Theorem 1.7 If $X_1, X_2, \ldots, X_n$ is a random sample from a $N\left(\mu_X, \sigma^2\right)$ population and $Y_1, Y_2, \ldots, Y_m$ is a random sample (independent of the first sample) from a $N\left(\mu_Y, \sigma^2\right)$ population, then

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2),$$

where

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

is the *pooled sample variance*.

Proof Using algebra to manipulate the fraction into a familiar form,

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\frac{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}} = \frac{\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\frac{\frac{(n-1)S_X^2}{\sigma^2}+\frac{(m-1)S_Y^2}{\sigma^2}}{n+m-2}}} = \frac{Z}{\sqrt{\chi^2(n+m-2)/(n+m-2)}}.$$

The numerator has the standard normal distribution by Theorem 1.6 in the special case when the population variances are equal, which is assumed for this result. Theorem 1.4 is used to show that

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2(n-1) \qquad \text{and} \qquad \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2(m-1)$$

are independent chi-square random variables. Since the sum of independent chi-square random variables has the chi-square distribution using a result from probability theory,

$$\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2(n+m-2).$$

So, the denominator in the expression is the square root of a chi-square random variable with $n+m-2$ degrees of freedom divided by its degrees of freedom. Finally, the ratio of a standard normal random variable divided by the square root of an independent chi-square random variable divided by its degrees of freedom has the t distribution with $n+m-2$ degrees of freedom using a result from probability theory. □

The primary purpose of Theorem 1.7 is to determine whether the difference between two sample means, that is $\bar{x} - \bar{y}$, is "statistically significant," as illustrated in the next example.

Example 1.36 One of the purposes of measuring the maximum breadth of the $n = 84$ Etruscan skulls from Example 1.34 was to determine whether the Etruscans were native Italians or immigrants. To that end, an additional $m = 70$ modern Italian skulls were measured, as shown in Table 1.11. Were the Etruscans native Italians or immigrants?

The first step in the analysis of these two data sets is to construct an appropriate statistical graphic. One such graphic is an adaptation of the population pyramid from Example 1.8. For that population pyramid, the population for each age was displayed for men and women. In this particular case, we want to adjust the pyramid so that the

133	138	130	138	134	127	128	138	136	131	126	120
124	132	132	125	139	127	133	136	121	131	125	130
129	125	136	131	132	127	129	132	116	134	125	128
139	132	130	132	128	139	135	133	128	130	130	143
144	137	140	136	135	126	139	131	133	138	133	137
140	130	137	134	130	148	135	138	135	138		

Table 1.11: Maximum skull breadths for $m = 70$ modern Italian adult males.

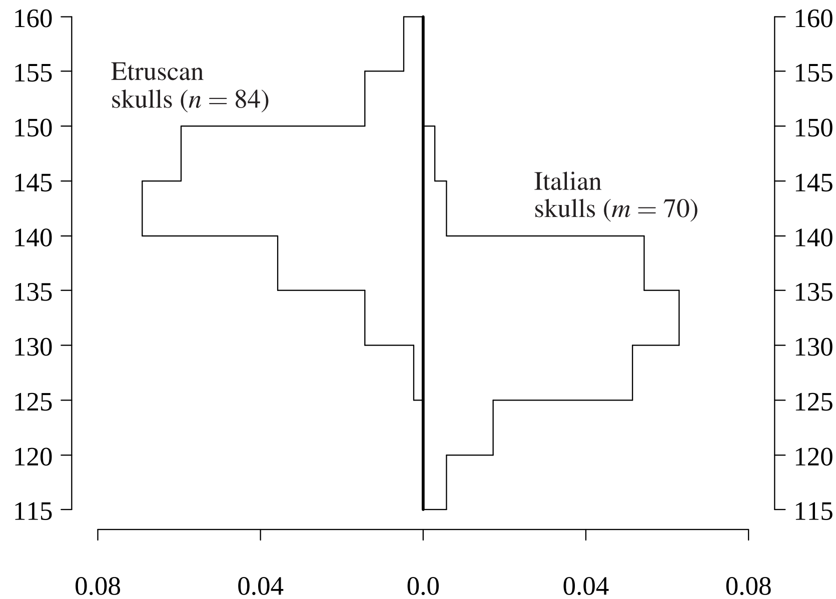

Figure 1.46: Normalized histograms of maximum skull breadth, in millimeters.

effect of the two sample sizes ($n = 84$ and $m = 70$) is not apparent. The pyramid in Figure 1.46 plots a histogram of the maximum skull breadth for the Etruscan skulls on the left and a histogram of the modern Italian skulls on the right. In this case, the line that is plotted for the histogram bins is the fraction of the observations falling into the bins. This way the area under each of the two histograms is equal. The grid lines that were included in the French population pyramid have been suppressed in order to highlight the shape of the two histograms.

Two conclusions can be drawn from Figure 1.46. First, the Etruscan skulls are larger than the modern Italian skulls on average. This is verified by computing the sample means:

$$\bar{x} = 144 \qquad \text{and} \qquad \bar{y} = 132.$$

Second, the variances of the two data sets appear to be roughly equal. But can these two conclusions which have been drawn from the statistical graphic be due to sampling variability, or is there a pattern here? Is the difference of 12 millimeters between the sample means of the two data sets a matter of sampling variability, or is there a "statistically significant" difference between the two populations means? This is where Theorem 1.7 comes into play.

Assume for the moment that our visual inspection of the two histograms leads us to conclude that the two populations have equal variances, that is $\sigma_X^2 = \sigma_Y^2$, where X corresponds to the Etruscan skulls and Y corresponds to the modern Italian skulls. If that is the case, and we also assume that the data are random samples drawn from two independent normal populations (which is also plausible given the shape of the two histograms), then all of the assumptions are in place to invoke Theorem 1.7. One device that statisticians use is to begin by assuming that there is no difference between the population means of the two normal populations, that is $\mu_X = \mu_Y$. The pooled variance is

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} = \frac{(84-1)35.6470 + (70-1)33.0619}{84+70-2} = 34.4735,$$

which is computed with the R statements

```
x = scan("etruscan.d")
y = scan("italian.d")
n = length(x)
m = length(y)
p = ((n - 1) * var(x) + (m - 1) * var(y)) / (n + m - 2)
```

This allows us to compute the value of the test statistic defined in Theorem 1.7, which is

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{143.7738 - 132.4429}{\sqrt{34.4735\left(\frac{1}{84} + \frac{1}{70}\right)}} = 11.9,$$

which is computed with the additional R statement

```
(mean(x) - mean(y)) / sqrt(p * (1 / n + 1/ m))
```

This statistic can be thought of as one value that is drawn from a t distribution with $n+m-2 = 84+70-2 = 152$ degrees of freedom. So the question that is now being asked is whether this observation of a t random variable with 152 degrees of freedom is unusual. The t distribution with 152 degrees of freedom is nearly identical to a standard normal distribution because of the large number of degrees of freedom. So it is clear that the value of 11.9 is way out in the far right-hand tail of the distribution. In fact, it is so far out in the right-hand tail that we have strong evidence that our assumed hypothesis of equal means is false. There is overwhelming statistical evidence to support the hypothesis that the Etruscans immigrated to modern-day Italy. This does assume that skull size is not changing over time, which must be verified separately.

Calculating the t statistic arises so often in statistics that it is built into an R function named `t.test`. An additional flag must be added in order to tell R that the variances between the two populations are assumed to be equal. The R code given below calculates the statistic 11.9 automatically.

```
x = scan("etruscan.d")
y = scan("italian.d")
t.test(x, y, var.equal = TRUE)
```