

Критерії достовірності оцінок



...або як встановити, наскільки відмінності між вибірками «правильно» відображують відмінності між генеральними сукупностями, з яких вони отримані

План:

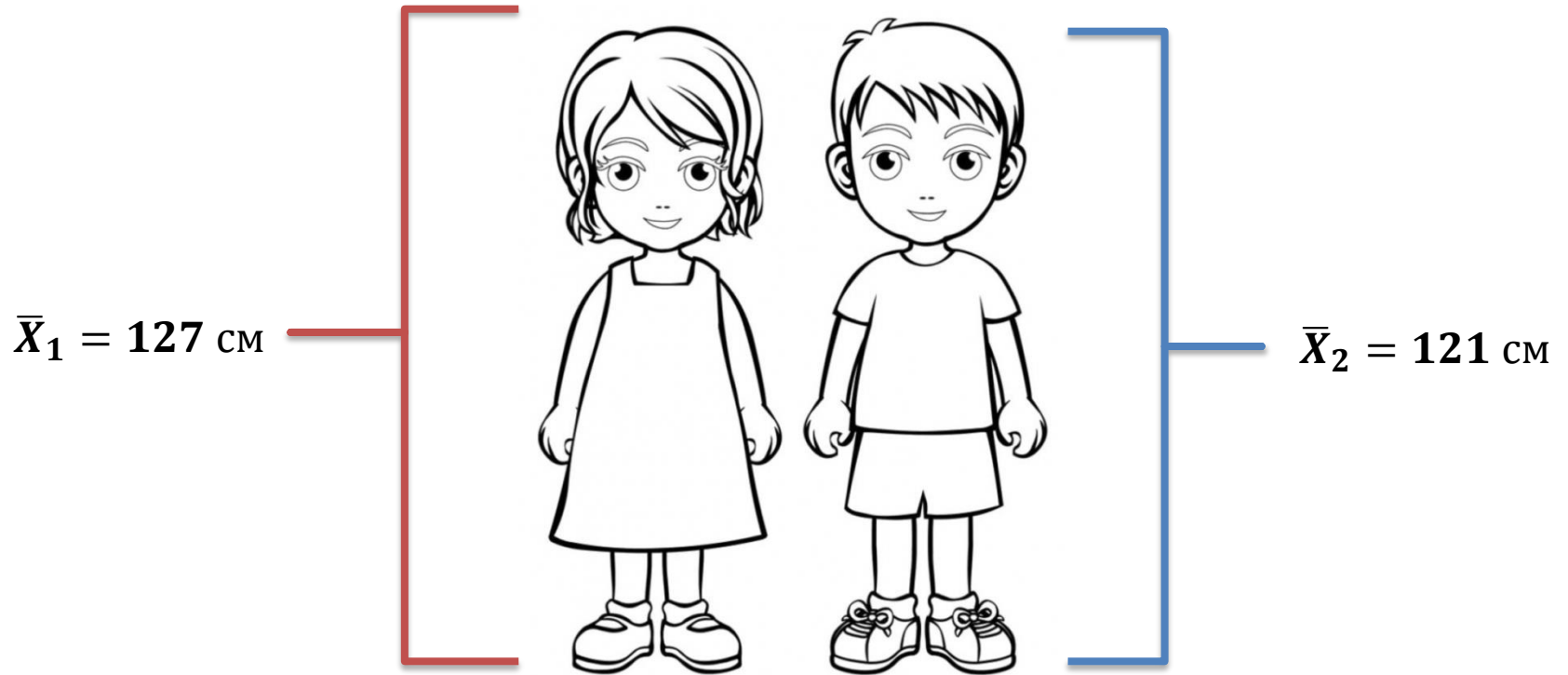
1. Загальна сутність порівняльних оцінок генеральних параметрів.
2. Достовірність вибіркової різниці. Нульова і альтернативна гіпотези. Поняття критерію достовірності.
3. Основні завдання, які виконують статистичні критерії.
4. Класифікація критеріїв достовірності. Параметричні та непараметричні критерії.
5. Переваги та недоліки параметричних та непараметричних критеріїв.
6. Схема вибору статистичного критерію залежно від визначеної мети статистичного аналізу.

