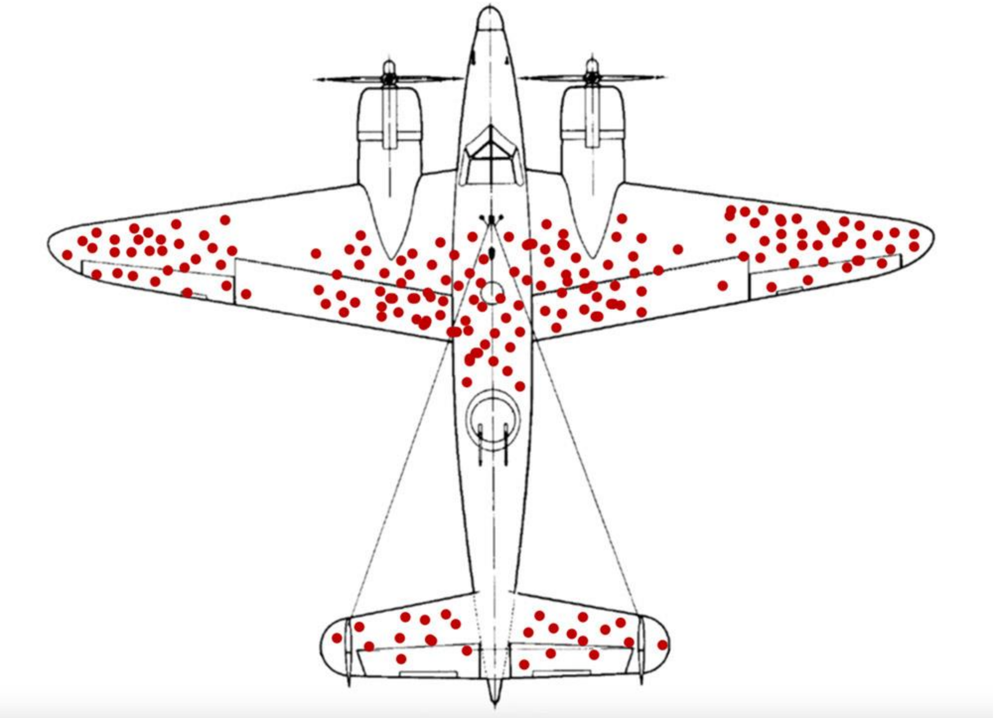
ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Помилка вцілілого

Дані: Вчені проаналізували літаки, що повернулися з бою, і нанесли на карту всі кульові отвори. Найбільше їх було на **крилах та фюзеляжі**.

Гіпотеза H0: Зміцнювати треба ті частини, де ми бачимо найбільше пробоїн (крила та фюзеляж).

Рішення Абрахама Вальда: Математик Абрахам Вальд відкинув цю гіпотезу. Він зрозумів, що дані мають критичну ваду: вони зібрані лише з літаків, що **повернулися**.



Він висунув **альтернативну гіпотезу**: якщо літак повернувся з дірками в крилах, значить, він може літати з такими пошкодженнями. А де ж дірки у тих літаків, що впали? Очевидно, вони отримали влучання в двигун та кабіну пілота.

Результат: Броню встановили на ті частини, де у літаків, що повернулися, не було жодної пробоїни.

Проблема 2000 року (Y2K). Перевірка гіпотези про граничні умови

Проблема: У багатьох старих комп'ютерних системах (особливо на мейнфреймах) для економії пам'яті рік зберігався у двозначному форматі (наприклад, "99" замість "1999").

Гіпотеза (H0): Коли настане 1 січня 2000 року, системи інтерпретують рік як "00", що буде сприйнято як 1900 рік. Це призведе до збоїв у розрахунках дат (нарахування відсотків, терміни придатності, розклади транспорту).

Дії: Світова IT-спільнота витратила мільярди доларів та роки праці, щоб перевірити цю гіпотезу. Програмісти вручну переглядали мільйони рядків старого коду, виправляючи формат зберігання дат

Результат: Перехід відбувся відносно гладко. Серйозних глобальних катастроф не сталося.

Корабель Vasa (помилка в специфікаціях та дизайні)

Проблема: Шведський король Густав II Адольф замовив будівництво нового, найпотужнішого військового корабля.

Гіпотеза (H0): Додавання другого ряду важких гармат (що було інновацією на той час) зробить корабель непереможним у бою.

