

What are some tutorials for conducting an Independent Samples t Test using the SAS software?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What are some tutorials for conducting an Independent Samples t Test using the SAS software?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=150514>

The Independent Samples t Test is a statistical analysis used to compare the means of two independent groups. Conducting this test using the SAS software can provide accurate and efficient results. To learn how to perform an Independent Samples t Test using SAS, there are various tutorials available online. These tutorials offer step-by-step instructions and examples on how to set up and run the test, interpret the results, and make any necessary adjustments. Some recommended tutorials for conducting an Independent Samples t Test using SAS include those provided by SAS Institute, SAS Support Community, and SAS Tutorials on YouTube. These tutorials can help individuals gain a better understanding of the test and improve their skills in using SAS for statistical analysis.

Independent Samples t Test

The Independent Samples t Test compares the means of two independent groups in order to determine whether there is statistical evidence that the associated population means are significantly different. The Independent Samples t Test is a parametric test.

This test is also known as:

Independent t Test
Independent Measures t Test
Independent Two-sample t Test
Student t Test
Two-Sample t Test
Uncorrelated Scores t Test
Unpaired t Test
Unrelated t Test

The variables used in this test are known as:

Dependent variable, or test variable
Independent variable, or grouping variable

Common Uses

The Independent Samples *t* Test is commonly used to test the following:

Statistical differences between the means of two groups
Statistical differences between the means of two interventions
Statistical differences between the means of two change scores

Note: The Independent Samples t Test can only compare the means for two (and only two) groups. It cannot make comparisons among more than two groups. If you wish to compare the means across more than two groups, you will likely want to run an ANOVA.

Data Requirements

Your data must meet the following requirements:

Dependent variable that is continuous (i.e., interval or ratio level)
Independent variable that is categorical *and* has exactly two categories
Cases that have values on both the dependent and

independent variables Independent samples/groups (i.e., independence of observations)

There is no relationship between the subjects in each sample. This means that:

Subjects in the first group cannot also be in the second group No subject in either group can influence subjects in the other group No group can influence the other group Violation of this assumption will yield an inaccurate p value Random sample of data from the population Normal distribution (approximately) of the dependent variable for each group

Non-normal population distributions, especially those that are thick-tailed or heavily skewed, considerably reduce the power of the test Among moderate or large samples, a violation of normality may still yield accurate p values Homogeneity of variances (i.e., variances approximately equal across groups)

When this assumption is violated and the sample sizes for each group differ, the p value is not trustworthy. However, the Independent Samples t Test output also includes an approximate t statistic that is not based on assuming equal population variances. This alternative statistic, called the Welch t Test statistic¹, may be used when equal variances among populations cannot be assumed. The Welch t Test is also known as Unequal Variance t Test or Separate Variances t Test. No outliers

Note: When one or more of the assumptions for the Independent Samples t Test are not met, you may want to run the nonparametric Mann-Whitney U Test instead.

Researchers often follow several rules of thumb:

Each group should have at least 6 subjects, ideally more. Inferences for the population will be more tenuous with too few subjects. A *balanced* design (i.e., same number of subjects in each group) is ideal. Extremely unbalanced designs increase the possibility that violating any of the requirements/assumptions will threaten the validity of the Independent Samples t Test.

¹ Welch, B. L. (1947). The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1-2), 28-35.

Hypotheses

The null hypothesis (H₀) and alternative hypothesis (H₁) of the Independent Samples t Test can be expressed in two different but equivalent ways:

H₀: $\mu_1 = \mu_2$ ("the two population means are equal")

H₁: $\mu_1 \neq \mu_2$ ("the two population means are not equal")

OR

H₀: $\mu_1 - \mu_2 = 0$ ("the difference between the two population means is equal to 0")

$H_1: \mu_1 - \mu_2 \neq 0$ ("the difference between the two population means is not 0")

where μ_1 and μ_2 are the population means for group 1 and group 2, respectively. Notice that the second set of hypotheses can be derived from the first set by simply subtracting μ_2 from both sides of the equation.

Test Statistic

The test statistic for an Independent Samples t Test is denoted t . However, there are different formulas for the test statistic and degrees of freedom, based on whether or not we assume that the two groups have equal variances.

SAS produces both forms of the test, so both forms of the test are described here. **Note that the null and alternative hypotheses are identical for both forms of the test statistic.**

There are two differences between the Pooled and Satterthwaite t tests: how the test statistic is calculated, and the degrees of freedom used to determine the significance of the test statistic. The test statistic for the Pooled t test uses pooled variances, and the degrees of freedom are n_1+n_2-2 . The test statistic for the Satterthwaite t test utilizes un-pooled variances, and the degrees of freedom use a special correction formula called the Satterthwaite equation. Why does it matter? When the groups have unequal variances, that introduces uncertainty into the results; the way we take this into account is by using a more conservative estimate for the degrees of freedom and test statistic.

Equal variances assumed (Pooled)

When the two independent samples are assumed to be drawn from populations with identical population variances (i.e., $\sigma_1^2 = \sigma_2^2$), the test statistic t is computed as:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

with

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Where

\overline{x}_1 = Mean of first sample

\overline{x}_2 = Mean of second sample

n_1 = Sample size (i.e., number of observations) of first sample

n_2 = Sample size (i.e., number of observations) of second sample

s_1 = Standard deviation of first sample

(s_{2}) = Standard deviation of second sample

(s_{p}) = Pooled standard deviation

The calculated t value is then compared to the critical t value from the t distribution table with degrees of freedom $df = n_1 + n_2 - 2$ and chosen confidence level. If the calculated t value is greater than the critical t value, then we reject the null hypothesis.

