

Проверка на хипотези
при две извадки

Лекция 8

Inferences Based on Two Samples:
Tests of Hypothesis

Сравнение на две средни
от независими извадки

Тест на хипотези за $\mu_1 - \mu_2$

Едностраниен тест	Двустраниен тест
$H_0: (\mu_1 - \mu_2) = D_0$	$H_0: (\mu_1 - \mu_2) = D_0$
$H_a: (\mu_1 - \mu_2) < D_0$ [или $H_a: (\mu_1 - \mu_2) > D_0$]	$H_a: (\mu_1 - \mu_2) \neq D_0$

Където D_0 = тестваната стойност на разликата между двете средни

Критерий:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Отхвърляме нулевата хипотеза ако $|z| > z_\alpha$

Параметри

Параметър	Съдържание	Вид на данните
$\mu_1 - \mu_2$	Разлика на средни	Количествени
$p_1 - p_2$	Разлика на относителни дялове	Качествени/ категорийни

Изисквания

1. Случайни, независими извадки
 2. Обема на всяка извадка трябва да е поне 30
- При тези условия ЦГТ е в сила.

Сравнение на средни
от две независими извадки

При големи извадки се използва Z-критерий.

При малки извадки се използва t-критерий.

Пример 1.

Данни: две независими извадки
Напр. две различни диети за отслабване:
X е намалението на теллото в кг.

$$n_1 = 100; \bar{X}_1 = 9; s_1 = 5$$

$$n_2 = 200; \bar{X}_2 = 7; s_2 = 4$$

Въпрос: Има ли разлика в резултатите от двете различни диети?

Стойности на Z

α	Z	α	Z
Едностранна		Двустранна	
0.05	1.645	0.10	1.645
0.025	1.96	0.05	1.96
0.01	2.33		
0.005	2.575	0.01	2.575

7

Пример 2 продължение

Тест на хипотези за $\mu_1 - \mu_2$

Едностраниен тест

$$H_0: (\mu_1 - \mu_2) = 5$$

$$H_a: (\mu_1 - \mu_2) > 5$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(9-1) - 5}{\sqrt{\frac{20^2}{100} + \frac{24^2}{200}}} = 1.14$$

$$\alpha = 0.05 \text{ Едностраниен тест, } Z = 1.645; \quad 1.14 < 1.645$$

Не можем да отхвърлим нулевата хипотеза.

10

Пример 1 продължение

Тест на хипотези за $\mu_1 - \mu_2$

Двустраниен тест

$$H_0: (\mu_1 - \mu_2) = 0$$

$$H_a: (\mu_1 - \mu_2) \neq 0$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(9-7) - 0}{\sqrt{\frac{5^2}{100} + \frac{4^2}{200}}} = 3.51$$

$$\alpha = 0.05 \text{ Двустраниен тест, } Z = 1.96; \quad 3.51 > 1.96$$

Отхвърляме нулевата хипотеза.

Разликата между двете диети е статистически значима.

8

Сравнение на две средни от зависими извадки (Paired)

	Извадка 1	Извадка 2	Разлика	
	Тест 1	Тест 2	d = Тест 2 - Тест 1	
1	Иван	72	77	5
2	Георги	68	74	6
3	Петър	76	83	7
4	Таня	68	73	5
5	Ваня	84	84	0
6	Валентина	68	87	9
7	Красимира	61	66	5
8	Гергана	76	70	-6

Въпрос: Дали студентите средно взето са подобрили успеха си от Тест 1 до Тест 2?

11

Пример 2.

Данни: две независими извадки

Напр. две различни диети за отслабване:

X е намалението на теглото в кг.

$$n_1 = 100; \quad \bar{X}_1 = 9; \quad s_1 = 20$$

$$n_2 = 200; \quad \bar{X}_2 = 1; \quad s_2 = 24$$

Въпрос: Дали разликата между двете диети е по-голяма от 5 килограма?