Що порівнюють біологи?

- дані дослідної та контрольної груп у дослідженні;
- ступінь забруднення тих чи інших ділянок;
- популяції за чисельністю та структурою;
- продуктивність різних озер;
- морфологічні особливості різних людей, тварин, рослин;
- ...



Чи відрізняються хлопчики від дівчат за зростом?



Чи достовірні спостережувані відмінності між вибірками або вони обумовлені лише випадковими причинами?

Статистична достовірність

Достовірність – це властивість вибіркової різниці (відмінність середніх, дисперсій 2-х вибірок) правильно із заданою вірогідністю відобразити генеральну різницю (відмінність генеральних середніх і дисперсій).
Вибіркова різниця може бути **достовірною (статистично значимою)** або **недостовірною (випадковою, статистично незначимою)**.

Нульова (H_0) та альтернативна (H_a) гіпотези

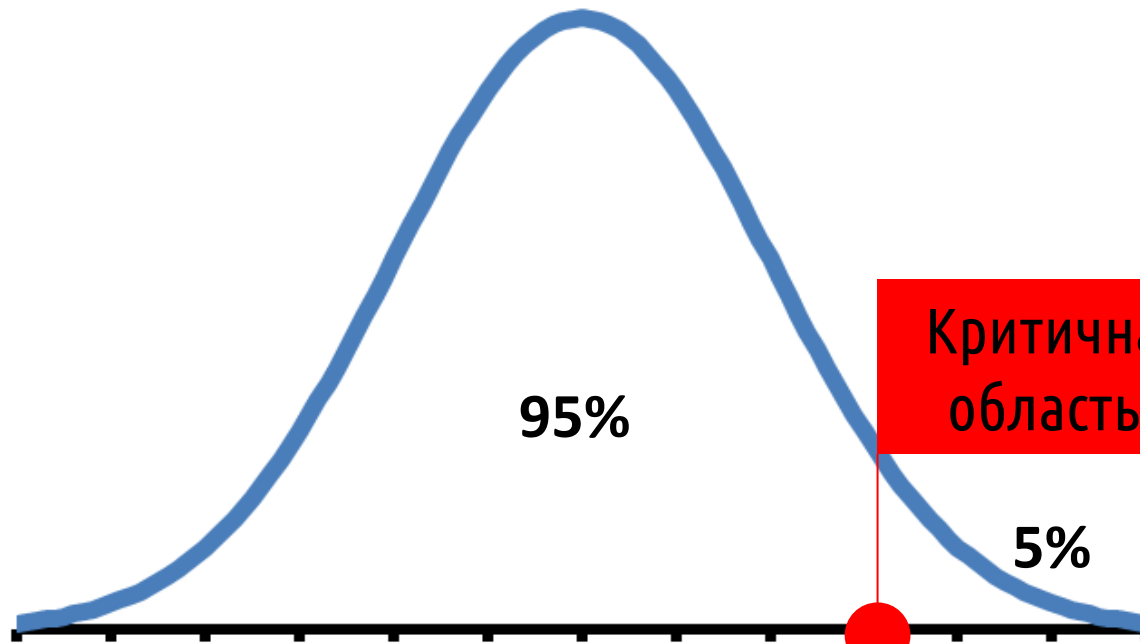
Нульова гіпотеза (H_0) – відмінності між вибірковими характеристиками випадкові, недостовірні.

Альтернативна гіпотеза (H_a) – відмінності між вибірковими характеристиками достовірні, тобто реально спостерігаються між генеральними параметрами у генеральних сукупностях, з яких отримані вибірки.

