

Кореляція:

```

Pearson's product-moment correlation

data: crim and unemp1
t = 0.38291, df = 11, p-value = 0.7091
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.4657268  0.6261090
sample estimates:
      cor
0.1146885

```

Рисунок 6 – Кореляція показників *crim* та *unemp1*

Дивно, але спостерігаємо низьку залежність цих показників.

Отже, провівши дослідження, ми помітили, що найбільше на кримінальність впливає кількість населення в місті.

Список літератури

1. Рівень криміналу в містах України [Електронний ресурс]. Режим доступу – <https://sovet.ua/novosti/2019/03/04/rejting-gorodov-ukrainyi-2019>
2. Середня заробітна плата в містах [Електронний ресурс]. Режим доступу – <https://www.work.ua/stat/?setlp=ua>
3. Кількість населення міст [Електронний ресурс]. Режим доступу - http://ukrmap.org.ua/Naselenie_ukr.htm

УДК 004.82: 004:85

*Степанюк О. С., студентка 2 курсу спеціальності 122 «Комп'ютерні науки»
Нескородєва Т. В., к.т.н., доцент, доцент кафедри інформаційних технологій*

АНАЛІЗ ДАНИХ ПРО РИНОК ВЖИВАНИХ АВТОМОБІЛІВ В УКРАЇНІ

Донецький національний університет імені Василя Стуса, м. Вінниця

У сучасному світі жодна людина не уявляє свого життя без зручного транспорту. При виборі вживаного авто люди керуються основною його перевагою – ціною. Але як залежить ціна від технічних показників авто? Щоб дати відповідь на це питання ми провели дослідження на основі даного набору даних.

Його актуальність обумовлена наступними факторами: зростання ринку вживаних авто, необхідність визначення факторів, що суттєво і несуттєво впливають на продаж. За даними «УкрАвтопром» [1] за 2018 рік в Україні первинну реєстрацію пройшло 116,8 тисячі вживаних автомобілів, що вдвічі більше ніж роком раніше. На противагу цьому, попит на нові автівки залишився без змін і склав 81,8 тисячі.

Подібні дослідження проводили такі вітчизняні вчені, як Юринець О.В [2], Савич О.П[3], а також були описані у підручнику зі статистичного навчання[4].

Аналіз проведемо за допомогою пакету R. Для дослідження був використаний набір даних Car Sale Advertisement [5], що містить 9576 спостережень за 10 показниками:

- car - автомобіль: марка виробника
- price - ціна: ціна продавця в рекламі (у USD)
- body - кузов: тип кузова автомобіля
- mileage - пробіг: як зазначено в рекламі ('000 км)
- engV - округлий об'єм двигуна ('000 куб. см)
- engType: тип палива
- registration - реєстрація: автомобіль зареєстрований в Україні чи ні
- year - рік виробництва
- model - модель: конкретна назва моделі
- drive - привід: тип приводу

Побудувавши графік залежності типу палива від ціни (рис.1), ми зробили висновок, що авто з типом палива дизель є найдорожчими на ринку, а з газом – мають найменшу вартість.

Щоб з'ясувати варіацію ціни в залежності від різноманітних характеристик автомобіля, побудуємо багатофакторну регресійну модель, у якій предикторами будуть виступати x_1 - car(марка), x_2 - body(тип кузова), x_3 - mileage(пробіг), x_4 - engV(об'єм двигуна), x_5 - engType (тип палива), x_6 - registration(реєстрація), x_7 - year(рік виробництва), x_8 - model(модель), x_9 - drive(тип приводу).

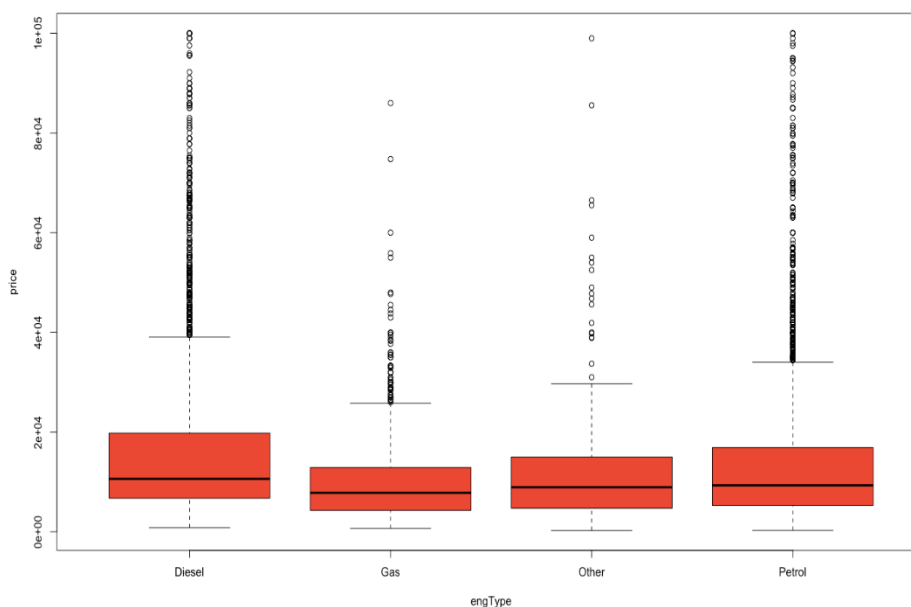


Рисунок 1 – Залежність ціни авто від типу палива

На рисунку 2 відображений звіт даної регресійної моделі. Асимптотична значимість (значення вірогідності, Р-значення) показує, наскільки варіація предиктора пояснює варіацію відгука. Р-значення x_3 , x_4 , x_7 , x_9 близькі до нуля, що вказує на істотний зв'язок між типом палива, пробігом, роком виготовлення,

типом привода та ціною автомобіля. Р-значення інших предикторів занадто великі, тому зв'язок між ними та ціною не прослідковується. Коефіцієнт детермінації (R^2) вказує на якість та тісноту зв'язків у моделі. Його значення варіюється у межах від 0 до 1. $R^2 = 0.4673$, що означає, що 46.73% дисперсії змінної ціни пояснюється вказаними предикторами.

