Chapter 10
t-Test

Sampling distributions are frequency distributions of a statistic, which are generated to make an inference between the sample statistic and its corresponding population parameter. The average (mean) statistical value in the sampling distribution is the expected value of the population parameter. The variance of the sampling distribution is used to calculate the standard error of the statistic. The standard error of the statistic is a function of sample size and is used to calculate confidence intervals around the sample statistic, which provides the basis for testing statistical hypotheses and determining the interval in which the population parameter falls.

The distribution theory that led to a solution of the problem of estimating the population mean, μ, when the population variance was unknown, σ^2, was due to William S. Gossett, a chemist, who in 1908 wrote under the pseudonym of "Student." William S. Gossett, who worked for a brewery, in Dublin, Ireland determined that when the sample size was large, the sampling distribution of the z-statistic was normal; however, when the sample size was small, the sampling distribution was leptokurtic or peaked. He referred to this slightly non-normal distribution as the t-distribution. W. S. Gossett further discovered that as he increased the sample sizes, the sampling distribution became normal, and therefore the t-values equaled the z-values. W. S. Gossett signed his pivotal work, "Student," and today small sample tests of mean differences are referred to as the "student t-test," or simply the t-test. Sir Ronald Fisher, using the early work of "Student," extended his ideas into modern day analysis of variance techniques, which is discussed in the next chapter.

One Sample t-Test

The sampling distribution of the mean can be used to determine the expected value of the population mean and variance when they are unknown. The Central Limit Theorem supports this assumption because as sample size increases, the sampling distribution of the mean becomes a normal distribution with the mean $=\mu$, and

variance $= \sigma^2/N$. Knowledge of this permits one to test whether the sample mean is statistically different from a population mean. The test is called a **one-sample t-test**, which is computed by:

$$t = \frac{\overline{X} - \mu}{\sqrt{\dfrac{S^2}{N}}} = \frac{\overline{X} - \mu}{S/\sqrt{N}}$$

Suppose a class of 25 students was given instruction using a new, enriched mathematics curriculum and then took a nationally standardized math achievement test. If the new, enriched mathematics curriculum leads to better learning in the population, then this sample of 25 students should perform better on the math achievement test than students in the larger population who learned mathematics under the traditional mathematics curriculum. Assume the 25 students had a mean math achievement test score of 110 with a standard deviation of 15. The mean math achievement test score for students in the population taking the traditional math curriculum was 100. What is the probability that a random sample of this size would have a sample mean of 110 or higher? This research question can be tested using a one-sample t-test. The one-sample t-test is computed as:

$$t = \frac{110 - 100}{\sqrt{\dfrac{15^2}{25}}} = \frac{10}{\sqrt{\dfrac{225}{25}}} = \frac{10}{3} = 3.33$$

If we were to examine the probability area in the tails of the normal distribution for a z-value of 3.33, we would find a probability area equal to 0.0004 ($0.4996 + 0.0004 = 0.5000$). However, since the probability area in the tails of the normal distribution are not the same as the t-distribution, given small sample sizes, we need to examine the probability area in the tails of the t-distribution *not* the normal distribution. We therefore use the t-table in the appendix (Table A1), where we find the degrees of freedom, *df*, (df = n−1 = 24) and a probability area depending upon a directional (one-tail) or non-directional (two-tailed) test. For *df* = 24 and *p* = 0.0005 (one-tail test), we find a tabled t-value equal to 3.745. Since the computed *t* = 3.33 is less than the tabled *t* = 3.745, we retain the null hypothesis. We would conclude that there is no difference in the mean math achievement test scores at the *p* = 0.0005 level of statistical significance. Because the t-distribution is leptokurtic or peaked for small samples, the probabilities in the tails of the t-distribution are greater than the probabilities in the normal distribution. When W. S. Gossett discovered this, he made an adjustment for small sample sizes by entering larger tabled t-values from the sampling distribution of the t-values. By adjusting the t-values, he compensated for the greater probability area under the sampling distributions. If you examine the last row in the t-table in the appendix, you will find t-values that are equal to the z-values under the normal distribution because when sample sizes are large $t = z$. In practice, we use the t-test rather than the z-test for testing means because it applies to both small and large samples.

Independent t-Test

The sampling distribution of the difference between the means of two independent samples provides the basis for the testing of a mean difference hypothesis between two groups. Taking a random sample from one population, a second random sample from another population, and then computing the difference in the means, provides the basis for the sampling distribution. The process is repeated several times and the mean differences are graphed in a frequency distribution. Under the assumption of no difference between the means (null hypothesis), the mean of this sampling distribution is zero.

