Міністерство освіти і науки України
Ніжинський державний університет імені Миколи Гоголя

Факультет природничо-географічних і точних наук
Кафедра інформаційних технологій і аналізу даних

Освітня програма: Середня освіта (Математика)

Спеціальність: 014 Середня освіта (Математика)

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на здобуття освітнього ступеня *магістр*

Непараметричні методи в аналізі данних

студентки **Карнаш Катерини Едуардівни**

Науковий керівник:
Зінченко Надія Мусіївна,
доктор фізико-математичних наук, професор

Рецензенти:
Казачков Іван Васильович,
доктор технічних наук, професор

Яблочніков Сергій Леонтійович,
доктор педагогічних наук, професор

Допущено до захисту: ____ _______ 2020 р.
Завідувач кафедри

проф. __________________ Казачков І.В.

Ніжин – 2020

АНОТАЦІЯ

Карнаш К.Е. Непараметричні методи в аналізі даних. – Рукопис.

Кваліфікаційна робота на здобуття освітнього ступеня магістра за спеціальністю 014.04 – Середня освіта (Математика). – Ніжинський державний університет імені М. Гоголя. Ніжин, 2020.

Проаналізовано теоретичне підґрунтя та специфіку практичного застосування найпоширеніших непараметричних критеріїв перевірки гіпотез: критеріїв згоди та однорідності. Основна увага приділена критеріям згоди Колмогорова та χ^2, серед критеріїв однорідності: критеріям Колмогорова-Смірнова, χ^2 та ранговим. Особливості застосування рангових критеріїв в педагогічних дослідженнях і в освітніх вимірюваннях розглянуто більш детально та проілюстровано на конкретних прикладах з педагогічного досвіду (педагогічної практики).

Ключові слова: непараметричні критерії, критерії згоди, критерії однорідності, рангові критерії, нульова гіпотеза, альтернативна гіпотеза.

ANNOTATION

Karnash K.E. Non-parametric methods in data analysis. - Manuscript.

Qualifying work for a master's degree in specialty 014.04 - Secondary education (Mathematics). - Nizhyn Gogol State University. Nizhyn, 2020.

This work deals with the theoretical basis and specific propertis of practical applications of the most common non-parametric tests for hypothesis testing, i.e. the goodness of fit and homogeneity tests. The main attention is focused on the Kolmogorov- and χ^2- goodness of fit tests; among the tests of homogeneity: on Kolmogorov-Smirnov test, χ^2 and rank tests. The importent features of applications of the rank tests in the pedagogical researches and in the educational measurements are discussed in details and illustrated by concrete examples from pedagogical practice.

Keywords: non-parametric test, goodness of fit test, homogeneity test, rank test, null hypothesis, alternative hypothesis.

ЗМІСТ

ВСТУП

Непараметричні методи - це частина математичної статистики, що охоплює методи і алгоритми обробки статистичних даних, які дозволяють зробити змістовні статистичні висновки, зокрема, оцінити базові характеристики та перевірити статистичні гіпотези за наявними даними (результатами спостережень) при досить «слабких» обмеженнях на апріорну інформацію стосовно ймовірнісних розподілів результатів спостережень (експериментів).

Вже сама назва «Непараметричні методи» підкреслює їх суттєву відмінність від класичних «параметричних» методів, в яких явно чи неявно припускається гауссівський розподіл результатів спостережень. Особливості непараметричних методів в їх незалежності від невідомого теоретичного розподілу. Вони спираються на більш широкі і менш обмежені властивості розподілу ймовірностей: часто це лише статистична незалежність спостережень і неперервність їх розподілів. В багатьох випадках можливо відмовитися і від припущення про неперервність.

Такий підхід має історичні корені: в минулому (30-і роки 20 століття) він виник як альтернатива класичним методам заснованих на гаусівському (нормальному) розподілі. «Непараметричні» - як назва для нових методів, підкреслювала їх універсальну застосовність до неперервних розподілів.

Спочатку непараметричні методи призначалися для перевірки статистичних гіпотез, в першу чергу, гіпотез про вид розподілу. Це ,так звані, «критерії згоди». Перші фундаментальні результати в цьому напрямку отримали А.Н. Колмогоров і Н.В. Смірнов. Пізніше (у 50-60 роки) важливі результати стосовно властивостей критеріїв Смірнова - Колмогорова отримали Б.В. Гнєденко і В.С. Королюк (академіки НАН України). З часом поняття «непараметричні методи» значно розширилося за рахунок різноманітних рангових критеріїв. Було виявлено, що рангові статистичні критерії можна застосовувати і для оцінки невідомих параметрів статистичних моделей, і, назагал, в аналізі категоризованих (класифікованих) даних.

В даний час непараметричні методи (в першу чергу рангові) утворюють розгалужену систему обробки статистичних даних та за своїми можливостями не поступається класичним методам. Перевагою непараметричних методів є широта застосування, стійкість статистичних висновків щодо грубих помилок, точність непараметричних гіпотез, математична простота статистичних правил і віповідних алгоритмів.

Методи непараметричної статистики – порівняно нова математична дисципліна, яка нині активно розвивається. Універсальність, простота і точність непараметричних гіпотез привертають усе більше увагу дослідників.

Непараметричні методи широко застосовуються в експериментальній і соціальної психології, в маркетингу, соціології, теорії надійності, політичних дослідженнях, гендерних дослідженнях, освітніх вимірюваннях, медицині тощо.

Актуальність проблематики, пов’язаної з непараметричними критеріями, зумовлена їх широкою сферою застосувань, зокрема, в педагогічних дослідженнях і в освітніх вимірюваннях.

Об’єктом дослідження є непараметиричні статистичні критерії, зокрема, критерії згоди і критерії однорідності.

Предмет дослідження - особливості застосування непараметричних критеріїв до задач статистичного аналізу даних.

Метою дослідження є обґрунтування теоретичної і практичної ролі непараметричних критеріїв.

Гіпотеза. Непараметричні критерії універсальні та прості в застосуванні, за їх допомогою можна оцінювати результати педагогічних досліджень.

Для досягнення поставленої мети необхідно виконати такі завдання:

- Проаналізувати та описати найбільш застосовані непараметричні критерії;
- Розглянути застосування деяких критеріїв на конкретних прикладах;
- Провести власне дослідження, отримані результати оцінити використавши описані критерії;

- Проаналізувати отримані результати.

Для виконання поставлених задач були використані такі методи:

1) Теоретичні: проаналізовано наукову літературу, розглянуто особливості застосування непараметричних критеріїв.
2) Практичні: підібрано та розв'язано приклади використання деяких критеріїв згоди і однорідності, проведено дослідження на основі педагогічного досвіду.

Наукова новизна полягає в тому, що було продемонстровано застосування непараметричних критеріїв на конкретних прикладах, взятих із педагогічної практики, та показано можливість використання даних критеріїв в освітніх вимірюваннях.

Структура кваліфікаційної роботи. Робота складається з таких частин: вступ, три розділи, висновок, список використаної літератури та додатки.

Перший розділ присвячений критеріям згоди, в ньому наведені найбільш поширені критерії. Процедури застосування критеріїв проілюстровані на конкретних прикладах.

В другому розділі розглянуто критерії однорідності: особливості їх застосування, переваги та недоліки, а також приклади застосування.

В третьому розділі проілюстровані приклади застосування непараметричних критеріїв однорідності в задачах пов'язаних з педагогічною діяльністю.

Список використаної літератури містить 19 джерел.

РОЗДІЛ 1

НЕПАРАМЕТРИЧНІ КРИТЕРІЇ. КРИТЕРІЇ ЗГОДИ

1.1. Критерії перевірки статистичних гіпотез. Основні поняття

Статистична гіпотеза - це будь-яке припущення, що стосується невідомого розподілу випадкових величин (елементів), що описують досліджуваний об'єкт чи процес, більш формально можна сказати, що статистична гіпотеза – це припущення, що стосується невідомого розподілу. Перевірка припущень робиться на основі конкретних даних (експериментальних даних, спостережень, статистичної звітності). Розрізняють нульову і альтернативну гіпотези. Нульова гіпотеза (далі її будемо позначати H_0) - гіпотеза, що підлягає перевірці. Альтернативна гіпотеза - кожна допустима гіпотеза, відмінна від нульової, далі будемо позначати H_1, якщо вона одна, якщо декілька H_1, H_2 і т.д.

Процедура перевірки гіпотез схематично має такий вигляд. На основі експериментальних даних (спостережень) розраховується деяка функція, оскільки кожне спостереження є випадковою величиною, то і функція є випадковою величиною. Ця функція носить назву статистики критерію, позначають її C. Вдається довести, що при виконанні гіпотези H_0 вибрана статистика має якийсь конкретний добре досліджений розподіл, з яким можна працювати. Далі процедура критерію має такий вигляд: по конкретним даним обчислюється значення статистики C (обчислене емпіричне значення статистики) і порівнюється з критичним, яке табульовано. Рішення про те, що дані підтверджують чи не підтверджують висунуту гіпотезу приймається в залежності від того, чи перевищує обчислене (емпіричне, розраховане) значення критичне (табульоване), чи ні. При цьому допускаються похибки. Похибкою першого роду називають ймовірність відхилити гіпотезу H_0, коли вона справджується. Похибкою другого роду вважають ймовірність не прийняття гіпотези H_0, коли вона не справджується. Одночасно зменшити похибку першого і другого роду неможливо. Проте, при роботі з конкретними даними зосереджуються в основному на похибці першого роду і вибирають певні граничні значення. Для похибки першого роду вибирають, як правило, верхню границю, що носить назву « рівень значущості» $(\alpha = 1 - P)$, і приймають рівень значущості рівним 0.05 або 0.01. Еквівалентний підхід - вибір надійності висновків $P = 0.95$ або $P = 95\%$.

Історично перші критерії перевірки гіпотез з математичної точки зору стосувалися параметрів нормального (гауссівського) розподілу. Це добре відомі t-критерій Стьюдента та F-критерій Фішера, які стосувалися перевірки гіпотез щодо середнього (математичного сподівання) і дисперсії для стандартного нормального закону. В цьому не має нічого дивного, оскільки результати спостережень(експериментів) дуже часто мають нормальний (гаусівський) розподіл в силу дії центральної граничної теореми , яка стверджує, що вплив великої кількості маленьких факторів в результаті дають гаусівський або нормальний розподіл. [16]

Вказані критерії (як і інші, які стосуються параметрів конкретного ймовірнісного розподілу) прийнято називати параметричними критеріями.
Гаусівський розподіл зустрічається дуже часто, але не завжди результати спостережень мають такий розподіл. Тому виникла необхідність в статистичних критеріях, що мають більш широку сферу застосувань. Такі критерії стали називати непараметричними, а в англомовній літературі їх прийнято називати «вільними від розподілу» (free of distribution).