Note that this form of the independent samples T test statistic assumes equal variances.

Because we assume equal population variances, it is OK to "pool" the sample variances (sp). However, if this assumption is violated, the pooled variance estimate may not be accurate, which would affect the accuracy of our test statistic (and hence, the p-value).

Equal variances not assumed (Satterthwaite)

When the two independent samples are assumed to be drawn from populations with unequal variances (i.e., $\sigma_1^2 \neq \sigma_2^2$), the test statistic t is computed as:

$$t = \frac{\overline{x}_{1} - \overline{x}_{2}}{\sqrt{\frac{s_{1}^2}{n_{1}} + \frac{s_{2}^2}{n_{2}}}}$$

where

(\bar{x}_{1}) = Mean of first sample

(\bar{x}_{2}) = Mean of second sample

(n_{1}) = Sample size (i.e., number of observations) of first sample

(n_{2}) = Sample size (i.e., number of observations) of second sample

(s_{1}) = Standard deviation of first sample

(s_{2}) = Standard deviation of second sample

The calculated t value is then compared to the critical t value from the t distribution table with degrees of freedom

$$df = \frac{\left(\frac{s_{1}^2}{n_{1}} + \frac{s_{2}^2}{n_{2}} \right)^2}{\frac{1}{n_{1}-1} \left(\frac{s_{1}^2}{n_{1}} \right)^2 + \frac{1}{n_{2}-1} \left(\frac{s_{2}^2}{n_{2}} \right)^2}$$

and chosen confidence level. If the calculated t value > critical t value, then we reject the null hypothesis.

Note that this form of the independent samples T test statistic does not assume equal variances. This is why both the denominator of the test statistic and the degrees of freedom of the critical value of t are different than the equal variances form of the test statistic.

Folded F Test for Equality of Variances

Recall that the Independent Samples t Test requires the assumption of *homogeneity of variance* -- i.e., both groups have the same variance. SAS includes a test for the homogeneity of variance, called the **Folded F Test**, whenever you run an independent samples T test.

The hypotheses for the folded F test are:

H0: $\sigma_1^2 = \sigma_2^2$ ("the population variances of group 1 and 2 are equal")

H1: $\sigma_1^2 \neq \sigma_2^2$ ("the population variances of group 1 and 2 are not equal")

This implies that if we reject the null hypothesis of the Folded F Test, it suggests that the variances of the two groups are not equal; i.e., that the homogeneity of variances assumption is violated. (Source: [SAS 9.2 User's Guide, Second Edition](#))

You will use the results of the Folded F test to determine which output from the Independent Samples t test to rely on: Pooled or Satterthwaite. If the test indicates that the variances are equal across the two groups (i.e., p-value large), you will rely on the Pooled output when you look at the results for the Independent Samples t Test. If the test indicates that the variances are not equal across the two groups (i.e., p-value small), you will need to rely on the Satterthwaite output when you look at the results of the Independent Samples t Test.

Data Set-Up

Your data should include two variables (represented in columns) that will be used in the analysis. The independent variable should be categorical, and should have exactly two groups. The independent variable's type can be numeric or string, as long as there are only two categories. (Missing values do not count as a category.) The dependent variable should be continuous (i.e., interval or ratio), and must therefore be numeric. Each row of the dataset should represent a unique subject or case.

The following screenshot shows a selection of variables (not exhaustive) from the sample dataset that could be used in an Independent Samples t Test:

	Gender	Athlete	State	Height	MileMinDur	English
1	0	1	In state	64.5	0:08:14	64
2	0	0	In state	74.0	0:07:21	74
3	0	0	In state	68.5	.	87
4	1	0	In state	67.5	0:08:34	90
5	0	0	In state	75.0	0:07:30	71
6	1	1	Out of state	73.5	0:06:43	68
7	0	0	In state	76.0	.	84
8	0	0	In state	76.0	0:09:23	78
9	.	0	In state	66.5	0:10:53	81

In this example, the variables Gender, Athlete, and State would be acceptable for use as independent variables in the Independent Samples *t* Test. Gender and Athlete are numeric, with data values 0 and 1. State is a string variable, with data values "In state" and "Out of state".

The variables Height, MileMinDur, and English would be acceptable for use as dependent variables in the Independent Samples *t* Test. Variables Height and English are both numeric. Variable MileMinDur is a duration variable, which is a special type of numeric variable in SAS. (However, if MileMinDur was read into SAS as a character variable, it will need to be converted to a duration variable before using it in the *t* test.)

SAS can only make use of cases that have nonmissing values for the independent and the dependent variables, so if a case has a missing value for either variable, it can not be included in the test. Additionally, if you try to use a variable with more than two categories as the independent variable, SAS will return an error.

