В.М. Турчин,
Є.В. Турчин

ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ ОСНОВНІ ПОНЯТТЯ

Навчально-методичний посібник

Дніпро
Видавництво "Ліра"
2024

УДК 519.2 (075.8)
Т89

Рецензенти:
канд. пед. наук, доц. Кірман В.К.
канд. фіз.-мат. наук, доц. Ткаченко М.Є.

Т89 Турчин В.М., Турчин Є.В.

Навчально-методичний посібник “Перевірка статистичних гіпотез. Основні поняття”. —Д.: Вид-во “Ліра”, 2024. — 41 с.

У навчально-методичного посібнику розглянуто загальні питання перевірки статистичних гіпотез. На низці прикладів продемонстровані основні ідеї перевірки статистичних гіпотез. Розв'язання супроводжуються програмами на мові R.

Для студентів ДНУ спеціальності 112 “Статистика”.

Рекомендовано Вченою Радою механіко-математичного факультету Дніпровського національного університету імені Олеся Гончара для студентів спеціальності 112 “Статистика”, протокол № 9 від 21.05. 2024 р.

Навчальне видання

Валерій Миколайович Турчин
Євген Валерійович Турчин

**Навчально-методичний посібник
“Перевірка статистичних гіпотез.
Основні поняття”**

Друкується за авторською редакцією.

Підписано до друку 24. Формат 84x108$\frac{1}{32}$. Папір офсетний. Друк цифровий. Ум. друк. арк. 2. Тираж 30 пр.

Зам. № . Друкарня ПП «Ліра ЛТД», вул. Наукова, 5, м. Дніпро, 49107. Свідоцтво про внесення до Державного реєстру серія ДК № 6042 від 26.02.2018 р.

1 Критерій, функція потужності критерію

Задача перевірки статистичних гіпотез. Часто задачі математичної статистики формулюються так:

Маємо наслідок $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ стохастичного експерименту, що полягає у спостереженні випадкової величини $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ зі значеннями в $\mathbb{R}^n$. Щодо розподілу випадкової величини $\xi = (\xi_1, \xi_2, \ldots \ldots, \xi_n)$, або, що те саме, вибірки, у кращому випадку відомо тільки те, що він належить до деякого класу розподілів $\mathcal{P}$. Далі клас $\mathcal{P}$ вважатимемо параметричним:

$$\mathcal{P} = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^s\}.$$

За значенням $\xi(\omega)$, якого набула випадкова величина ξ, необхідно вибрати її розподіл із класу $\mathcal{P}$.

Діятимемо у такий спосіб. Із класу $\mathcal{P}$ можливих розподілів випадкової величини ξ вибираємо деякий розподіл G — "кандидатуру" на розподіл ξ. Інакше кажучи, стосовно розподілу випадкової величини ξ висуваємо гіпотезу (припущення): розподілом ξ є G. Далі проводимо експеримент — добуваємо реалізацію $\xi(\omega)$ випадкової величини ξ і за значенням $\xi(\omega)$, якого набула ξ, доходимо висновку: випадкова величина ξ може мати своїм розподілом G чи G не може бути розподілом ξ. При цьому діятимемо так, як діяв свого часу Сократ, коли хотів спростувати твердження свого опонента. А саме, якщо гіпотеза "G є розподілом ξ" — суперечить наслідку експерименту, то її слід відхилити (G не може бути розподілом випадкової величини ξ), у супротивному разі — ні (гіпотезу не варто відхиляти, тобто G може бути розподілом ξ).

Сформульована задача називається задачею перевірки статистичних гіпотез.

Домовленості й означення. У теорії перевірки статистичних гіпотез прийнято такі означення і домовленості.

Гіпотези щодо розподілів випадкових величин називають *статистичними*.

Вибір розподілу (чи класу розподілів) із сукупності

$$\mathcal{P} = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^s\}$$

можливих розподілів ξ називатимемо *вибором основної (нульової) гіпотези* стосовно розподілу випадкової величини ξ. Після вибору нульової гіпотези всі інші гіпотези стосовно розподілу випадкової величини ξ стають альтернативними (конкуруючими) до нульової. (Оскільки сукупність $\mathcal{P}$ можливих розподілів ξ є параметричною, то і альтернативні гіпотези утворюють параметричну сукупність, позначатимемо її так: H_θ, $\theta \in \Theta$, $\theta \neq 0$.)

Проілюструємо введені поняття на прикладі біномного розподілу.

Монету, ймовірність випадання герба θ якої невідома, у незалежний спосіб підкидають n разів. Число гербів, що при цьому випало, виявилося рівним ξ. Чи можна вважати, що монета симетрична?

Сформулюємо цю задачу в термінах перевірки статистичних гіпотез.

Розподілом випадкової величини ξ може бути будь-який розподіл із класу $\mathcal{P} = \{P_\theta, \theta \in (0;1)\}$, де

$$P_\theta(k) = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n;\ \theta \in (0;1).$$

Оскільки ми цікавимося симетричністю монети, то як кандидатуру на розподіл випадкової величини ξ із сукупності розподілів $\mathcal{P}$ вибираємо розподіл

$$P_{1/2}(k),\ k = 0, 1, \ldots, n.$$

Тим самим вибрана нульова гіпотеза H_0: розподілом ξ є $P_{1/2}(k)$, $k = 0, 1, \ldots, n$. Конкуруючими до H_0 є гіпотези H_θ: розподілом випадкової величини є

$$P_\theta(k),\ \theta \in (0, 1/2) \cup (1/2, 1).$$

Питання щодо симетричності монети зводиться до перевірки гіпотези H_0.

Гіпотеза відносно розподілу, яка однозначно його визначає, називається *простою*; якщо гіпотеза не визначає розподіл однозначно, вона називається *складною*.

Наприклад, гіпотеза H_0: випадкова величина ξ має розподіл

$$P_{\theta_0}(k) = C_n^k {\theta_0}^k (1-\theta_0)^{n-k},\ k = 0, 1, \ldots, n,$$

де θ_0 — фіксоване, n відоме, проста. Гіпотеза H: випадкова величина ξ має розподіл

$$P_\theta(k) = C_n^k \theta^k (1-\theta)^{n-k},\ k = 0, 1, \ldots, n,$$

де $\theta \in (0; 1/2)$, складна.

Формулюючи задачу перевірки статистичних гіпотез, як нульову гіпотезу (для наочності та простоти) ми вибрали просту гіпотезу: розподілом випадкової величини ξ є G, де G — цілком визначений розподіл.

Зазначимо, що математична статистика не дає рекомендацій щодо вибору нульової гіпотези, цей вибір визначається поставленою задачею і дослідником.

Критерій. Нехай ξ — випадкова величина зі значеннями в $\mathbb{R}^n$. Відносно невідомого розподілу ξ висувається гіпотеза H_0: розподілом ξ є G; альтернативна гіпотеза: G не є розподілом ξ.

Необхідно перевірити гіпотезу H_0, тобто дійти висновку: G може бути розподілом випадкової величини ξ (казатимемо "гіпотеза H_0 не відхиляється") або G не може бути розподілом випадкової величини ξ (казатимемо "гіпотеза H_0 відхиляється").

Згідно з нашим загальним підходом, проводимо стохастичний експеримент: добуваємо вибірку — реалізацію $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ випадкової величини $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$. Далі, якщо гіпотеза H_0 суперечить результату експерименту, вона відхиляється, у супротивному разі — не відхиляється. (Зазначимо, що крім реалізації $\xi(\omega)$ вибірки ξ, ми не маємо нічого, що несло б інформацію про розподіл ξ.) Щоб можна було дійти висновку про відхилення або невідхилення гіпотези H_0, необхідно визначити множину $S \subset \mathbb{R}^n$ тих наслідків $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ експерименту, яким гіпотеза суперечить. Маючи таку множину, гіпотезу H_0 відхилятимемо, якщо $\xi(\omega)$ потрапляє до S, і не відхилятимемо у супротивному разі.

Означення. Борелеву множину S вибіркового простору таку, що при $\xi \in S$ гіпотеза H_0 відхиляється, а при $\xi \notin S$ — не відхиляється, називатимемо *критичною множиною* (*критичною областю*) або *критерієм для перевірки гіпотези* H_0.

Далі, природно, постає задача (це основна задача перевірки статистичних гіпотез): як вибрати критичну множину $S \subset \mathbb{R}^n$ для перевірки даної гіпотези (борелевих множин S у просторі $\mathbb{R}^n$ багато і, ясна річ, що не всі вони, як критерії для перевірки гіпотези H_0, однаково хороші).

Щоб вказати шляхи розв'язання цієї задачі, спочатку розглянемо так звані помилки першого й другого роду, що неминучі під час перевірки статистичних гіпотез.

Помилки першого й другого роду. Нехай H_0 — нульова гіпотеза стосовно розподілу випадкової величини ξ, $\xi(\omega)$ — реалізація ξ, S — борелева множина у просторі $\mathbb{R}^n$. Перевірятимемо H_0, користуючись множиною S як критичною, а саме, якщо $\xi(\omega)$ потрапляє до S, то гіпотезу H_0 відхиляємо, у супротивному разі — ні.

При цьому можливі такі ситуації: гіпотеза H_0 справджується або не справджується (апріорі ми цього не знаємо), реалізація $\xi(\omega)$ випадкової величини ξ потрапила до S або ні. Розглянемо їх докладніше.