Особливість непараметричних методів, на відміну від класичних методів полягає в тому, що не потрібно прив’язуватися до конкретного розподілу. Серед процедур статистичного аналізу даних помітне місце займають такі два класи задач: оцінювання і перевірка гіпотез. Далі зосередимося на задачах пов’язаних з перевіркою статистичних гіпотез, а саме з критеріями згоди і критеріями однорідності.

1.2. Критерії згоди

Критерій згоди - це статистичне правило, побудоване на основі спостережень, за яким перевіряється гіпотеза про те, що досліджувана випадкова величина підпорядковується тому чи іншому розподілу $F(x)=F(x,\theta)$, де $\theta \in R^k, k \geq 1$ - параметр (вектор параметрів).Тобто перевіряється відповідность емпіричного і гіпотетичного (теоретичного) законів розподілу.

Висувають дві гіпотези: нульову H_0 та альтернативну H_1. Нульова гіпотеза стверджує, що різниця між емпіричним і гіпотетичним законами (наприклад нормальним) незначна (вона передбачає відсутність значущих відмінностей між ними), а отже, наявність узгодженості емпіричного і гіпотетичного розподілів. Альтернативна ж гіпотеза передбачає існування значущих відмінностей, а отже, відсутність узгодженості емпіричного і гіпотетичного розподілів.

Таким чином, критерій згоди повинен підтвердити або відхилити нульову гіпотезу.

1.2.1. Критерій Колмогорова

Найбільш розповсюдженим вважають критерій згоди Колмогорова. Нехай $x=(x_1,x_2,...,x_n)$– вибірка невідомого неперервного розподілу $F_n(x)$ обсягу n; x_i - вибіркове спостереження, де $i=\overline{1,n}$; $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$- варіаційний ряд; а $x_{(i)}$ - i - та впорядкована статистика, де $i=\overline{1,n}$.

Відносно розподілу $F_n(x)$ висуваєтеся гіпотеза $H_0: F_n(x)=F(x)=F(x,\theta)$,

де $\theta=(\theta_1,\theta_2,...,\theta_n) \in R^k, k \geq 1$ повністю відомий, тобто вектор параметрів повністю відомий.

Нехай $F_n(x)=\frac{1}{n}\sum_{i=1}^{n} I_{\{x_i<x\}}$ – емпірична функція розподілу. Виникає необхідність перевірити дану гіпотезу H_0 та зробити висновок відхиляти її чи ні. За альтернативну гіпотезу приймають $H_1: F_n(x)\neq F(x)$.

Для перевірки гіпотези використовують статистику Колмогорова

$$D=D_n=\sup_x\left|F_n(x)-F(x)\right|=\max\left(D_n^+;D_n^-\right),$$

де $D_n^+=\sup_x\left\{F_n(x)-F(x)\right\}$; $D_n^-=\sup_x\left\{F(x)-F_n(x)\right\}$. D_n^+, D_n^- - статистики Смірнова.

Для розрахунків зручно використовувати такі формули:

$$D_n^+=\max_{1\le i\le n}\left\{\frac{i}{n}-F(x_i)\right\},\ D_n^-=\max_{1\le i\le n}\left\{F(x_i)-\frac{(i-1)}{n}\right\}.$$

Коли гіпотеза справджується $F_n(x)=F(x)$ статистика D_n набуває значення, яке є дуже малим, в протилежному випадку при великих значеннях статистики гіпотеза відхиляється.

Існують точні та асимптотичні розподіли статистики D_n, D_n^+, D_n^-

$$P\left(D_n^+\le a\right)=\begin{cases} 0, a<0 \\ 1-\sum\limits_{j=1}^{[n-an]} C_n^j\left(\frac{j}{n}+a\right)\left(1-\frac{j}{n}-a\right), 0<a<1 \\ -1, a\ge 1\end{cases}$$

Асимптотичний розподіл визначається граничними теоремами Колмогорова і Смірнова

$$\lim_{n\to\infty}P\left(\sqrt{n}D_n<a\right)=K(a)=1-2\sum_{k=1}^{n}(-1)^{k-1}\exp\left(-2k^2a^2\right), a>0;$$

$$\lim_{n\to\infty}P\left(\sqrt{n}D_n^+<a\right)=\lim_{n\to\infty}P\left(\sqrt{n}D_n^-<a\right)=S(a)=1-\exp\left(-2a^2\right), a>0.$$

Якщо $K_n(a)$ - розподіл статистики $\sqrt{n}D_n$, то для будь-якого n $K_n(a)>K(a)$ і максимальна похибка від заміни точного розподілу на асимптотичний при $n\ge 60$ має порядок $0.8\cdot 10^{-2}$.

Для точного і асимптотичного розподілу статистик D_n, D_n^+, D_n^-, розраховані табличні значення $S(a)$ і $D(a)$ [1,3,13,16] та відповідні критичні значення , вони наведені в додатку 1.

Складемо алгоритм перевірки гіпотези H_0:

1. Дані вибірки необхідно впорядкувати в порядку зростання, отже перейти від вибірки до варіаційного ряду $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$;
2. Знайти $F(x_i)$;ё
3. Обчислити значення статистик Смірнова – Колмогорова;
4. Обчислити ймовірність $p = P(D_n > t) = 1 - K(D_n)$,

використовуючи таблиці для точного або асимптотичного розподілу статистики Колмогорова (або знайти відповідне критичне значення $d_{кр} = d_\alpha$) порівняти D_n та $d_{кр}$.

5. Якщо отримане значення p велике (більше за вибраний рівень значущості α) або D_n – мале, $D_n < d_{кр}$ то гіпотеза H_0 про розподіл $F(x,\theta)$ підтверджується.

   Якщо значення p - мале (менше за вибраний рівень значущості α) або D_n – велике $D_n > d_{кр}$, то гіпотезу H_0 відхиляють.

   Інколи гіпотезу H_0 відхиляють і при дуже малих значеннях D_n.

Зауважимо, що використання критерію Колмогорова можливе лише тоді, коли гіпотетичний розподіл – неперервний та повністю визначений (тобто параметри відомі).

Розглянемо застосування критерію на конкретних прикладах.

Приклад 1.1

Результати $n = 10$ спостережень [1] наведені в табл. 1.1 в другому стовпчику. Чи узгоджується з наведеними даними гіпотеза про те, що досліджувана величина має рівномірний розподіл на відрізку $[0,36]$?

Для розв'язання цієї задачі вирішили скористатися критерієм Колмогорова з надійністю висновків 0.95 (з рівнем значущості $\alpha = 0.05$)

Теоретична функція розподілу має вигляд

$$F_0(x_i) = \begin{cases} 0, x < 0 \\ \frac{x}{36}, 0 \le x \le 36 \\ 1, x > 36 \end{cases}.$$

Таблиця 1.1

i	x_i	$F_n(x_i) = \frac{i}{n}$	$t_i = F_0(x_i)$	$\frac{i}{n} - t_i$	$\frac{i-1}{n}$	$t_i - \frac{i-1}{n}$
1	0	0.1	0	0.1	0	0
2	1	0.2	$1/36 = 0.0278$	0.1722	0.1	–0.0722
3	2	0.3	$2/36 = 0.0556$	0.2444	0.2	–0.1444
4	3	0.4	$3/36 = 0.0883$	0.3167	0.3	–0.2167
5	4	0.5	$4/36 = 0.1111$	0.3889	0.4	–0.2889
6	7	0.6	$7/36 = 0.1944$	0.4056	0.5	–0.3056
7	16	0.7	$16/36 = 0.444$	0.2006	0.6	–0.1006
8	18	0.8	$18/36 = 0.5$	0.3	0.7	–0.2
9	20	0.9	$20/36 = 0.5556$	0.3444	0.8	–0.2444
10	23	1	$23/36 = 0.6389$	0.3611	0.9	–0.2611

Щоб скористатися критерієм Колмогорова слід обчислити значення таких статистик:

$$\left|D_n^+\right| = \left|D_{10}^+\right| = \max_{(1 \le i \le 10)} \left|F_n\left(x_{(i)}\right) - F_0\left(x_{(i)}\right)\right| = \max_{(1 \le i \le 10)} \left|(i/n) - t_i\right|,$$

$$\left|D_n^-\right| = \left|D_{10}^-\right| = \max_{(1 \le i \le 10)} \left|F_0\left(x_{(i)}\right) - F_n\left(x_{(i-1)}\right)\right| = \max_{(1 \le i \le 10)} \left|t_i - (i-1)/n\right| ,$$

$$D_n = \max\left(\left|D_n^+\right|, \left|D_n^-\right|\right).$$

Необхідні проміжні розрахунки наведені див. табл. 1.1.

Отже маємо: $\left|D_n^+\right| = 0.4056$, $\left|D_n^-\right| = 0.3056$, $D_n = \max\left(\left|D_n^+\right|, \left|D_n^-\right|\right) = 0.4056$.

Значення статистики D_n необхідно порівняти з критичним, що знаходиться за таблицями для точного розподілу статистики Колмогорова в залежності від кількості даних n і рівня значущості α. В данному випадку $d_{кр} = d_{(n,\alpha)} = d_{(10,0.05)} = 0.4093$.

Порівнявши отримані значення бачимо, що $D_n < d_{кр}$ $(0.4056 < 0.4093)$; отже, немає підгрунття для відхилення гіпотези про рівномірний розподіл на $[0,36]$, данні підтверджують цю гіпотезу.

Якщо ж нема таблиць для точного розподілу можна скористатися таблицями для граничного (асимптотичного) розподілу і модифікованою формою $\tilde{D}_n$ для D статистики

$$\tilde{D}_n = D_n\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) = 1.3454.$$

Критичне значення тепер залежить лише від рівня значущості (надійності висновків) і при $\alpha = 0.05$ дорівнює $d_{кр} = d_{(\alpha)} = 1.36$. Знову бачимо, що $\tilde{D}_n < d_{кр}$ тобто данні підтверджують гіпотезу про рівномірний розподіл.

Приклад 1.2

Результати $n = 10$ спостережень наведені в табл. 1.2 в другому стовпчику. Чи узгоджується з наведеними даними гіпотеза про те, що досліджувана величина має показниковий (експоненційний) розподіл з параметром 1?

Застосувати рівень значущості $\alpha = 0.05$.