Using PROC TTEST for the Independent Samples *t* Test

When conducting an Independent Samples *t* Test, the general form of PROC TTEST is:

```
PROC TTEST DATA=dataset-name ALPHA=.05;
VAR dependent-variable-name(s);
CLASS independent-variable-name(s);
RUN;
```

In the PROC TTEST statement, the DATA option specifies the name of your dataset. The optional ALPHA option specifies the desired significance level. By default, PROC TTEST uses ALPHA=.05 (i.e., 5% significance), but you can set it to ALPHA=.01 for 1% significance, or ALPHA=.10 for 10% significance, etc.

The VAR statement is where you specify the dependent variable(s) -- that is, the continuous

numeric variable -- to use in the test. If you are specifying more than one dependent variable, simply separate the names of the variables using spaces.

The `CLASS` statement is where you specify the independent variable -- that is, the categorical variable -- to use in the test. You may only specify one `CLASS` variable at a time; if you try to specify more than one `CLASS` variable, the procedure will not run.

If you specify more than one `VAR` variable, you will get back more than one t test result. Specifically, SAS will produce t tests comparing the means of each `VAR` variable between the groups of the `CLASS` variable.

When using `PROC TTEST` for an independent samples t -test, your independent variable must be specified using the `CLASS` statement, not the `BY` statement. Using the `BY` statement will partition your data into subsets based on the `BY` variable and run One Sample t Tests on those subsets. Using the `CLASS` statement will compare the means of the `CLASS` variable group using an Independent Samples t Test.

Example: Independent samples T test when variances are equal

Problem Statement

In our sample dataset, students reported their writing placement test scores, and whether or not they were male or female. Suppose we want to know if the average writing score is different for males versus females. This involves testing whether the sample means for writing scores among males and females in your sample are statistically different (and by extension, inferring whether the means for writing scores in the population are significantly different between these two groups). You can use an Independent Samples t Test to compare the mean writing scores for males and females.

Before the Test

State the Null and Alternative Hypotheses

The hypotheses for this example can be expressed as:

H_0 : $\mu_{\text{males}} = \mu_{\text{females}}$ ("the mean writing scores in the population of males is identical to the mean writing score in the population of females")

H_1 : $\mu_{\text{males}} \neq \mu_{\text{females}}$ ("the two population means are not equal")

where μ_{males} and μ_{females} are the population means for males and females, respectively.

Before we perform our hypothesis tests, we should decide on a *significance level* (denoted α). The

significance level is the threshold we will use to decide whether a test result is significant. For this example, let's use $\alpha = 0.05$, or 5%.

Data Set-Up

In the sample data, we will use two variables: Gender and Writing. The variable Gender has values of either "1" or "0" which correspond to females and males, respectively. It will function as the independent variable in this t test. The variable Writing is a numeric variable, and it will function as the dependent variable. In SAS, the first few rows of data look like this (if variable and value labels have been applied):

	ID Number	Gender	Score on Writing placement test
1	43783	Male	81.44
2	20278	Male	73.27
3	20389	Male	84.24
4	22820	Female	73.16
5	24559	Female	80.45
6	28980	Male	77.48
7	33312	Male	70.79
8	40274	Male	65.89

Exploratory Data Analysis + Check the Assumptions

Recall that the Independent Samples t Test has several assumptions that we must take into account:

The dependent variable should be normally distributed in both groups. The variance of the dependent variable should be the same in both groups; if it isn't, use the alternate version of the test statistic.

