THE SIGNIFICANCE TEST FOR \( r \) IS DERIVED FROM THE TEST STATISTIC \( t \)

The textbook Box 16.4 showed that testing a hypothesis about \( r \) is another application of the Equation (9.1) general formula for the test statistic: test statistic = (sample statistic - population parameter)/(standard error of the sample statistic).

The population parameter (specified by the null hypothesis) is \( \rho \), the population correlation coefficient; the sample statistic that point-estimates \( \rho \) is \( r \); and the denominator of the test statistic must therefore be \( s_r \), the standard error of the correlation coefficient.

The **standard error of \( r \)** can be defined analogously to the standard error of the mean \( s_\bar{X} \). Suppose we take a sample of size \( n \), measure each subject on variables \( X \) and \( Y \), and compute \( r \) for those data. Then we take another sample of size \( n \) and again compute \( r \), and another sample (and another \( r \)), and so on, indefinitely often. The standard deviation of the resulting distribution of all the \( r \) values is \( s_r \), the standard error of \( r \). If the null hypothesis is true, then it can be shown that

\[
s_r = \sqrt{\frac{1 - r^2}{n - 2}}
\]  

(16.7)

For our height and weight data, \( s_r = \sqrt{(1 - .634^2)/(10 - 2)} = \sqrt{.075} = .274 \).

The test statistic can be shown to be \( t \), so following the Equation (9.1) model, \( t = (r - \rho)/s_r \), but because \( \rho = 0 \) by the null hypothesis, the test statistic becomes simply, as we saw,

\[
t = \frac{r}{s_r} \quad [df = n - 2]
\]  

(16.4)

We can sketch the distribution of the sample statistic \( r \) and the test statistic \( t \) as shown in Figure 1. The distribution on the upper axis shows the values we might expect \( r \) to take if we computed the correlation coefficients of a long series of independent samples of size ten drawn from a population in which the null hypothesis \( \rho = 0 \) is true. The lower axis shows \( t \) as defined by Equation (16.4).

Choosing a .05 level of significance, we find from the textbook’s Table A.2 that the critical value of \( t \) with \( df = 10 - 2 = 8 \) is \( t_{.05; 8} \) (two-tailed) = 2.306. The critical value of \( r \) is then \( r_{cv} = \rho \pm t_{cv}(s_r) = 0 \pm 2.306(.274) = \pm .632 \). We show these values on the appropriate axes in Figure 1 and shade the rejection regions that lie beyond them.
We show the observed value of the statistic $r_{\text{obs}} = .634$ on the upper axis of Figure 1, and we compute $t_{\text{obs}} = r_{\text{obs}} / s_r = .634 / .274 = 2.314$ and show it on the lower axis. Because the observed values exceed the critical values (just barely!), we reject $H_0$, just as when using Table A.8.

Note that if you try this process with a different correlation coefficient, you may arrive at a different critical value $r_{cv}$. For example, if $r_{\text{obs}}$ had been .4, $s_r$ would have been $\sqrt{(1 - .4^2) / (10 - 2)} = .324$, so the critical values would have been $\pm 2.306(.324) = \pm .747$ instead of $\pm .632$. The larger the $r_{\text{obs}}$, the smaller the $s_r$. As $r_{\text{obs}}$ moves closer and closer to $r_{cv}$, $r_{cv}$ moves closer and closer to $r_{\text{obs}}$, with the result that the hypothesis test based on $t$ will reject $H_0$ in exactly those instances where the test based on Table A.8 rejects $H_0$.

**EXERCISE**

(Answers below)

1. Use the $t$ test procedure described here to determine the smallest positive correlation coefficient that would be significantly different from 0 (non-directional test) in these cases. [Hint: Use an iterative method: Make a guess of $r$ and see whether it is significant; use the results of the first guess to generate a second (better) guess of $r$; and continue until you have a guess of $r$ that is just barely significant (accurate to within .01).]

(a) $n = 10$ and $\alpha = .05$
(b) $n = 30$ and $\alpha = .05$
(c) $n = 120$ and $\alpha = .05$
(d) $n = 30$ and $\alpha = .01$
(e) Compare your answers with the critical values of Pearson’s $r$ in Table A.8. Could Table A.8 have been generated using the procedure of this problem?
Answers

1. (a) $df = 8; t_{cv} = 2.306; r = .64$
   (b) $df = 28; t_{cv} = 2.048; r = .37$
   (c) $df = 118; t_{cv} = 1.98; r = .18$
   (d) $df = 28; t_{cv} = 2.763; r = .47$
   (e) Yes