1° Гіпотеза H_0 справджується. Реалізація $\xi(\omega)$ не потрапила до S, тому згідно з критерієм S гіпотеза H_0 не відхиляється.

2° Гіпотеза H_0 не справджується. Реалізація $\xi(\omega)$ не потрапила до S, і, отже, згідно з критерієм S гіпотеза H_0 не відхиляється.

3° Гіпотеза H_0 справджується. Реалізація $\xi(\omega)$ потрапила до S, і тому згідно з критерієм S гіпотеза H_0 відхиляється.

4° Гіпотеза H_0 не справджується. Реалізація $\xi(\omega)$ потрапила до S, і отже згідно з критерієм S гіпотеза H_0 відхиляється.

Із чотирьох ситуацій дві (1° та 4°) задовільні, і дві (2° і 3°) — незадовільні. У ситуаціях 2° і 3° ми припускаємося помилки. При цьому помилки у ситуації 2° (гіпотеза H_0 не відхиляється, коли вона не справджується) і в ситуації 3° (гіпотеза H_0 відхиляється, коли вона справджується) істотно відрізняються.

Проілюструємо відмінність між помилками у ситуаціях 2° і 3° на прикладі перевірки медичного препарату на токсичність біологічними методами.

Досліджуючи препарат на токсичність, певну його дозу вводять піддослідним тваринам (кроликам) і реєструють число летальних кінців (зазначимо, що це число є випадковою величиною). Необхідно за реалізацією $\xi(\omega)$ випадкової величини ξ (числом летальних кінців) дійти висновку щодо токсичності препарату. Зрозуміло, що коли число $\xi(\omega)$ велике, препарат слід вважати токсичним, у супротивному разі — ні.

Задачу дослідження препарату на токсичність біологічними методами можна сформулювати в термінах перевірки статистичних гіпотез. А саме, стосовно токсичності препарату висуваються гіпотези: H_0 — препарат токсичний і H_1 — препарат нетоксичний. Необхідно перевірити гіпотезу H_0, тобто відхилити її або не відхилити. Вибір між цими діями здійснюється за реалізацією $\xi(\omega)$ випадкової величини ξ — числа летальних кінців: якщо $\xi(\omega) \in S = \{x : x \leq k\}$, то гіпотеза відхиляється, у супротивому разі — ні (S — критерій для перевірки гіпотези H_0).

Як і в кожній задачі перевірки статистичних гіпотез у задачі, що розглядається, можливі помилки:

гіпотеза H_0 не справджується, але згідно з критерієм вона не відхиляється;

гіпотеза H_0 справджується, але згідно з критерієм вона відхиляється.

Подивимося, які наслідки цих помилок і яка їхня "ціна".

1. Нехай зроблена помилка: гіпотеза H_0 не справджується, але згідно з критерієм не відхиляється. Твердження "гіпотеза H_0 не справджується" у задачі, що розглядається, означає, що препарат нетоксичний (не є небезпечним для здоров'я пацієнтів), а твердження "H_0 не відхиляється" означає, що препарат класифікується як токсичний. Таким чином, нетоксичний препарат згідно з критерієм класифікується як токсичний і повертається постачальнику (для пероблення або знищення). Наслідки помилки такого роду — зростання вартості товару (ціна помилки — фінансові збитки).

2. Нехай зроблена помилка: гіпотеза H_0 справджується, але згідно з критерієм відхиляється. Твердження "гіпотеза H_0 справджується" означає, що препарат токсичний (небезпечний для здоров'я пацієнтів). Твердження "гіпотеза H_0 відхиляється" означає, що препарат класифікується як нетоксичний (не є небезпечним для здоров'я пацієнтів). Таким чином, токсичний препарат (небезпечний для здоров'я пацієнтів) відповідно до критерію класифікується як нетоксичний (не є небезпечним для здоров'я пацієнтів) і йде у продаж. Наслідком помилки такого роду може стати смерть пацієнта, який вживає цей препарат (ціна помилки — летальний кінець для пацієнта).

Цей приклад показує, що описані вище помилки істотно різняться за своєю ціною.

Означення. Помилка, яка полягає у тому, що гіпотеза H_0 відхиляється, коли вона справджується, називається *помилкою першого роду.*

Помилка, яка полягає у тому, що гіпотеза H_0 не відхиляється, коли вона не справджується, називається *помилкою другого роду.*

Вибір нульової гіпотези. Як зазначалося, нульову гіпотезу із сукупності всіх можливих гіпотез ми вибираємо самі. Цей вибір тісно пов'язаний з помилками, яких ми припускаємося, перевіряючи гіпотези. Як правило, за нульову гіпотезу вибираємо ту, для якої важливіше уникнути помилки, що полягає у відхиленні цієї гіпотези, коли вона справджується. (Про помилки, пов'язані з перевіркою гіпотез, які полягають у тому, що гіпотезу H_θ $(\theta \in \Theta)$ ми відхиляємо, коли вона справджується, можна говорити і не вибравши нульову гіпотезу.)

Рівень значущості критерію, функція потужності критерію. Чи можна побудувати критичну множину для перевірки гіпотези, яка б не призводила до помилок? Ні, не можна. Адже, якою б не була критична множина $S \neq \emptyset$, значення $\xi(\omega)$ випадкової величини ξ (наслідок стохастичного експерименту) може потрапити до S, коли гіпотеза H_0 справджується, при цьому буде зроблена помилка першого роду. Значення $\xi(\omega)$ може потрапити і до $\overline{S}$, коли H_0 не справджується (при цьому буде зроблена помилка другого роду). І оскільки побудувати критерій для перевірки гіпотези H_0, який би не призводив до помилок, неможливо в принципі, ми, природно, намагатимемося будувати такі критерії, які б гарантували мінімальну частоту помилок при їх використанні.

Нехай H_0 — основна гіпотеза, H_1 — конкуруюча (для наочності припустимо, що вони прості), S — критерій для перевірки H_0. Користуючись критерієм S, ми можемо припуститися помилок двох типів: H_0 справджується, але згідно з критерієм S відхиляється (помилка першого роду); H_0 не справджується, але згідно з критерієм S не відхиляється (помилка другого роду). Імовірність помилки першого роду дорівнює ймовірності значенню ξ потрапити до критичної множини S, коли гіпотеза H_0 справджується, тобто $\mathsf{P}\{\xi \in S|H_0\}$ ($\mathsf{P}\{\xi \in S|H_0\}$ стисло записуватимемо так: $\mathsf{P}(S|H_0)$). Імовірність помилки другого роду дорівнює ймовірності вибірковому значенню потрапити до множини $\overline{S}$, коли справджується гіпотеза H_1, тобто $\mathsf{P}\{\xi \in \overline{S}|H_1\}$ ($\mathsf{P}\{\xi \in \overline{S}|H_1\}$ стисло записуватимемо так: $\mathsf{P}(\overline{S}|H_1)$).

Зазначимо, що $\mathsf{P}\{\xi \in S|H_0\}$ — це ймовірність події $\{\xi \in S\}$, обчислена за припущення, що гіпотеза H_0 справджується; вона не має нічого спільного з умовною ймовірністю.

Імовірності помилок першого та другого роду однозначно визначаються критичною множиною S. Тому, вибираючи S, скажімо, за умови, що ймовірність помилки першого роду мала, водночас одержимо ймовірність помилки другого роду, яка визначається вибраною критичною множиною S і буде такою, якою вийде. Можна вибрати S і за умови, що ймовірність помилки другого роду мала, але при цьому ймовірність помилки першого роду однозначно визначиться вибраною множиною S. Отже, вибрати S так, щоб одночасно були контрольовані ймовірності помилок першого та другого роду, не вдасться. А оскільки важливіше уникнути помилки першого роду (її ціна вища), то перша вимога до критичної множини S така: імовірність помилки першого роду $\mathsf{P}(S|H_0)$ має бути малою. (Це означає, що використовуючи критерій S у довгій серії експериментів, гіпотезу H_0, коли вона справджується, відхилятимемо зрідка.) Формалізуємо цю вимогу.

Фіксуємо мале α і вибираємо критичну множину S так, щоб імовірність помилки першого роду не перевищувала α:

$$\mathsf{P}(S|H_0) \le \alpha.$$

Якщо останню нерівність задовольняє не одна множина S, то остаточний вибір критерію здійснюється так, щоб імовірність $\mathsf{P}(S|H_1)$ відхилення гіпотези H_0, коли вона не справджується, була максимальною. Для цього множину S вибираємо якомога "ширшою".

Означення. Число α, яке обмежує зверху ймовірність помилки першого роду, називається *рівнем значущості.*

Якщо критична множина S задовольняє умову

$$\mathsf{P}(S|H_0) \le \alpha,$$

то казатимемо, що S відповідає рівню значущості α.

З а у в а ж е н н я. Питання про те, яким має бути рівень значущості, не є статистичною задачею. Здебільшого за рівень значущості приймають числа 0,10; 0,05; 0,01. Чим серйозніші наслідки помилки першого роду, тим меншим має бути рівень значущості.

Означення. Імовірність $\mathsf{P}(S|H_1)$ відхилити основну гіпотезу, коли справджується альтернативна гіпотеза H_1, називається *потужністю критерію S.*

Якщо альтернативна гіпотеза складна, причому коли вона справджується, справджується одна з простих гіпотез H_θ, $\theta \in \Theta$, $\theta \neq 0$, то для кожного $\theta \in \Theta$, $\theta \neq 0$, можна обчислити $\mathsf{P}(S|H_\theta)$.