Що сталося: Король постійно втручався у процес будівництва, вимагаючи більше гармат та багатшого оздоблення. Майстри, які будували корабель за старими кресленнями, не змогли правильно розрахувати стабільність корабля з новим, набагато важчим озброєнням, розташованим вище ватерлінії. Щоб компенсувати вагу гармат, вони поклали недостатньо баласту в трюм.

Результат: Під час свого першого ж виходу в море у 1628 році корабель проплив лише кілька сотень метрів. Коли дмухнув легкий вітерець, корабель нахилився і він затонув.

Основна термінологія

Статистична гіпотеза - це припущення про параметри або розподіл вибірки, яке підлягає перевірці за допомогою статистичних методів

Основні типи статистичних гіпотез: нульова гіпотеза (H0) та альтернативна гіпотеза (H1)

Нульова гіпотеза (H0) - це гіпотеза, що стверджує відсутність ефекту або різниці в даних

Альтернативна гіпотеза (H1) - це гіпотеза, що стверджує наявність ефекту або різниці в даних

Мета перевірки статистичних гіпотез - визначити, чи є достатні докази для відхилення нульової гіпотези

Статистичні критерії використовуються для оцінки ймовірності того, що спостережувані дані відповідають нульовій гіпотезі

Помилка першого роду - це відхилення нульової гіпотези, коли вона насправді є правильною

Помилка другого роду - це прийняття нульової гіпотези, коли насправді вона є неправильною

Основна термінологія

Потужність критерію - це ймовірність правильно відхилити нульову гіпотезу, коли вона є неправильною.
На потужність критерію впливають розмір вибірки, рівень значущості та величина ефекту.
Рівень значущості (α) - це ймовірність помилки першого роду, зазвичай встановлюється на рівні 0.05 або 0.01.
Найбільш поширені критерії: t-критерій Стьюдента, критерій χ^2, ANOVA, критерій Манна-Уїтні.
t-критерій Стьюдента використовується для порівняння середніх значень двох груп.
Критерій χ^2 використовується для перевірки незалежності між категоріальними змінними.
ANOVA (дисперсійний аналіз) використовується для порівняння середніх значень більше ніж двох груп.

Основна термінологія

Критична область - це область значень статистики, в якій нульова гіпотеза відхиляється.
p-значення - це ймовірність отримати спостережувані дані, якщо нульова гіпотеза є правильною. Якщо p-значення менше рівня значущості, нульову гіпотезу відхиляють
Двостороння перевірка гіпотез перевіряє можливість відхилення в обидва боки від нульової гіпотези
Одностороння перевірка гіпотез перевіряє можливість відхилення лише в одному напрямку

Основні етапи: формулювання гіпотез, вибір критерію, обчислення статистики, порівняння з критичною областю

Задача про числове значення математичного очікування при невідомій дисперсії

Тип задачі	**Що перевіряємо**	**Рекомендований критерій**
Порівняння середніх	Чи швидше працює код?	t-критерій Стьюдента
Порівняння частот	Чи чесна гральна кость?	Критерій χ2 (Пірсона)
Аналіз зв'язку	Чи залежить кількість помилок від мови програмування?	Кореляція (Пірсона/Спірмена)
Зміщена вибірка	Хто не потрапив у статистику?	Логічний аналіз (Вальд) / Корекція зміщення

Необхідно оцінити середнє значення (математичне очікування) генеральної сукупності на основі вибірки, **не знаючи** при цьому справжньої дисперсії (міри розсіювання) цієї сукупності

Для такої задачі зазвичай використовується **t-критерій Стьюдента**, де використовується вибіркова дисперсія як оцінка невідомої генеральної дисперсії

Порівняння продуктивності двох веб-серверів

Необхідно перевірити, чи працює новий веб-сервер (Сервер Б) швидше за старий (Сервер А). Справжній середній час відгуку чи дисперсії для жодного з серверів не відомий. Робиться по 30 випадкових запитів до кожного сервера і вимірюється час відгуку (в мілісекундах).

Гіпотеза H0: Середній час відгуку Сервера А дорівнює середньому часу відгуку Сервера Б ($\mu_A = \mu_B$).

Гіпотеза H1: Середній час відгуку Сервера Б менший за середній час відгуку Сервера А ($\mu_B < \mu_A$).