Загальне формулювання	Статистичне судження	Стандартний поріг тесту	
«Статистично значимий», «Достовірний», «Невипадковий»	Нульова гіпотеза відхилена	$p < 0,05$	$ \Phi Z \geq K_3$
«Статистично незначимий», «Недостовірний», «Випадковий»	Нульова гіпотеза не може бути відхилена	$p > 0,05$	$ \Phi Z < K_3$

Схема використання критеріїв достовірності

Область прийняття рішення



H_0 приймається

H_0 відхиляється

Критична точка

Існують два підходи до оцінки результатів статистичного тесту

1. Розрахунок фактичного значення (ΦZ) критерію та порівняння його з критичним значенням (KZ) на обраному рівні довірчої імовірності (α):

- ❑ Якщо $|\Phi Z| \geq KZ$, то H_0 відкидається, доводиться достовірність різниці.
- ❑ Якщо $|\Phi Z| < KZ$, то H_0 приймається, доводиться недостовірність різниці.

2. Розрахунок рівня статистичної значимості (p -рівня) та порівняння його з обраним рівнем довірчої імовірності (α):

- ❑ Якщо $p < 0.05$, то H_0 відкидається, доводиться достовірність різниці (при $\alpha = 0,05$).
- ❑ Якщо $p > 0.05$, то H_0 приймається, доводиться недостовірність різниці.

Приклади порівняння груп за середнім значенням у програмах статистичного аналізу

	A	B	C	D
1	Guminorm	PhytoMAX		
2		27	28	
3		20	19	
4		16	18	
5		18	21	
6		22	24	
7		19	20	
8		23	25	
9		21	27	
10		17	29	
11		19	21	
12	Двухвыборочный t-тест с различными дисперсиями			
13				
14		<i>Guminorm</i>	<i>PhytoMAX</i>	
15	Среднее	20,2	23,2	
16	Дисперсия	10,4	15,511111	
17	Наблюдения	10	10	
18	Гипотетическая разность средних	0		
19	df		17	
20	t-статистика	-1,8637096		
21	t критическое двухстороннее	2,10981558		
22				

Microsoft Excel 2010

```
Console Terminal x
~/
R version 3.4.4 (2018-03-15) -- "Someone to Lean On"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> Guminorm <- c(27, 20, 16, 18, 22, 19, 23, 21, 17, 19)
> PhytoMAX <- c(28, 19, 18, 21, 24, 20, 25, 27, 29, 21)
> t.test(Guminorm, PhytoMAX)

      Welch Two Sample t-test

data:  Guminorm and PhytoMAX
t = -1.8637, df = 17.326, p-value = 0.0794
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.391296  0.391296
sample estimates:
mean of x mean of y
 20.2      23.2

> |
```

R 3.4.4

Основні завдання, які виконують статистичні критерії

Використання критеріїв достовірності дозволяє вирішувати 3 основні завдання:

1. Здійснювати оцінку відповідності між емпіричним розподілом показника, що вивчається, і відомим теоретичним розподілом - **підгонка розподілу**.
2. Проводити порівняння двох вибірок і визначення достовірності відмінностей середніх величин і дисперсій цих вибірок.
3. Встановлювати приналежність окремого значення змінної, що вивчається, до вибірки (генеральній сукупності) - **вibraковування викидів** або значень, що виходять за межі вибірки.

Класифікація критеріїв достовірності

I. Залежно від типу порівняння :

1. Критерії, що визначають міру відповідності емпіричного розподілу відомому теоретичному, називаються **критеріями згоди**.
2. Критерії, що визначають достовірність відмінностей 2-х вибірок за певними характеристиками (середнє або дисперсія), називаються **критеріями відмінності**.
3. Критерії, що визначають приналежність викидів до вибірки (генеральної сукупності), називаються **критеріями виключення**.

Залежно від способу розрахунку критерію:

1. Параметричні критерії – розраховуються на основі параметрів вибіркової сукупності (на основі середнього значення, дисперсії, стандартної помилки і т. п.).