Означення. Функція

$$\beta(\theta) = \mathsf{P}(S|H_\theta),\ \theta \in \Theta,\ \theta \neq 0,$$

яка для кожного $\theta \in \Theta$ дорівнює імовірності відхилити основну гіпотезу H_0, коли справджується гіпотеза H_θ, називається *функцією потужності критерію.*

Нехай S — критерій для перевірки нульової гіпотези H_{θ_0} проти альтернативи $H_1 = \{H_\theta,\ \theta \in \Theta\}$. Користуючись критерієм, ми гіпотезу H_{θ_0} відхилятимемо, якщо ξ потрапило у S, і не відхилятимемо у супротивному разі.

Критерій S "автоматично" задає на множині Θ функцію потужності критерію S:

$$\beta(\theta) = \mathsf{P}(S|H_\theta),\ \theta \in \Theta.$$

Доозначимо функцію $\beta(\theta)$ на множині $\Theta \cup \{\theta_0\}$ так: $\beta(\theta_0) = \mathsf{P}(S|H_{\theta_0})$ (як і раніше, будемо називати $\beta(\theta)$ функцією потужності критерію). Значення $\beta(\theta) = \mathsf{P}(S|H_\theta)$ для кожного $\Theta \cup \{\theta_0\}$ дорівнює імовірності відхилити нульову гіпотезу H_{θ_0}. Зокрема, значення $\beta(\theta_0) = \mathsf{P}(S|H_{\theta_0})$ функції $\beta(\theta)$ у точці θ_0 дорівнює імовірності помилки першого роду.

2 Приклади

Приклад 1 (вибірковий контроль). Призначена для продажу партія виробів кількістю $N = 10\,000$ штук проходить вибірковий контроль на якість. Постачальник упевнений, що частка дефектних виробів дорівнює $1\,\%$ (або менше), і бажає, щоб кожного разу, коли частка дефектних виробів становить $1\,\%$, імовірність того, що партія витримає контроль, дорівнювала 0,9. Покупець вважає, що партію доцільно закупити навіть тоді, коли частка дефектних виробів перевищуватиме $1\,\%$, але $6\,\%$ дефектних виробів він вважає гранично допустимою часткою і бажає, щоб контроль виявляв партії із $6\,\%$-м вмістом дефектних виробів з імовірністю 0,95. Постачальник і покупець домовилися здійснювати контроль так: із партії виробів добувають випадкову вибірку обсягом n. Якщо при цьому кількість $\xi(\omega)$ дефектних виробів у вибірці виявиться малою (менше деякого l), то партія витримує контроль і закуповується, у супротивному разі— ні.

Якими мають бути обсяг n вибірки з партії виробів і значення l?

Дати відповідь на ці питання, сформулювавши і розв'язавши поставлену задачу як задачу перевірки статистичних гіпотез.

Розв'язання. Висновок про закупівлю або відхилення партії виробів робитимемо за реалізацією $\xi(\omega)$ випадкової величини ξ — кількості дефектних виробів із n вибраних.

Вироби до вибірки обсягом n вибиратимемо послідовно. Оскільки обсяг N партії великий ($N = 10\,000$), а n порівняно до N мале, то можна вважати, що випадкова величина ξ має біномний розподіл з параметрами $(n; p)$:

$$\mathsf{P}\{\xi = k\} = P_p(k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n$$

(параметр p — імовірність вибору дефектного виробу — невідомо). Інакше кажучи, сукупністю можливих розподілів випадкової величини ξ є

$$\mathcal{P} = \{P_p : p \in (0; 1)\}.$$

Зазначимо, що розподіл із сукупності $\mathcal{P}$ однозначно визначається значенням параметра p і навпаки.

За реалізацією $\xi(\omega)$ випадкової величини ξ необхідно дійти висновку щодо розподілу ξ або, що те саме, щодо значення параметра p. Становлять інтерес такі значення параметра p: $p_1 = 0{,}01$ і $p_2 = 0{,}06$ (їм відповідає 1 %-й і 6 %-й вміст дефектних виробів у партії).

Відповідно до викладеної методики із сукупності можливих гіпотез $\mathcal{P} = \{P_p, p \in (0; 1)\}$ необхідно вибрати нульову, а потім перевірити її. Оскільки між покупцем і постачальником існує угода: "якщо кількість $\xi(\omega)$ дефектних виробів у вибірці мала, то партія витримує контроль і закуповується, у супротивному разі — ні", то за нульову виберемо гіпотезу "ξ має біномний розподіл з параметрами $(n; 0{,}06)$" (альтернативна гіпотеза $P_p \in \{P_p : p \in (0; 0{,}06)\}$. Невідхилення нульової гіпотези означає класифікацію партії як такої, що містить 6 % (або більше) дефектних виробів, із подальшою відмовою закупити партію. (Якщо при цьому якусь із партій з малим відсотком дефектних виробів буде забраковано, то це вже проблема постачальника.) Відхилення нульової гіпотези означає класифікацію партії як такої, що містить менше 6 % дефектних виробів, і закупівлю партії. (Якщо при цьому якусь із партій з великим відсотком дефектних виробів не буде забраковано, а отже, буде закуплено, покупець, ясна річ, зазнає збитків.)

Для перевірки H_0 як критичну природно вибрати множину вигляду $S = \{k : k \leq l\}$. Ця множина визначається числом l і обсягом вибірки n.

За домовленістю між покупцем і постачальником множина $S = \{k : k \leq l\}$, а фактично числа n і l, мають бути такими, щоб виконувалися наведені далі умови. Партія з 6 %-м вмістом дефектних виробів повинна виявлятися з імовірністю 0,95, тобто

$$\mathsf{P}(\overline{S}|H_0) \geq 0{,}95,$$

або, що те саме,

$$\mathsf{P}(S|H_0) = \mathsf{P}\{\xi \in S|H_0\} = P_0\{\xi \leq l\} =$$

$$= \sum_{k=0}^{l} C_n^k (0{,}06)^k (0{,}94)^{n-k} \leq \alpha = 0{,}05.$$

Остання нерівність — формалізований запис вимог покупця, відповідно до яких партії з 6 %-м вмістом браку мають виявлятися з імовірністю, не меншою ніж 0,95.

Значення $\beta(p) = \mathsf{P}(S|H_p)$ функції потужності критерію S у точці $p = 0,01$ має бути не меншим ніж 0,9:

$$\beta(0{,}01) = \mathsf{P}(S|H_{0{,}01}) = \mathsf{P}\{\xi \in S|H_{0{,}01}\} =$$

$$= \mathsf{P}\{\xi \leq l|H_{0{,}01}\} = \sum_{k=0}^{l} C_n^k (0{,}01)^k (0{,}99)^{n-k} \geq 0{,}9.$$

Остання нерівність є формальним записом побажань постачальника, згідно з якими партії з 1 %-м вмістом браку мають витримувати контроль з імовірністю, не меншою ніж 0,9.

Таким чином, шукані n і l мають задовольняти співвідношення

$$\sum_{k=0}^{l} C_n^k (0{,}06)^k (0{,}94)^{n-k} \leq 0{,}05; \qquad (1)$$

$$\sum_{k=0}^{l} C_n^k (0{,}01)^k (0{,}99)^{n-k} \geq 0{,}9. \tag{2}$$

У задачі, що розглядається, біномний розподіл

$$B_{n,p}(m) = C_n^m p^m (1-p)^{n-m},\ m = 0, 1, \ldots, n,$$

можна апроксимувати пуассоновим, оскільки параметр p малий (імовірність p вибору дефектного вибору мала), а n — обсяг вибірки — великий. Підставою для цього є теорема Пуассона. Нагадаємо її зміст.

Нехай $np \to \lambda > 0$, коли $n \to \infty$. Тоді для кожного фіксованого m, $m = 0, 1, \ldots$,

$$\lim_{n\to\infty} C_n^m p^m (1-p)^{n-m} = \frac{\lambda^m}{m!} e^{-\lambda}.$$

Щоб похибка апроксимації біномного розподілу пуассоновим була незначною, n має бути не меншим кількох десятків (краще сотень), а добуток np має задовольняти нерівності

$$1 \leq np \leq 10.$$

Власне кажучи, ми з самого початку за розподіл випадкової величини ξ — кількості дефектних виробів у вибірці — могли б узяти пуассонів розподіл (оскільки ймовірність успіху мала, а кількість випробувань велика).

Таким чином, виходитимемо з того, що n і l мають задовольняти нерівності

$$\sum_{k=0}^{l} \frac{\lambda_0^k}{k!} e^{-\lambda_0} \leq 0{,}05,\quad \lambda_0 = n \cdot 0{,}06; \tag{3}$$

$$\sum_{k=0}^{l} \frac{\lambda_1^k}{k!} e^{-\lambda_1} \geq 0{,}90,\quad \lambda_1 = n \cdot 0{,}01 \tag{4}$$

(порівняйте з нерівностями (1) і (2)). Визначаючи n і l, ми, природно, намагатимемося вибирати n якомога меншим.

Подивимося, чи можна для $n = 100$ визначити l так, щоб співвідношення (3) і (4) справджувалися. Якщо ні, вибираємо n більшим, і так діємо доти, доки знайдемо n і l, які задовольнятимуть нерівності (3) і (4).

