

ПОРІВНЯННЯ МОЖЛИВОСТЕЙ РІЗНИХ ПРОГРАМНИХ ЗАСОБІВ ДЛЯ ПОБУДОВИ НЕЛІНІЙНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ У ДОСЛІДЖЕННЯХ ПАРФУМЕРНО-КОСМЕТИЧНИХ ЗАСОБІВ

Коновалова К.С.

Науковий керівник: Нессонова М.М.

Національний фармацевтичний університет, Харків, Україна

k.380997113402@gmail.com

Вступ. Що таке регресія та який аналіз з нею пов'язаний? Регресія – залежність значення випадкової величини (змінної) від однієї або декількох інших випадкових величин (змінних). Регресійним аналізом називають метод статистичного аналізу залежності випадкової величини від змінних, що розглядаються в регресійному аналізі як не випадкові величини. При цьому мається на увазі підбір конкретної формули залежності цільової змінної (Y) від незалежних змінних (X), які пояснюють варіабельність випадкової величини Y, а також оцінювання точності (адекватності) цієї формули досліджуваній залежності. У дослідженнях парфумерно-косметичних засобів регресійний аналіз може стати у нагоді, наприклад, при створенні моделі, яка допоможе прогнозувати зміни у цільовому показнику дієвості засобу в залежності від відомих (контрольованих) параметрів.

У сучасному світі є безліч комп'ютерних програм, які можуть з легкістю відтворити регресійну залежність за допомогою спеціальних вбудованих формул, функцій та звести всі дані в електронні таблиці, графіки, діаграми. Значуще місце на світовій арені займають такі програмні засоби: Microsoft Office Excel, Paleontological Statistics Software for Education and Data Analysis (скорочено Past), Statistics Kingdom, Libre Office.

Мета дослідження. У ході наукової роботи треба дослідити можливості кількох програмних засобів здійснення парного нелінійного регресійного аналізу в дослідженнях парфумерно-косметичних засобів, порівняти їх між собою. Знайти дані на основі, яких виконати розрахунки. Визначити та показати за допомогою яких вбудованих функцій можна автоматично дізнатися ефективність дослідження. Зробити висновки щодо значущості та компетентності аналізу. Визначити, який з програмних засобів зручніший у використанні за інший.

Матеріали та методи. В якості програмних засобів для порівняння можливостей щодо проведення нелінійного регресійного аналізу нами було обрано статистичну програму Paleontological Statistics Software for Education and Data Analysis (скорочено PAST) і статистичний онлайн калькулятор Statistics Kingdom (<https://www.statskingdom.com>).

Програма Past – це безкоштовна статистична програма, що працює без установки, має простий інтерфейс англійською мовою. Дані легко переносяться через буфер обміну з електронних таблиць інших програмних засобів або вписуються відповідно в самій програмі. Недоліки програми пов'язані з її простотою. Вибір графічних налаштувань невеликий, результати розрахунків не зберігаються в особливому та окремому файлі; їх можна роздрукувати або скопіювати в окремий додаток на ПК. Має обмежений вибір методів аналізу, але зростає з кожним оновленням.

Statistics Kingdom – статистичний калькулятор у мережі Інтернет, що забезпечує не тільки прості обчислення, але також перевіряє припущення у аналізі, обчислює потужність аналізу та надає інтерпретацію результатів. Інтернет-калькулятори підтримують не тільки статистичні дані та р-значення, але й більше результатів, таких як розмір ефекту, потужність та рівень точності. Якщо одна з перевірок не вдається, інструмент рекомендує інші рішення. В тому числі має інструменти для проведення регресійного аналізу.

Дослідження можливостей зазначених програмних засобів щодо здійснення нелінійного регресійного аналізу проведено на модельних даних дослідження дії крему із залученням 100 добровольців, яким вимірювалася еластичність шкіри на різних термінах застосування косметичного засобу. При цьому визначалась задача підбору найкращої нелінійної регресійної моделі залежності еластичності шкіри від терміну використання крему.