T-критерій Ст'юдента

T-критерій Ст'юдента (t-test) – заснований на вибірковому середньому значенні. Якщо виникає завдання порівняти дві вибірки за середніми значеннями, то фактичне значення критерію Ст'юдента розраховується у класичному варіанті як відношення різниці 2-х вибіркових середніх до стандартної помилки цієї різниці:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

При цьому абсолютно очевидно, що чим більше різниці між середніми значеннями 2-х вибірок і чим менше стандартна помилка цієї різниці, тим більша вірогідність того, що ці вибірки достовірно розрізняються між собою за середньою тенденцією.

Для кожного подібного фактичного значення критерію можна розрахувати фактичний p-рівень значущості, порівняти його з критичним рівнем (0.05) і визначити достовірність відмінностей.

Подібна схема ухвалення рішення зберігається для усіх інших критеріїв достовірності.

F-критерій Фішера

F-критерій Фішера (F-test) – заснований на вибірковій дисперсії. Фактичне значення F-критерію розраховується як відношення більшої вибіркової дисперсії до меншої:

$$F = \frac{S_1^2}{S_2^2}$$

2. Непараметричні критерії

Непараметричні критерії – розраховуються на основі частоти зустрічальності значень ознаки, або рангів.

□ Рангові критерії:

- Критерій Мана-Утні (U) (Mann-Whitney test)
- Критерій Вілкоксона (T) (Wilcoxon test)
- Критерій знаків (Z) (Sign test)
- Критерій серій Вальда-Вольфовица (S) (Wald-Wolfowitz test)

□ Частотні критерії:

- Критерій χ^2 («хі-квадрат»), або критерій згоди Пірсона, (Chi-Square test) - є сумою квадратів відхилень емпіричних частот (f) від обчислених теоретичних частот розподілу (f'), віднесену до теоретичних частот:

$$\chi^2 = \sum_{i=1}^n \frac{(f - f')^2}{f'}$$

Параметричні критерії

Перевага - велика «потужність» у порівнянні з непараметричними критеріями. Потужність критерію – його здатність більш безпомилково відкидати нульову гіпотезу, якщо вона невірна.

Недоліки:

1. Можуть застосовуватися тільки при нормальному розподілі порівнюваних змінних.
2. Більш чутливі до малого об'єму вибірки, ніж непараметричні критерії.
3. Використовуються тільки для кількісних змінних.

Непараметричні критерії

Переваги:

1. Можуть застосовуватися незалежно від закону розподілу порівнюваних змінних, тобто у випадках, коли розподіл змінної відрізняється від нормального.
2. Використовуються для кількісних, порядкових і якісних даних.
3. Добре працюють при малих об'ємах вибірки.

Недолік - дають грубішу оцінку відмінностей вибірок порівняно з параметричними критеріями.

Вибір критерію також залежить від типу вибірки

Критерії для незалежних вибірок:

- t-критерій Ст'юдента;
- критерій Мана-Уїтні;
- критерій серій Вальда-Вольфовіца.

Критерії для залежних вибірок:

- парний t-критерій Ст'юдента;
- критерій Вілкоксона;
- критерій знаків.

Завдання: ДОВЕСТИ ВІДМІННІСТЬ МІЖ ЕМПІРИЧНИМ ТА ТЕОРЕТИЧНИМ ЧАСТОТНИМИ РОЗПОДІЛАМИ

Критерій	Тип критерію	Умови і особливості застосування критерію
Критерій Пірсона χ^2-квадрат	Критерії згоди (нормальності)	<ol style="list-style-type: none"> 1. Об'єм вибірки має бути великим ($n > 50$). 2. Непридатний для малих вибірок ($n < 10 - 20$). 3. Необхідно, щоб частоти значень ознаки в крайніх класах були > 5. 4. Застосовується при перевірці як безперервних, так і дискретних законів розподілу.
Критерій Колмогорова-Смирнова (d) (Kolmogorov-Smirnov test)		Має більшу потужність для безперервних змінних.
Критерій Шапіро-Уїлка (W) (Shapiro-Wilk's test)		<ol style="list-style-type: none"> 1. Вважається найбільш потужним критерієм, якщо об'єм вибірки невеликої ($8 < n < 50$). 2. Потужніший при визначенні відмінностей в асиметрії розподілу, чим в ексцесі.