При $n = 100$ маємо $\lambda_0 = 100 \cdot 0{,}06 = 6$; $\lambda_1 = 100 \times \times 0{,}01 = 1$. З таблиці розподілу Пуассона знаходимо, що нерівність (3) задовольняється, коли $l \le 1$, а (4) — коли $l \ge 2$. Тому для $n = 100$ не існує l, яке б задовольняло нерівності (3) і (4).

При $n = 150$ маємо $\lambda_0 = 150 \cdot 0{,}06 = 9$; $\lambda_1 = 150 \times \times 0{,}01 = 1{,}5$. Для таких λ_0 і λ_1 нерівність (3) задовольняється, коли $l \le 4$, а нерівність (4) — коли $l \ge 3$. Спробуємо зменшити n. Нехай $n = 130$. Тоді $\lambda_0 = 130 \cdot 0{,}06 = = 7{,}8$; $\lambda_1 = 130 \cdot 0{,}01 = 1{,}3$; нерівність (3) задовольняється, коли $l \le 3$, а нерівність (4) — коли $l \ge 3$. Таким чином, для $n = 130$ і $l = 3$ обидві нерівності (3) і (4) задовольняються. І отже, шуканою критичною множиною (критерієм) є

$$S = \{k \colon k \le 3\}.$$

Для цього критерію ймовірність помилки першого роду

$$\mathsf{P}(S|H_0) = P_0(S) = \sum_{k=0}^{3} \frac{\lambda_0^k}{k!} e^{-\lambda_0} = \sum_{k=0}^{3} \frac{7{,}8^k}{k!} e^{-7{,}8} = 0{,}048.$$

Значення функції потужності критерію $\beta(p) = P(S|H_p)$ у точці $p = 0{,}01$:

$$\beta(0{,}01) = \mathsf{P}(S|H_{0{,}01}) = P_{0{,}01}(S) =$$

$$= \sum_{k=0}^{3} \frac{\lambda_1^k}{k!} e^{-\lambda_1} = \sum_{k=0}^{3} \frac{1{,}3^k}{k!} e^{-1{,}3} = 0{,}957.$$

Імовірність помилки першого роду, що дорівнює 0,048, можна інтерпретувати так. Використовуючи описаний критерій, на 100 партій виробів, які містять 6 % браку, близько п'яти партій класифікуватимуться як такі, що містять менше 6 % браку і, отже, витримуватимуть контроль.

Значення потужності критерію $P(S|H_{0{,}01}) = 0{,}957$ інтерпретується так: на 100 партій виробів із 1 %-м вмістом браку, близько 96 партій будуть класифікуватися як такі, що містять 1 % браку і, отже, витримиватимуть контроль.

Для розв'язування задач можна користуватися таблицями математичної статистики з книжки [1].

Приклад 2 (про контроль чистоти води). Для потреб деякого хімічного виробництва необхідно, щоб вода, яка використовується, була чистою, тобто містила незначну кількість бактерій. Вода, на одиницю об'єму якої в середньому припадає менше однієї бактерії, вважається придатною для зазначеного хімічного виробництва. Якщо ж середня кількість бактерій на одиницю об'єму води дорівнює одній або більше, то вода непридатна для використання.

Звичайна методика контролю води на чистоту така. Береться 10 (у загальному випадку n) проб води одиничного об'єму. Потім кожну з цих проб додають у колби із живильним середовищем, які тримають за температури, сприятливої для росту бактерій. Якщо проба забруднена, тобто містить принаймні одну бактерію, то колонія бактерій росте і розчин, який спочатку був прозорий, стає каламутним. Про чистоту води судять за кількістю ξ забруднених проб: якщо ця кількість не перевищує певного числа l, то вода вважається чистою; у супротивному разі — забрудненою, тоді провадиться додаткове очищення води і вона знову досліджується на чистоту. Додаткове очищення води, природно, вимагає додаткового труда і витрат проте, якщо вода, що використовується, містить значну кількість бактерій, то зумовлені цим збитки значно більші.

Яким має бути l, щоб вода з середнім вмістом бактерій, що дорівнює одній бактерії на одиницю об'єму, виявлялася з імовірністю 0,99? Знайти l, сформулювавши поставлену задачу як задачу перевірки статистичних гіпотез.

Нехай тепер рівень значущості й число n проб води не фіксовані. Дослідити як змінюється функція потужності критерію залежно від його рівня значущості і від n, побудувати графіки функції потужності критерію.

З а у в а ж е н н я. Як розподіл кількості бактерій в одиничному об'ємі води з середнім вмістом λ бактерій на одиницю об'єму природно прийняти пуассонів розподіл з параметром λ.

Р о з в' я з а н н я. При тестуванні води на забруднення насамперед важливо не пропустити забруднену воду. Тому за H_0 візьмемо гіпотезу "середня кількість бактерій на одиницю об'єма води = 1". Альтернативна гіпотеза

$$H_1 = \{H_\lambda, \lambda \in (0;1)\},$$

— "середня кількість бактерій на одиницю об'єма води менше 1". Тут H_λ — гіпотеза "середня кількість бактерій на одиницю об'єма води становить λ". Зазначимо, що альтернативна гіпотеза є складною і однобічною. Ми вибираємо її однобічною, причому саме такою, оскільки нам необхідно отримати відповідь на питання "чи можна вважати, що середня кількість бактерій на одиницю об'єма води менше 1".

Про чистоту води судять за кількістю ξ забруднених проб. Якщо ξ не перевищує певного l, то воду класифікуємо як чисту, якщо ж $\xi > l$, то воду класифікуємо як забруднену. Тому за критичну множину для перевірки гіпотези $H_0\colon \lambda = 1$ проти альтернативи $H_1\colon \lambda < 1$ природно розглянути множину S вигляду

$$S = \{k\colon\ k \le l\}.$$

Якщо $\xi \in S$ — кількість забруднених проб мала, то гіпотезу H_0 відхиляємо, воду класифікуємо як чисту (придатну для виробництва), у супротивному разі — ні.

Число l (а разом з ним і множину $S = \{k : k \le l\}$) виберемо так, щоб критерій забруднену воду (з $\lambda = 1$) виявляв з імовірністю 0,99, тобто

$$\mathsf{P}\{\xi \in \overline{S}|H_0\} \ge 0{,}99.$$

Докладніше,

$$\mathsf{P}\{\xi > l|H_0\} \ge 0{,}99,$$

або, що те ж саме,

$$\mathsf{P}\{\xi \le l|H_0\} \le 0{,}01.$$

Цей критерій має рівень значущості 0,01.

1. Знайдемо l. Почнемо з випадку, коли береться $n = 10$ проб води. Випадкова величина ξ (кількості забруднених проб з n) має біномний розподіл з параметрами $(10; p)$, де p — імовірність успіху, тобто імовірність того, що вода у даній пробі помутніє.

Чому дорівнює p? Проба помутніє, якщо у одиничному об'ємі води є хоча б одна бактерія. За умовою кількість бактерій на одиницю об'єма є пуассоновою випадковою величиною з параметром λ. А тому ймовірність того, що у одиничному об'ємі є хоча б одна бактерія, дорівнює

$$p = 1 - \mathsf{P}\{\eta = 0\} = 1 - e^{-\lambda}\frac{\lambda^0}{0!} = 1 - e^{-\lambda}.$$

Зокрема, для $\lambda = \lambda_0 = 1$ маємо

$$p = p_0 = 1 - e^{-\lambda_0} = 1 - e^{-1}.$$

Число l вибираємо якомога більшим, але так, щоб

$$\mathsf{P}\{\xi \le l | H_0\} \le 0{,}01,$$

докладніше,

$$\mathsf{P}\{\xi \le l | H_0\} = \sum_{k=0}^{l} C_{10}^k p_0^k (1 - p_0)^{10-k} \le 0{,}01 \qquad (5)$$

(l вибираємо якомога більшим щоб “максимізувати” функцію потужності критерію).

Обчислюємо послідовно суми

$$\sum_{k=0}^{l} C_{10}^k p_0^k (1 - p_0)^{10-k}$$

для $l = 0, 1, 2, \dots$ доти, поки не знайдемо таке l, що

$$\sum_{k=0}^{l} C_{10}^k p_0^k (1 - p_0)^{10-k} \le 0{,}01,$$

а значення

$$\sum_{k=0}^{l+1} C_{10}^k p_0^k (1-p_0)^{10-k} > 0{,}01.$$

Це l і є шуканим.

Для обчислення біноміальних імовірностей $B_{n;p}(k)$ зручно використовувати R:

`dbinom(k, size = n, prob = p)`.

Для $l = 0$ і $l = 1$ значення $\mathsf{P}\{\xi \le l|H_0\} \le 0{,}001$; для $l = 2$ значення $\mathsf{P}\{\xi \le l|H_0\} = 0{,}0069$; для $l = 3$ значення $\mathsf{P}\{\xi \le l|H_0\} = 0{,}0345$. Тому найбільшим l, що задовольняє умову (5), є $l = 2$.

Графік функції потужності

$$\beta(\lambda) = \mathsf{P}\{\xi \le 2|H_\lambda\}, \quad \lambda \in (0;1),$$

критерію $S = \{k : k \le 2\}$ зображено на рис. 1. У табл. 1 наведено значення $\beta(\lambda)$ у деяких точках з $(0;1)$.

