

УДК 004.82: 004:85

*Бевз Д. М., студент 2 курсу
спеціальності 122 «Комп'ютерні науки»
Нескородєва Т. В. к.т.н., доцент, доцент
кафедри інформаційних технологій*

АНАЛІЗ ФАКТОРІВ ВИЖИВАННЯ ЛЮДИНИ В ЕКСТРИМАЛЬНИХ УМОВАХ НА ПРИКЛАДІ НАБОРУ ДАНИХ «ТІТАНІС»

Донецький національний університет імені Василя Стуса, м. Вінниця

На сьогодні в умовах збільшення частоти стихійних лих, технологічних, катастроф, епідемій локального і глобального масштабу актуальною є проблема аналізу факторів виживання людини в екстремальних умовах.

Розглянемо приклад такого аналізу на основі даних технологічної катастрофи. Ми використовували набір даних Титаніка [1], який поставляється з вбудованим [2,3] в пакеті R набором даних «Titanic» та містить історичні записи всіх пасажирів, які піднялися на борт Титаніка. Нижче наводиться короткий опис 12 змінних в наборі даних:

- PassengerId – серійний номер пасажирів.
- Survived – включає двійкове значення 0 та 1. Пасажир не вижив – 0, пасажир вижив – 1.
- Pclass – квитковий клас | квиток 1-го, 2-го або 3-го класу.
- Name – ім'я пасажирів.
- Sex – чоловічий чи жіночий.
- Age – вік у роках – ціле число.
- SibSp – кількість братів чи сестер / подружжя – братів, сестер і / або дружини / чоловіка.
- Parch – кількість батьків / дітей - мати / батько і / або дочка, син.
- Ticket – номер квитка.
- Fare – пасажирський тариф.
- Cabin – номер каюти.
- Embarked – Порт Погрузки С – Cherbourg, Q – Queenstown, S – Southampton.

Об'єктами спостереження є пасажирів Титаніка. Отримаємо описову статистику для змінних набору даних «Титанік» (titanic) (показано на рис. 1).

Зараз ми проаналізуємо моделі виживання і перевіримо фактори. Для простоти було обрано тільки 2 такі змінні: стать та клас Pclass.

Які фактори впливали на кількість людей, що вижили в залежності від статі та класу? Коли ми говоримо про набір даних Титаніка, виникає перше запитання: «Скільки людей вижило?». Побудуємо гістограму, щоб продемонструвати відповідь на це питання.

PassengerId	Survived	Pclass	Name	Sex	Age	Sibsp
Min. : 1.0	Min. :0.0000	Min. :1.000	Abbing, Mr. Anthony	female:314	Min. : 0.42	Min. :0.000
1st Qu.:223.5	1st Qu.:0.0000	1st Qu.:2.000	Abbott, Mr. Rossmore Edward	male :577	1st Qu.:20.12	1st Qu.:0.000
Median :446.0	Median :0.0000	Median :3.000	Abbott, Mrs. Stanton (Rosa Hunt)		Median :28.00	Median :0.000
Mean :446.0	Mean :0.3838	Mean :2.309	Abelson, Mr. Samuel		Mean :29.70	Mean :0.523
3rd Qu.:668.5	3rd Qu.:1.0000	3rd Qu.:3.000	Abelson, Mrs. Samuel (Hannah Wozosky)		3rd Qu.:38.00	3rd Qu.:1.000
Max. :891.0	Max. :1.0000	Max. :3.000	Adahl, Mr. Mauritz Nils Martin (other)		Max. :80.00	Max. :8.000
					NA's :177	

Parch	Ticket	Fare	Cabin	Embarked
Min. :0.0000	1601 : 7	Min. : 0.00	:687	: 2
1st Qu.:0.0000	347082 : 7	1st Qu.: 7.91	B96 B98 : 4	C:168
Median :0.0000	CA. 2343: 7	Median : 14.45	C23 C25 C27: 4	Q: 77
Mean :0.3816	3101295 : 6	Mean : 32.20	G6 : 4	S:644
3rd Qu.:0.0000	347088 : 6	3rd Qu.: 31.00	C22 C26 : 3	
Max. :6.0000	CA 2144 : 6	Max. :512.33	D : 3	
	(other) :852		(other) :186	

Рисунок 1 – Звіт щодо описової статистики набору даних «Титанік»

Зробимо це більш зрозумілим, використовуючи перевірку відсотків.

```
/prop.table(table(titanic$Survived))
> prop.table(table(titanic$Survived))
```

```
      0      1
0.6161616 0.3838384
```

Тільки 38,38% пасажирів, які піднялися на борт «Титаніка», вижили.

