

Двухэтапная проверка равенства средних с помощью критериев Стюдента и Уэлша

Прикладная статистика
(практика, Е.Рябенко)

К.О. Неклюдов

МФТИ(ГУ)

24 февраля 2014 г.

Исследуются две независимые нормально распределённые выборки с неизвестными дисперсиями и мат. ожиданиями:

$$X^{n1} = x_{11}, \dots, x_{1n_1}; x_{1i} \sim N(\mu_1, \sigma_1^2)$$

$$X^{n2} = x_{21}, \dots, x_{2n_2}; x_{2i} \sim N(\mu_2, \sigma_2^2)$$

$$H_0 : \mu_1 = \mu_2$$

t-критерий Стьюдента (используется при равенстве дисперсий):

$$T = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Критерий Уэлша (используется при различных дисперсиях):

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (1)$$

Для критерия Стьюдента справедливо: когда большей выборке соответствует большая дисперсия, ошибка первого рода снижается относительно заданного уровня значимости, и наоборот.

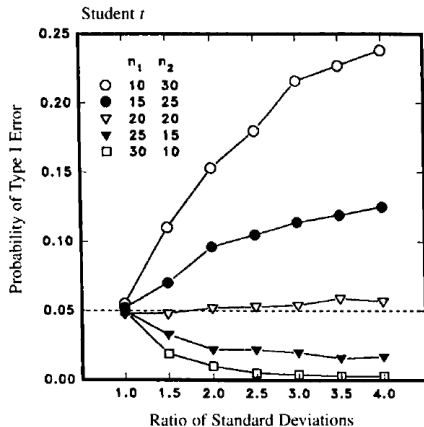


FIGURE 1. Probability of a Type I error of the Student t test as a function of the ratio of population standard deviations (σ_1/σ_2) for various sample sizes.

Для критерия Уэлша:

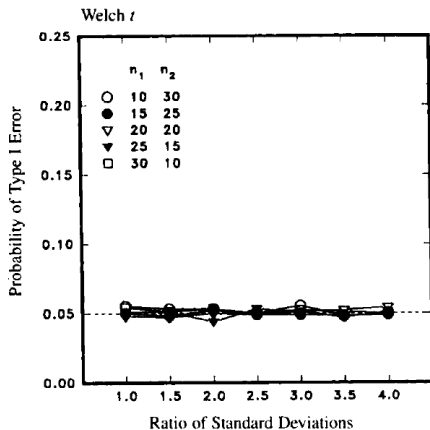
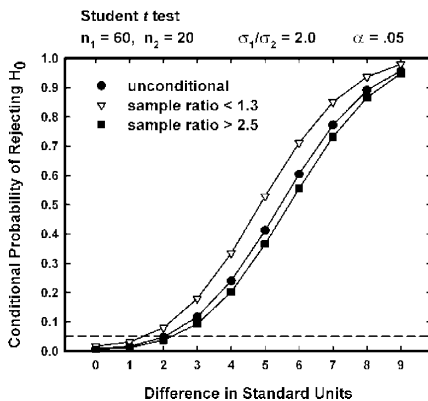
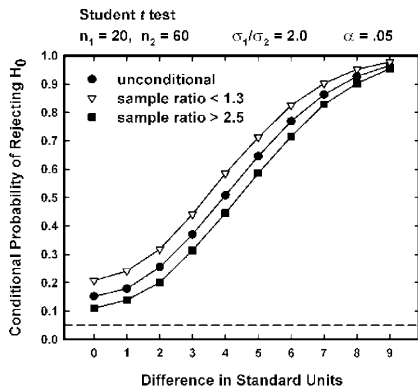
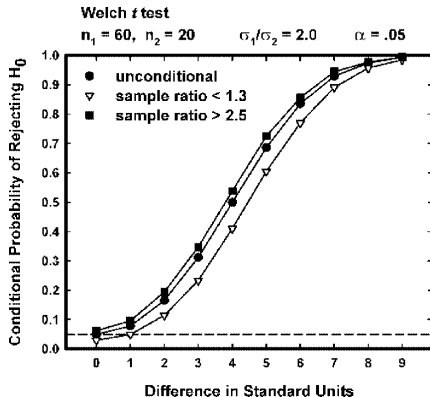
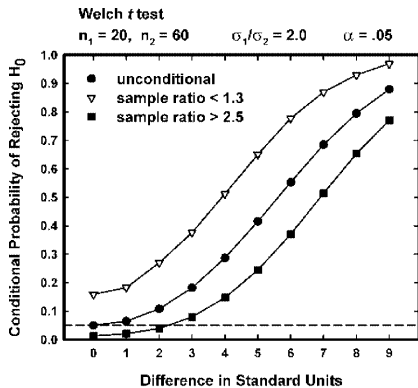


FIGURE 3. Probability of a Type I error of the Welch separate-variances t test as a function of the ratio of population standard deviations (σ_1/σ_2) for various sample sizes.

Вычислительный эксперимент





Пример

X_1	X_2	X_1	X_2
25	5	20	14
31	12	22	15
23	21	20	29
34	12	20	13
37	16	24	14
16	13	21	10
	12	17	4
	7	22	24
	3	17	15
	7		
	27		
	8		

$\mu_1 - \mu_2 = 1.000$	$\mu_1 - \mu_2 = 1.000$
$\sigma_1/\sigma_2 = 2.000$	$\sigma_1/\sigma_2 = 2.000$
$s_1 = 7.789$	$s_1 = 2.291$
$s_2 = 6.868$	$s_2 = 7.314$
$s_1/s_2 = 1.134$	$s_1/s_2 = 3.192$
$\bar{X}_1 - \bar{X}_2 = 15.750$	$\bar{X}_1 - \bar{X}_2 = 5.000$
t -Student = 2.022	t -Student = 2.182
Critical value = 2.120	Critical value = 2.120
t -Welch = 2.293	t -Welch = 0.684
Critical value = 2.262	Critical value = 2.228

Знаменатель в критерии Стьюдента

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \left(\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\right)} + \epsilon_1$$

Знаменатель в критерии Уэлша

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} + \epsilon_2$$

Отношение дисперсий влияет на значение знаменателей, а соответственно на значение статистики.