Використовується t-критерій Стьюдента, щоб порівняти вибіркові середні, оскільки справжні генеральні дисперсії часу відгуку не відомі

Формулювання гіпотези

У контексті комп'ютерних наук перевірка гіпотез про середнє значення **μ** найчастіше зустрічається під час оцінювання продуктивності систем або стабільності алгоритмів.

Сформулюємо нульову гіпотезу $\boldsymbol{H}_0$ про те, що невідомий параметр μ дорівнює заданому числу μ_0, тобто $\boldsymbol{H}_0 : \mu = \mu_0$. Альтернативну гіпотезу $\boldsymbol{H}_1$ можна сформулювати трьома способами:

1) $\boldsymbol{H}_1 : \mu \neq \mu_0$;
2) $\boldsymbol{H}_1 : \mu > \mu_0$;
3) $\boldsymbol{H}_1 : \mu < \mu_0$.

Перевірка гіпотези

Для обчислень вводиться критерій ϕ (у цьому прикладі цей критерій має розподіл Стьюдента)

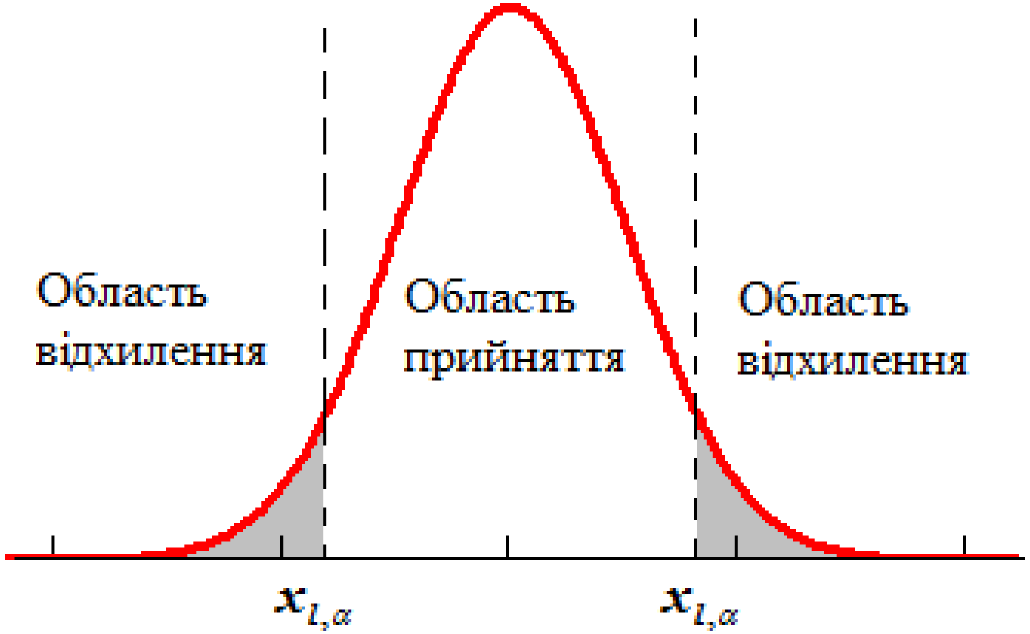
Області значень оцінок випадкової величини, у яких гіпотеза відхиляється, називається **областю відхилення** (**критичної областю**); область значень оцінок випадкової величини, у яких гіпотеза приймається, називається **областю прийняття гіпотези**

На малюнку представлений графік функції щільності ймовірності випадкової величини, на якому зазначені точки і для яких виконані умови:

$$P(\varphi < x_{l,\alpha}) = 0{,}5\alpha \text{ и } P(\varphi > x_{r,\alpha}) = 0{,}5\alpha$$

$$P(\varphi < x_{l,\alpha}) = 0{,}5\alpha \text{ и } P(\varphi > x_{r,\alpha}) = 0{,}5\alpha$$

α – рівень значущості: якщо ймовірність події не перевищує α, то подія малоймовірна, тобто якщо обчислене значення φ буде за інтервалом $(x_{l,\alpha}, x_{r,\alpha})$, всі підстави засумніватися в тому, що справжнє значення параметра μ дорівнює μ_0, і в цьому випадку гіпотезу про те, що $\mu = \mu_0$ слід відхилити. Якщо φ попадає в інтервал $(x_{l,\alpha}, x_{r,\alpha})$, то гіпотеза про те, що $\mu = \mu_0$ може бути прийнятою.