Таблиця 1.2

i	x_i	$F_n(x_i) = \frac{i}{n}$	$\exp(-x_i)$	$1-\exp(-x_i) = t_i = F_0(x_i)$	$\frac{i}{n} - t_i$	$\frac{i-1}{n}$	$t_i - \frac{i-1}{n}$

1	0.01	0.1	$\exp(-0.01)=0.990$	0.01	0	0	0.01
2	0.17	0.2	$\exp(-0.17)=0.844$	0.156	0.044	0.1	0.56
3	0.25	0.3	$\exp(-0.25)=0.779$	0.221	0.079	0.2	0.021
4	0.76	0.4	$\exp(-0.76)=0.467$	0.533	-0.113	0.3	0.213
5	1.01	0.5	$\exp(-1.01)=0.364$	0.636	-0.136	0.4	0.236
6	1.05	0.6	$\exp(-1.05)=0.35$	0.65	-0.05	0.5	0.15
7	1.5	0.7	$\exp(-1.5)=0.2231$	0.7769	-0.0769	0.6	0.1769
8	1.71	0.8	$\exp(-1.71)=0.181$	0.819	-0.019	0.7	0.119
9	2.35	0.9	$\exp(-2.35)=0.095$	0.905	-0.005	0.8	0.105
10	2.5	1	$\exp(-2.5)=0.082$	0.918	0.082	0.9	0.182

Як і в прикладі 1.1 , щоб скористатися критерієм Колмогорова слід обчислити значення таких статистик:

$$\left|D_n^+\right| = \left|D_{10}^+\right| = \max_{(1\le i\le 10)}\left|F_n\left(x_{(i)}\right) - F_0\left(x_{(i)}\right)\right| = \max_{(1\le i\le 10)}\left|(i/n) - t_i\right|,$$

$$\left|D_n^-\right| = \left|D_{10}^-\right| = \max_{(1\le i\le 10)}\left|F_0\left(x_{(i)}\right) - F_n\left(x_{(i-1)}\right)\right| = \max_{(1\le i\le 10)}\left|t_i - (i-1)/n\right|,$$

$$D_n = \max\left(\left|D_n^+\right|, \left|D_n^-\right|\right).$$

Необхідні проміжні розрахунки наведені див. табл. 1.2. Маємо:

$$\left|D_n^+\right| = 0.136\,,\ \left|D_n^-\right| = 0.236\,, D_n = \max\left(\left|D_n^+\right|, \left|D_n^-\right|\right) = 0.236\,.$$

Значення статистики D_n треба порівняти з критичним, що знаходиться за таблицями для точного розподілу статистики Колмогорова в залежності від кількості даних n і рівня значущості α. В данному випадку це те ж саме $d_{кр} = d_{(n,\alpha)} = d_{(10,0.05)} = 0.4093$.

Знову бачимо, що $D_n < d_{кр}$ $(0.236 < 0.4093)$ отже, немає підгрунття для відхилення гіпотези про показниковий розподіл $1-\exp(-x)$.

1.2.2. Критерії ω^2, A^2

Розглянемо критерій ω^2 (Крамера-Мізеса-Смірнова). Використання даного критерію ефективне при великій кількості даних $(n \geq 15)$.Основа критерію - це середньоквадратична міра відстані між емпіричною та гіпотетичною функціями розподілу:

$$\omega^2 = n\int_{-\infty}^{\infty}\left\{F_n(x) - F(x)\right\}^2 dF(x).$$

Для зручності розрахунків використовують формулу

$$\omega^2 == \sum_{i=1}^{n}\left\{z_{(i)} - \frac{2i-1}{2n}\right\}^2 + \frac{1}{12n},$$

де $z_{(i)} = F\left(x_{(i)}\right), x_{(i)}$ i–та впорядкована статистика, де $i = \overline{1,n}$. Граничний розподіл $\lim\limits_{n\to\infty} P\left(\omega^2 \leq x\right) = a(x)$. Якщо гіпотеза H_0 справедлива при $n \to \infty$, то закон розподілу не залежить від виду функції $F(x)$і прямує до розподілу «омега-квадрат» $\alpha(x)$.

Для деяких значень n є обчислені точні розподіли ω при гіпотезі H та рівні значущості α. В цьому випадку отримане значення порівнюється з табличним, і якщо отримане значення менше або рівне табличному, то нульова гіпотеза не відхиляється. Існують таблиці для асимптотичного розподілу та відповідних критичних значень статистики ω^2, вони представленні в роботах[3, 13].

Статистика критерія A^2 Андерсена – Дарлінга має вигляд

$$A^2 = n\int_{-\infty}^{\infty} \frac{\left\{F_n(x) - F(x)\right\}^2}{F(x)\left(1-F(x)\right)} dF(x).$$

Для зручності розрахунків використовують формулу

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{\ln z_{(i)} + \ln\left(1 - z_{(n+1-i)}\right)\right\} = -n - \frac{1}{n}\sum_{i=1}^{n}\left\{(2i-1)\ln z_{(i)} + (2n+1-2i)\ln\left(1-z_{(i)}\right)\right\}$$

Існують таблиці для асимптотичного розподілу та відповідних критичних значень статистики A^2, вони наведені в роботах [3, 13].

Методика застосування та обмеження для даних критеріїв аналогічні критерію Колмогорова.

Розглянемо модифікацію статистик критеріїв Колмогорова-Смірнова та ω^2 для вибірок з невеликим обсягом. Розподіл статистик D_n, D_n^+, D_n^- та ω^2 зі збільшенням обсягу n швидко наближається до граничного. Застосовувати граничний(асимптотичний) розподіл рекомендовано, якщо $n > 100$, але існують прості модифікації цих розподілів, які можна використовувати при невеликій кількості даних n. Ці модифікації представленні в табл. 1.3.

Таблиця 1.3

Модифікація статистик Колмогорова, ω^2

Базова статистика	Модифікована статистика
	Верхній(правий) розподіл
$\sqrt{n}D_n^+$	$\tilde{D}_n^+ = D_n^+\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right)$
$\sqrt{n}D_n^-$	$\tilde{D}_n^- = D_n^-\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right)$

$\sqrt{n}D_n$	$\tilde{D}_n = D_n\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right)$
ω_n^2	$\tilde{\omega}^2 = \frac{\omega_n^2 - \frac{0.4}{n} + \frac{0.6}{n^2}}{1 + \frac{1}{n}}$
	Нижній(лівий) розподіл
$\sqrt{n}D_n$	$\tilde{D}_n = D_n\left(\sqrt{n} + 0.275 - \frac{0.04}{\sqrt{n}}\right)$
ω_n^2	$\tilde{\omega}^2 = \frac{\omega_n^2 - \frac{0.03}{n} + \frac{0.6}{n^2}}{1 + \frac{0.5}{n}}$

При використанні відповідних критеріїв на основі модифікованих статистик $\tilde{D}_n^+, \tilde{D}_n^-, \tilde{D}_n, \tilde{\omega}^2$ можна користуватися лише граничним розподілом для $\sqrt{n}D_n^+, \sqrt{n}D_n^-, \sqrt{n}D_n, \omega_n^2$ відповідно[1].

Розглянемо випадок застосування критеріїв згоди Колмогорова-Смірнова та ω^2 коли параметри невідомі. Нехай T - загальне позначеня розглянутих вище критеріїв Колмогорова, ω^2 і A^2, t - значення критерію, обраховане за емпіричними даними. Нехай $p\frac{1}{2}$ - ймовірність

$$p = P(T > t) = 1 - \psi(t) = 1 - \psi(t, \theta),$$

де ψ - функція розподілу статистики критерію, де параметр θ - відомий, значення $p^* = P(T > t) = 1 - \psi^*(t) = 1 - \psi\left(t, \hat{\theta}\right)$ обраховане для функції розподілу з оціненими параметрами $\hat{\theta}$. Якщо отримане значення p - мале, то p^* - ще менше, а значить гіпотезу відхиляють.

Існують такі підходи для перевірки згоди з розподілом, коли параметри невідомі:

1. Якщо вибірка велика, то половину спостережень можна використати для оцінки параметрів, другу половину – для перевірки згоди.
2. Якщо вибірка велика, то половину спостережень можна використати для оцінки параметрів, потім всю вибірку – для перевірки згоди.
3. У випадку параметрів маштабу і зсуву можна використовувати критерії Колмогоровава, A^2, ω^2 при цьому критичні значення розподілу статистик не залежать від значень параметрів, проте залежать від типу гіпотетичного розподілу F.

Модифікації критеріїв Колмогоровава, A^2, ω^2 у випадку невідомих значень параметрів маштабу і зсуву, зокрема, для нормального і експоненційного розподілів вивчалися, наприклад, в роботах:[1, 11].

Далі детальніше розглянемо застосування критеріїв згоди Колмогорова-Смірнова та ω^2 у випадку невідомих параметрів (зокрема, при нормальному розподілі). Тобто, коли гіпотетична функція розподілу відома з точністю до параметрів і вони оцінюються за вибіркою, граничні розподіли статистик D_n, D_n^+, D_n^- та ω^2 вже не будуть вільними від розподілу. У цьому їх недолік в порівнянні з критерієм χ^2. Проте коли оцінці підлягають параметри зсуву і масштабу, то розподіл буде залежати тільки від форми розподілу $F(x)$, а не від його параметрів. Незалежність граничних розподілів від параметрів зсуву і масштабу дає можливість побудувати, наприклад на основі зазначених статистик, критерії перевірки нормального характеру розподілу.

Перевірка нормального (гаусівського) закону розподілу на основі критерію ω^2.

Нехай потрібно перевірити гіпотезу належності невідомої функції розподілу до класу нормального закону розподілу. Можна виділити три випадки.

1. Середнє (математичне сподівання) m невідомо, дисперсія σ відома. Значення m оцінюється середнім

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

2. Дисперсія невідома, середнє невідоме. Значення σ^2 оцінюється статистико

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - m\right)^2 .$$

3. Середнє і дисперсія невідомі. Значення m і σ^2 оцінюються відповідно

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i ; s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 .$$

Для кожного з трьох випадків існують таблиці граничних розподілів для статистики ω^2, які наведені в додатку 2.

Для розподілу статистики Колмогорова критичні значення для перевірки відповідності гауссівському (нормальному) закону у випадку невідомих значень математичного сподівання і дисперсії для різних n і рівнях значущості $\alpha = 0.05$ та $\alpha = 0.01$ наведені в таблиці, додаток 3.

1.2.3. Критерій χ^2 Пірсона

Критерій χ^2 Пірсона вважають найпопулярнішим та найбільш універсальним критерієм. Даний критерій дозволяє перевірити гіпотезу $H_0 : F_n(x) = F(x) = F(x,\theta)$, у випадках коли повністю відома теоретична функція розподілу $F(x) = F(x,\theta)$, або коли невідомий k-вимірний параметр $\theta = (\theta_1, \theta_2, ..., \theta_n)$ і значення $\theta_1, \theta_2, ..., \theta_n$ оцінюється за наявними спостереженнями (вибірковими даними). Критерій можна використовувати для неперервних і дискретних теоретичних розподілів.

Розглянемо випадок коли теоретична функція розподілу повністю відома $F(x) = F(x,\theta)$.

Для використання критерію χ^2 данні $x_1, x_2, ..., x_n$ необхідно згрупувати в l інтервалів $\Delta_i, i = \overline{1,l}$. Тоді припустимо, що : v_i - кількість даних, що потрапили в i-інтервал $\Delta_i, i = [a_{i-1}, a_i], \overline{1,l}$, $n = \sum_{i=1}^{l} \nu_i$ - загальна кількість спостережень, $p_i = F(a_i) - F(a_{i-1})$- ймовірність попадання в інтервал Δ_i. Статистика критерію χ^2 має вигляд:

$$\hat{\chi}^2 = \chi^2(l-1) = \sum_{i=1}^{l} \frac{(\nu_i - np_i)}{np_i} = \sum_{i=1}^{l} \frac{n}{p_i}\left(\frac{\nu_i}{n} - p_i\right)^2$$

і асимптотично (при $n \to \infty$) має розподіл χ^2 з $l-1$ степенями свободи.