Рівняння регресії буде виглядати наступним чином:

$$y = -1638000 - 9.899 \times x_1 - 555.7 \times x_2 - 39.32 \times x_3 + 8215 \times x_4 - 817.3 \times x_5 + 4.7779 \times x_6 + 812 \times x_7 + 2.8 \times x_8 + 3003 \times x_9$$

```
> summary(lm.fit)

Call:
lm(formula = price ~ car + body + mileage + engV + engType +
    registration + year + model + drive, data = Car_new)

Residuals:
    Min       1Q   Median       3Q      Max
-70659  -5503  -1585   2745 253258

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.638e+06  5.452e+04  -30.039 < 2e-16 ***
car          -9.899e+00  6.180e+00  -1.602  0.109
body        -5.557e+02  1.037e+02  -5.360  8.53e-08 ***
mileage     -3.932e+01  1.957e+00 -20.088 < 2e-16 ***
engV        8.215e+03  1.763e+02  46.604 < 2e-16 ***
engType     -8.173e+02  1.232e+02  -6.631  3.54e-11 ***
registration 4.779e+03  6.652e+02  7.184  7.35e-13 ***
year        8.120e+02  2.713e+01  29.936 < 2e-16 ***
model       2.800e+00  7.028e-01  3.985  6.80e-05 ***
drive       3.003e+03  2.349e+02  12.783 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13650 on 8162 degrees of freedom
Multiple R-squared:  0.4678,    Adjusted R-squared:  0.4673
F-statistic: 797.3 on 9 and 8162 DF,  p-value: < 2.2e-16
```

Рисунок 2 – Звіт по регресійній моделі

Отже, побудована регресійна модель виявила значний вплив предикторів mileage (пробіг), engV (об'єм двигуна), year (рік виробництва), drive(тип привода) на цінову політику різних виробників автомобілів. Також за допомогою вбудованих функцій пакету R було визначено, як ранжуються за ціною авто з різними типами палива: найдорожчі – дизель, найдешевші – газ.

Список літератури

1. УкрАвтопром[Електронний ресурс]. Режим доступу – <http://ukrautoprom.com.ua/uk/v-2018-godu-ukraincy-vvezli-i-zaregistrovali-1168-tys-legkovyx-avto-s-probegom>
2. Юринець О. В.- «Тенденції та перспективи розвитку автомобілебудівної галузі України у післякризовий період» /, О. Я. Марущак, О. В. Юринець // Вісник Національного університету "Львівська політехніка. – 2013.
3. Савич О. П. – «Маркетинг автозапчастин»: дис. ... канд. екон. наук. – К. : Держ. вищ. навч. закл. "Київ. нац. екон. ун-т ім. Вадима Гетьмана", 2012..
4. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani – "An Introduction to Statistical Learning with Applications in R". – 2009.

5. *Набір даних Car Sale Advertisement [Електронний ресурс]. Режим доступу – <https://www.kaggle.com/antfarol/car-sale-advertisements/data#>*

УДК 004.82: 004:85

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

ПЕРЕДБАЧЕННЯ КІЛЬКІСТІ ІНФІКОВАНИХ COVID-19 В УКРАЇНІ

Донецький національний університет імені Василя Стуса, м. Вінниця

Під час пандемії коронавірусної хвороби 2019 (COVID-19) важливо не панікувати і зважливо оцінювати ситуацію. Переглядаючи новини, ми часто чуємо прогнози аналітиків щодо кількості інфікованих, смертей та одужань. Часто ці прогнози є не дуже оптимістичними, найгірший результат передбачення пріоритетний, ніж найкращий.

На 23.04.2020 в Україні офіційно зафіксовано 7170 випадків коронавірусної хвороби COVID-19. Актуальність передбачення кількості інфікованих досить велика. Кожен з нас хоче знати, що чекає українців в найближчому майбутньому. Саме тому я вирішив проаналізувати дані про інфікованих на коронавірус в Україні та зробити деякі висновки.

Для дослідження був використаний набір даних COVID19 Global Forecasting [1] та мова програмування Python. Із набору я обрав дані, які стосуються лише України і почав розвідувальний аналіз (EDA).

Графік залежності кількості інфікованих від кількості днів має вид експоненціальної залежності – на початку періоду швидкість зростання не велика, в наступні періоди швидко зростає. Порівняємо швидкість інфікування з іншими країнами, а саме Італією, Іспанією та Великобританією. На 45 день від першого випадку інфікування для кожної країни кількість інфікованих була така:

- Україна – 4662 особи
- Італія – 27980 осіб
- Іспанія – 11748 осіб
- Великобританія – 1551 особа

Отже, можна відзначити позитивну користь від оперативно введених обмежень в Україні на пересування, навчання, масові скупчення тощо.

Для прогнозування оберемо дві моделі: просту лінійну регресію та логістичну регресію. Ці моделі є у модулі scikit-learn.

Для того, щоб побудувати просту лінійну регресію, потрібно нормалізувати дані за допомогою натурального логарифму. Нормалізувавши, розділяємо дані на навчальну та контрольну вибірки і тренуємо модель. Похибка