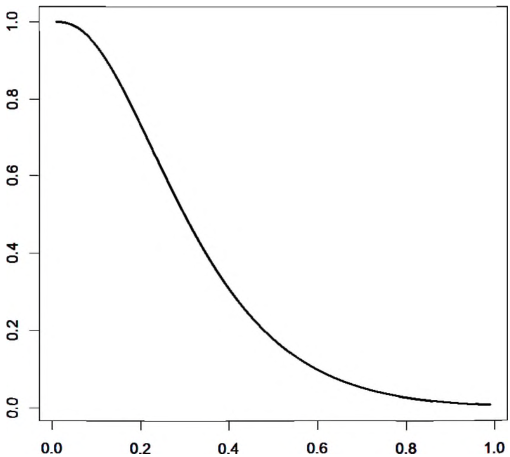

Рис. 1: Графік функції потужності критерію, $n = 10, l = 2$

Критерій $S = \{k : k \leq 2\}$ забруднену воду з $\lambda = 1$ класифікує як забруднену з імовірністю $1 - 0{,}007 = 0{,}993$ (на 100 партій забрудненої води близько 99 бракуватимуться).

Табл. 1: Значення функції потужності критерію, $n = 10, l = 2$

λ	0,1	0,2	0,3	0,4	0,5
$\beta(\lambda)$	0,938	0,734	0,498	0,308	0,178
λ	0,6	0,7	0,8	0,9	1,0
$\beta(\lambda)$	0,098	0,052	0,027	0,014	0,007

Проте досить чиста вода ($\lambda = 0{,}2$) буде витримувати контроль з імовірністю 0,734; вода з $\lambda = 0{,}4$ витримуватиме контроль з імовірністю 0,308; а вода з $\lambda = 0{,}5$ витримуватиме контроль з імовірністю 0,178 (тобто у середньому тільки у 18 випадках на 100). Дивись також табл. 1.

Користуючись критерієм $S = \{k : k \leq 2\}$, досить чисту воду будемо класифікувати як забруднену і направляти на додаткове очищення. Останнє означає додаткові витрати.

2. Збільшимо число проб з $n = 10$ до $n = 20$. З тих же міркувань, що і у пункті 1, знаходимо критичну множину для перевірки H_0. Для $n = 20$ маємо

$$S = \{k : k \leq 7\}.$$

Графік функції потужності критерію $S = \{k : k \leq 7\}$

$$\beta(\lambda) = \mathsf{P}\{\xi \leq 2 | H_\lambda\}, \quad \lambda \in (0; 1),$$

зображено на рис. 2. Значення $\beta(\lambda)$ у деяких точках наведені у табл. 2.

Табл. 2: Значення функції потужності критерію, $n = 20, l = 7$

λ	0,1	0,2	0,3	0,4	0,5
$\beta(\lambda)$	1,000	0,982	0,879	0,674	0,440
λ	0,6	0,7	0,8	0,9	1,0
$\beta(\lambda)$	0,249	0,125	0,057	0,024	0,0097

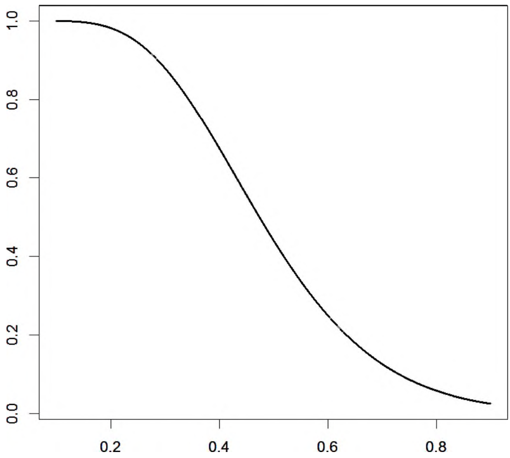

Рис. 2: Графік функції потужності критерію. $n = 20, l = 7$

У порівнянні з критерієм $S = \{k : k \leq 2\}$ для $n = 10$ проб критерій $S = \{k : k \leq 7\}$ для $n = 20$ збільшує імовірність класифікації води з $\lambda = 0{,}2$ як чистої з 0,734 до 0,982; води з $\lambda = 0{,}4$ як чистої з 0,308 до 0,674; а води з $\lambda = 0{,}5$ як чистої з 0,178 до 0,440 (див. табл. 1 і табл. 2).

3. Збільшимо число проб до $n = 50$. При цьому критична множина для перевірки H_0 набуває вигляду

$$S = \{k : k \leq 23\}.$$

Графік функції потужності критерію $\beta(\lambda) = \mathsf{P}\{S|H_\lambda\}$, $\lambda \in (0; 1)$, зображено на рис. 3, а значення $\beta(\lambda)$ у деяких точках наведені у табл. 3.

Табл. 3: Значення функції потужності критерію. $n = 50, l = 23$

λ	0,1	0,2	0,3	0,4	0,5
$\beta(\lambda)$	1,000	1,000	0,999	0,981	0,866
λ	0,6	0,7	0,8	0,9	1,0
$\beta(\lambda)$	0,607	0,318	0,126	0,039	0,0097

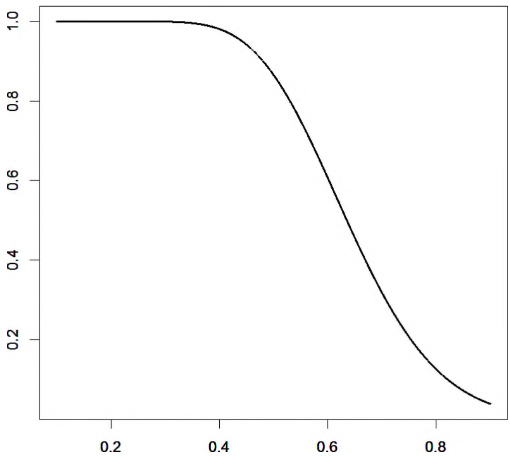

Рис. 3: Графік функції потужності критерію. $n = 50, l = 23$

Критерій $S = \{k : k \leq 23\}$ з імовірністю близькою до 1 класифікує воду з $\lambda = 0{,}1; 0{,}2; 0{,}3; 0{,}4$ як чисту (див. табл. 3). Забруднену воду з $\lambda = 1$ класифікує як забруднену з імовірністю $1 - 0{,}0097 = 0{,}9903$. А воду з $\lambda = 0{,}8; 0{,}9$ із забрудненням, близьким до критичного $\lambda = 1$ класифікує як чисту з невеликою імовірністю: $0{,}126; 0{,}039$ (див. табл. 3). Вода "середньої" чистоти — з $\lambda = 0{,}6$; $\lambda = 0{,}7$ класифікується "невпевнено", що цілком природно (див. табл. 3 та рис. 3).

Хоча гіпотеза H_0: $\lambda = 1$ перевіряється проти альтернативи H_1: $\lambda \in (0; 1)$, формально можна обчислювати значення функції потужності критерію $\beta(\lambda)$ і у точках $\lambda > 1$ (що відповідає "дуже" забрудненій воді).

Скажімо, у точці $\lambda = 1{,}2$ значення $\beta(1{,}2) = 0{,}0004$. Це означає, що вода з $\lambda = 1{,}2$ класифікується як чиста з імовірністю $0{,}0004$, а отже з імовірністю $1 - 0{,}0004 = 0{,}9996$ класифікується як забруднена. Таким чином, критерій S "правильно працює" і для значень $\lambda > 1$, класифікуючи воду з $\lambda > 1$ як забруднену з імовірністю, близькою до 1.

