Comparison of Means: Paired Data

A computer scientist is investigating the usefulness of two design languages in improving programming tasks. Twelve expert programmers are asked to code a standard function in each language. The scientist wants to compare the mean programming times in the two design languages. To do so, he computes D, the difference between the programmer's programming time using language 1 and that using language 2. Assuming that these differences follow a normal distribution, he constructs an hypothesis test for the mean difference in programming time, μ_D. The statistical hypotheses are $H_0 : \mu_D = 0$ and $H_a : \mu_D \neq 0$. The data (found in SASDATA.PROGRAM_TIMES) are:

	LANGUAGE		
PROGRAMMER	1 (y_{1i})	2 (y_{2i})	DIFF (d_i)
1	17	18	−1
2	16	14	2
3	21	19	2
4	14	11	3
5	18	23	−5
6	24	21	3
7	16	10	6
8	14	13	1
9	21	19	2
10	23	24	−1
11	13	15	−2
12	18	20	−2

For these data, $\overline{d} = 0.6667$, and $s_d = 2.9644$, so the standardized test statistic is

$$t^* = \frac{0.6667}{2.9644/\sqrt{12}} = 0.7790,$$

so

$$p_- = Pr(t_{11} \leq 0.7790) = 0.7738,\ p^+ = Pr(t_{11} \geq 0.7790) = 0.2262,$$

and the desired p-value is

$$p\pm = 2\min(0.7738, 0.2262) = 0.4524.$$

Based on this, one would not reject H_0 in favor of H_a at any reasonable level of significance (such as 0.05 or even 0.10), and thus he concludes that there is little evidence of a difference in mean programming time.