

Comparing Group Means: T-tests and One-way ANOVA Using Stata, SAS, and SPSS*

Hun Myoung Park

This document summarizes the methods of comparing group means and illustrates how to conduct t-tests and one-way ANOVA using Stata 10, SAS 9, and SPSS 15.

1. Introduction
2. One sample T-test
3. Paired T-test: Dependent Samples
4. Comparing Independent Samples with Equal Variances
5. Comparing Independent Samples with Unequal Variances
6. Comparison Using the One-way ANOVA, GLM, and Regression
7. Comparing Proportions of Binary Variables
8. Conclusion

1. Introduction

T-tests and analysis of variance (ANOVA) are widely used statistical methods to compare group means. For example, the independent sample t-test enables you to compare annual personal income between rural and urban areas and examine the difference in the grade point average (GPA) between male and female students. Using the paired t-test, you can also compare the change in outcomes before and after a treatment is applied.

For a t-test, the mean of a variable to be compared should be substantively interpretable. Technically, the left-hand side (LHS) variable to be tested should be interval or ratio scaled (continuous), whereas the right-hand side (RHS) variable should be binary (categorical). The t-test can also compare the proportions of binary variables. The mean of a binary variable is the proportion or percentage of success of the variable. When sample size is large, t-tests and z-test for comparing proportions produce almost the same answer.

1.1 Background of the T-test: Key Assumptions

The t-test assumes that samples are randomly drawn from normally distributed populations with unknown population variances.¹ If such assumption cannot be made, you may try nonparametric methods. The variables of interest should be *random variables*, whose values change randomly. A constant such as the number of parents of a person is not a random variable. In addition, the occurrence of one measurement in a variable should be independent of

* The citation of this document should read: “Park, Hun Myoung. 2008. *Comparing Group Means: T-tests and One-way ANOVA Using STATA, SAS, and SPSS*. Technical Working Paper. The University Information Technology Services (UITs) Center for Statistical and Mathematical Computing, Indiana University.”

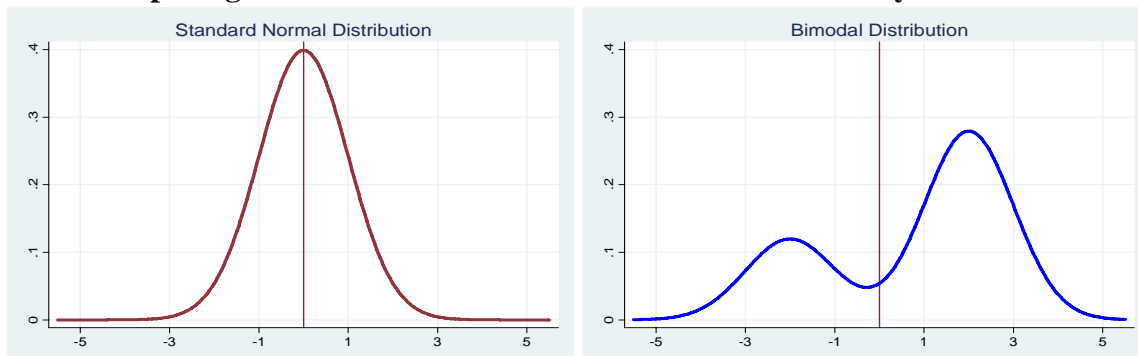
¹ If population variances are known, you may try the z-test with σ^2 ; you do not need to estimate variances.

the occurrence of others. In other word, the occurrence of an event does not change the probability that other events occur. This property is called *statistical independence*. Time series data are likely to be statistically dependent because they are often autocorrelated.

T-tests assume *random sampling* without any selection bias. If a researcher intentionally selects some samples with properties that he prefers and then compares them with other samples, his inferences based on this non-random sampling are neither reliable nor generalized. In an experiment, a subject should be randomly assigned to either the control or treated group so that two groups do not have any systematic difference except for the treatment applied. When subjects can decide whether or not to participate (non-random assignment), however, the independent sample t-test may under- or over-estimate the difference between the control and treated groups. In this case of self-selection, the propensity score matching and treatment effect model may produce robust and reliable estimates of mean differences.

Another, yet closely related to random sampling, key component is *population normality*. If this assumption is violated, a sample mean is no longer the best measure (unbiased estimator) of central tendency and t-test will not be valid. Figure 1 illustrates the standard normal probability distribution on the left and a bimodal distribution on the right. Even if the two distributions have the same mean and variance, we cannot say much about their mean difference.

