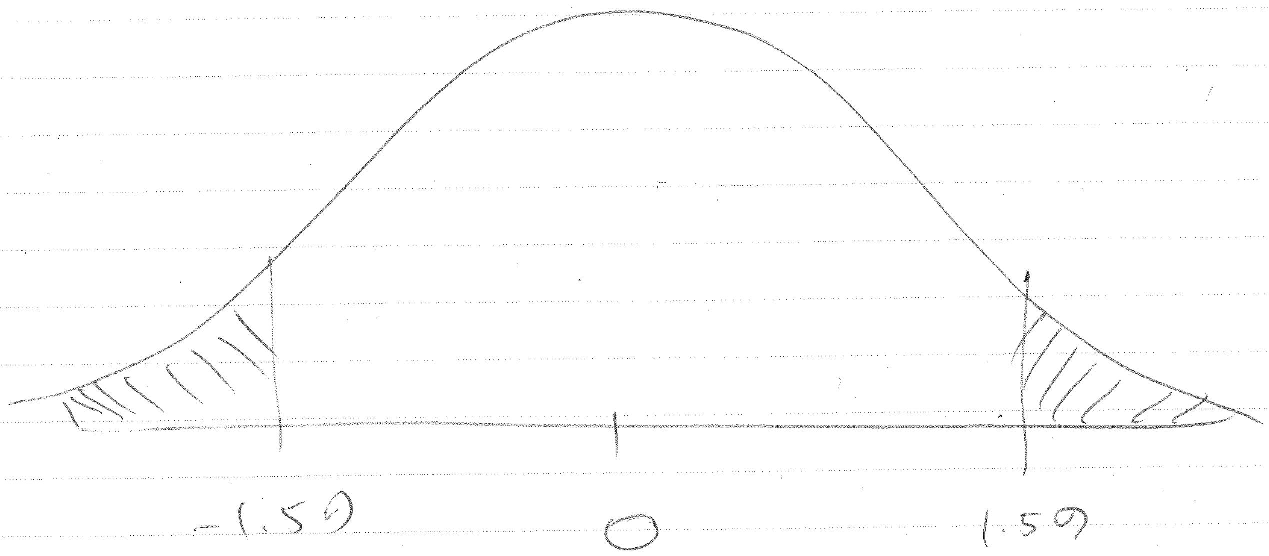

-1.59
0
1.59

In practice, the common variance σ^2 is not always known, but even then, the problem can be solved using a variant of the t-test, as follows. The common variance is estimated by the pooled estimate

$$s_p{}^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} \tag{3}$$

If σ is replaced by S_p (the random variable corresponding to s_p) in (2), then the resulting test statistic is

$$\tau = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \tag{4}$$

The test is based on the observed value of τ, i.e. $\tau_{\text{obs}} = (\bar{x} - \bar{y}) / \left(s_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}\right)$. The denominator in (4) is not a constant but a random variable, so that the statistic τ is not normally distributed. In fact, τ can be shown to have the t_ν distribution with $\nu = n_x + n_y - 2$ degrees of freedom.

The size of the P-value can now be gauged from the $t_{n_x+n_y-2}$ tables. As in the one-sample case, if the alternative hypothesis is $H_1 : \mu_x \neq \mu_y$, then a large value of $|\tau_{\text{obs}}|$ provides evidence against H_0, whereas for instance for $H_1 : \mu_x < \mu_y$, a large *negative* value of τ_{obs} provides this evidence.

Example

A cross-sectional study was designed to assess the effect of the use of oral contraceptives (OC's) on blood pressure in women. Prior to collecting the data, a pilot study (a small-scale version of the study) was analysed using small samples.

For the pilot study, independent random samples of 10 OC users and 14 non-users were selected from women who were aged 35-39, pre-menopausal and non-pregnant. The results (systolic blood pressure in mm Hg) were as follows.

Users	x	115.7	125.9	122.9	125.2	139.3	141.4	141.3	123.0	135.7	124.0
Non-users	y	111.8	120.1	117.7	119.5	131.0	132.8	132.7	117.7	128.1	118.6
		138.7	94.7	127.2	126.5						

Is there strong evidence from these samples of a difference in blood pressure for the two groups of women?

Solution

From the data,

$$\bar{x} = 129.44,\ \bar{y} = 122.65,\ s_x = 9.14843,\ s_y = 11.01515 \text{ and, using (3), } s_p = 10.2925$$

Thus the observed value of the test statistic is $\tau_{\text{obs}} = \frac{129.44-122.65}{10.2925\sqrt{\frac{1}{10}+\frac{1}{14}}} \approx 1.59$. From the tables for t_{22}, it follows that the P-value is greater than 0.1. Therefore the data do not indicate any difference in blood pressure due to OC's. □

- *Aside* To test for the equality of the means of more than two groups, a different (but related) method is required. See §5.