

3. Дубейковський В. І., Ефективне моделювання з СА ERwin® Process Modeler: BPwin; AllFusion Process Modeler: практичний посібник, 2009 – 121 с.
4. Січко Т. В. Методи моделювання бізнес-процесів підприємств засобами системного аналізу. Галицький економічний вісник. 2016. № 2. С. 190–201.

УДК 004.82:004:85

*Кучер М.О., студент 1 курсу
магістратури, спеціальність 122
«Комп'ютерні науки»
Нескородєва Т.В. к.т.н., доцент, доцент
кафедри інформаційних технологій*

АНАЛІЗ ФАКТОРІВ ВПЛИВАЮЧИХ НА РІВЕНЬ ЩАСТЯ НАСЕЛЕННЯ МЕТОДАМИ СТАТИСТИЧНОГО НАВЧАННЯ

Донецький національний університет імені Василя Стуса, м. Вінниця

На сьогодні все більше країн слідкують за рівнем щастя своїх громадян, оскільки розвинуті країни вже давно легко забезпечують первинні потреби. У зв'язку є необхідність розуміння тих факторів, що впливають на рівень щастя безпосередньо. Актуальним напрямком в цьому контексті є застосування аналізу наявних в нас даних за допомогою R[1,2].

Розглянемо набір даних «World Happiness Report 2021»[3] та спробуємо проаналізувати причини та зв'язки тих чи інших параметрів, які впливають на рівень щастя в країні.

Наведемо короткий опис змінних в даному наборі:

- Country name – назва країни
- Regional indicator – назва регіону країни
- Ladder score – оцінка рівня щастя
- Standard error of ladder score – стандартна помилка оцінки рівня щастя
- Lowerwhisker та upperwhisker – верхні та нижні оцінки (вуса) діаграми розмаху
- Logged GDP per capita – ВВП на душу населення
- Social support – рівень соціальної підтримки
- Healthy life expectancy – здорові роки життя (індекс Саллівана)
- Freedom to make life choices – рівень загальних свобод
- Generosity – щедрість
- Perceptions of corruption – відсоток корупції, тощо.

Набір даних включає 20 стовбців та 149 строк, деякі стовбці – якісні, зокрема «Regional indicator», потрібно змінити його, деякі стовбці, крім того, варто виключити з датасету, такі як «Country name».

Ключовим параметром є оцінка рівня щастя, саме аналіз впливу на неї інших змінних ми будемо проводити.

Для переведення якісної змінної назви регіону – треба скористатись фіктивним кодуванням і перетворити її в декілька бінарних стовбців в яких закодована назва регіону. Після цього можна спробувати проаналізувати дані за допомогою лінійної регресії.

```

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -0.90023    1.00986   -0.891  0.37430
Commonwealth.of.Independent.States -0.36017    0.20070   -1.795  0.07500 .
East.Asia                     -0.21930    0.24563   -0.893  0.37358
Latin.America.and.Caribbean    0.12894    0.17346    0.743  0.45860
Middle.East.and.North.Africa   -0.27899    0.18192   -1.534  0.12750
North.America.and.ANZ          0.34917    0.30510    1.144  0.25450
South.Asia                    -0.72817    0.24800   -2.936  0.00392 **
Southeast.Asia                -0.63100    0.22855   -2.761  0.00658 **
Sub.Saharan.Africa            -0.30538    0.22526   -1.356  0.17749
Western.Europe                 0.33538    0.19296    1.738  0.08451 .
Logged.GDP.per.capita          0.26751    0.08690    3.079  0.00253 **
Social.support                 1.94955    0.65425    2.980  0.00343 **
Healthy.life.expectancy        0.01409    0.01550    0.909  0.36495
Freedom.to.make.life.choices    2.26657    0.50127    4.522  1.34e-05 ***
Generosity                    0.49703    0.31801    1.563  0.12045
Perceptions.of.corruption      -0.32848    0.31022   -1.059  0.29159
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4954 on 133 degrees of freedom
Multiple R-squared:  0.8087,    Adjusted R-squared:  0.7872
F-statistic: 37.49 on 15 and 133 DF,  p-value: < 2.2e-16

```

Рисунок 1 – Результат виведення даних отриманих лінійною регресією

З цієї лінійної регресії видно, що нульову гіпотезу можна відкинути для свободи життєвого вибору, рівня соціальної підтримки та ВВП на душу населення.

Що цікаво – також є сильна негативна кореляція між рівнем щастя та регіоном, якщо мова йде про південну Африку чи південно-східну Азію.

Оскільки з наших даних не можна точно сказати чому саме ці 2 регіони є «депресивними», будемо це враховувати, але дані про регіон видалимо, щоб вони не заважали далі при аналізі, крім того, якщо керівники держави можуть вплинути на такі показники, як ВВП на душу населення чи особисті свободи громадян, то на регіон знаходження – ні. Тому враховувати його при аналізі – не зовсім коректно, адже для підвищення рівня щастя населення ці дані ми використати не зможемо.

