```
> t.test(log.wages.low, log.wages.high, paired=TRUE)
        Paired t-test

data:  log.wages.low and log.wages.high
t = -4.5516, df = 74, p-value = 2.047e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3930587 -0.1537032
sample estimates:
mean of the differences
             -0.2733810
```

Here the observed difference in means is clearly statistically significant and a 95% confidence interval for the difference is $(-0.393, -0.154)$. In an exercise, one will be asked to rerun the `t.test` function on this dataset without using the `paired = TRUE` option and comment on the difference in the confidence intervals for the difference in means using the two options.

Exercises

6.1 (Gender of marathoners). In 2000, the proportion of females who competed in marathons in the United States was 0.375. One wonders if the proportion of female marathoners has changed in the ten-year period from 2000 to 2010. One collects the genders of 276 people who competed in the 2010 New York City Marathon – in this sample, 120 were women.

a. If p denotes the proportion of 2010 marathoners who are female, use the `prop.test` function to test the hypothesis that $p = 0.375$. Store the calculations of the test in the variable `Test`.
b. From the components of `Test`, construct a 95% interval estimate for p.
c. Using the function `binom.test`, construct an exact-test of the hypothesis. Compare this test with the large-sample test used in part (a).

6.2 (Ages of marathoners). The datafile "nyc.marathon.txt" contains the gender, age, and completion time (in minutes) for 276 people who completed the 2010 New York City Marathon. It was reported that the mean ages of men and women marathoners in 2005 were respectively 40.5 and 36.1.

a. Create a new dataframe "women.marathon" that contains the ages and completion times for the women marathoners.
b. Use the `t.test` function to construct a test of the hypothesis that the mean age of women marathoners is equal to 36.1.
c. As an alternative method, use the `wilcox.test` function to test the hypothesis that the median age of women marathoners is equal to 36.1. Compare this test with the t-test used in part (b).
d. Construct a 90% interval estimate for the mean age of women marathoners.

6.3 (Ages of marathoners, continued). From the information in the 2005 report, one may believe that men marathoners tend to be older than women marathons.

a. Use the `t.test` function to construct a test of the hypothesis that the mean ages of women and men marathoners are equal against the alternative hypothesis that the mean age of men is larger.
b. Construct a 90% interval estimate for the difference in mean ages of men and women marathoners.
c. Use the alternative Mann-Whitney-Wilcoxon test (function `wilcox.test`) to test the hypothesis that the ages of the men and ages of the women come from populations with the same location parameter against the alternative that the population of ages of the men have a larger location parameter. Compare the result of this test with the t-test performed in part (a).

6.4 (Measuring the length of a string). An experiment was performed in an introductory statistics class to illustrate the concept of measurement bias. The instructor held up a string in front of the class and each student guessed at the string's length. The following are the measurements from the 24 students (in inches).

```
22 18 27 23 24 15 26 22 24 25 24 18
18 26 20 24 27 16 30 22 17 18 22 26
```

a. Use the `scan` function to enter these measurements into R.
b. The true length of the string was 26 inches. Assuming that this sample of measurements represents a random sample from a population of student measurements, use the `t.test` function to test the hypothesis that the mean measurement μ is different from 26 inches.
c. Use the `t.test` function to find a 90% confidence interval for the population mean μ.
d. The t-test procedure assumes the sample is from a population that is normally distributed. Construct a normal probability plot of the measurements and decide if the assumption of normality is reasonable.

6.5 (Comparing snowfall of Buffalo and Cleveland). The datafile "buffalo.cleveland.snowfall.txt" contains the total snowfall in inches for the cities Buffalo and Cleveland for the seasons 1968-69 through 2008-09.

a. Compute the differences between the Buffalo snowfall and the Cleveland snowfall for all seasons.
b. Using the `t.test` function with the difference data, test the hypothesis that Buffalo and Cleveland get, on average, the same total snowfall in a season.
c. Use the `t.test` function to construct a 95% confidence interval of the mean difference in seasonal snowfall.

6.6 (Comparing Etruscan and modern Italian skulls). Researchers were interested if ancient Etruscans were native to Italy. The dataset "Etruscan-Italian.txt" contains the skull measurements from a group of Etruscans and modern Italians. There are two relevant variables in the dataset: `x` is the skull measurement and `group` is the type of skull.

a. Assuming that the data represent independent samples from normal distributions, use the `t.test` function to test the hypothesis that the mean Etruscan skull measurement μ_E is equal to the mean Italian skull measurement μ_I.
b. Use the `t.test` function to construct a 95% interval estimate for the difference in means $\mu_E - \mu_I$.
c. Use the two-sample Wilcoxon procedure implemented in the function `wilcox.test` to find an alternative 95% interval estimate for the difference $\mu_E - \mu_I$.

6.7 (President's heights). In Example 1.2, the height of the election winner and loser were collected for the U.S. Presidential elections of 1948 through 2008. Suppose you are interested in testing the hypothesis that the mean height of the election winner is equal to the mean height of the election loser. Assuming that this data represent paired data from a hypothetical population of elections, use the `t.test` function to test this hypothesis. Interpret the results of this test.