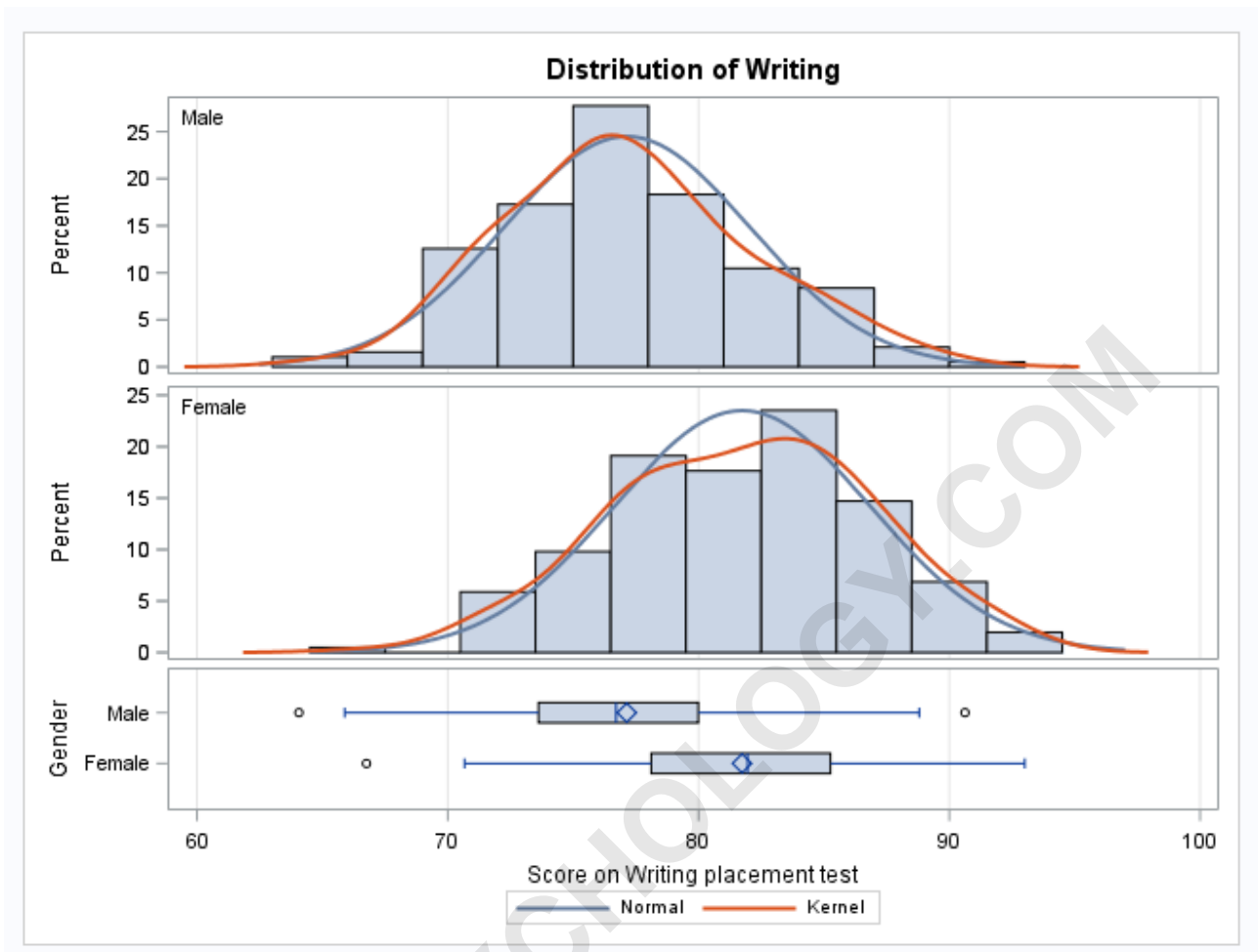
So before we jump into the Independent Samples t Test, it is a good idea to look at descriptive statistics and graphs to get an idea of what to expect, and to see if the assumptions of the test have been reasonably met. To do this, we'll want to look at the means and standard deviations of Writing for males and females, as well as graphs that compare the distribution of Writing for males versus females. `PROC TTEST` automatically runs descriptive statistics and graphs for us, but we can also use `PROC MEANS` to produce descriptive statistics by group:

```
PROC MEANS DATA=sample;
VAR Writing;
CLASS Gender;
RUN;
```

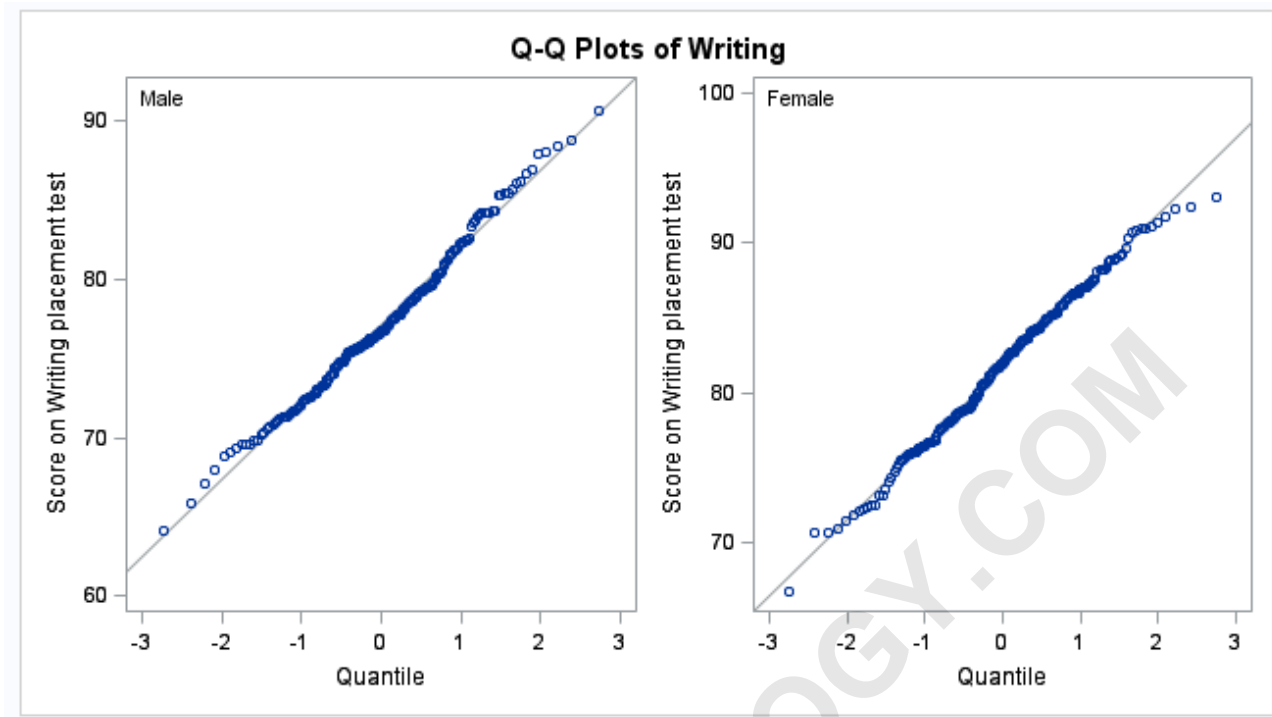
Analysis Variable : Writing Score on Writing placement test						
Gender	N Obs	N	Mean	Std Dev	Minimum	Maximum
Male	204	191	77.1382723	4.8877155	64.0600000	90.6200000
Female	222	204	81.7343627	5.0907804	66.7600000	93.0100000