Гіпотезу H_0 відхиляють при великих значеннях χ^2. Критичне значення (критичний рівень) $\chi^2_{кр} = z_\alpha$ визначається з умови $P\left(\chi^2_{l-1} < z_\alpha\right) = 1-\alpha$, тобто є квантилем розподілу χ^2 з $l-1$ степенями свободи.

Процедура статистичної перевірки гіпотези H_0 в даному випадку складається з таких етапів:

1. Увесь діапазон значень досліджуваної випадкової величини ξ необхідно розбити на ряд інтервалів групування $\Delta_i = [a_{i-1}, a_i], i = \overline{1,l}$ необов'язково однакової довжини. Це розбиття на інтервали необхідно підпорядкувати наступним умовам:
   a) загальна кількість інтервалів i повинна бути не менше 8[1];
   b) в кожен інтервал групування має потрапити не менше $7-10$ вибіркових значень ξ, причому бажано, щоб в різні інтервали потрапила приблизно однакова кількість точок;
   c) якщо діапазон досліджуваної випадкової величини ξ вся числова пряма, то крайні інтервали групування будуть напівпрямими.
2. Підрахувати ν_i - кількість даних, що попали в і- інтервал $\Delta_i = [a_{i-1}, a_i], i = \overline{1,l}$
3. Обчислити теоретичні ймовірністі попадання в інтервали Δ_i, а саме
$$p_i = F(a_i) - F(a_{i-1}).$$
4. Обчислити значення статистики χ^2 за формулою:

$$\chi^2 = \chi^2(l-1) = \sum_{i=1}^{l} \frac{(v_i - np_i)^2}{np_i} = \sum_{i=1}^{l} \frac{n}{p_i}\left(\frac{v_i}{n} - p_i\right)^2 .$$

5. Скориставшись розподілом χ^2 з $l-1$степенями свободи, для заданого рівня значущості α знайти відповідне критичне значення $\chi^2_{кр} = z_\alpha$.Порівняти значення статистики χ^2 та $\chi^2_{кр}$.
6. Зробити висновки: дані підтверджують гіпотезу H_0 про модельний розподіл $F(x,\theta)$, при $\chi^2 < \chi^2_{кр}$. Гіпотезу H_0про модельний розподіл $F(x,\theta)$ відхиляють при $\chi^2 > \chi^2_{кр}$. Іноді гіпотезу H_0відхиляють і при дуже малих значеннях χ^2.

Інший спосіб перевірки гіпотези H_0: щоб не порівнювати χ^2та $\chi^2_{кр}$, необхідно обчислити (якщо рівень значущості α відомий, знайти за таблицями для розподілу χ^2з $l-1$степенями свободи, таблиці представлені в роботі [6]) ймовірність $P = P\left(\chi^2 > \chi^2_{кр}\right)$. Якщо P велике (більше за вибраний рівень значущості α) то вважають, що данні підтверджують гіпотезу H_0про модельний розподіл $F(x,\theta)$ якщо ж P- мале (менше за вибраний рівень значущості α), то данні відхиляють гіпотезу H_0.

Розглянемо випадок, коли k параметрів теоретичного розподілу оцінюються за вибіркою. Процедура статистичної перевірки гіпотези H_0 аналогічна, що й у випадку коли теоретична функція розподілу повністю відома, слід лише виконати корекцію степеней свободи з урахуванням кількості параметрів, значеня яких оцінюється за вибіркою.

Процедура застосування критерію χ^2 у випадку, коли k параметрів теоретичного розподілу оцінюються за вибіркою:

1. На основі вибіркових даних (спостережень) $x_1, x_2, ..., x_n$ знаходяться оцінки $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k\right)$ невідомих параметрів $\theta = \left(\theta_1, \theta_2, ..., \theta_k\right)$, від яких залежить функція розподілу $F(x, \theta)$;

2. Увесь діапазон значень досліджуваної випадкової величини ξ розбивають на l інтервалів групування $\Delta_i = \left[a_{i-1}, a_i\right], i = \overline{1, l}$ і підраховують v_i - кількість даних, що попали в і-інтервал Δ_i;

3. Обчислюють теоретичні ймовірності попадання в інтервали Δ_i:

$$p_i = p_i\left(\hat{\theta}\right) = F\left(a_i, \hat{\theta}\right) - F\left(a_{i-1}, \hat{\theta}\right).$$

4. Обчислюють значення статистики χ^2 за формулою

$$\chi^2 = \chi^2\left(l - k - 1\right) = \sum_{i=1}^{l} \frac{\left(v_i - np_i\right)^2}{np_i} = \sum_{i=1}^{l} \frac{n}{p_i}\left(\frac{v_i}{n} - p_i\right)^2 .$$

5. Скориставшись розподілом χ^2 з $l - k - 1$ степенями свободи, для заданого рівня значущості α знаходять відповідне критичне значення $\chi^2_{кр} = z_\alpha$ і порівнюють значення статистики χ^2 та $\chi^2_{кр}$ якщо $\chi^2 < \chi^2_{кр}$ данні підтверджують гіпотезу H_0 про модельний розподіл $F(x, \theta)$; якщо $\chi^2 > \chi^2_{кр}$ гіпотезу H_0 про модельний розподіл $F(x, \theta)$ відхиляють.

Теоретичним підгрунтям цієї процедури є гранична теорема щодо асимптотики модифікованої статистики хі-квадрат, згідно з якою, у випадку, коли оцінки $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k\right)$ є оцінками максимальної вірогідності в поліноміальній схемі випробувань, що задовольняють ще деяким додатковим умовам, статистика

$$\chi^2 = \chi^2\left(l - k - 1\right) = \sum_{i=1}^{l} \frac{\left(v_i - np_i\left(\hat{\theta}\right)\right)^2}{np_i\left(\hat{\theta}\right)} = \sum_{i=1}^{l} \frac{n}{p_i\left(\hat{\theta}\right)}\left(\frac{v_i}{n} - p_i\left(\hat{\theta}\right)\right)^2$$

асимптотично (при $n \to \infty$) має розподіл χ^2 з $l - k - 1$ степенями свободи.

У випадку, коли оцінки $\hat{\theta}=\left(\hat{\theta}_1,\hat{\theta}_2,...,\hat{\theta}_k\right)$ - це звичайні оцінки максимальної вірогідності, статистика χ^2 вже не буде асимптототично розподіленою за законом χ^2, проте дійсний розподіл χ^2 знаходиться між $\chi^2\left(k-1\right)$ і $\chi^2\left(k-l-1\right)$, і при зростанні k різниця між ними стає настільки малою, що на неї можна не зважати. При відносно малих k варто впевнитися, що χ^2 більше за обидва критичних значення $\chi^2_{кр}\left(l-k-1\right)$ та $\chi^2_{кр}\left(l-1\right)$.

Переваги і недоліки: можна використовувати для неперервних і дискретних теоретичних розподілів; необхідно багато даних, оскільки перевірка відбувається по згрупованим даним.

Приклад 1.3. Параметри відомі, ідея розв'язання взята з роботи [16].

У табл. 1.4 наведено $n=500$ спостережень про час безвідмовної роботи деякого обладнання. Данні згруповані в $l=12$ інтервалів. Є припущення , що T має рівномірний розподіл на відрізку $\left[0,12\right]$. Необхідно перевірити цю гіпотезу, скориставшись критерієм хі-вадрат.

Таблиця 1.4

Номер Інтервалу	інтервал	Кількість відмов $v_i = n_i$	np_i	$\left(v_i - np_i\right)^2$
1	$\left[0,1\right)$	41	41.7	$\left(41-41.7\right)^2=0.49$
2	$\left[1,2\right)$	34	41.7	$\left(34-41.7\right)^2=59.29$
3	$\left[2,3\right)$	54	41.7	$\left(54-41.7\right)^2=151.29$
4	$\left[3,4\right)$	39	41.7	$\left(39-41.7\right)^2=7.29$
5	$\left[4,5\right)$	49	41.7	$\left(49-41.7\right)^2=53.29$
6	$\left[5,6\right)$	45	41.7	$\left(45-41.7\right)^2=10.89$

7	$[6,7)$	41	41.7	$(41-41.7)^2=0.49$
8	$[7,8)$	33	41.7	$(33-41.7)^2=75.69$
9	$[8,9)$	37	41.7	$(37-41.7)^2=22.09$
10	$[9,10)$	41	41.7	$(41-41.7)^2=0.49$
11	$[10,11)$	47	41.7	$(47-41.7)^2=28.09$
12	$[11,12)$	39	41.7	$(39-41.7)^2=7.29$
$\sum$		500		416.68

Щоб скористатися критерієм слід, знайти теоретичні ймовірності попадання значень T в i-ий інтервал групування

$$p_i = p_i\left(\hat{\theta}\right) = F\left(a_i,\hat{\theta}\right) - F\left(a_{i-1},\hat{\theta}\right);\ p_i = P_{(i\le T\le i+1)} = F_{i+1} - F_i = \frac{(i-1)-i}{12} = \frac{1}{12};$$

$$(v_i - np_i)^2 = 500/12 = 41.7\,,$$

ці дані записані в 4-му стовпчику. Далі знаходимо $(v_i - np_i)^2$ і значення статистики

$$\chi^2 = \chi^2(l-1) = \chi^2(11) = \sum_{1}^{12}\frac{(v_i - np_i)^2}{np_i} = \frac{416.68}{41.7} = 9.99\,.$$

Отримане значення необхідно порівняти з критичним, яке знаходиться за таблицями для розподілу χ^2 з $l-1=11$ степенями свободи в залежності від рівня значущості α. В данному випадку покладемо $\alpha = 0.05$. Тоді

$$\chi^2_{кр} = \chi^2(l-1,\alpha) = \chi^2(11,0.95) = 19.68\,.$$

Бачимо, що $\chi^2 < \chi^2_{кр}$ $(9.99 < 19.68)$ отже, немає підгрунття для відхилення гіпотези про рівномірний розподіл на $[0,12]$, данні підтверджують цю гіпотезу.

Приклад 1.4

Дано 53 даних. З’ясувати, чи можна вважати, що ці спостереження мають нормальний розподіл, користуючись критерієм χ^2:

178,170,167,165,164,164,163,161,157,175,170,167,164,164,163,161,158,161,158,151,172,167,166,163, 163,161,157,174,168,166,165,164,164,162,160,155,173,168,166,165,164,164,162,160, 154,173,167,166,165,164,164,164,165.