A typical research situation in which one would use the independent t-test might involve one group of employees receiving sales training and the second group of employees *not* receiving any sales training. The number of sales for each group is recorded and averaged. The null hypothesis would be stated as, "The average sales for the two groups are equal." An alternative hypothesis would be stated as, "The group receiving the sales training will on average have higher sales than the group that did not receiving any sales training." If the sample data for the two groups were recorded as (1) Sales Training: mean=50, standard deviation=10, n=100, and (2) No Sales Training: mean=40, standard deviation=10, n=100; the independent t-test would be computed as:

$$ t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}} $$

The numerator of the independent t-test formula computes the difference between the means of the two independent groups. The denominator of the independent t-test formula computes the standard error of the statistic (standard deviation of the sampling distribution), which is computed as:

$$ S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} $$

This equation uses information from both samples to estimate a common population variance, which is referred to as "pooling the variances" or "averaging" the sample variances. The square root of this "pooled variance" estimate is the standard error of the statistic (or sampling distribution of the statistic).

The independent t-test for the mean difference between the two groups would be computed as follows:

$$ t = \frac{50 - 40}{1.41} = 7.09 $$

Referring to the table of t-values in the appendix, we find that the computed t-value=7.09 exceeds the tabled t-value of 3.291, for df=198 (df=∞) and p=0.0005 (one-tailed test). This is considered a statistically significant finding because a mean

difference this large is beyond a chance occurrence. We would conclude that the group who received sales training on average had higher sales than the group that received no sales training. This result is an important finding if the company is spending a substantial amount of money on its employees for sales training and wished to know if the sales training was effective in increasing sales.

Dependent t-Test

Drawing one random sample, taking two measures on each person, subtracting these two related measures, and graphing the difference generates the sampling distribution of the difference between the means of two related samples. The expected average difference between the related measures is zero under the null hypothesis. The **dependent t-test** is sometimes referred to as the paired t-test or correlated t-test because it uses two sets of scores on the same individuals.

A typical research situation that uses the dependent t-test involves a repeated measure design with one group. For example, a psychologist is studying the effects of a certain motion picture upon attitudes toward violence. The psychologist hypothesizes that viewing the motion picture will cause the students' attitudes to be more violent. A random sample of ten students is given an attitude toward violence inventory before viewing the motion picture. Next, the ten students view the motion picture, which contains graphic violence portrayed as acceptable behavior. The ten students are then given the attitude toward violence inventory after viewing the motion picture. The average attitude toward violence score for the students before viewing the motion picture was 67.5, but after viewing the motion picture was 73. There are ten pairs of scores, so ten score differences are squared and summed to calculate the sum of square differences. The standard error of the dependent t-test is the square root of the sum of squared differences divided by N (N−1).

The dependent t-test to investigate whether the students' attitudes toward violence changed after viewing the motion picture would be computed as:

$$t = \frac{\bar{D}}{S_D}$$

The numerator in the formula is the average difference between the post- and pre-mean scores on the attitude toward violence inventory, which is 73−67.5=5.5. The denominator is calculated as:

$$S_D = \sqrt{\frac{\Sigma(D - \bar{D})^2}{N(N-1)}}$$

The repeated measure sample data are given below

Student	Pre	Post	D	D²
1	70	75	+5	25
2	60	70	+10	100
3	85	80	−5	25
4	50	65	+15	225
5	65	75	+10	100
6	80	70	−10	100
7	90	95	+5	25
8	70	80	+10	100
9	40	55	+15	225
10	65	65	0	0
Calculations:	$\overline{D}_1 = 67.5$	$\overline{D}_2 = 73.0$	$\Sigma D = 55$	$\Sigma D^2 = 925$

There are ten pairs of student scores

NOTE: $\Sigma(D - \overline{D})^2 = \Sigma D^2 - (\Sigma D)^2 / N = 925 - (55)^2 / 10 = 622.5$

$S_D = 2.63$ and $\overline{D} = 5.5$

The dependent t-test is calculated as: $t = 5.5/2.63 = 2.09$. This value is compared to a tabled t-value for $df = 9$ and $p = 0.01$ (probability value), which is $t = 2.82$. Since the computed $t = 2.09$ does not exceed the tabled $t = 2.82$, we retain the null hypothesis. We conclude that watching the motion picture did not change students' attitudes toward violence.

Notice in calculating the dependent t-test that the sum of the differences between the pre- and post-test scores took into account whether the difference was positive or negative. Also notice that a computation formula makes the calculation of the sum of squares for the difference scores much easier to calculate. The importance of the standard error of the sampling distribution in determining whether the sample statistic is significant should be apparent.

The sampling distribution of a statistic is important in making inferences about the corresponding population parameter. The standard error of the sampling distribution of a statistic is important in testing whether the sample statistic is statistically different from the population parameter. The t-distribution is not normally distributed for small sample sizes. The t-values in the t-distribution become equal to z-values in the normal distribution when sample sizes are infinitely large. Different sampling distributions are created for different statistical tests. The one-sample t-test is useful for testing whether a sample mean from a population is statistically different from the population mean. The two sample independent t-test is useful for testing whether the means of two independent groups are identical. The single sample dependent t-test is useful for testing whether the means before and after some intervention in a single group are statistically different. A table of t-values has been created for various sample sizes (degrees of freedom) and probability areas under the t-distribution.