PROC MEANS tells us several important things. First, there were 204 males and 222 females in the dataset, but only 191 males and 204 females reported a writing score. (This is important to know, because PROC TTEST can only use cases with nonmissing values for both gender and writing score. So our effective sample size for this test is $191+204 = 395$, which is less than the total number of rows in the sample dataset (435).) Second, the mean writing score for males is 77.14 points, while the mean writing score for females is 81.73 points. This is a difference of more than four points. Third, the standard deviations for males' and females' writing scores are very similar: 4.88 for males, and 5.09 for females.

For graphs, we can use the two graphs that PROC TTEST produces for an independent samples t test:



The first graph contains histograms (top 2 panels) and boxplots (bottom panel) comparing the distributions of males' writing scores and females' writing scores. From the histograms, we can see that the distribution of writing scores for both the males and the females are roughly symmetric, but the distribution of females' writing scores is "shifted" slightly to the right of the males. From the boxplots, we can see that the total length of the boxplots and the inter-quartile range (distance between the 1st and 3rd quartiles, i.e. the edges of the boxes) is similar for males and females. This is what we would expect to see if the two groups had the same variance. By contrast, when we look at the center lines in the boxplot (which represent the median score), we see that they do not line up: the center line for the females' box plot is to the right of the center line for the males' boxplot. Additionally, the diamond shape in each box plot represents the mean score; we see that the mean score for the females is to the right of the mean score for the males. If the two groups had the same mean, we would expect these center lines and/or diamonds to "line up" vertically.



The second graph contains Q-Q plots of the writing scores for males (left panel) versus females (right panel). The Q-Q plots produced by PROC TTEST can be used to check if a variable's observed values are consistent with what we would expect them to be if the variable was truly normally distributed. To read a Q-Q plot, we look to see if the dots (the observed values) match up with the expected values for a normal distribution (the diagonal line). If the points fall along the line, then the values are consistent with what we would expect them to be if the data were truly normally distributed. In this case, we see that the values in the middle of the range are consistent with a normal distribution, for both males and females. Both groups have slight deviations from normality in the tails. Therefore, the normality assumption required for the independent samples t test appears to be satisfied.

Running the Test

SAS Program

```
PROC TTEST DATA=work.sample ALPHA=.05;  
VAR Writing;  
CLASS Gender;  
RUN;
```

Output

Tables

Four tables appear in the `PROC TTEST` output; let's go through each one in the order they appear.

Gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
Male	191	77.1383	4.8877	0.3537	64.0600	90.6200
Female	204	81.7344	5.0908	0.3564	66.7600	93.0100
Diff (1-2)		-4.5961	4.9936	0.5028		

The first table contains descriptive statistics for both groups, including the valid sample size (n), mean, standard deviation, standard error (s/sqrt(n)), minimum, and maximum. Much of this we already saw in the `PROC MEANS` output, but this table also contains the computed difference between the two means. In this case, the first mean (male) was 4.5961 points lower than the second mean (females). In plain English, this means that, on average, females scored over 4 points higher on their writing placement test than males. Keep in mind that the independent samples t test is testing whether or not this difference is statistically different from zero.

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Male		77.1383	76.4407	77.8359	4.8877	4.4418	5.4339
Female		81.7344	81.0316	82.4371	5.0908	4.6401	5.6392
Diff (1-2)	Pooled	-4.5961	-5.5846	-3.6076	4.9936	4.6676	5.3690
Diff (1-2)	Satterthwaite	-4.5961	-5.5833	-3.6089			

The second table contains confidence limits for the group means, confidence limits for the group standard deviations, and confidence limits for the difference in the means (which is what we're interested in). Notice that there are two different confidence interval formulas for the difference. The first, Pooled, assumes that both groups have the same variance in writing scores. The second, Satterthwaite, does not make this assumption (i.e., it takes into account that one group has a different variance in writing scores than the other). We know from our exploratory data analysis that males and females have similar standard deviations, so we should look at the Pooled confidence interval. The 95% confidence interval for the difference in the writing scores is (-5.5846, -3.6076).

The third table contains the actual t-test results, and the fourth table contains the "Equality of Variances" test results:

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	393	-9.14	<.0001
Satterthwaite	Unequal	392.75	-9.15	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	203	190	1.08	0.5703

Previously, we had used informal methods (descriptive statistics and graphs) to check if the two groups had the same variance in writing scores. However, we can do a "formal" hypothesis test to check if the two variances are approximately equal, using the Folded F test in the "Equality of Variances" table. This can help us decide whether we should use the Pooled or Satterthwaite result. The null hypothesis of the Folded F test is that the variances are equal; the alternative is that the variances are not equal. Because the p-value is greater than alpha (.05), we fail to reject the null hypothesis, and conclude that the variance of writing scores is equal for these two groups. Therefore, we will use the Pooled version of the independent samples t test.

