Реферат

**на тему:**

**“Критерій х2 Пірсона”**

Критерій незалежності хі-квадрат Пірсона призначений для перевірки гіпотези про незалежність двох ознак, що задають рядки і стовпці таблиці спряженості. Статистика цього критерію

де сума береться по всіх клітках таблиці спряженості. Вона збігається зі статистикою [критерия согласия хи-квадрат](http://www.uran.donetsk.ua/~masters/2002/fvti/spivak/library/book2/book2.htm#chi_square_test_for_goodness_of_fit)\*, специфіка складається лише в способі обчислення очікуваних зустрічальностей: eij=ricj/N, де ri – сума зустрічальностей у i-й рядку, cj – сума зустрічальностей у j-м стовпці.

Критерій згоди хі-квадрат використовується для перевірки гіпотези про збіг емпіричного і теоретичного розподілів дискретних випадкових величин. Критерій ґрунтується на порівнянні спостережених і очікуваних (теоретичних) [встречаемостей](http://www.uran.donetsk.ua/~masters/2002/fvti/spivak/library/book2/book2.htm#frequency). [Статистика критерия](http://www.uran.donetsk.ua/~masters/2002/fvti/spivak/library/book2/book2.htm#test_statistics) дорівнює сумі квадратів різниць між спостереженими й очікуваними зустрічальностями, ділених на очікувані зустрічальності , де oi – спостережена зустрічальність i-й градації, а ei – її очікувана зустрічальність. Зверніть увагу: значення статистики залежить від обсягу вибірки.

Розглянемо одну з основних задач математичної статистики -задачу про перевірку правдоподібності гіпотез. Перед дослідником завжди поставав питання: як установити, чи суперечать досвідчені дані гіпотезі про те, що СВ\_Х розподілена за деяким законом. Для відповіді на це питання користаються так називаними критеріями згоди. Одним з таких критеріїв є критерій c 2 - Пірсона. У чому його суть? Пірсон запропонував розрахувати теоретичні частоти реалізації СВ\_Х, що підкоряється гіпотезі, що перевіряється, про закон розподілу , і порівняти їх з  емпіричними за визначеним критерієм. Якщо критерій задовольняється, то гіпотеза про передбачуваний закон розподілу СВ\_Х не відкидається, якщо критерій не задовольняється, те гіпотеза про передбачуваний закон розподілу СВ\_Х відкидається і дослідник повинний висувати нову гіпотезу про закон розподілу СВ\_Х (тобто переглянути свої погляди на природу досліджуваного явища).
Припустимо, що зроблено n незалежних досвідів, у кожнім з який СВ\_Х прийняла визначене значення. Ці значення занесені в таблицю:



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | x1 | x2 | ... | xk-1 | xk |
| n | n1 | n2 | ... | nk-1 | nk |
| P\* | p1\* | p2\* | ... | pk-1\* | pk\* |

Тут  – частота події. Ми висуваємо гіпотезу Н0, що складається в тім, що СВ\_Х має розподіл



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | x1 | x2 | ... | xk-1 | xk |
| n | n1’ | n2’ | ... | nk-1’ | nk’ |
| P\* | p1\* | p2\* | ... | pk-1\* | pk\* |

Щоб перевірити правдоподібність цієї гіпотези, треба вибрати якусь міру розбіжності статистичного розподілу з гіпотетичним. Як міру розбіжності береться сума квадратів відхилення статистичних імовірностей від гіпотетичних, узятих з деякими "вагами" сj: Коефіцієнти сj уводяться тому, що відхилення, що відносяться до різних значень pi, не можна вважати рівноправними: те саме по абсолютній величині відхилення  може бути малозначним, якщо імовірність pj велика, і дуже помітним, якщо вона мала. Пірсон довів, що якщо прийняти , те при великому числі досвідів n закон розподілу величини R має дуже прості властивості: він практично не залежить від закону розподілу СВ\_Х и мало залежить від числа досвідів n, а залежить тільки від числа значень випадкової величини (СВ\_Х) k і при збільшенні n наближається до розподілу c 2. При такому виборі коефіцієнтів cj міра розбіжності R звичайно позначається c *2*набл: чи з обліком того, що , одержимо .

Величина R підкоряється розподілу c 2 і залежить від параметра r, називаного "числом ступенів волі". При даному критерії число ступенів волі дорівнює числу значень СВ\_Х k мінус число незалежних умов ("зв'язків"), накладених на частоти р\*.

Проста лінійна кореляція (Пірсона r). Кореляція Пірсона (далі називана просто *кореляцією*) припускає, що дві розглянуті перемінні обмірювані, принаймні, у [интервальной шкале](http://pcbeck.ihep.su/work_web/statistica/statbook/glossary/gloss_i.html#Interval Scale) (см. [*Элементарные понятия статистики*](http://pcbeck.ihep.su/work_web/statistica/statbook/esc.html)). Вона визначає ступінь, з яким значення двох перемінних "пропорційні" один одному. Важливо, що значення коефіцієнта кореляції не залежить від масштабу виміру. Наприклад, кореляція між ростом і вагою буде однієї і тієї ж, незалежно від того, проводилися виміри в *дюймах* і чи фунтах у *сантиметрах* і *кілограмах*. *Пропорційність* означає просто *лінійну залежність*. Кореляція висока, якщо на графіку залежність "можна представити" прямою лінією (з позитивним чи негативним кутом нахилу).

Проведена пряма називається *прямою* регресії чи прямою, побудованою *методом найменших квадратів*. Останній термін зв'язаний з тим, що сума *квадратів* відстаней (обчислених по осі Y) від крапок, що спостерігаються, до прямої є мінімальної. Помітимо, що використання *квадратів* відстаней приводить до того, що оцінки параметрів прямої сильно реагують на викиди.

Як інтерпретувати значення кореляцій. Коефіцієнт кореляції Пірсона (*r*) являє собою міру лінійної залежності двох перемінних. Якщо звести його в квадрат, то отримане значення [коэффициента детерминации](http://pcbeck.ihep.su/work_web/statistica/statbook/glossary/gloss_k.html#Coefficient of Determination) r2) представляє частку варіації, загальну для двох перемінних (іншими словами, "ступінь" чи залежності зв'язаності двох перемінних). Щоб оцінити залежність між перемінними, потрібно знати як "величину" кореляції, так і її значимість.

**Використана література:**

1. Вища математика для ВУЗів. – Харків, 2000.
2. Высшая математика. – Одесса, 1992.