні незалежні між собою

H1: існує залежність

Завантажимо дані з файлу ексель для роботи та нарисуємо графік точок даних разом із випадковими відхиленнями від рівномірного розподілу на проміжку [-0.5, 0.5] (інакше графік не має сенсу):

```
> x <- read_excel("~/Downloads/regrasympt/values.xls", col_names=TRUE)
> n <- nrow(x)
> plot((x$WorkImp)+runif(n,-0.5,0.5),(x$Happyness)+runif(n,-0.5,0.5),cex=0.3)
```



Бачимо, що найчастіше, щодо важливості роботи, люди відповідали від 0.5 до 2.5, а наскільки вони щасливі – від 1.5 до 3.5.

Далі відфільтруємо дані для м. Києва, і щоб всі значення були додатні. Побудуємо таблицьку спряженості ознак:

```
> x <- x[x$Region==804011,]  
> x <- x[(x$WorkImp>0)&(x$Happyness>0),c("WorkImp","Happyness")]  
> tbl <- table(x)  
> tbl
```

	Happyness			
WorkImp	1	2	3	4
1	10	59	36	8
2	2	24	14	2
3	1	5	2	3
4	0	3	3	0

Якщо залежності не виявлено, то розподіли людей по змінній Happyness при фіксованій змінній WorkImp мають бути приблизно однаковими для всіх значень WorkImp. Тобто всі рядочки таблиці повинні бути приблизно пропорційними (відрізнятись лише сталими множниками). Те ж має виконуватись і для стовпчиків. Проглядаючи таблицьку бачимо, що строгої пропорційності немає, але і надзвичайно сильних відхилень не помітно. Застосуємо тест хі-квадрат:

```
> chisq.test(tbl)
```

Pearson's Chi-squared test

```
data: tbl  
X-squared = 9.3169, df = 9, p-value = 0.4086
```

Функція підрахувала статистику $\chi^2_{emp} = 9.3169$, визначила кількість ступенів свободи $df = 9$ і знайшла досягнутий рівень значущості $p\text{-value} = 0.4086$. Отже, приймаємо H_0 .

Оскільки у таблиці спряженості зустрічаються комірки з частотами меншими п'яти, то покладатися на результат стандартного тесту хі-квадрат не можна. Використаємо техніку імітаційного наближеного розрахунку $p\text{-value}$:

```
> set.seed(4)  
> chisq.test(tbl,simulate.p.value=T)
```

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

```
data: tbl  
X-squared = 9.3169, df = NA, p-value = 0.3893
```

Бачимо, що p-value вийшло 0.3893. ($p\text{-value} = 0.3893 > 0.05 = \alpha$) \Rightarrow знову треба прийняти основну гіпотезу (H_0) про незалежність між важливістю роботи та рівнем задоволення життям (щастям) – WorkImp і Happiness.