Going back to the third table, we see that there are two versions of the *t* test: Pooled (which assumes equal variances) and Satterthwaite (which does not assume equal variances). The columns of the table, from left to right, are:

Method and *Variances* show which formula and variance assumptions are used for that test *df* is the degrees of freedom *t Value* is the test statistic, determined using the appropriate formula *Pr > |t|* is the p-value corresponding to the given test statistic and degrees of freedom

Based on the Folded F test, we decided to use the Pooled version of the test statistic. To determine if the result is significant or not, we compare the Pooled p-value ($p < .001$) against our chosen significance level alpha (.05). Since the p-value is smaller than alpha, we reject the null hypothesis.

Decision and Conclusions

Since $p < .0001$ is less than our chosen significance level $\alpha = 0.05$, we can reject the null hypothesis, and conclude that males and females had a statistically significant difference in their average writing scores.

Based on the results, we can state the following:

There was a significant difference in mean writing scores between males and females ($t_{393} = -9.14, p < .05$). The average writing score for females was over 4 points greater than the average writing score for males (95% confidence interval -5.5846, -3.6076).

Example: Independent samples T test when variances are not equal

Problem Statement

In our sample dataset, students reported their typical time to run a mile, and whether or not they were an athlete. Suppose we want to know if the average time to run a mile is different for athletes versus non-athletes. This involves testing whether the sample means for mile time among athletes and non-athletes in your sample are statistically different (and by extension, inferring whether the means for mile times in the population are significantly different between these two groups). You can use an Independent Samples t Test to compare the mean mile time for athletes and non-athletes.

Before the Test

State the Null and Alternative Hypotheses

The hypotheses for this example can be expressed as:

H0: $\mu_{\text{athlete}} = \mu_{\text{non-athlete}}$ ("the mean mile time in the population of athletes is identical to the mean mile time in the population of non-athletes")

H1: $\mu_{\text{athlete}} \neq \mu_{\text{non-athlete}}$ ("the two population means are not equal")

where μ_{athlete} and $\mu_{\text{non-athlete}}$ are the population means for athletes and non-athletes, respectively.

Additionally, we should decide on a *significance level* (typically denoted using the Greek letter alpha, α) before we perform our hypothesis tests. The significance level is the threshold we use to decide whether a test result is significant. For this example, let's use $\alpha = 0.05$.

Data Set-Up

In the sample data, we will use two variables: Athlete and MileMinDur. The variable Athlete has values of either "1" or "0" which correspond to athletes and non-athletes, respectively. It will function as the independent variable in this *t* test. The variable MileMinDur is a numeric duration variable (h:mm:ss), and it will function as the dependent variable. In SAS, the first few rows of data

look like this (if variable and value labels have been applied):

	ID Number	Are you an athlete?	Mile run time (duration)
1	43783	Non-athlete	.
2	20278	Non-athlete	0:06:21
3	20389	Non-athlete	0:07:00
4	22820	Non-athlete	0:12:44
5	24559	Athlete	0:06:25
6	28980	Athlete	0:07:24
7	33312	Athlete	0:05:47
8	40274	Athlete	0:05:40
9	40390	Athlete	0:06:51

Exploratory Data Analysis + Check the Assumptions

As before, we will look at descriptive statistics and graphs to get an idea of the differences in the groups' distributions, means, and variances. This time, we will compare the means and standard deviations of MileMinDur for the athletes and non-athletes, as well as graphs that compare the distribution of MileMinDur for athletes versus non-athletes.

```
PROC MEANS DATA=sample MAXDEC=1;
VAR MileMinDur;
CLASS Athlete;
RUN;
```