Завдання: ДОВЕСТИ ВІДМІННІСТЬ ДВОХ СЕРЕДНІХ АРИФМЕТИЧНИХ ДЛЯ ОДНІЄЇ ОЗНАКИ

Критерій	Тип критерію	Умови і особливості застосування критерію
t-критерій Ст'юдента	Параметричний критерій відмінності	<ol style="list-style-type: none">1. Нормальність розподілу порівнюваних змінних.2. Недостовірна відмінність дисперсій порівнюваних змінних. Якщо ця умова не дотримується, то в MS Excel і Statistica необхідно розраховувати t-критерій Ст'юдента для вибірок з нерівними дисперсіями.

Завдання: ДОВЕСТИ ВІДМІННІСТЬ ДВОХ ДИСПЕРСІЙ ДЛЯ ОДНІЄЇ ОЗНАКИ

Критерій	Тип критерію	Умови і особливості застосування критерію
F-критерій Фішера	Параметричний критерій відмінності	Нормальність розподілу порівнюваних змінних.

Завдання: ДОВЕСТИ ВІДМІННІСТЬ ДВОХ ВИБІРОК В ЦІЛОМУ

Критерій	Тип критерію	Умови і особливості застосування критерію
Критерій Мана-Уїтні	Непараметричні рангові критерії відмінності для незалежних вибірок	<ol style="list-style-type: none">1. Найбільш потужна непараметрична альтернатива t-критерію Стьюдента для незалежних вибірок.2. Найдоцільніше застосовувати для малих вибірок ($3 < n < 60$).3. У 2-х порівнюваних вибірках не повинно бути співпадаючих значень або таких збігів повинно бути дуже мало.
Критерій Вальда-Вольфовіца		<ol style="list-style-type: none">1. Може застосовуватись для малих вибірок.2. У досить великих вибірках вловлює відмінності практично будь-якого типу: у центральній тенденції, у дисперсії, у характеру розподілу. У вибірках малого обсягу реагує в основному на зрушення у центральній тенденції.

Завдання: ДОВЕСТИ ВІДМІННІСТЬ ДВОХ ВИБІРОК В ЦІЛОМУ

Критерій	Тип критерію	Умови і особливості застосування критерію
Критерій знаків	Непараметричні рангові критерії відмінності для залежних вибірок	<ol style="list-style-type: none">1. Має більшу чутливість до вибірок середнього і великого обсягу.2. Враховує лише спрямованість змін у кожній парі значень ознаки.3. Може бути застосований як до порядкових так і якісних даних.4. Якщо порівнюються кількісні дані, то застосування критерію знаків можливе тільки для попередньої оцінки відмінностей 2-х вибірок.
Критерій Вілкоксона		Є потужнішим критерієм порівняно з критерієм знаків, оскільки враховує не лише знак різниці між пов'язаними значеннями 2-х вибірок, але і величину цієї різниці.

Рекомендована література

1. Атраментова Л.О. Біометрія : підруч. для студ. вищ. навч. закл. / Л.О. Атраментова, О. М. Утєвська. – Х. : Ранок, 2007. – 176 с.
2. Акімова О.В. Статистика в малюнках та схемах : навчальний посібник / О.В. Акімова, О.С. Дубинська. – К.: Центр учбової літератури, 2007. – 168 с.
3. Ивантер Э.В., Коросов А.В. Элементарная биометрия / Э.В. Ивантер, А.В. Коросов. – Петрозаводськ : Изд-во ПетрГУ, 2010. – 104 с.

Дякую за увагу