Figure 1. Comparing the Standard Normal and a Bimodal Probability Distributions

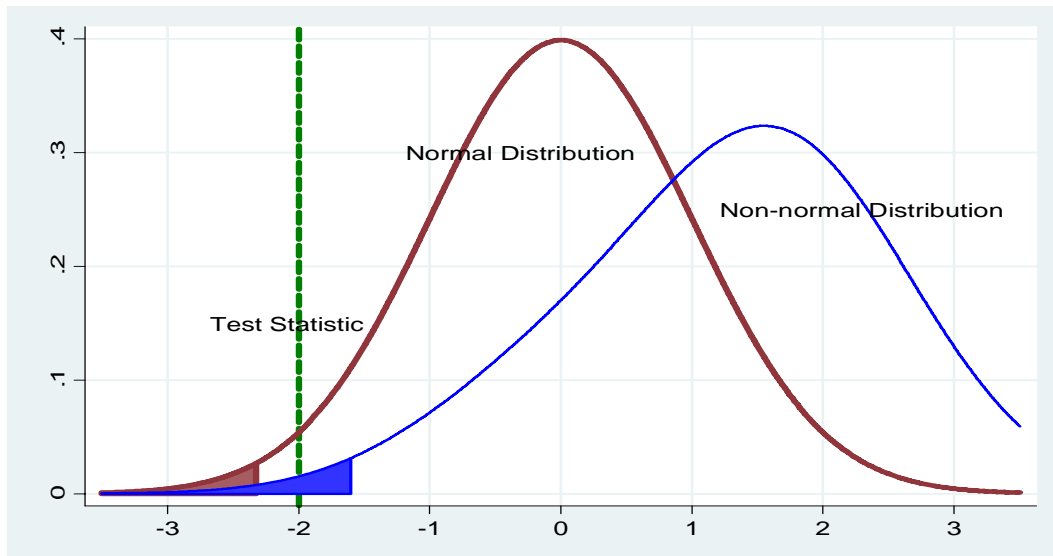


The violation of normality becomes more problematic in the *one-tailed test* than the *two-tailed* one (Hildebrand et al. 2005: 329). Figure 2 shows how the violation influences *statistical inferences*.² The left red curve indicates the standard normal probability distribution with its 1 percent one-tailed rejection region on the left. The blue one is for a non-normal distribution with the blue 1 percent *rejection region* (critical region). The test statistic indicated by a vertical green line falls in the rejection region of the skewed non-normal distribution but does not in the red shaded area of the standard normal distribution. If the populations follow such a non-normal distribution, the one-tailed t-test based on the normality does not mistakenly reject the null hypothesis.

² A statistical inference in general follows five steps: (1) formulating the null and research hypothesis (one-tailed or two-tailed), (2) choosing a (subjective) significance level (α) or forming rejection regions, (3) computing the (objective) test statistic or its p-value, (4) comparing the test statistic with the critical value or the p-value with the significance level, (5) finally draw a conclusion (reject or do not reject the null hypothesis).

Due to the Central Limit Theorem, the normality assumption is not as problematic as imagined in the real world. The Theorem says that the distribution of a sample mean (e.g., \bar{y}_1 and \bar{y}_2) is approximately normal when its sample size is sufficiently large. When $n_1 + n_2 \geq 30$, in practice, you do not need to worry too much about normality.

Figure 2. Inferential Fallacy When the Normality Assumption Is Violated



When sample size is small and normality is questionable, you might draw a histogram, P-P plot, and Q-Q plots or conduct the Shapiro-Wilk W ($N \leq 2000$), Shapiro-Francia W ($N \leq 5000$), Kolmogorov-Smirnov D ($N > 2000$), and Jarque-Bera tests. If the normality assumption is violated, you might try such nonparametric methods as the Kolmogorov-Smirnov Test, Kruscal-Wallis Test, or Wilcoxon Rank-Sum Test.