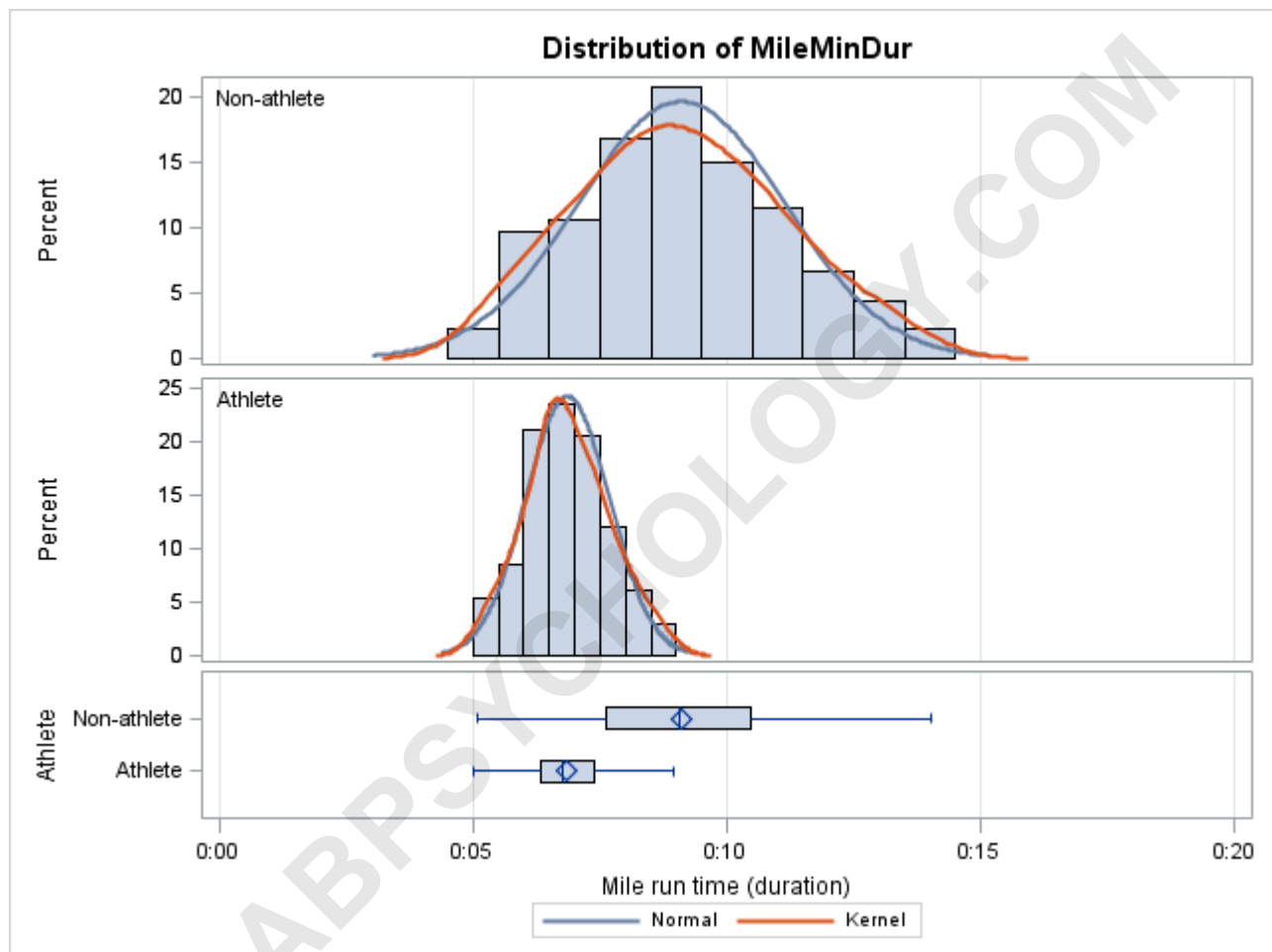
Analysis Variable : MileMinDur Mile run time (duration)						
Are you an athlete?	N Obs	N	Mean	Std Dev	Minimum	Maximum
Non-athlete	251	226	546.1	121.7	305.0	842.0
Athlete	184	166	411.3	49.4	301.0	536.0

From this table, we can see that:

Although there were 251 non-athletes and 184 athletes in the dataset, not all of them had observations for their mile run times. We can only use cases with nonmissing values for both Athlete and MileMinDur in the test, so for the purposes of the test, there are 226 non-athletes and 166 athletes. The mean mile time for non-athletes is 546.1 seconds (9 minutes, 6.1 seconds), while

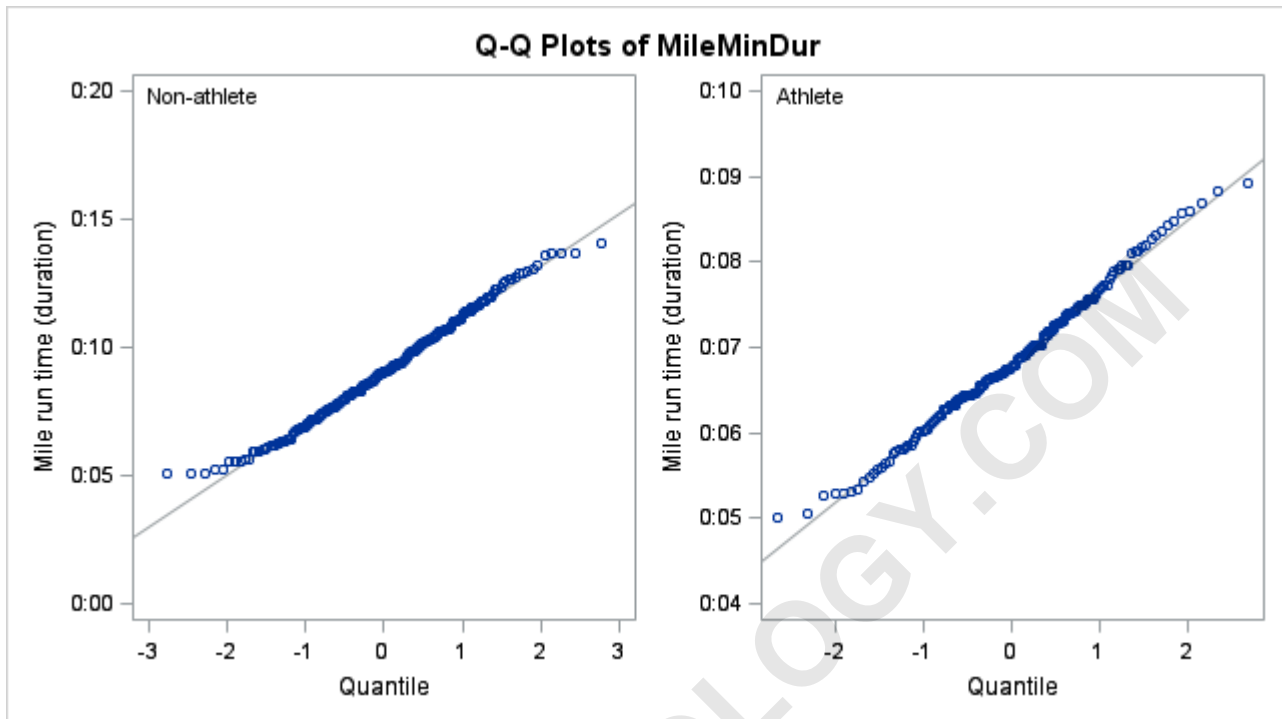
the mean mile time for athletes is 411.3 seconds (6 minutes, 51.3 seconds). This is a difference of more than two minutes. The standard deviation for the non-athletes' mile times is just over two minutes (121.7 seconds), while the standard deviation of the athletes' mile times is under one minute (49.44 seconds).