STUDENT R Program

The STUDENT program shows the versatility of the t-test function in R by utilizing it for a one sample t-test, a two sample independent t-test, and a dependent t-test. This is accomplished by changing parameters in the **t-test** function. The greatest complexity occurs in determining the values for the user-defined variables based on the type of test. How the different variables used for each test is described in the program. Once these variables are defined, two samples are drawn (only the second sample is actually used for the one sample t-test) from normal distributions with the prescribed means and standard deviations. Next, the type of test desired determines which block of code is used to perform the test and produce the output. For the one sample t-test, only *sampleTwo* is passed to the function along with a **mu** equal to *meanOne*. The results of the t-test are saved to the *tResults* object for later use and output appropriate to the one sample test is appended to the *output-Text* vector. For the two sample independent t-test, both samples are passed along with the **paired** keyword set to false. Again, appropriate output is created for this test. Finally, the dependent t-test looks exactly like the independent t-test except **paired** is set to true. All tests have the keyword **alternative** set to **two.sided** because the alternate hypothesis in all cases is for a difference in either direction, instead of a directional hypothesis. After the specific test functions are performed, the output from the t-test that is common to all test types is extracted from the *tResults* object by use of **$statistics** and **$p.value** components. Once these values are appended to the output, the output vector is formatted into a matrix and the matrix is printed.

STUDENT Program Output

One Sample t-test

```
One sample t-test       Sample size= 20
Sample Mean= 53.52      Sample Std Dev= 10.13
Population Mean= 50      Population Std Dev= 10
t-test= 1.55            p-value= 0.137
```

Independent t-test

```
Two sample independent t-test    Sample size= 20
Sample One Mean= 49.42           Sample One Std Dev= 10.33
Sample Two Mean= 47.9            Sample Two Std Dev= 13.44
t-test= 0.4                      p-value= 0.691
```

Dependent t-test

```
Dependent t-test           Number of paired scores= 20
Pre Mean= 48.83            Post Mean= 50.16
Sum of Differences= 26.68  Sum Squared Diff = 2557.61
t-test= 0.52               p-value= 0.611
```

t Exercises

1. Run the STUDENT program for a one-sample t-test. Record the values in the
   table.
   Use the following values:

```
sampleSize <- 30
meanOne <- 50    (In one-sample t-test equals population mean.)
stdDevOne <- 20 (In one-sample t-test equals population s.d.)
meanTwo <- 60
stdDevTwo <- 10
testType <- 1
```

One-sample t-test	Sample size =
Sample Mean =	Sample SD =
Population Mean =	Population SD =
t-test =	p-value =

   a. Is the sample mean statistically different from the population mean?
      YES _____ NO _____
   b. Was the one-sample t-test calculated correctly? YES _____ NO _____
   c. Would you obtain the same one-sample t-test using the sample standard devia-
      tion as the population standard deviation? YES _____ NO _____
   d. Interpret the results of the one-sample t-test. What would you conclude?
      _____
      _____

2. Run the STUDENT program for the two sample independent t-test. Record the
   values in the table. Use the following settings:

```
sampleSize <- 30
meanOne <- 50
stdDevOne <- 20
meanTwo <- 60
stdDevTwo <- 10
testType <- 2
```

Two sample independent t-test	Sample size =
Sample One Mean =	Sample One SD =
Sample Two Mean =	Sample Two SD =
t-test =	p-value =

   a. Are the two sample means statistically different? YES _____ NO _____
   b. Was the two sample independent t-test calculated correctly? YES _____
      NO _____
   c. Would you obtain different results if you ran the STUDENT program a
      second time?
      YES _____ NO _____
      Why? _____

d. Interpret the results of the two sample independent t-test. What would you conclude?

_____

_____

3. Run the STUDENT program for the dependent t-test. Record the values in the table below. Use the following settings:

```
sampleSize <- 30
meanOne <- 50
stdDevOne <- 20
meanTwo <- 60
stdDevTwo <- 10
testType <- 3
```

Dependent t-test	Number of paired scores =
Pre Mean =	Post Mean =
Sum of differences =	Sum Diff Squared =
t-test =	p-value =

a. Was the group statistically different from pre to post? YES _____ NO _____
b. Was the dependent t-test calculated correctly? YES _____ NO _____
c. Would you obtain different results if you ran the STUDENT program a second time?
   YES _____ NO _____
   Why? _____
d. Interpret the results of the dependent t-test. What would you conclude?

_____

_____

True or False Questions

t-Test

T   F   a. The sampling distribution of a statistic is useful for making an inference about the population parameter.

T   F   b. The standard error of a statistic provides the basis for testing hypotheses.

T   F   c. The t-distribution is normally distributed.

T   F   d. The one-sample t-test may or may not be significant depending upon which probability value is selected from the t-table.

T   F   e. The t-values and z-values become equal as sample size increases.

T   F   f. A z-test is preferred over the t-test for hypothesis testing.

T   F   g. The independent t-test is used to test whether two independent sample means are statistically different.

T   F   h. The dependent t-test is used to test whether two groups differ after receiving some type of intervention.