1.2 T-test and Analysis of Variance

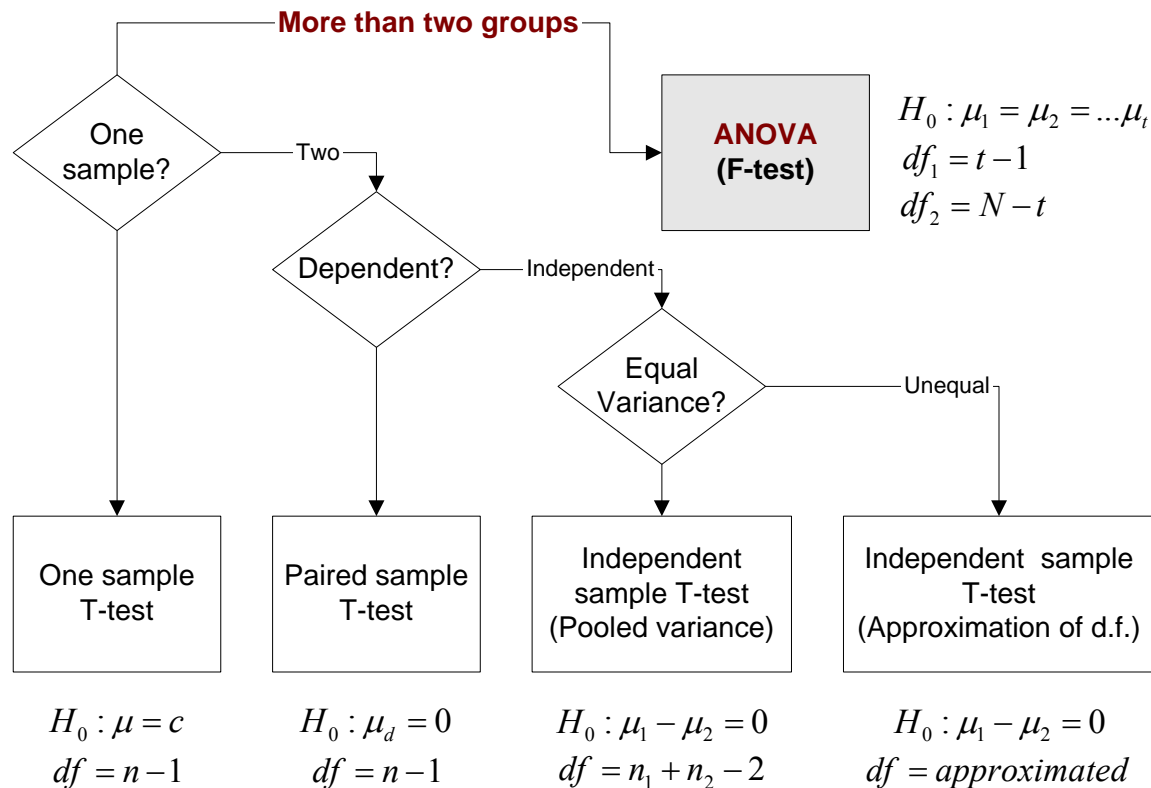
The t-test can be conducted on a one sample, paired samples, and independent samples. The one sample t-test checks if the population mean is different from a hypothesized value (oftentimes zero). If you have two samples, which are not independent but paired, you need to compute differences of individual matched pairs. A typical example is outcome measurements of pre- and post- treatment. The paired t-test examines if the mean of the differences (effect of treatment) is discernable from zero (no effect). Therefore, the underlying methods of one-sample t-test and paired t-test are in fact identical.

If two samples are taken from different populations and their elements are not paired, the independent sample t-test compares the means of two samples.³ In a GPA data set of male and female students, for example, the GPA of the first male student is nothing to do with that of the first female student. When two samples have the same population variance, the independent

³ Hildebrand et al. (2005: 361) call it the independence of “between samples” and differentiate it from the independence of “within samples,” which is the statistical independence of observations.

samples t-test uses the pooled variance when computing standard error.⁴ Otherwise, individual variances need to be used instead, and degrees of freedom should be approximated. In both cases, the null hypothesis is two samples have the same mean. Figure 3 illustrates these four types of t-tests.

Figure 3. T-tests and Analysis of Variance



While the independent sample t-test is limited to comparing the means of two groups, the one-way ANOVA (Analysis of Variance) can compare more than two groups. ANOVA use F statistic to test if all groups have the same mean. Therefore, the t-test is considered a special case of the one-way ANOVA. When comparing means of two groups (one degree of freedom), the t statistic is the square root of the F statistic of ANOVA ($F=t^2$).

Table 1. Comparison between the Independent Sample T-test and One-way ANOVA

	Independent Sample T-test	One-way ANOVA
LHS (Dependent)	Interval or ratio variable	Interval or ratio variable
RHS (Independent)	Binary variable	Categorical variable
Null Hypothesis	$\mu_1 = \mu_2$	$\mu_1 = \mu_2 = \mu_3 \dots$
Prob. Distribution	T distribution	F distribution
Sample size	Balanced/unbalanced	Balanced/unbalanced

⁴ The folded form F test examines if two samples have an equal variance in a statistical sense.

These analyses do not necessarily posit any causal relationship between the left-hand and right-hand side variables. Whether data are balanced does not matter in the t-test and one-way ANOVA.⁵ Table 1 compares the independent sample t-test and one-way ANOVA.