Looking at the graphs from the PROC TTEST output:



The top two panels in the histogram show the distribution of the mile run times for athletes and non-athletes, respectively. We can see that the distribution of mile times for both the athletes and the non-athletes are roughly symmetric, but the data range for the non-athletes is much larger. In the bottom panel, we see comparative boxplots of the same data. From the boxplots, we can see that the total length of the boxplots and the inter-quartile range (distance between the 1st and 3rd quartiles, i.e. the edges of the boxes) is much larger for non-athletes than athletes. If the variances of these two groups were indeed equal, we would expect the total length of the boxplots to be about the same for both groups. However, from this boxplot, it is clear that the spread of observations for non-athletes is much greater than the spread of observations for athletes. Already, we can estimate that the variances for these two groups are quite different. (We can confirm this

later using the Folded F test in the PROC TTEST output.)



The second graph contains Q-Q plots of the mile run times for non-athletes and athletes. As before, we check to see if the data values (the dots) match up to the values expected from a normal distribution (the line). In this case, we see that the values in the middle of the range are consistent with a normal distribution, for both athletes and non-athletes. Both groups have slight deviations from normality in the tails.

Running the Test

SAS Program

```
PROC TTEST DATA=work.sample ALPHA=.05;
VAR MileMinDur;
CLASS Athlete;
RUN;
```

Output

Tables

Four tables appear in the PROC TTEST output.

Athlete	N	Mean	Std Dev	Std Err	Minimum	Maximum
Non-athlete	226	546.1	121.7	8.0930	305.0	842.0
Athlete	166	411.3	49.4455	3.8377	301.0	536.0
Diff (1-2)		134.8	97.8475	10.0019		

The first table contains descriptive statistics for both groups, including the valid sample size (n), mean, standard deviation, standard error (s/sqrt(n)), minimum, and maximum. Much of this we already saw in the PROC MEANS output, but this table also contains the computed difference between the two means. In this case, the first mean (non-athletes) was 134.8 seconds larger than the second mean (athletes). In plain English, this means that, on average, it took the non-athletes about 2 minutes and 14.8 seconds longer than the athletes to complete their mile run. Keep in mind that the independent samples t test is testing whether or not this difference is statistically different from zero.

Athlete	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Non-athlete		546.1	530.2	562.0	121.7	111.4	134.0
Athlete		411.3	403.7	418.9	49.4455	44.6375	55.4235
Diff (1-2)	Pooled	134.8	115.1	154.5	97.8475	91.4364	105.2
Diff (1-2)	Satterthwaite	134.8	117.2	152.4			

The second table contains confidence limits for the group means, confidence limits for the group standard deviations, and confidence limits for the difference in the means. We specifically want to focus on the confidence intervals for the difference in the means. Notice that there are two different confidence interval formulas for the difference. The first, Pooled, assumes that both groups have the same variance in mile run time. The second, Satterthwaite, does not make this assumption (i.e., it takes into account that one group has a different variance in mile run time than the other). We know from our exploratory data analysis that the athletes and non-athletes have different variances, so we should look at the Satterthwaite confidence interval. The 95% confidence interval for the difference in the mile run times is (117.2, 152.4).

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	390	13.48	<.0001
Satterthwaite	Unequal	315.79	15.05	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	225	165	6.05	<.0001

Tables 3 and 4 contain the independent samples t test and Folded F test, respectively. This time, we had ample graphical evidence of unequal variances between the groups, so we can use the Folded F test to see if the difference in the variances is significant. Recall that the null hypothesis of this test is that the variances are equal; the alternative is that the variances are not equal. Because the p-value is less than alpha (.05), we reject the null hypothesis, and conclude that the variance of the mile run times is different for these two groups. Because of this, we will use the Satterthwaite version of the test.

Going back to table 3, we now compare the Satterthwaite t-test's p-value ($p < .001$) against our chosen significance level alpha (.05). Since the p-value is smaller than alpha, we reject the null hypothesis.

Decision and Conclusions

Since $p < .0001$ is less than our chosen significance level $\alpha = 0.05$, we can reject the null hypothesis, and conclude that the mean mile time for athletes and non-athletes is significantly different.

Based on the results, we can state the following:

There was a significant difference in mean mile time between non-athletes and athletes ($t_{315.846} = 15.047$, $p < .001$). The average mile time for athletes was 2 minutes and 14 seconds faster than the average mile time for non-athletes (95% confidence interval 117.2, 152.4).

Tutorial Feedback