Згрупуємо дані: $x_{\min} = 151$; $x_{\max} = 178$ - це границі. Відрізок $(151,178)$ довжиною $27 \approx 30$ розб’ємо на 6 інтервалів. Довжину інтервалу оберемо $\Delta = 5, l = 6$. Оскільки параметри розподілу невідомі - знайдемо їх: $x = 164.4$, $S = 5.14$. Обчислимо значення

$$p_i = F(x_i) - F(x_{i-1}) = \Phi\left(\frac{x_i - x}{S}\right) - \Phi\left(\frac{x_{i-1} - x}{S}\right) = \Phi\left(\frac{x_i - 164.4}{5.14}\right) - \Phi\left(\frac{x_{i-1} - 164.4}{5.14}\right).$$

Отримані значення представлені у вигляді табл. 1.5

Таблиця 1.5

Номер інтерва-лу	Інтервал $[x_{i-1}, x_i)$	$v_i = n_i$	np_i	$\approx np_i$	p_i	p_i
1	$[151,156)$	3	2.7286	3	$\Phi(-8.4/5.14) - \Phi(-13.4/5.14)$	0.051
2	$[156,161)$	10	10.811	11	$\Phi(-3.4/5.14) - \Phi(-8.14/5.14)$	0.204
3	$[161,166)$	26	19.529	20	$\Phi(1.6/5.14) - \Phi(-3.4/5.14)$	0.368
4	$[166,171)$	8	14.701	15	$\Phi(6.6/5.14) - \Phi(1.6/5.14)$	0.278
5	$[171,176)$	5	4.6028	5	$\Phi(11.6/5.14) - \Phi(6.6/5.14)$	0.088

6	$\left[176{,}181\right)$	1	0.6274	1	$\Phi(16.6/5.14)-\Phi(11.6/5.14)$	0.011
		53				

Вирішили об'єднати деякі інтервали. Нехай далі $l=4$. Знаходимо $(v_i-np_i)^2$ та значення статистики

$$\chi^2=\chi^2(l-k-1)=\chi^2(4-2-1)=\sum_1^4\frac{(v_i-np_i)^2}{np_i}=5.14.$$

Обраховані значення представлені у табл. 1.6.

Таблиця 1.6

Номер Інтервалу	Інтервал $\left[x_{i-1},x_i\right)$	$v_i=n_i$	np_i	$\approx np_i$	v_i-np_i	$\frac{(v_i-np_i)^2}{np_i}$
1	$\left[151{,}161\right)$	13	13.539	14	-1	$1/14=0.071$
2	$\left[161{,}166\right)$	26	19.529	20	6	$36/20=1.8$
3	$\left[166{,}171\right)$	8	14.701	15	-7	$49/15=3.266$
4	$\left[171{,}181\right)$	6	5.230	6	0	0
		53				$5.137=5.140$

Розраховане значення статистики 5.14 необхідно порівняти з критичним, яке знаходиться за таблицями для розподілу χ^2 з $(l-k-1)=1$ степенями свободи в залежності від рівня значущості α. В данному випадку покладемо $\alpha=0.05$, $(p=0.95)$ і отримаємо $\chi^2_{кр}=\chi^2(l-k-1,\alpha)=\chi^2(1{,}0.95)=3.84$.

Бачимо, що $\chi^2 > \chi^2_{кр}\left(5.14 > 3.84\right)$, отже, розподіл не можна вважати нормальним, данні не підтверджують цю гіпотезу.

Висновок до розділу 1

Критерії згоди дозволяють розв’язувати задачі, в яких необхідно перевірити гіпотезу про те, що досліджувана випадкова величина підпорядковується тому чи іншому розподілу ймовірностей. Для цього обчислюють статистику, на основі якої роблять висновок про сформовану гіпотезу, порівнюючи отримане значення з критичним, що знаходиться за таблицями. Найбільш вживаними критеріями згоди є критерій Колмогорова і критерій χ^2. Кожен з них має свої недоліки і переваги. Так, для статистики Колмогорова відомий точний і асимптотичний розподіли, що дало змогу розрахувати критичні значення як для «великих», так і для «малих» виборок, ці критичні значення наведені в спеціальних таблицях, які легко знайти в різноманітних джерелах. Проте сфера застосування критерія Колмогорова дещо обмежена: його можна застосовувати лише у випадку неперервних повністю відомих теоретичних розподілів. Критерій χ^2 вільний від цих обмежень, але для його коректного застосування необхідно мати багато даних, оскільки перевірка відбувається по згрупованим даним. Процедури застосування критеріїв згоди проілюстровані на конкретних прикладах цього розділу. В додатках наведені критичні значення.

РОЗДІЛ 2

КРИТЕРІЇ ОДНОРІДНОСТІ

Перевірка однорідності двох і більше виборок.

Загальна постановка задачі:

Нехай маємо $k \geq 2$ незалежних вибірок, що містять $n_1, n_2, ..., n_k$ незалежних спостережень $\left(x_1, ..., x_{n_1}\right); \left(x_2, ..., x_{n_2}\right); ... \left(x_k, ..., x_{n_k}\right)$.

Гіпотеза однорідності: вибіркам, які відібрані з однієї генеральної сукупності відповідають однакові функції розподілу. Отже, $H_0 : F_1\left(x\right) = F_2\left(x\right) = \ldots = F_k = F\left(x\right)$, де $F_i\left(x\right)$– функція розподілу i-тої генеральної сукупності. Найпоширеніший випадок при $k = 2$, розглянемо його.

2.1. Критерії однорідності, які базуються на емпіричних функціях розподілу

Нехай $k = 2$, тобто розглядається випадок критеріїв однорідності двох вибірок. $\left(x_1, ..., x_{n_1}\right); \left(y_1, ..., y_{n_2}\right)$ - дві вибірки обсягом n_1 та n_2 з функціями розподілу $F_1\left(x\right)$ та $F_2\left(x\right)$; $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n_1)}$; $y_{(1)} \leq y_{(2)} \leq ... \leq y_{(n_2)}$– відповіді варіаційні ряди, а $F_1^{(n_1)}\left(x\right)$та $F_1^{(n_2)}\left(x\right)$ - емпіричні функції розподілу. Будемо вважати, що функції розподілу $F_1\left(x\right)$ та $F_2\left(x\right)$ неперервні.

В даному випадку перевіряється гіпотеза $H_0 : F_1\left(x\right) = F_2\left(x\right) = F\left(x\right)$, як альтернативу розглядають $\mathrm{H}_1 : F_1\left(x\right) \neq F_2\left(x\right)$.

Часто використовують двовибіркові аналоги статистик Смірнова-Колмогорова, а саме:

$$D^+_{n_1,n_2} = \sup_x \left\{F_1^{(n_1)}\left(x\right) - F_2^{(n_2)}\left(x\right)\right\} = \max_{1 \leq j \leq n_2} \left\{\frac{j}{n_2} - F_1^{(n_1)}\left(y_{(j)}\right)\right\} = \max_{1 \leq j \leq n_1} \left\{F_2^{(n_2)}\left(x_{(i)}\right) - \frac{\left(i-1\right)}{n_1}\right\}$$

$$D^-_{n_1,n_2} = \sup_x \left\{F_2^{(n_2)}\left(x\right) - F_1^{(n_1)}\left(x\right)\right\} = \max_{1 \leq j \leq n_2} \left\{F_1^{(n_1)}\left(y_{(j)}\right) - \frac{j-1}{n_2}\right\} = \max_{1 \leq j \leq n_1} \left\{\frac{i}{n_1} - F_2^{(n_2)}\left(x_{(i)}\right)\right\},$$

$$D_{n_1,n_2} = \max\left\{D^{+}_{n_1,n_2}; D^{-}_{n_1,n_2}\right\} \qquad (3.1)$$

У випадку коли гіпотеза H_0 підтверджується розподіли статистик $D^{+}_{n_1,n_2}$ та $D^{-}_{n_1,n_2}$ рівні, тому далі розглядають статистику D_{n_1,n_2}.

Нехай гіпотеза H_0 вірна, $n_2 \le n_1$, $n_2 \to \infty$, а $n_0 = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$. Тоді $\sqrt{n_0} D^{+}_{n_1,n_2}$; $\sqrt{n_0} D^{-}_{n_1,n_2}$; $\sqrt{n_0} D_{n_1,n_2}$, мають ті ж самі розподіли, що і одновибіркові статистики, тобто $S(a)$ та $K(a)$ з розділу 2.1

$$\lim_{n\to\infty} P\left(\sqrt{n} D_n < a\right) = K(a) = 1 - 2\sum_{k=1}^{n} (-1)^{k-1} \exp\left(-2k^2 a^2\right), a > 0;$$

$$\lim_{n\to\infty} P\left(\sqrt{n} D_n^{+} < a\right) = \lim_{n\to\infty} P\left(\sqrt{n} D_n^{-} < a\right) = S(a) = 1 - \exp\left(-2a^2\right), a > 0.$$

При малих n_1 і n_2 $(n_1 = n_2 = n)$ маємо:

$$P\left(D^{+}_{(n)} \le \frac{k}{n}\right) = 1 - \frac{C_{2n}^{n+k+1}}{C_{2n}^{n}};\ P\left(D_{(n)} \le \frac{k}{n}\right) = 1 - \sum_{i=-r}^{r} (-1)^i \frac{C_{2n}^{n+i(k+1)}}{C_{2n}^{n}}, r = \left[\frac{n}{k+1}\right].$$

Багато цікавих фактів стосовно двовибіркових статистик Смірнова-Колмогорова наведено в [1, 10].

Таблиці точних розподілів для статистик Смірнова-Колмогорова в загальному випадку наведені у роботах [3,13, 15].

Для статистики $\sqrt{n_0} D_{n_1,n_2}$ критична область (область відхилення гіпотези H_0) - область великих значень статистики $\sqrt{n_0} D_{n_1,n_2} > d_{кр} = d_{\alpha}$, де d_{α}- критична точка розподілу статистики при рівні значущості α.

Для статистики $D^{+}_{n_1,n_2}$ $\left(D^{-}_{n_1,n_2}\right)$ критична область - це також область великих значень. Але статистика $D^{+}_{n_1,n_2}$ $\left(D^{-}_{n_1,n_2}\right)$ може набувати від'ємних значень. Це означає, що одна з цих емпіричних функцій більша за іншу на всьому інтервалі спостереження, а при великих значеннях n_1 та n_2 це не сумісно з нульовою гіпотезою про рівність граничних функцій розподілу. В такому випадку вважають, що $n_1 = n_2$ тоді

$$P\left(D^{+}_{(n_1)} \le 0\right) = 1 - \frac{C^{n_1+1}_{n_1}}{C^{n_1}_{2n_1}} = \frac{1}{n_1},$$

де $D^{+}_{(n_1)} = D^{+}_{n_1,n_2}$. При $n_1 \ge 20$ ця ймовірність менша 0.05. Тому якщо $n_1, n_2 > 20$ область від'ємних значень відповідних статистик вважають критичною для прийняття нульової гіпотези.