Наведемо лістинг на R.

```
Test_Pow <- function(lamb){
  #### n, l - глобальні змінні
  p <- 1 - exp(-lamb)
  bin_probs_krit_arr <- dbinom(0:l,
                   size = n, prob = p)
  Test_Pow_value <-
               sum(bin_probs_krit_arr)
  return(Test_Pow_value)  }


lamb0 <-  1
p0 <- 1 - exp(-lamb0)

###  Почнемо з n=10 і малих l
n <- 10
l <- 0
(bin_probs_krit_arr <- dbinom(0:l,
               size = n, prob = p0))
(Prob_TypeI_err <-
              sum(bin_probs_krit_arr) )
###

l=1
( bin_probs_krit_arr <- dbinom(0:l,
                 size = n, prob = p0)  )
( Prob_TypeI_err <-  sum(bin_probs_krit_arr) )
###



## ------------------------------------
##     ------- n=10,  l=2  ------------
###
n <- 10
l <-  2
( bin_probs_krit_arr <- dbinom(0:l,
          size = n, prob = p0))
( Prob_TypeI_err <-
                        sum(bin_probs_krit_arr))
###  0.00686
```

```
####    Графік функції потужності критерію
lamb_arr <- seq(from=0.01, to=0.99,
                                by=0.0001)
T_Power_arr <- sapply(lamb_arr, Test_Pow)
plot(lamb_arr, T_Power_arr, cex=0.3)
####


###  ---- n=10, l=2, таблиця -----
##
lamb_1_arr <- seq(from=0.1, to=0.9, by=0.1)
T_Power_1_arr <- sapply(lamb_1_arr,
                            Test_Pow)
tabl_1 <- rbind(lamb_1_arr, T_Power_1_arr)
print(tabl_1)


## __________________
###  n=10, l=3
##
l <- 3
bin_probs_krit_arr <- dbinom(0:l,
         size = n, prob = p0)
( Prob_TypeI_err <-  sum(bin_probs_krit_arr)  )
### n=10, l=3:  Вже занадто велике


### _____________________________________
##
####        Збільшимо n:   n=20
n <- 20

l=3
bin_probs_krit_arr <- dbinom(0:l,
          size = n, prob = p0)
( Prob_TypeI_err <-  sum(bin_probs_krit_arr)  )
### n=20, l=3:  1.31e-05

l=5
bin_probs_krit_arr <- dbinom(0:l,
```

```
         size = n, prob = p0)
( Prob_TypeI_err <-  sum(bin_probs_krit_arr))
### n=20, l=5:  0.0006

l=7
bin_probs_krit_arr <- dbinom(0:l,
             size = n, prob = p0)
(Prob_TypeI_err <-  sum(bin_probs_krit_arr))


###  ----- n=20, l=7 ---------

####    Графік ф. потужності критерію
n <- 20
l <-  7
lamb_arr <- seq(from=0.1, to=0.9,
                                by=0.002)
T_Power_arr <- sapply(lamb_arr, Test_Pow)
plot(lamb_arr, T_Power_arr, cex=0.3)


###  Табл. значень ф. потужності критерію:
lamb_1_arr <- seq(from=0.1, to=0.9, by=0.1)
T_Power_1_arr <- sapply(lamb_1_arr, Test_Pow)
tabl_1 <- rbind(lamb_1_arr, T_Power_1_arr)
print(tabl_1)


### ____________________________________
##
####   III)   ще  збільшимо n:   n=50

n <- 50
l <-  5
bin_probs_krit_arr <- dbinom(0:l, size = n,
                        prob = p0)
( Prob_TypeI_err <-  sum(bin_probs_krit_arr)  )
### n=50, l=5:  6.528e-15

l=17
bin_probs_krit_arr <- dbinom(0:l, size = n,
                        prob = p0)
```

```
( Prob_TypeI_err <-  sum(bin_probs_krit_arr)  )
### n=50, l=17:  2.624e-05



### ------------------------------
### ------- n=50, l=23   ---------
##
n <- 50
l <-  23
bin_probs_krit_arr <- dbinom(0:l,
          size = n, prob = p0)
( Prob_TypeI_err <-  sum(bin_probs_krit_arr)  )
### n=50, l=23:
##  Імовірність помилки 1-го роду = 0.0097



####    Графік ф. потужності критерію
cat("n=", n, "\n")
cat("l=", l, "\n")
lamb_arr <- seq(from=0.1, to=0.9, by=0.0002)
T_Power_arr <- sapply(lamb_arr, Test_Pow)
plot(lamb_arr, T_Power_arr, cex=0.3)



####  ------ Таблиця,  n=50, l=23  ----------
n <- 50
l <-  23
lamb_1_arr <- seq(from=0.1, to=0.9, by=0.1)
T_Power_1_arr <- sapply(lamb_1_arr, Test_Pow)
tabl_1 <- rbind(lamb_1_arr, T_Power_1_arr)
print(tabl_1)
```

Приклад 3 (булочки з ізюмом).[1] Державним стандартом установлено, що при випіканні солодких булочок на 1000 виробів має припадати 10 000 родзинок (у середньому 10 на одну булочку). У нас, однак, є сумнів, що всі родзинки використано за призначенням (їх могли, при-

[1]Задача запозичена з книжки В.М. Тутубаліна.

наймні частково, використати на інші потреби), і ми хочемо перевірити, чи це так. Для цього купуємо одну булочку й рахуємо кількість родзинок у ній: виявилося, що в булочці $\xi(\omega)$ родзинок. За кількістю $\xi(\omega)$ родзинок у булочці треба дійти висновку, чи всі родзинки використано за призначенням.

Розв'язати поставлену задачу, сформулювавши її як задачу перевірки статистичних гіпотез.

Як зміниться висновок, коли родзинки підраховують у двох булочках?

Дослідити як змінюється функція потужності критерію залежно від його рівня значущості, побудувати графіки функцій потужності критерію.

Розв'язання. Почнемо розв'язувати задачу з випадку, коли ізюм підраховується в одній булочці.

У задачі фактично мова йде про частку θ родзинок, які використані за призначенням, а саме, $\theta = 1$ чи $\theta < 1$?

Позначимо через H_θ гіпотезу "частка θ родзинок, яка пішла у булочки, становить θ", де $\theta \in (0;1)$.

За нульову гіпотезу розглянемо гіпотезу $H_0 : \theta = 1$ — усі родзинки пішли у булочки. Альтернативна до H_0 гіпотеза — $H_{(0;1)} = \{H_\theta,\ \theta \in (0;1)\}$. Альтернативна гіпотеза $H_{(0;1)}$ складна — коли $H_{(0;1)}$ справджується, вірною є одна з простих гіпотез $H_\theta, \theta \in (0;1)$.

Гіпотезу $H_0 : \theta = 1$ (усі родзинки пішли у булочки) природно відхиляти, якщо число ξ родзинок у купленій булочці мале — не перевищує певного l ($\xi \le l$) і не відхиляти у супротивному разі, тобо коли $\xi > l$. Тому в якості критерію для перевірки гіпотези H_0 проти альтернативи $H_{(0;1)}$ природно вибрати множину S вигляду

$$S = \{k\colon\ k \le l\},$$

де число l вибираємо так, щоб забезпечити даний рівень значущості α (щоб імовірність помилки першого роду не перевищувала α).

Розподіл кількості родзинок ξ у булочці можна вважати пуассоновим. Справді, коли гіпотеза H_θ справджується, число родзинок, що пішли у булочки, становить $\theta \cdot 10\,000$. Кожна з $\theta \cdot 10\,000$ ізюминок може або потрапити у одну з куплену з $1\,000$ булочок, або не потрапити, причому імовірність потрапити дорівнює $1/1000$. Таким чином, маємо послідовність з $\theta \cdot 10\,000$ незалежних випробувань з імовірністю успіху $1/1000$. За великого числа випробувань $\theta \cdot 10\,000$ і малої імовірності успіху $1/1000$ можна вважати (на основі теореми Пуассона), що число успіхів (число родзинок у булочці) має пуассонів розподіл з параметром

$$\theta \cdot 10\,000 \cdot (1/1000) = 10 \cdot \theta,$$

тобто

$$P\{\xi = k\} = \frac{(10 \cdot \theta)^k}{k!} e^{-10 \cdot \theta}, \quad k = 0, 1, 2, \ldots$$

Коли гіпотеза $H_0 : \theta = 1$ справджується, за розподілом ξ знайдемо l як найбільше число, для якого

$$P\{\xi \in S | H_0\} = P\{\xi \le l | H_0\} = \sum_{k=0}^{l} \frac{10^k}{k!} e^{-10} \le \alpha,$$

а разом з цим отримаємо критерій $S = \{k : \ k \le l\}$ для перевірки гіпотези $H_0 : \theta = 1$ з рівнем значущості α.

Нехай $\alpha = 0{,}10$. Обчислюватимемо послідовно суми

$$\sum_{k=0}^{l} \frac{10^k}{k!} e^{-10}$$

для $l = 0, 1, 2, \ldots$ доти, поки не знайдемо таке l, що

$$\sum_{k=0}^{l} \frac{10^k}{k!} e^{-10} \le 0{,}10,$$

а значення

$$\sum_{k=0}^{l+1} \frac{10^k}{k!} e^{-10} > 0{,}10.$$

Це l і буде шуканим.

Для обчислення $(\lambda^k/k!)e^{-\lambda}$, $k = 0, 1, 2, \ldots$, зручно користуватися `R`:

`dpois(x=`k`, lambda=`λ`)`.

Для $l = 5$ маємо

$$\sum_{k=0}^{5} \frac{10^k}{k!} e^{-10} = 0{,}067 < 0{,}1,$$

а значення

$$\sum_{k=0}^{6} \frac{10^k}{k!} e^{-10} = 0{,}13 > 0{,}1.$$

Таким чином, $l = 5$.

Отже, $S = \{k : \ k \le 5\}$ — критерій (що відповідає рівню значущості 0,10) для перевірки гіпотези $H_0 : \theta = 1$ проти альтернативи $H_{(0;1)} = \{H_\theta,\, \theta \in (0;1)\}$.

Функція потужності критерію S:

$$\beta(\theta) = \mathsf{P}(S|H_\theta), \quad \theta \in (0;1),$$

докладніше,

$$\beta(\theta) = P\{\xi \le 5|H_\theta\} = \sum_{k=0}^{5} \frac{(10\cdot\theta)^k}{k!} e^{-10\cdot\theta}, \quad \theta \in (0;1).$$

Графік функції потужності критерію $\beta(\theta)$ зображено на рис. 4. Значення $\beta(\theta)$ у деяких точках наведені у табл. 4.

Табл. 4: Значення функції потужності критерію (ізюм підраховується в одній булці)

θ	0,1	0,2	0,3	0,4	0,5
$\beta(\theta)$	0,999	0,983	0,916	0,785	0,616
θ	0,6	0,7	0,8	0,9	1,0
$\beta(\theta)$	0,446	0,301	0,191	0,116	0,067

Дамо частотну інтерпретацію отриманих результатів. Згідно з табл. 4 імовірність виявити нестачу, якщо частка використаного за призначенням ізюму становить 0,1, дорівнює 0,999 (така нестача виявляється критерієм майже завжди). Якщо частка використаного ізюму становить 0,3, то критерій виявляє це з імовірністю 0,916 (десь у 91 випадку на 100). А ось якщо частка використаного ізюму дорівнює 0,5, то така нестача виявляється тільки у 61 випадку на 100; нестача ж, коли 0,9 ізюму йде за призначенням, виявляється тільки у 11 випадках на 100. Таким чином, критерій досить добре виявляє значну нестачу ізюму, незначна нестача виявляється критерієм значно рідше.