Розглянемо гіпотезу H_0 и першу альтернативну гіпотезу.

Якщо $\hat{x} = (x_1, x_2, \ldots, x_n)$ – вибірка з нормального розподілом і величини $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ та $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ – оцінки математичного очікування та дисперсії, то при перевірці гіпотези $H_0 : \mu = \mu_0$ використовується критерій:

$$\varphi = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}},$$

Який при виконання гіпотези H_0 має розподіл Стьюдента с числом ступенів свободи *n-1*.

Розподіл Стьюдента симетричний, тому достатньо обчислити

$$x_{r,\alpha}, \ x_{l,\alpha} = -x_{r,\alpha}.$$

При заданому рівні значущості α обчислюється значення $x_{r,\alpha}$ як розв'язок рівняння:

$$F_{n-1}(x_{r,\alpha}) = 1 - 0.5\alpha,$$

де $F_{n-1}(x_{r,\alpha})$ – функція розподілу Стьюдента с *n-1* степенями свободи.

Коли критична область знайдена, можна обчислити за вибіркою значення критерію φ і перевірити, чи воно потрапляє в критичну область. Якщо так, то гіпотеза приймається, інакше – відхиляється.

Ступені свободи

Ступені свободи - це кількість значень у підсумковому розрахунку статистики, які можуть змінюватися вільно, не порушуючи встановлених математичних обмежень

Нехай необхідно обрати 5 чисел, середнє значення яких має дорівнювати **10**.

•Перше число обираєте будь-яке (наприклад, 8).

•Друге - будь-яке (12).

•Третє - будь-яке (5).

•Четверте - будь-яке (15).

•А ось **п'яте число вже не можна обрати вільно**. Щоб середнє було рівно 10, сума всіх чисел має бути 50. Отже, останнє число жорстко визначене попередніми виборами (50 - 8 - 12 - 5 - 15 = 10).

У цьому прикладі було **4 ступені свободи** (n - 1). Останнє значення «пожертвувало» своєю свободою, щоб задовольнити обмеження (середнє значення).

$N := 20$

$X := \mathrm{READPRN}(\text{"c:\textbackslash MNV\textbackslash data.txt"})$

$X^T =$

	0	1	2	3	4	5	6	7	8	9
0	2.27	2.48	3.55	2.75	3.09	4.26	2.29	3	4.11	...

Точкові оцінки математичного очікування та середньоквадратичного відхилення

$xmean := \mathrm{mean}(X) = 2.775$ $S := \mathrm{Stdev}(X) = 1.034$

Сформулюємо нульову та першу альтернативну гіпотези:

H0:μ=μ0

H1:μ≠μ0

$\mu0 := 2.5$

$\alpha := 0.1$

$Xright := qt\left(1 - \frac{\alpha}{2}, N - 1\right)$

$Xright = 1.729$

$Xleft := -Xright = -1.729$

Границі критичної області

$\varphi := \dfrac{xmean - \mu0}{\sqrt{\dfrac{S^2}{N}}}$

$\varphi = 1.19$ Значення критерію

Гіпотеза H0 приймається

$xmean := mean(X) = 2.775$ $\quad$ $S := Stdev(X) = 1.034$

Сформулюємо нульову та другу альтернативну гіпотези:

H0:μ=μ0

H1:μ>μ0

$\mu 0 := 2.3$

$\alpha := 0.1$

$Xright := qt(1 - \alpha, N - 1)$ $\quad$ $Xright = 1.328$

$Xleft := -Xright = -1.328$

Границі критичної області

$$\varphi := \frac{xmean - \mu 0}{\sqrt{\frac{S^2}{N}}}$$

$\varphi = 2.055$ Значення критерію

$xmean := mean(X) = 2.775$ $\quad$ $S := Stdev(X) = 1.034$

Сформулюємо нульову та третю альтернативну гіпотези:

H0:μ=μ0

H1:μ<μ0

$\mu0 := 3.2$

$\alpha := 0.1$

$Xright := qt(1 - \alpha, N - 1)$ $\quad$ $Xright = 1.328$

$Xleft := -Xright = -1.328$

Границі критичної області

$$\varphi := \frac{xmean - \mu0}{\sqrt{\frac{S^2}{N}}}$$

$\varphi = -1.839$ Значення критерію