Зауваження. На практиці часто застосовують розподіли з дискретними випадковими величинами або з групованими даними. В таких випадках для статистик $D^{+}_{n_1,n_2}$, D_{n_1,n_2} використовують формули (3.1), але рівень значущості нульової гіпотези буде менше заданого. Тобто ймовірність відхилення нульової гіпотези по згрупованим даним при однаковому обсягу менше, ніж по незгрупованим.

Коли число вибірок $k > 2$ і обсяг вибірок однаковий $n_1 = n_2 = \ldots = n_k = n$, застосовують узагальнений критерій Смірнова-Колмогорова, який базується на статистиках

$$D^{k}_{n} = \max_{i,j} \sup_{x} \left| F^{(n)}_{i}(x) - F^{(n)}_{j}(x) \right|.$$

Для випадку, коли $k = 3$ розподіл статистики та відповідні критичні значення наведені в роботі [13].

2.2. Критерій однорідності χ^2

Критерій однорідності χ^2 використовують, коли дані згруповані. Нехай є $k \ge 2$вибірок обсягом $n_1, n_2, \ldots, n_k$, і дані кожної вибірки згруповані в r інтервалів групування. Перевіряється гіпотеза $H_0 : F_1(x) = F_2(x) = F(x)$, як альтернативу розглядають $\mathrm{H}_1 : F_1(x) \ne F_2(x)$.

Позначимо: v_{ij} - кількість спостережень з j-ої вибірки, які потрапили в i-й інтервал

$$v_{i.}=\sum_{j=1}^{k}v_{ij}\ ;\ v_{.j}=\sum_{j=1}^{k}v_{ij}\ ;\ n=n_1+n_2+...+n_k\,.$$

Статистикою критерія є величина:

$$\chi^2=\chi^2\left(\left(k-1\right)\left(r-1\right)\right)=n\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{\left(v_{ij}-v_{i.}v_{.j}/n\right)^2}{v_{i.}v_{.j}}=n\sum_{i=1}^{r}\sum_{j=1}^{k}\frac{v_{ij}^2}{v_{i.}v_{.j}}-1.$$

При $k=2$ формула набуває вигляду

$$\chi^2=\chi^2\left(r-1\right)=n_1\cdot n_2\sum_{i=1}^{r}\frac{\left(\frac{v_{i_1}}{n_1}-\frac{v_{i_2}}{n_2}\right)^2}{v_{i_1}+v_{i_2}}\,.$$

При виконанні гіпотези H_0 статистика $\widehat{\chi^2}$ асимптотично має розподіл χ^2 з $(k-1)(r-1)$ степенями свободи (для $k=2$ степінь свободи $(r-1)$).

Зауважимо, що для використання критерія однорідності χ^2 необхідна велика кількість даних, оскільки перевірка робиться по згрупованим даним.

Приклад 2.1. Аналогічний приклад розглянуто в [1].

В табл. 2.1 (2 – 4 стовп.) наведені данні щодо заробітної плати фахівців 2-х галузей промисловості (по 100 працівників кожної галузі).

Необхідно перевірити гіпотезу про однорідність цих 2-х вибірок (розподіли заробітної плати для 2-х галузей однакові). Дані згруповані у 8 інтервалів групування.

Для розв’язання скористаємось критерієм χ^2. За умовою задачі $n_1=n_2=100$; $r=8$, $k=2$. Нехай u_i - кількість працівників першої промисловості, v_i - кількість працівників другої промисловості. Необхідні обчислення наведені в таблиці.

Таблиця 2.1

i	Місячна зарплата	u_i	v_i	u_i+v_i	u_i-v_i	$(u_i-v_i)^2$	$\frac{(u_i-v_i)^2}{u_i+v_i}$
1	$1300-1500$	4	1	5	3	9	1.8
2	$1500-1700$	4	1	5	3	9	1.8
3	$1700-2000$	15	8	23	7	49	2.13
4	$2000-2500$	51	43	94	8	64	0.68
5	$2500-3000$	22	34	56	-12	144	2.57
6	$3000-3500$	3	7	10	-4	16	1.6
7	$3500-4000$	1	3	4	-2	4	1
8	$4000-5000$	0	3	3	-3	9	3
$\sum$		$\underline{100}$	$\underline{100}$				$\underline{14.582}$

Обчислимо значення статистики χ^2. Скористаємось формулою

$$\chi^2 = \chi^2(r-1) = n_1 \cdot n_2 \sum_{i=1}^{r} \frac{\left(\frac{u_i}{n_1} - \frac{v_i}{n_2}\right)^2}{u_i+v_i} = 14.58 .$$

Отримане значення необхідно порівняти з критичним, яке знаходимо за таблицями для розподілу χ^2 з $r-1=7$ степенями свободи в залежності від рівня значущості α. В даному випадку покладемо $\alpha = 0.05$. Тоді $\chi^2_{кр} = \chi^2_{кр}(r-1,\alpha) = 14.08$.

Бачимо, що $\chi^2 > \chi^2_{кр}$ $(14.58 > 14.07)$, отже, немає підґрунтя для прийняття гіпотези про однорідність вибірок (цю гіпотезу варто відхилити, дані не підтверджують цю гіпотезу).

2.3. Рангові критерії однорідності. Критерій Вілкоксона - Манна - Уітні

В основі рангових критеріїв однорідності лежить використання номерів спостереження у варіаційному ряді, отримане після впорядкування об'єднаної вибірки обсягом n.Номер, який отримує спостереження x_i в упорядкованій вибірці називають його рангом і позначають R_i, функцію від рангу позначають $\varphi(R_i)$ і часто називають «міткою».

Нехай $(x_1,...,x_{n_1});(y_1,...,y_{n_2})$ – дві вибірки обсягом n_1 та n_2 з функціями розподілу $F_1(x)$ та $F_2(x)$. Об'єднаємо їх, поклавши $x_{n_1+j}=y_i$, впорядкуємо і отримаємо новий варіаційний ряд $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$; $n=n_1+n_2$.

Розглянемо лінійні рангові критерії, статистики яких мають вигляд:

$$K=\sum_{i=1}^{n_j}\phi(R_i),$$

де додавання виконується за елементами тільки першої вибірки або тільки другої вибірки. Далі, для визначеності, вважаємо, що додавання ведеться тільки за першою вибіркою. Перевіряється гіпотеза: $H_0: F_1(x)=F_2(x)$, як альтернативу розглядають $H_1: F_1(x)=F_2(x-\Delta)$, причому Δ: або $\Delta \ne 0$, або $\Delta > 0$, або $\Delta < 0$, де $F_2(x)$ більш менш відома, тобто обидві функції розподілу однакові за виключенням зсуву Δ, тому їх часто називають «критеріями зсуву», хоча і застосовують для перевірки однорідності і у більш загальних випадках.

Критерій Вілкоксона-Манна-Уітні є найпоширенішим критерієм однорідності. Статистика цього критерію має вигляд:

$$S=\sum_{i=1}^{n_1}R_i.$$

Мітками є самі ранги спостереження. Інколи статистику S називають статистикою суми рангів.

Часто використовують еквівалентну статистику:

$$U=S_*-\frac{1}{2}n_1(n_1+1).$$

Якщо гіпотеза H_0 підтверджується, то

$$ES_* = \frac{1}{2}(n_1 + n_2 + 1)n_1; \; EU = \frac{1}{2}n_1 n_2; \; VarS_* = EU = \frac{1}{12}(n_1 + n_2 + 1)n_1 n_2.$$

Критичні значення можна знайти, наприклад, в роботі [16]. Також часто використовують еквівалентну статистику

$$U = \frac{n_1 n_2 + n_1(n_1 + 1)}{2 - S_*}.$$

Критичні значення для статистики табульовані і наведені в [13,17].

Обрахувати статистику також можна використовуючи таку еквівалентну статистику, яка знаходиться за алгоритмом (в даному випадку алгоритм складений для двох вибірок):

1. Дані вибірок об'єднують у один ряд та розташовують їх у порядку зростання, присвоюють кожному числу ранг.
2. Знаходять суму рангів (R_1 та R_2). Обчислюють емпіричне значення критерія за формулою: $U_{емп} = n_1 n_2 + \frac{n_x(n_x + 1)}{2} - R_x$, де n_1 – кількість елементів першої вибірки; n_2– кількість елементів другої вибірки; $R_x = \max(R_1, R_2)$.
3. Порівнюють обчислене значення $U_{емп}$та знайдене у таблиці $U_{кр}$. Якщо $U_{кр} > U_{емп}$, нульову гіпотезу відхиляють.

При $n_1, n_2 \to \infty$ розподіл статистики S (і U) збігається до нормального, причому збіжність дуже швидка, і користуватися асимптотичною нормальністю варто вже при $n_1, n_2 > 8$.

Критерій Вілкоксона - Манна - Уітні вважають локально найбільш потужним для розподілів логістичного типу, та простим у використанні.

Існують певні обмеження даного критерію: вибірки, результати яких порівнюються мають бути незалежними, мінімальний обсяг вибірок – по 3 одиниці, але якщо в одній вибірці 2 одинці, у іншій має бути мінімум 5.

Існує і інша форма критерія Вілкоксона-Манна-Уітні. Її застосовують, коли є n пар спостережень (x_i, y_i), (т.т. вибірки залежні), тоді знаходять різниці $d_i = x_i - y_j$ та впорядковують їх абсолютні величини $|d_i| = d_i^* : d_{(1)}^* \le d_{(2)}^* \le \ldots \le d_{(n)}^*$ і знаходять відповідні ранги $|R_i|$ потім кожному рангу приписують знак відповідної різниці $sign(d_i)$. Нехай T_n - сума додатних рангів. Розподіл T_n, тобто ймовірності $P(T_n \le a)$ табульовано. Перевіряється гіпотеза $H_0 : F_1(x) = F_2(x)$ проти альтернативи $H_1 : P(X > Y) = b > 0$. Якщо a^* - значення статистики T_n, обчислине за вибіркою, то H_0 відхиляють, якщо $P(T_n \ge a^*) \ge 1 - \alpha$, де α - рівень значущості.