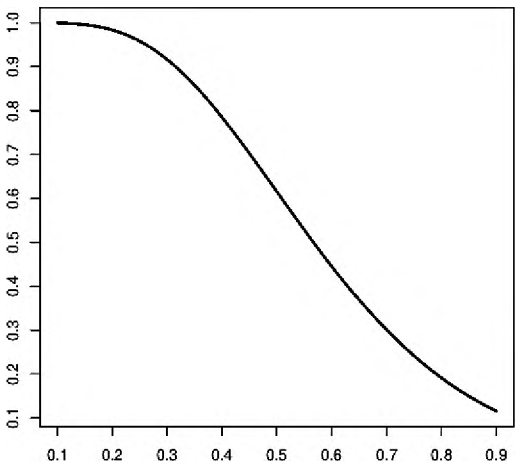

Рис. 4: Графік функції потужності критерію (ізюм підраховується в одній булці)

Значення імовірності $P\{S|H_0\} = 0{,}067$ помилки першого роду означає, що в середньому у 7 випадках на 100 ми будемо підозрювати, що не весь ізюм пішов у булочки, хоча це не так.

Дамо тепер відповідь на те ж саме питання "Чи всі родзинки йдуть за призначенням?", якщо родзинки підраховуються у двох булочках.

Нехай ξ — кількість родзинок у двох булочках. Випадкову величину ξ можна подати у вигляді суми незалежних випадкових величин η і ζ:

$$\xi = \eta + \zeta,$$

де η — кількість родзинок у першій булочці, ζ — у другій. η і ζ можна вважати незалежними — кількість булочок є великою. Кожна з випадкових величин η і ζ має пуассонів розподіл з параметром $10 \cdot \theta$, а тому випадкова величина $\xi = \eta+\zeta$ має розподіл Пуассона з параметром $10\cdot\theta+10\cdot\theta = 20 \cdot \theta$.

Нульова гіпотеза — $H_0 : \theta = 1$, альтернативна гіпотеза — $H_{(0;1)} = \{H_\theta,\ \theta \in (0;1)\}$. Критична множина S для перевірки H_0 проти альтернативи $H_{(0;1)}$ має вигляд

$$S = \{k \colon\ k \le l\}.$$

Для рівня значущості $\alpha = 0{,}10$ ціле число l вибираємо як найбільше число, для якого

$$\mathsf{P}\{S|H_0\} = \mathsf{P}\{\xi \le l|H_0\} = \sum_{k=0}^{l} \frac{20^k}{k!} e^{-20} \le 0{,}10.$$

Таким l є 13, оскільки

$$\sum_{k=0}^{13} \frac{(20)^k}{k!} e^{-20} = 0{,}066 < 0{,}1,$$

$$\sum_{k=0}^{14} \frac{(20)^k}{k!} e^{-20} = 0{,}105 > 0{,}1.$$

Функція потужності критерію:

$$\beta(\theta) = \mathsf{P}\{S|H_\theta\} = \mathsf{P}\{\xi \le 13|H_\theta\} =$$

$$= \sum_{k=0}^{13} \frac{(20 \cdot \theta)^k}{k!} e^{-20\cdot\theta},$$

$\theta \in (0;1)$.

Графік функції потужності критерію $\beta(\theta)$ зображено на рис. 5, а значення функції потужності критерію у деяких точках наведені у табл. 5.

Табл. 5: Значення функції потужності критерію (ізюм підраховується у двох булках)

θ	0,1	0,2	0,3	0,4	0,5
$\beta(\theta)$	1,000	1,000	0,996	0,966	0,864
θ	0,6	0,7	0,8	0,9	1,0
$\beta(\theta)$	0,682	0,464	0,275	0,143	0,066

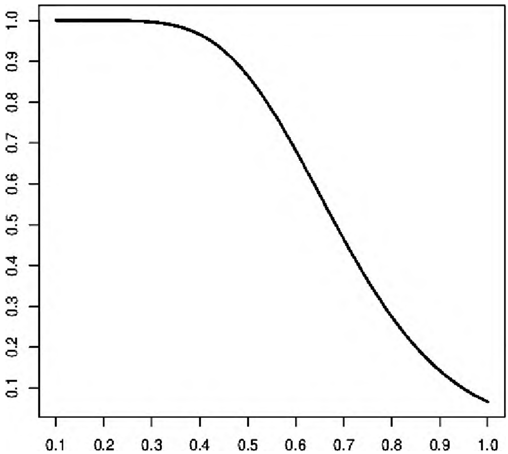

Рис. 5: Графік функції потужності критерію (ізюм підраховується у двох булках)

Згідно з табл. 5 і у порівнянні з табл. 4 видно, що критерій $S = \{k : k \leq 13\}$ виявляє нестачу ізюму суттєво краще — якщо частка використаного ізюму становить 0,1; 0,2; 0,3; 0,4, то це виявляється практично завжди (див. табл. 5). Помітно частіше нестача виявляється і при більших значеннях θ. Скажімо, нестача коли $\theta = 0{,}8$ критерієм $S = \{k : k \leq 5\}$ виявляється у 19 випадках на 100, а критерій $S = \{k : k \leq 13\}$ виявляє ту ж саму нестачу у 28 випадках на 100.

Графік функції потужності $\beta(\theta) = \mathsf{P}\{S|H_\theta\}$, $\theta \in (0;1)$, критерію $S = \{k: k \le 13\}$ "проходить вище" графіка функції потужності $\beta(\theta) = \mathsf{P}\{S|H_\theta\}$, $\theta \in (0;1)$, критерію $S = \{k\colon k \le 5\}$ (див. табл. 4, 5 і рис. 4, 5).

Наведемо лістинг.

```
###  Функція потужності критерію,
##   1 булочка
Test_Pow <- function(lamb1) {
   t_pow <-  sum(dpois(0:5, lambda = lamb1))
   return(t_pow)
} ###


### Функція потужності критерію,
##   2 булочки
Test_Pow_2buns <- function(lamb1) {
   t_pow <-  sum(dpois(0:13, lambda = lamb1))
   return(t_pow)
} ###



### ------   ДЛЯ ОДНІЄЇ БУЛОЧКИ ----------
##  Імовірність помилки першого роду:
sum(dpois(0:5, lambda = 10))
###  0.067
sum(dpois(0:6, lambda = 10))
##  0.130  - вже більше 0.10

## -- Графік ф. потужності кр-рію --
lamb_grid <- seq(from=0.1, to=0.9, by=0.001)
t_pow_arr <- sapply(10*lamb_grid,
                    Test_Pow)
plot(lamb_grid, t_pow_arr, cex=0.2)


###  Таблиця значень
##   ф. потужності критерію
lamb_1_arr <- seq(from=0.1, to=1, by=0.1)
T_Power_1_arr <- sapply(10*lamb_1_arr,
```

```
                        Test_Pow)
tabl_1 <- rbind(lamb_1_arr, T_Power_1_arr)
print(tabl_1)


## ------------------------------
### -----   ДЛЯ 2-х БУЛОЧОК -----
#
##  Імовірність помилки першого роду:
sum(dpois(0:13, lambda = 20))
###  0.066
sum(dpois(0:14, lambda = 20))
###  0.105

###  Графік ф. потужності кр-рію
lamb_grid <- seq(from=0.1, to=1.0, by=0.001)
t_pow2_arr <- sapply(2*10*lamb_grid,
                         Test_Pow_2buns)
plot(lamb_grid, t_pow2_arr, cex=0.2)

#### Таблиця значень
##   ф. потужності критерію (2 булочки)
lamb_1_arr <- seq(from=0.1, to=1, by=0.1)
T_Power_1_arr <- sapply(20*lamb_1_arr,
                            Test_Pow_2buns)
tabl_1 <- rbind(lamb_1_arr, T_Power_1_arr)
print(tabl_1)
```

Приклад 4 (контроль міцності дроту). Для сталевого дроту, що йде на виготовлення канатів, середнє граничне значення розтягаючого зусилля має бути не меншим ніж 6720 кг/см^2. Серія послідовно проведених випробувань показала, що середнє граничне розтягаюче зусилля, яке витримує дріт, можна вважати нормально розподіленим, а стандартне відхилення становить 220 кг/см^2. З метою контролю величини граничного розтягаючого зусилля з кожної партії дроту, що надходить на завод, перевіряють 20 екземплярів. Результати одного з таких випробувань наведені нижче:

6300, 6870, 6720, 6980, 6780, 6780, 6760, 6720, 6630, 6650, 6900, 7130, 6690, 6750, 6560, 6700, 6930, 6720, 6950, 6960.

Чи можна на підставі цих даних стверджувати, що для даної партії дроту середнє граничне розтягаюче зусилля не менше ніж 6720 кг/см^2?