1.3 T-tests in Stata, SAS, and SPSS

Stata has the `.ttest` (or `.ttesti`) command to conduct t-tests. The `.anova` and `.oneway` commands perform one-way ANOVA. Stata also has `.prtest` (or `.prtesti`) to compare proportions of binary variables. The `.ttesti` and `.prtesti` commands are useful when only aggregated information (i.e., the number of observations, means or proportions, and standard deviations) is available.

In SAS, the TTEST procedure conducts various t-tests and the UNIVARIATE and MEANS procedures have options for the one sample t-test. SAS also has the ANOVA, GLM, and MIXED procedures for the one-way ANOVA. The ANOVA procedure can handle balanced data only, while the GLM and MIXED can analyze both balanced and unbalanced data. However, whether or not data are balanced does not matter in the t-test and one-way ANOVA.

SPSS has T-TEST, ONEWAY, GLM (or UNIANOVA), and MIXED commands for the t-test and one-way ANOVA. Table 2 summarizes related Stata commands, SAS procedures, and SPSS commands.

Table 2. Related Procedures and Commands in Stata, SAS, and SPSS

	Stata 10	SAS 9	SPSS 15
Normality Test	<code>.swilk; .sfancia</code>	UNIVARIATE	EXAMINE
Equal Variance Test	<code>.oneway</code>	TTEST	T-TEST
Nonparametric Method	<code>.ksmirnov; .kwallis</code>	NPARIWAY	NPARTESTS
T-test	<code>.ttest; .ttesti</code>	TTEST; MEANS	T-TEST
ANOVA	<code>.anova; .oneway</code>	ANOVA	ONEWAY
GLM*		GLM; MIXED	GLM; MIXED
Comparing Proportions	<code>.prtest; prtesti</code>	(point-and-click)	

* The Stata `.glm` command (generalized linear model) is not used for t-tests.

1.4 Data Arrangement

Figure 4 contrasts two types of data arrangement. The first data arrangement has a variable to be tested and a grouping variable to classify groups (0 or 1). The second, appropriate especially for paired samples, lists two variables to be tested. The two variables in this type are not, however, necessarily paired.

SAS and SPSS require the first type of data arrangement for the independent sample t-test and the second type for the paired t-test. Accordingly, data may need to be manipulated properly before conducting t-test with these software packages. In contrast, Stata can handle both types flexibly using options. It is generally recommended that data are arranged in the second type

⁵ When the number of observations varies across groups, we call it unbalanced data

when using Stata. Notice that the numbers of observations across groups are not necessarily equal (balanced).

Figure 4. Data Arrangement for the T-test

Variable	Group
x	0
x	0
x	0
x	0
...	...
y	1
y	1
y	1
...	...

Data Arrangement I

Variable1	Variable2
x	Y
x	y
x	Y
x	...
...	

Data Arrangement II

2. One sample T-Test

Suppose we obtain n measurements y_1 through y_n that were randomly selected from a normally distributed population with unknown parameters μ and σ^2 . One example is the SAT scores of 100 undergraduate students who were randomly chosen.

The one sample t-test examines whether the unknown population mean μ differs from a hypothesized value c . This is the null hypothesis of the one sample t-test, $H_0 : \mu = c$.⁶ The t statistic is computed as follows.

$$t = \frac{\bar{y} - c}{s_{\bar{y}}} \sim t(n-1)$$

where y is a variable to be tested, $\bar{y} = \frac{\sum y_i}{n}$ is the mean of y , $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$ is the variance

of y , $s_{\bar{y}} = \frac{s}{\sqrt{n}}$ is the standard error of \bar{y} , and n is the number of observations. The t statistic

follows Student's t probability distribution with $n-1$ degrees of freedom (Gosset 1908).

Here we are testing if the population mean of the death rate from lung cancer is 20 per 100,000 people at the .01 significance level. The null hypothesis of this two-tailed test is $H_0 : \mu = 20$.