Далі слід використати 3 параметри, для яких р-величина нижче ніж 0.01, в нашому випадку значення для таких змінних: 0.00253, 0.00343 та 1.34e-05.

```

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.6143    0.4171  -6.268 3.96e-09 ***
Logged.GDP.per.capita  0.4361    0.0645   6.761 3.12e-10 ***
Social.support    2.3424    0.6698   3.497 0.000625 ***
Freedom.to.make.life.choices 2.6849    0.4662   5.759 4.88e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5602 on 145 degrees of freedom
Multiple R-squared:  0.7334,    Adjusted R-squared:  0.7279
F-statistic: 133 on 3 and 145 DF,  p-value: < 2.2e-16

```

Рисунок 2 – Результат виведення даних отриманих лінійною регресією з найбільш корелюючими параметрами

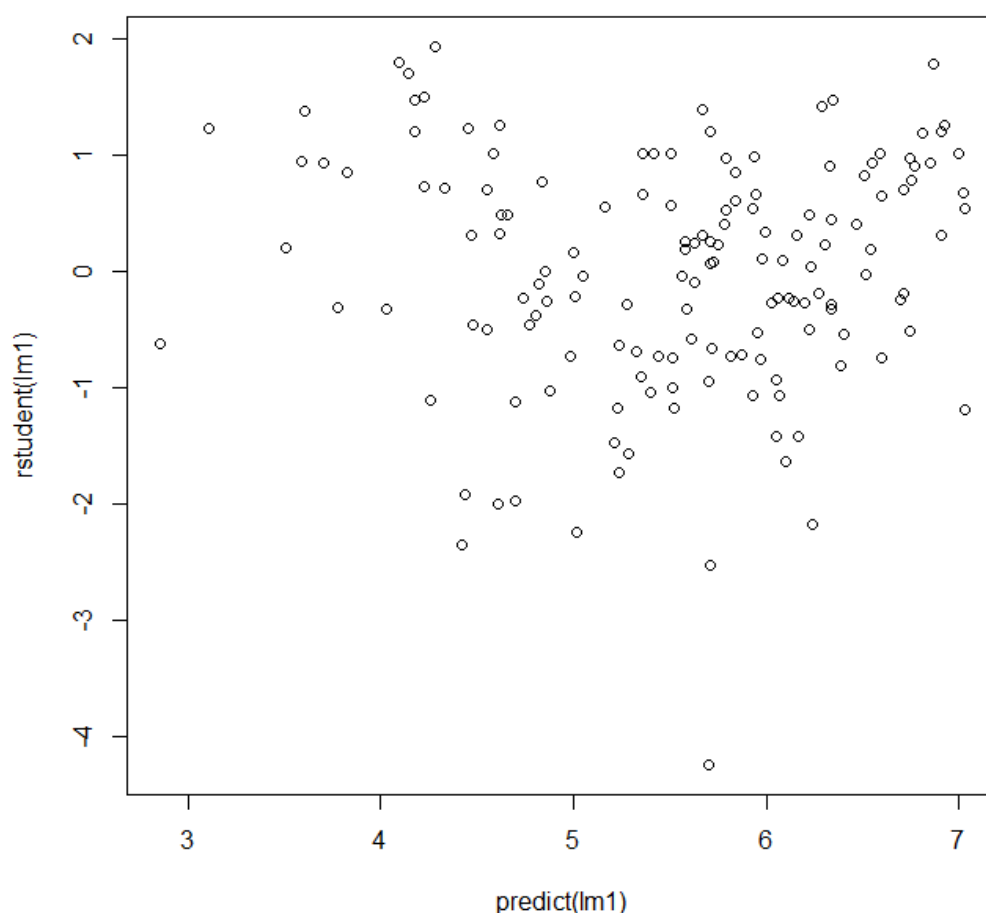


Рисунок 3 – Результат виведення залишків Стюдента

Залишки Стюдента розташовані в межах ± 3 з одним викидом поза цих меж, більшість залишків зосереджені в межах ± 1.5 .

Як видно з двох останніх рисунків – 3 обрані параметри досить добре характеризують рівень щастя, тож для його покращення країни мають зосередитись саме на них.

При тому, найбільше вплив має особиста свобода та соціальна підтримка, а ВВП на душу населення має дещо нижчий вплив.

Отже, побудована модель демонструє залежність рівня щастя від рівня особистих свобод, соціальної підтримки та ВВП на душу населення і якщо ці показники підвищуються – зростає і щастя населення. За допомогою пакету R продемонстровано дані, що свідчать про цю залежність.

Список використаних джерел

1. Тревор Хасті, Роберт Тибширані, Гарет Джеймс, Даніела Віттен – «Введение в статистическое обучение с приложениями в R» - Москва: ДМК-Пресс, 2016. - 456 с.
2. Burns P. The R Inferno. - lulu.com; Second Edition (January 12, 2012). - 154 с.
3. World Happiness Report 2021 [Електронний ресурс] // Online Journal for Research and Education. – 2021. – Режим доступу до ресурсу: <https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021?select=world-happiness-report-2021.csv>

УДК 004.82:004.85

*Огороднік М. О., студентка 3 курсу спеціальності 122 «Комп'ютерні науки»
Нескордова Т. В., к.т.н., доцент, доцент кафедри інформаційних технологій*

АНАЛІЗ ДАНИХ ПРО МОБІЛЬНІ ДОДАТКИ APPLE IOS ЗАСОБАМИ МОВИ R

Донецький національний університет імені Василя Стуса, м. Вінниця

Сьогодні в епоху інформаційного світу мобільні телефони стали невід'ємним елементом життя людей. Тому один з найперспективніших продуктів цифрового ринку є мобільні додатки, які є окремим програмним забезпеченням для кожної операційної системи будь-якого мобільного пристрою.

Велика частина мобільних пристроїв функціонують на базі операційної системи Apple IOS, тому розглянемо набір даних аналітики по мобільних додатках Apple IOS [1] для прогнозування та планування наступних застосунків розробниками.

Аналіз даних проведемо за допомогою засобів пакету R. Для дослідження був використаний набір даних Mobile App Statistics (Apple iOS app store), що містить 7197 спостережень на основі 17 показників:

- **id** – ідентифікатор програми;
- **track_name** – назва програми;
- **size_bytes** – розмір (у байтах);
- **currency** – тип валюти;