Розв'язати задачу, сформулювавши її у термінах перевірки статистичних гіпотез:

— вибрати нульову гіпотезу й альтернативну;

— з'ясувати, у чому полягають помилки першого і другого роду;

— запропонувати критерій для перевірки нульової гіпотези (а разом з ним і рівень значущості);

— обчислити значення функції потужності критерію у точках $6000 + 50i$, $i = 10, 11, \dots, 15$;

— яка ймовірність забракувати партію, для якої середнє граничне розтягаюче зусилля дорівнює $6500, 6550, 6600$;

— якою має бути кількість зразків для контролю, щоб партія із середнім граничним розтягаючим зусиллям на разрив рівним 6600 виявлялась з імовірністю 0,99;

— як буде змінюватися потужність критерію зі зміною n; що зміниться, якщо для контролю вибирається 10 зразків, 30 зразків;

— дати частотну інтерпретацію отриманим результатам.

Р о з в' я з а н н я. За нульову гіпотезу розглянемо гіпотезу H_0 — вибірка отримана з $N_{a_0;\sigma^2}$, $a_0 = 6720$. Оскільки наша задача — виявити дріт, для якого середнє граничне розтягаюче зусилля дорівнює менше 6720 кг/см^2, то за альтернативу будемо розглядати однобічну альтернативу H_1: $a < a_0$. За такого вибору нульової та альтернативної гіпотез помилка першого роду буде полягати у тому, що дріт із середнім граничним розтягаючим зусиллям на разрив 6720 кг/см^2 бракуватиметься — тобто класифікуватиметься як дріт із середнім граничним розтягаючим зусиллям на разрив меншим 6720 кг/см^2. Помилка другого роду буде полягати у тому, що дріт із середнім граничним розтягаючим зусиллям на разрив, що дорівнює a ($a < 6720$), класифікуватиметься як дріт , що витримує середнє граничне розтягаюче зусилля на розрив 6720 кг/см^2 — тобто витримуватиме контроль.

Для перевірки гіпотези H_0: $a = a_0$ $(a_0 = 6720)$ проти однобічної альтернативи H_1: $a < a_0$ скористаємося однобічним критерієм: якщо

$$(\bar{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1},$$

то гіпотеза H_0 відхиляється, у супротивному разі — ні (рівень значущості критерію дорівнює α).

У задачі, що розглядається, $n = 20$; $\bar{\xi} = 6775$; $\sigma = 220$. Рівень значущості α виберемо рівним 0,05, при цьому $N_{0{,}05;0;1} = 1{,}65$ ($N_{0{,}05;0;1}$ — верхня 0,05-межа $N_{0;1}$-розподілу — знайдено за таблицями нормального розподілу, див. [1]). Нормоване відхилення

$$(\bar{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} = (6775 - 6720)\Big/\frac{220}{\sqrt{20}} = 1{,}13 >$$

$$> -1{,}65 = -N_{0{,}05;0{,}1}.$$

Тому на 5 % рівні значущості гіпотеза H_0 — середнє граничне розтягаюче зусилля на розрив дорівнює 6720 кг/см^2, не відхиляється.

Значення функції потужності критерію

$$S = \left\{(x_1, x_2, \ldots, x_n):\ (\bar{x} - a_0)\Big/\frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1}\right\} \tag{6}$$

у точці a дорівнює

$$\beta(a) = \mathsf{P}\left\{(\bar{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1}\middle| H_a\right\} =$$

$$= \mathsf{P}\left\{(\bar{\xi} - 6720)\Big/\frac{220}{\sqrt{20}} < -N_{\alpha;0;1}\right\},$$

де $\bar{\xi} = \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ — оцінка параметра a, знайдена за вибіркою $\xi_1, \xi_2, \ldots, \xi_n$ з $N_{a;\sigma^2}$. Тому $\beta(a)$ можна переписати так:

$$\beta(a) = \mathsf{P}\left\{(\bar{\xi} - a_0)\Big/\frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1}\right\} =$$

$$= \mathsf{P}\left\{((\bar{\xi}-a)-(a_0-a))\Big/\frac{\sigma}{\sqrt{n}} < -N_{\alpha;0;1}\right\} =$$

$$= \mathsf{P}\left\{(\bar{\xi}-a)\Big/\frac{\sigma}{\sqrt{n}} < (a_0-a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right\} =$$

$$= N_{0;1}\left((a_0-a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right),$$

або, враховуючи, що $a_0 = 6720, \sigma = 220, n = 20$, так:

$$\beta(a) = N_{0;1}\left((a_0-a)\Big/\frac{\sigma}{\sqrt{n}} - N_{\alpha;0;1}\right) =$$

$$= N_{0;1}\left((6720-a)\Big/\frac{220}{\sqrt{20}} - 1{,}65\right)$$

($N_{0;1}(x)$ — значення функції нормального розподілу з параметрами $(0;1)$ у точці x).

Значення $\beta(a)$ функції потужності критерію S у деяких точках наведено у табл. 6. Графік функції потужності критерію (6) подано на рис. 6.

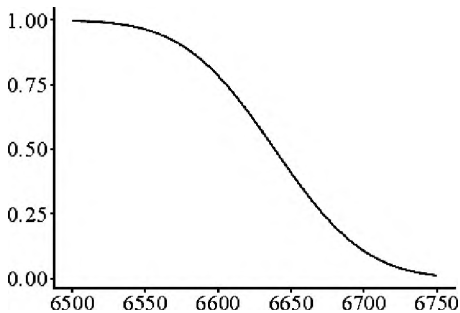

Рис. 6: Функція потужності критерію (6)

Табл. 6: Значення функції потужності критерію (6), $n = 20$

a	6500	6550	6600	6650	6700	6720	6750
$\beta(a)$	0,997	0,965	0,785	0,409	0,108	0,049	0,012

Імовірність забракувати партію (класифікувати її як таку, що має середнє граничне зусилля на розрив менше ніж 6720), для якої це значення насправді 6500, дорівнює $\beta(a) = \beta(6500) = 0{,}99$, для партії з $a = 6550$ та з $a = 6600$ ці ймовірності дорівнюють відповідно $\beta(6550) = 0{,}96$ і $\beta(6600) = 0{,}79$.

Кількість n зразків, необхідне для виявлення з імовірністю 0,99 партії з середнім граничним зусиллям на розрив 6600, знаходимо з співвідношення

$$\beta(6600) = N_{0;1}\left((6720 - 6600)\Big/\frac{220}{\sqrt{n}} - 1{,}65\right) = 0{,}99.$$

З останньої рівності, скориставшись таблицею нормального розподілу (див. [1]), маємо:

$$120\Big/\frac{220}{\sqrt{n}} - 1{,}65 = 2{,}33.$$

Звідси знаходимо, що необхідний обсяг вибірки має становити $n = 54$.

З ростом n за одного й того ж рівня значущості значення функції потужності критерію (для $a < 6720$) зростають (див. табл. 6, 7, 8).

Для $n = 10$ значення функції потужності критерію S у деяких точках наведено у табл. 7, а для $n = 30$ — у табл. 8.

Табл. 7: Значення функції потужності критерію (6), $n = 10$

a	6500	6550	6600	6650	6700	6720	6750
$\beta(a)$	0,936	0,788	0,532	0,261	0,087	0,049	0,019

Табл. 8: Значення функції потужності критерію (6), $n = 30$

a	6500	6550	6600	6650	6700	6720	6750
$\beta(a)$	0,999	0,996	0,910	0,534	0,125	0,049	0,008

Здобутим результатам можна дати таку частотну інтерпретацію (для $n = 20$).

Імовірність помилки першого роду 0,049 означає, що якщо проводити контроль із відбором 20 зразків, то на 100 партій із середнім граничним зусиллям на розрив $a = 6720$ кг/см^2 у середньому п'ять партій будуть бракуватися.

Значення $\beta(a) = \beta(6500) = 0{,}99$ функції потужності критерію у точці 6500 інтерпретується так. На 100 партій із середнім граничним зусиллям на розрив $a = 6500$ кг/см^2 у середньому 99 будуть бракуватися. Аналогічно інтерпретуються значення $\beta(a)$ в інших точках a.

Далі, хоча гіпотеза H_0: $a = 6720$ перевіряється проти альтернативи H_1: $a < 6720$, ми можемо формально обчислювати значення функції потужності критерію $\beta(a)$ і у точках $a > 6720$.

Наприклад, у точці $a = 6750$ значення $\beta(6750) = 0{,}012$ (див. табл. 6). Останнє означає, що за вірної гіпотези H_{6750} — середнє граничне зусилля на розрив дроту дорівнює 6750, гипотеза H_0 з імовірністю 0,012 відхилятиметься, і з імовірністю $1 - \beta(6750) = 0{,}988$ не відхилятиметься, тобто партія дроту з середнім граничним зусиллям на розрив, що дорівнює $a = 6750$ кг/см^2, витримуватиме контроль із імовірністю 0,988. Таким чином, наш критерій "правильно працює" і для значень параметра $a > 6720$ — з великою імовірністю не відхиляє якісний дріт (з $a > 6720$).

Література

[1] *Турчин В. М.* Теорія ймовірностей і математична статистика. Основні поняття, приклади, задачі. Дніпропетровськ: ІМА-прес, 2014. 556 с.

[2] *DeGroot M. H., Schervish M. J.* Probability and Statistics. Boston: Addison-Wesley, 2012. 893 p.

[3] *Ugarte M. D, Militino A. F., Arnholt A. T.* Probability and Statistics with R. Boca Raton: CRC Press, 2016. 930 p.

Зміст

1 Критерій, функція потужності критерію — **3**

2 Приклади — **11**