2.4. Критерій нормальних міток Фішера-Йєтса-Террі-Гефдінга

Статистика критерію

$$C = \frac{1}{n_1} \sum_{i=1}^{n_1} E(R_i, n), n = n_1 + n_2,$$

де $E(r,n)$ - математичне сподівання r-ї порядкової статистики у вибірці обсягом n зі стандартного нормального розподілу $N(0,1)$. Щільність розподілу r-ї порядкової статистики у вибірці обсягом n з розподілу зі щільністю:

$$f_{r,n}(x) = \frac{n!}{r!(n-r-1)} \{F(x)\}^{r-1} \{1 - F(x)\}^{n-r} f(x).$$

Тоді для стандартного нормального закону $N(0,1)$ маємо щільність розподілу r-ї порядкової статистики

$$\phi_{r,n}(x) = \frac{n!}{r!(n-r-1)} \{\Phi(x)\}^{r-1} \{1 - \Phi(x)\}^{n-r} e^{-x^2/2},$$

відповідно математичне сподівання r-ї порядкової статистики дорівнює

$$E(r,n)(x) = \frac{n!}{r!(n-r-1)} \int_{-\infty}^{\infty} x \{\Phi(x)\}^{r-1} \{1 - \Phi(x)\}^{n-r} e^{-x^2/2} dx.$$

Має місце також апрксимація

$$E(r,n)(x) \approx \Phi(x)^{-1}(r/(n+1)).$$

Якщо вірна гіпотез H_0, то статистика C має симетричний розподіл з

$$EC = 0; VarC = \frac{n_2}{n(n-1)n_1} \sum_{r=1}^{n} [E(r,n)]^2 .$$

Для великих $n_1, n_2 VarC \approx \frac{n_2}{n \cdot n_1}$. При $n_1, n_2 \to \infty$ розподіл статистики збігається до нормального.

Критерій Фішера-Йєтса-Террі-Гефдінга асимптотично оптимальний для розподілів нормального типу, має ту ж потужність, що і t-критерій Стьюдента.

2.5. Критерій Ван дер Вардена

Статистика критерію

$$V = \sum_{i=1}^{n_1} \Phi^{-1}\left(\frac{R_i}{n_1+n_2+1}\right) \text{ або } V^* = \frac{1}{n_1}\sum_{i=1}^{n_1} \Phi^{-1}\left(\frac{R_i}{n_1+n_2+1}\right),$$

де Φ^{-1} (.) - функція, обернена до Φ(.) – функції розподілу стандартного нормального закону.

Критерій V еквівалентний

$$C = \frac{1}{n_1}\sum_{i=1}^{n_1} E(R_i, n)$$

і асимптотично оптимальний для альтернатив з нормальною функцією розподілу. Розподіл статистики V симетричний, асимптотично нормальний, його математичне сподівання і дисперсія дорівнюють

$$EV = 0; VarV = \frac{n_2}{n\cdot(n-1)\cdot n_1}\sum_{i=1}^{n_1}\left[\Phi^{-1}\left(\frac{r}{n+1}\right)\right]^2 .$$

2.6. Медіанний критерій

Статистика критерію

$$M = \sum_{i=1}^{n_1} \frac{1}{2}\left[\sin g\left(R_i - \frac{1}{2}\left(n_1 + n_2 + 1\right)\right) + 1\right],$$

що дорівнює $\frac{1}{2}$ плюс кількість спостережень з 1-ї вибірки, які більші за медіану об'єднаної вибірки, тоді і тільки тоді, коли n непарне і медіана належить 1-й вибірці. При $n_1, n_2 \geq 10$ розподіл статистики асимптотично нормальний, математичне сподівання і дисперсія дорівнюють

$$EM = \frac{n_1}{2}\,;\; VarM = \begin{cases} \dfrac{n_1 n_2}{4\left(n_1 + n_2 - 1\right)}, \text{якщо}\, n_1 + n_2 - парне \\ \dfrac{n_1 n_2}{4\left(n_1 + n_2\right)}, \text{якщо}\, n_1 + n_2 - непарне \end{cases}.$$

Загальна схема застосування рангових критеріїв однорідності.

Для кожної із статистик K розділу нормована величина $A_K = \left(K - EK\right)/\sqrt{VarK}$ при H_0 має асимптотично стандартний нормальний розподіл $N(0,1)$.При застосуванні вище наведених критеріїв можна скористатися алгоритмом:

1. Вибірки об’єднуються і об’єднана вибірка впорядковується в порядку зростання значень спостереження.
2. За рангами першої (або другої) вибірки обчислюється величина статистики критерія K (це може бути значення або S , або V , або C , або M) потім обчислюється величина $A_K = \left(K - EK\right)/\sqrt{VarK}$.
3. Значення A_K порівнюється з критичним, яке дорівнює квантилю стандартного нормального закону (або обчислюється значення функції стандартного нормального розподілу $\Phi(A_K)$). В залежності від альтернативної гіпотези критична область відхилення гіпотези H_0 для рівня значущості α визначається нерівностями: $\left|A_K\right| < u_\alpha$ або $\left|A_K\right| \geq u_{1-\alpha}$ для

односторонніх альтернатив $\Delta < 0$ або $\Delta > 0$ відповідно; $|A_K| \geq u_{1-\alpha/2}$ – для двосторонніх альтернатив.

4. Якщо величина обсягу вибірок мала, то отримання більш точного результату використовують таблиці критичних значень відповідних статистик, які представлені в роботах [3, 15].

Серед розглянутих рангових критеріїв проти альтернатив зсуву при достатньо великих значеннях n найбільш потужним є критерій нормальних міток, а найменш потужним – Вілкоксона.

2.7. Двовибіркові критерії маштабу. Рангові критерії перевірки гіпотез про рівність дисперсій

Нехай $(x_1,...,x_{n_1});(y_1,...,y_{n_2})$– дві вибірки обсягом n_1 та n_2 з функціями розподілу $F_1(x)$ та $F_2(x)$. Об'єднаємо їх, поклавши $x_{n_1+j} = y_i$ впорядкуємо і отримаємо новий варіаційний ряд $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$; $n = n_1 + n_2$ та визначимо ранги спостережень вже в об'єднаній вибірці. Перевіряється гіпотеза $H_0 : F_1(x) = F_2(x)$ проти альтернативи

$$H_1 : F_1(x) = F_2(x/\tau) \text{ або } H_1 : F_1(x) = F_2((x-\mu)/\tau);\ F_2(x) = F_2(x-\mu),$$

де τ - параметр маштабу, що підлягає перевірці, μ - параметр, що заважає,тобто обидві функції розподілу (щільності) одного типу, мають однаковий зсув μ, при альтернативі різняться параметрами маштабу. Дуже часто параметр маштабу - це дисперсія σ^2(або середня квадратична похибка σ) , тоді альтернативна гіпотеза має вигляд:

$$H_1 : F_1(x) = F_2(x/\tau);\ \tau = \frac{\sigma_1}{\sigma_2} \neq 1$$

двостороння альтернатива, при одностороній альтернативі $\tau > 1$ або $\tau < 1$.

Для перевірки гіпотез про рівність дисперсій використовують критерій Клотца, що є аналогом критерія Ван дер Вардена, та критерій Муда, що є аналогом критерія Вілкоксона.

Висновок до розділу 2

Критерії однорідності - це критерії перевірки гіпотез, які дозволяють перевірити гіпотезу про те, що двом (або більше) вибіркам відповідають однакові функції розподілу.

Змістовно це може означати перевірку припущень про вплив (або відсутність впливу) деякого чинника (фактора) чи його рівнів на результуючу характеристику об’єкта дослідження (явища або процесу).

Один підхід до розв’язання цієї задачі базується на використанні різниць між емпіричними функціями розподілу, що відповідають різним вибіркам (критерій Смірнова-Колмогорова). Вказаний критерій досить точний і може застосовуватися для широкого кола альтернативних гіпотез, проте він має і свої обмеження: теоретичні (але невідомі) розподіли мають бути неперервними. Цього недоліка не має критерій однорідності χ^2, але він вимагає великої кількості даних. Альтернативний підхід полягає у застосуванні рангових критеріїв, вони прості у застосуванні, особливо у випадку малих вибірок, не вимагають складних обчислень, можуть бути застосовані не тільки за відсутності інформації і додаткових припущень щодо виду розподілу, але і у випадку, коли спостереження не є кількісними, а можуть бути тільки впорядковані, як це часто буває в медицині, соціології, психології та інших поведінкових науках. Серед розглянутих рангових критеріїв при достатньо великих обсягах виборок найбільш потужним є критерій нормальних міток, а найменш потужним – Вілкоксона, проте він найпростіший у використанні.

РОЗДІЛ 3

ПРИКЛАДИ ЗАСТОСУВАННЯ НЕПАРАМЕТРИЧНИХ КРИТЕРІЇВ ОДНОРІДНОСТІ В ОСВІТНІХ ВИМІРЮВАННЯХ (ПЕДАГОГІЧНИХ ДОСЛІДЖЕННЯХ)

В попередньому розділі було відмічено, що рангові критерії однорідності виявляються вельми зручними і корисними для аналізу даних в поведінкових науках (соціології, психології, педагогіці, освітніх вимірюваннях і т.п.) та гендерних дослідженнях. Нижче розглянемо приклади використання непараметричного/ рангового критерію Манна-Уітні для розв'язання конкретних задач математико-статистичного аналізу даних, пов'язаних з педагогічною діяльністю і отриманих (на базі мого професійного досвіду) підчас дослідження, проведеного серед учнів 4-х класів Ніжинського навчально-виховного комплексу №16 «Престиж». Було досліджено рівень знань навчального матеріалу з інформатики серед учнів 4-А та 4-Б класів. Вимірювання рівня знань відбувалося за допомогою тестування, яке складалось з 10 запитань. Метою дослідження було з'ясувати чи відрізняється рівень засвоювання знань між учнями 4-А та 4-Б класів, а також чи відрізняється рівень знань дівчат та хлопців. Результати дослідження продемонстровано в наступних прикладах.

Приклад 3.1

Сформовано дві групи(вибірки) учнів: дівчата(10 чоловік) та хлопці(10 чоловік). Необхідно дослідити чи дійсно рівень знань з інформатики у дівчат вищий ніж в хлопців. Результати тестування представлені в табл. 3.1.

Таблиця 3.1

Номер учня в списку	Рівень знань / кількість балів при тестуванні	
	Дівчата	Хлопці
1	10	7
2	8	9

3	8	5
4	5	5
5	9	6
6	10	8
7	6	4
8	7	5
9	8	7
10	7	6

В даному випадку основна гіпотеза H_0: рівень знань дівчат та хлопців однаковий, альтернативна H_1 - рівень знань дівчат вищий за рівень знань хлопців. Для перевірки гіпотези скористаємося непараметричним критерієм Манна-Уітні(див. п.2.3). Дані обох груп об'єднаємо у один ряд, розташувавши їх у порядку зростання. Значення об'єднаного ряду проранґуємо, тобто присвоємо кожному числу його ранг (табл. 3.2)

Таблиця 3.2

Номери місць в упорядкованому ряду	Кількість балів при тестуванні	Ранги
1	4	1
2	5	3.5
3	5	3.5
4	5	3.5
5	5	3.5
6	6	7
7	6	7
8	6	7
9	7	10.5
10	7	10.5

11	7	10.5
12	7	10.5
13	8	14.5
14	8	14.5
15	8	14.5
16	8	14.5
17	9	17.5
18	9	17.5
19	10	19.5
20	10	19.5

Далі знайдемо суму рангів для обох груп ($R_1 = 131.5$ та $R_2 = 78.5$). Обчислимо емпіричне значення критерія за формулою:

$$U_{емп} = n_1 n_2 + \frac{n_x(n_x+1)}{2} - R_x,$$

де n_1 – кількість досліджуваних в 1 групі; n_2 – кількість досліджуваних в 2 групі; $R_x = \max(R_1, R_2)$. Отже, $U_{емп} = 23.5$. Порівняємо обчислене значення $U_{емп}$ та знайдене у таблиці $U_{кр}$. Маємо $U_{кр}(0.05) = 27$, $U_{кр}(0.01) = 19$. Так як $U_{кр} > U_{емп}$ нульову гіпотезу відхиляємо, приймаємо гіпотезу H_1 з імовірністю 95%, тобто рівень знань дівчат вищий за рівень знань хлопців.