Отримані результати. В програмі PAST для аналізу доступно більше нелінійних моделей, ніж в табличних процесорах. За допомогою різних моделей апроксимації (лінійна, поліноміальна, логарифмічна, експоненціальна, моделі зростання von Bertalanffy і Гомперца, степенева, логістична та інші) дивимося на рівняння регресії.

У програмі PAST розробник не виділяє конкретного коефіцієнту детермінації, але аналогом слугує інформаційний критерій Акаїке, який показує ненадлишковість (не переускладненість) моделей. Сенс цього показнику – чим значення критерію менше, тим краще модель. Додатково у цій програмі можна обчислити 95% довірчі інтервали для коефіцієнтів обраної моделі, що дозволяють визначити значущість коефіцієнтів.

За допомогою засобів PAST для нашої задачі виділено як найкращі такі моделі: лінійну та модель Michaelis-Menten. Лінійна модель задається рівнянням $y = 0.40076 \cdot x + 52.107$; модель Michaelis-Menten має рівняння

$$y = \frac{71.736 \cdot x}{1.9263 + x}$$

Для інших моделей довірчі інтервали для деяких коефіцієнтів показували, що з імовірністю 95% цей коефіцієнт може виявлятися нульовим, саме в такому випадку цей коефіцієнт не буде значущим.

Якщо порівняти між собою дві найкращих моделі, то буде переважати лінійна; це пов'язано з тим, що значення інформаційного критерію Акаїке в неї набагато менше (710.04), ніж у моделі Michaelis-Menten (3181.5).

Онлайн калькулятор Statistics Kingdom дозволяє будувати нелінійні моделі на базі інструменту Multiple Regression Calculator, в якому доступні такі операції нелінійних перетворень для залежної (Y) та незалежної (X) змінних як натуральний і десятковий логарифм, квадратний корінь і піднесення у 2-й степінь. Комбінація цих перетворень у визначеному користувачем порядку дозволяє отримати досить широкий набір нелінійних моделей для порівняння і вибору найкращої. Калькулятор обчислює р-рівні та довірчі інтервали для коефіцієнтів регресійної моделі, значення коефіцієнту детермінації (R^2) та скоригованого R^2 .

За допомогою засобів онлайн калькулятора Statistics Kingdom для нашої задачі виділено як найкращі такі моделі: експоненціальну ($R^2=0.901805$) та модель, що включає корінь квадратний з незалежної змінної ($R^2=0.902262$). Рівняння експоненціальної моделі Statistics Kingdom визначив у формі $\ln(y) = 3.97 + 0.006 \cdot x$, що після нескладних перетворень переписується у вигляді $y = e^{3.97} \cdot e^{0.006 \cdot x} = 53.07 \cdot e^{0.006 \cdot x}$ і співпадає з результатами, отриманими у PAST. Звертає на себе увагу той факт, що PAST визначив коефіцієнти цього рівняння як незначущі, в той час як Statistics Kingdom показав їх значущість. Можливо, це пов'язано із різними підходами до обчислення довірчих інтервалів для коефіцієнтів, реалізованими у цих програмах. Рівняння другої моделі: $\sqrt{y} = 6.659 + 0.259 \cdot \sqrt{x}$, що переписується як $y = 44.34 + 3.45 \cdot \sqrt{x} + 0.067 \cdot x$. Подібної моделі PAST не дозволяє побудувати, але її точність в нашій задачі виявилася на високому рівні.

Висновки. Таким чином, у ході наукової роботи досліджено можливості кількох програмних засобів здійснювати парний нелінійний регресійний аналіз в дослідженнях парфумерно-косметичних засобів, порівняла їх між собою. Визначила та показала за допомогою яких вбудованих функцій можна автоматично дізнатися ефективність дослідження. Зробила висновки щодо значущості та компетентності аналізу, показуючи на прикладі різних моделей.

На мою думку, кращим програмним засобом є Past, тому що він працює без мережі Інтернет, не потребує установки, має дружній інтерфейс.