9

Решение:

Тъй като извадките не са независими е необходим нов подход. Създава се нова променлива d която е разликата между Тест 2 и Тест 1.

Новата променлива се тества като обикновен тест на хипотези за една извадка:

$$H_0: \mu_d = 0$$

$$H_a: \mu_d > 0$$

12

Терминология

- Този вид експеримент се нарича експеримент по двойки (**paired experiment**)
- Сравнението по двойки премахва разликите по време (дните между двата теста) и се фокусира на разликите между резултатите от двата теста
- Сравнението в рамките на групите се нарича блокиране (**blocking**)
- Това се нарича рандомизиран блок експеримент (**randomized block experiment**)

13

Пример 3. Разлики по двойки

	Извадка 1	Извадка 2	Разлика
Име	Тест 1	Тест 2	d=Тест 2 - Тест 1
1 Иван	72	77	5
2 Георги	68	74	6
3 Петър	76	83	7
4 Таня	68	73	5
5 Ваня	84	84	0
6 Валентина	68	87	9
7 Красимира	61	66	5
8 Гергана	76	70	-6

Въпрос: Дали студентите средно взето са подобрили успеха си от Тест 1 до Тест 2?

16

Тест на хипотези за разлики между средни на извадки по двойки

$\mu_d = \mu_1 - \mu_2$, големи извадки

Едностраниен тест

$H_0: (\mu_d) = D_0$
 $H_a: (\mu_d) < D_0$
 [или $H_a: (\mu_d) > D_0$]

Критерий

$$z = \frac{\bar{d} - D_0}{s_d / \sqrt{n_d}} \quad \text{същото както преди} \quad z = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Отхвърляме нулевата хипотеза ако $|z| > z_\alpha$

14

Пример 3 продължение

Голяма извадка (за примера не е вярно)

Изчисляваме за d: $\bar{d} = 3.9$; $s = 4.7$

Едностраниен тест

$H_0: (\mu_d) = 0$

$H_a: (\mu_d) > 0$

Критерий

$$Z = \frac{\bar{d} - D_0}{s_d / \sqrt{n_d}} = \frac{3.9 - 0}{4.7 / \sqrt{8}} = 2.35$$

$\alpha = 0.05$ едностранен $Z = 1.645$; $2.35 > 1.645$

Отхвърляме нулевата хипотеза.

17

Изисквания

1. Чиста случайна извадка
2. Обема на извадката е поне 30

15

Сравнение на два относителни дяла при независими извадки

И двете извадки са големи

Едностраниен тест

$H_0: (p_1 - p_2) = D_0$
 $H_a: (p_1 - p_2) < D_0$
 [или $H_a: (p_1 - p_2) > D_0$]

Критерий

$$z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

Отхвърляме нулевата хипотеза ако $|z| > z_\alpha$

18

Пример 4.

Данни: две независими извадки, една от студенти в НБУ и една от УНСС.

Въпрос “Харесвате ли Lady Gaga?”

$$n_1 = 100; \hat{p}_1 = 0.7$$

$$n_2 = 200; \hat{p}_2 = 0.6$$

Въпрос: Има ли разлика в резултатите от двата университета?

19

Пример 4 продължение

Големи извадки: $100 \cdot 0.7 > 5$ и $100 \cdot 0.3 > 5$; също $200 \cdot 0.6 > 5$ и $200 \cdot 0.4 > 5$

Двустранен тест

$H_0: (p_1 - p_2) = 0$

$H_a: (p_1 - p_2) \neq 0$

Критерий

$$z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = \frac{0.7 - 0.6 - 0}{\sqrt{\frac{0.7 \cdot 0.3}{100} + \frac{0.6 \cdot 0.3}{200}}} = 1.74$$

$\alpha = 0.05$ двустранен $Z = 1.96$; $1.74 < 1.96$

Не можем да отхвърлим нулевата хипотеза.

Няма разлика в отговорите на студентите от двата университета.

20