Приклад 3.2

Дослідити чи однаковий рівень знань з інформатики мають учні 4-А та 4-Б класів. Результати тестування представлені в табл. 3.3.

Таблиця 3.3

Номер учня в списку	Рівень знань / бали	
	4-А	4-Б
1	8	5
2	4	6

3	6	7
4	7	8
5	9	5
6	7	5
7	5	6
8	8	10
9	6	6
10	8	8
11	5	10
12	9	7
13	6	5
14	4	7
15	4	7
16	6	4
17	7	9
18	10	7
19	7	8
20	7	10
21	8	
22	6	
23	7	
24	7	

Основна гіпотеза H_0: рівень знань з інформатики учнів 4-А і 4-Б класів однаковий, альтернативна H_1 - рівень знань відрізнається. Перевірку виконаємо, застосовуючи критерій Манна-Уітні. Дані обох груп об'єднаємо у один ряд, розташувавши їх у порядку зростання. Значення об'єднаного ряду прорангуємо, тобто присвоємо кожному числу його ранг (табл.3.4)

Таблиця 3.4

Номери місць в упорядкованому ряду	Кількість тестових балів	Ранги
1	4	2.5
2	4	2.5
3	4	2.5
4	4	2.5
5	5	7.5
6	5	7.5
7	5	7.5
8	5	7.5
9	5	7.5
10	5	7.5
11	6	14.5
12	6	14.5
13	6	14.5
14	6	14.5
15	6	14.5
16	6	14.5
17	6	14.5
18	6	14.5
19	7	24.5
20	7	24.5
21	7	24.5
22	7	24.5
23	7	24.5
24	7	24.5
25	7	24.5
26	7	24.5
27	7	24.5

28	7	24.5
29	7	24.5
30	7	24.5
31	8	34
32	8	34
33	8	34
34	8	34
35	8	34
36	8	34
37	8	34
38	9	39
39	9	39
40	9	39
41	10	42.5
42	10	42.5
43	10	42.5
44	10	42.5

Знайдемо суму рангів для обох груп ($R_1 = 523$ та $R_2 = 467$). Обчислимо емпіричне значення критерія аналогічно першому прикладу. Порівняємо обчислене значення $U_{емп}$ та табличне $U_{кр}$. Маємо $U_{кр}(0.05) = 169$, $U_{кр}(0.01) = 140$. Оскільки, $U_{емп} > U_{кр}\,(257 > 169)$, підстави відхиляти нульову гіпотезу не має. Можемо зробити висновок, що рівень знань з інформатики учнів 4-А і 4-Б класів Ніжинського НВК №16 «Престиж» однаковий.

Висновок до розділу 3

Ранговий критерій Вілкоксона –Манна-Уітні є дуже простим і зручним у застосуванні, його використання не потребує складних обчислень і тим більше використання спеціалізованого ПЗ, необхідні критичні значення табульовані і їх легко знайти у доступних джерелах інформації.

ВИСНОВКИ

Популярність непараметричних методів пояснюється стійкістю висновків та простотою математичних засобів, саме тому вони вони набули широкого застосування в різних галузях. Особливість непараметричних методів - в тому, що не потрібно прив'язуватися до конкретного розподілу. Важливою складовою комплексу непараметричних процедур виявляються методи і алгоритми перевірки статистичних гіпотез; а серед задач, пов'язаних з перевіркою статистичних гіпотез, значне місце займають критерії згоди та критерії однорідності.

Критерії згоди дозволяють розв'язувати задачі, в яких необхідно перевірити гіпотезу про те, що досліджувана випадкова величина підпорядковується тому чи іншому розподілу ймовірностей. Для цього обчислюють статистику, на основі якої роблять висновок про сформовану гіпотезу, порівнюючи отримане значення з критичним, що знаходиться за таблицями. Найчастіше застосовують критерії Колмогорова та χ^2. Кожен з них має свої переваги і недоліки. Так, використання критерію Колмогорова можливе лише тоді, коли гіпотетичний розподіл - неперервний та повністю визначений (тобто параметри відомі). Критерій χ^2 можна використовувати для неперервних і дискретних теоретичних розподілів, але необхідно багато даних, оскільки перевірка відбувається по згрупованим даним. Процедури застосування критеріїв згоди проілюстровані на конкретних прикладах. Критичні значення наведені в додатках.

Критерії однорідності - це критерії перевірки гіпотез, які дозволяють перевірити гіпотезу про те, що двом (або більше) вибіркам відповідають однакові функції розподілу. Критерій Смірнова-Колмогорова базується на використанні різниць між емпіричними функціями розподілу, що відповідають різним вибіркам, але теоретичні (невідомі) розподіли мають бути неперервними. Для використання критерія однорідності χ^2 необхідна велика кількість даних, оскільки перевірка робиться по згрупованим даним. Рангові критерії прості у застосуванні, особливо у випадку малих вибірок, вони не вимагають складних обчислень, можуть бути застосовані не тільки за відсутності інформації і додаткових припущень щодо виду розподілу, але і у випадку, коли спостереження не є кількісними, а можуть бути тільки впорядковані, як це часто буває в медицині, соціології, психології та інших поведінкових науках. Серед розглянутих рангових критеріїв при достатньо великих обсягах виборок найбільш потужним є критерій нормальних міток, а найменш потужним – Вілкоксона (Уілкоксона), проте він найпростіший у використанні.

В роботі особлива увага була приділена застосуванню рангового критерія Вілкоксона –Манна-Уітні, оскільки він є дуже простим і зручним у застосуванні, а його критичні значення табульовані в багатьох джерелах. Тому його було засновано до оцінки результатів конкретних задач математико-статистичного аналізу даних, пов’язаних з педагогічною діяльністю і отриманих на базі мого професійного досвіду.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Айвазян С.А., Енюков И.О., Мешалкин Л.Д. Прикладная статистика: справ. изд. Москва: Финансы и статистика, 1983. 471 с.
2. Байхельт Ф., Франкен П. Надежность и техническое обслуживание. Математический подход: пер. с нем. Москва: Радио и связь, 1988. 392 с.
3. Большов Л.Н., Смирнов Н.В. Таблицы математической статистики. Москва: Наука, 1983. 416с.
4. Боснюк В.Ф. Математичні методи в психології: курс лекцій. Харків: НУЦЗУ,2016.URL:http://univer.nuczu.edu.ua/tmp_metod/1321/KURS_LEKC_J_Matematichni_metodi_v_psihologiji.pdf
5. Гаек Я., Шидак З. Теория рангових критериев. Москва:Наука, 1971. 376с.
6. Данилов В.Я. Теорія ймовірностей і математична статистика: навч.пос. Київ: Принт-Сервіс, 2013. 158с.
7. Жалдак М.І., Кузьміна Н.М., Михалін Г.О. Теорія ймовірностей і математична статистика: навч. посіб. 2-ге вид., доп. Полтава: Довкілля-К, 2009. 500с.
8. Карташов М.В. Імовірність, процеси, статистика: посіб. Київ: Видавничо-поліграфічний центр «Київський університет», 2008. 494с.
9. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. Москва: Физматлит, 2006. 816с.
10. Королюк В.С., Боровских Ю.В.. Аналитические проблеми асимптотики вероятностних распределений. Киев: Наукова думка, 1981.
11. Крамер Г. Математические методы статистики. URL: https://ikfia.ysn.ru/wp-content/uploads/2018/01/Kramer1975ru.pdf
12. Лемешко Б.Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьюторный подход: монография. Новосибирск: НГТУ, 2011. 888с.
13. Оуен Д.Б. Сборник статистических таблиц. Москва: Вычислительный центр АН СССР, 1966. 586с.

14. Руденко В. М. Математична статистика: навч. посібник. Київ: Центр учбової літератури, 2012. 304 с.
15. Таблиці функцій та критичних точок розподілів. Розділи: Теорія ймовірностей. Математична статистика. Математичні методи в психології. / уклад. М. М. Горонескуль. Харків: УЦЗУ,2009. 90с.
16. Турчин В.М. Теорія ймовірностей і математична статистика: навч. пос. Дніпропетровськ: IMA, 2014. 556 с.
17. Яременко Л.І., Лупан І.В. Кількісні методи в поведінкових науках: навч. пос. Кропивницький: Видавець Лисенко В.Ф., 2019. 224с.
18. Gail M.H., Green S.B. Critical values for the one-sided two-sample Kolmogorov-Smirnov statistic. J. of the Amer. Statist. Assoc.,1976. 757-760.
19. David H. A three-sample Kolmogorov-Smirnov test. Aon. Math. Statist., 1958. 842-851.

ДОДАТКИ

Додаток 1

Граничний розподіл статистики Колмогорова-Смирнова

a	$K(a)$
0.30	0.000009
0.38	0.0013
0.52	0.0503
0.58	0.1104
0.77	0.4064
0.83	0.5038
0.90	0.6073
1.08	0.8061
1.23	0.9030
1.36	0.9505
1.45	0.9702
1.63	0.990
1.96	0.999079
2.24	0.999912

Додаток 2

Граничний розподіл статистики ω^2

x	$a(x)$
0.02480	0.01
0.03865	0.06
0.04601	0.10
0.06222	0.16
0.06381	0.20
0.07860	0.30
0.11888	0.50
0.14663	0.60
0.18433	0.70
0.24124	0.80
0.34730	0.90
0.46136	0.95
0.54885	0.97
0.74346	0.99
1.16786	0.999

Додаток 3

Критичні значення $d_{кр}$ для критерія згоди Колмогорова

n	$\alpha = 0.05$	$\alpha = 0.01$
5	0.5633	0.6885
7	0.4834	0.5758
10	0.4093	0.4889
11	0.3912	0.4677
12	0.3754	0.4491
13	0.3614	0.4325
14	0.3489	0.4176
15	0.3376	0.4042
20	0.2941	0.3524
30	0.2417	0.2899
35	0.2243	0.2690
40	0.2101	0.2520
45	0.1984	0.2380
50	0.1884	0.2260
55	0.1798	0.2157
60	0.1723	0.2067
65	0.1657	0.1988
70	0.1598	0.1917
75	0.1544	0.1853
80	0.1469	0.1795
85	0.1452	0.1742
90	0.1412	0.1694
95	0.1375	0.1649
100	0.1340	0.1608
$n > 100$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$