Survival rate basis Gender (Основа рівня виживання «стать»).

Вважається, що в разі рятувальних операцій під час лих безпека жінки є пріоритетною. Що ж саме сталося тоді?

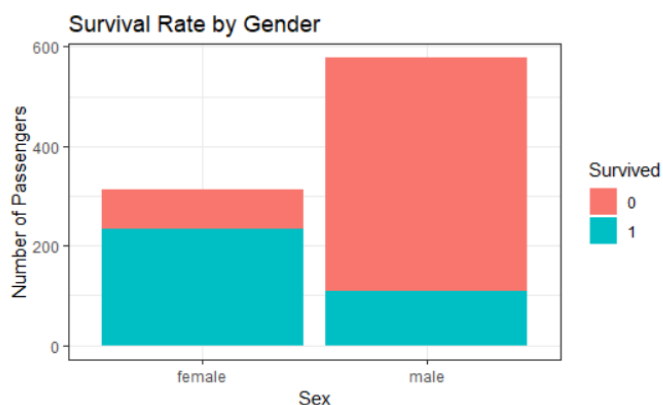


Рисунок 2 – Діаграма, яка порівнює виживання по статі описової статистики набору даних «Титанік»

Ми бачимо, що рівень виживання серед жінок був значно вище в порівнянні з чоловіками. Коефіцієнт пристосованості серед жінок становив близько 75%, тоді як у чоловіків він становив менше 20%.

Аналіз виживання на основі класу квитків (Pclass).

Було 3 сегмента пасажирів, в залежності від класу, в якому вони подорожували, а саме, 1 клас, 2 клас і 3 клас. Ми бачимо, що більше 50% пасажирів подорожували в 3-му класі.

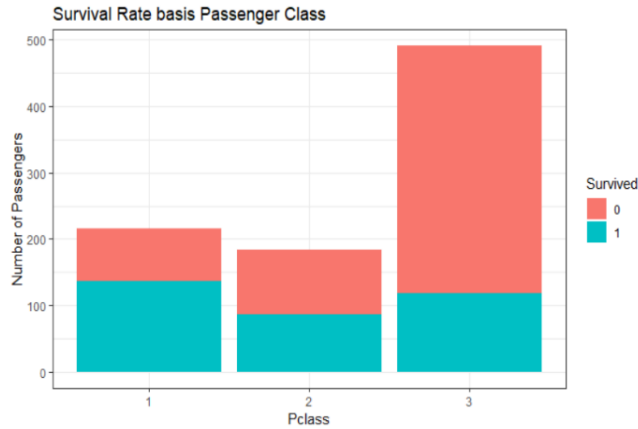


Рисунок 3 – Базова ставка виживання (Пасажирський клас) описової статистики набору даних «Титанік»

Пасажири 1-го і 2-го класів вижили непропорційно з більш ніж 60% виживання пасажирів 1-го класу, близько 45-50% 2-го класу і менш ніж 25% виживання тих, хто подорожує в 3-му класі.

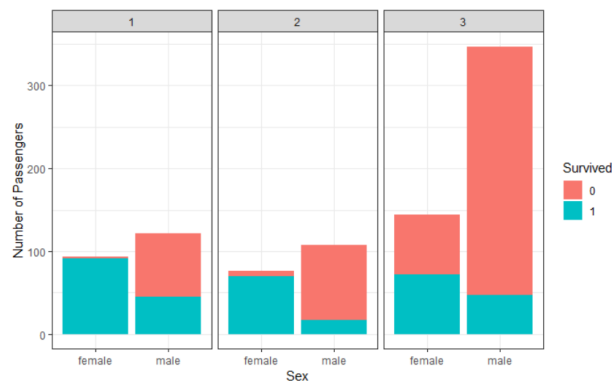


Рисунок 4 – Рівень виживання на основі класу квитків і статі (pclass) описової статистики набору даних «Титанік»

Висновок. Очевидно, що кількість людей, які вижили в 1 і 2 класах була найвищою. За винятком 1 дівчинки всі діти, які подорожували 1-м та 2-м класом, вижили. Показники кількості людей, які вижили були найнижчими для чоловіків, які подорожували в 3-му класі.

Список використаної літератури

1. «Титанік» статистика – [Електронний ресурс]. Режим доступу: <https://golos.io/@viromiro/titanik-statistika>
2. Тревор Хасті, Роберт Тибушрані, Гарет Джеймс, Даніела Віттен – «Введение в статистическое обучение с приложениями в R» - Москва: ДМК-Пресс, 2016. - 456 с.
3. Burns P. The R Inferno. - lulu.com; Second Edition (January 12, 2012). - 154 с.