2.1 One Sample T-test in Stata

The `.ttest` command conducts various forms of t-tests in Stata. For the one sample test, the command requires that a hypothesized value be explicitly specified. The `level()` option is to specify a confidence level as a percentage; if omitted, 95 percent by default is assumed. Note that the 99 percent confidence level is equivalent to the .01 significance level.

```
. ttest lung=20, level(99)
```

```
One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[99% Conf. Interval]
lung	44	19.65318	.6374133	4.228122	17.93529 21.37108

```

-----
      mean = mean(lung)                                t = -0.5441
Ho: mean = 20                                         degrees of freedom = 43

      Ha: mean < 20                                Ha: mean != 20                                Ha: mean > 20
Pr(T < t) = 0.2946                                Pr(|T| > |t|) = 0.5892                                Pr(T > t) = 0.7054

```

Stata first lists the descriptive statistics of the variable `lung`. The mean and standard deviation of the 44 observations are 19.6532 and 4.2281, respectively. The standard error is $.6374 = 4.2281 / \sqrt{44}$ and the 99 percent confidence interval of the mean is $\bar{y} \pm t_{\alpha/2} s_{\bar{y}} = 19.6532 \pm 2.695 * .6374$, where the 2.695 is the critical value of the two-tailed test with 43 (=44-1)

⁶ The hypothesized value c is commonly set to zero.

degrees of freedom at the .01 significance level. Finally, the t statistic is $-.5441 = (19.6532-20) / .6374$.

There are three t-tests at the bottom of the output above. The first and third are one-tailed tests, whereas the second is the two-tailed test. The first p-value .2946, for example, is for one-tailed test for the *research hypothesis* $H_a : \mu < 20$. Since this test is two tailed, you should read the second p-value for the *null hypothesis* $H_0 : \mu = 20$.

The t statistic of $-.5441$ and its large p-value of $.5892$ do not reject the null hypothesis that the population mean of the death rate from lung cancer is 20 at the .01 level. The average death rate may be 20 per 100,000 people ($p < .5892$). Notice that the hypothesized value 20 falls into the 99 percent confidence interval 17.9353-21.3711.

If you just have aggregated data (i.e., the number of observations, mean, and standard deviation of a sample), use the `.ttesti` command to replicate the t-test above. The hypothesized value 20 is specified at the end of the summary statistics. You may not omit the `level(99)` option.

```
. ttesti 44 19.65318 4.228122 20, level(99)
(output is skipped)
```

2.2 One Sample T-test Using the SAS TTEST Procedure

The TTEST procedure conducts various types of t-tests in SAS. The `H0` option specifies a hypothesized value, and `ALPHA` indicates a significance level. If omitted, zero and the .05 level by default are assumed. Let us first declare a library using the `LIBNAME` statement. A SAS data set is read though a SAS library, which is an alias of the collection of SAS data sets.⁷

```
LIBNAME masil 'c:\data\sas';
```

The `LIBNAME` statement defines a library `masil` directing to `c:\data\sas`. `DATA=masil.smoking` in the following TTEST statement tells SAS to read the data set `smoking` in the `masil` library.

```
PROC TTEST H0=20 ALPHA=.01 DATA=masil.smoking;
  VAR lung;
RUN;
```

The TTEST Procedure									
Statistics									
Variable	N	Lower CL		Upper CL		Lower CL		Upper CL	
		Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Dev	Std Err
lung	44	17.935	19.653	21.371	3.2994	4.2281	5.7989	0.6374	

⁷ SAS Enterprise Guide enables users to read SAS data sets without declaring SAS libraries. This Microsoft Windows application provides a handy point-and-click interface and thus makes SAS more like SPSS.

T-Tests			
Variable	DF	t Value	Pr > t
lung	43	-0.54	0.5892

The TTEST procedure reports descriptive statistics followed by the two-tailed t-test. The small t statistic does not reject the null hypothesis of the population mean 20 at the .01 level ($p < .5892$). If you have a summary data set containing the values of a variable `lung` and their frequencies, say `count`, use the FREQ statement.

```
PROC TTEST H0=20 ALPHA=.01 DATA=masil.smoking;
  VAR lung;
  FREQ count;
RUN;
(output is skipped)
```

2.3 One Sample T-test Using the SAS UNIVARIATE and MEANS Procedures

The SAS UNIVARIATE and MEANS procedures also perform the one sample t-test. The UNIVARIATE procedure is basically designed to produce a variety of descriptive statistics of a variable. Its MU0 option tells the procedure to perform a t-test using the hypothesized value specified. VARDEF=DF specifies a divisor (degrees of freedom) used in computing the variance (standard deviation).⁸ The NORMAL option examines if the variable is normally distributed.

```
PROC UNIVARIATE MU0=20 VARDEF=DF NORMAL ALPHA=.01 DATA=masil.smoking;
  VAR lung;
RUN;
```

The UNIVARIATE Procedure			
Variable: lung			
Moments			
N	44	Sum Weights	44
Mean	19.6531818	Sum Observations	864.74
Std Deviation	4.22812167	Variance	17.8770129
Skewness	-0.104796	Kurtosis	-0.949602
Uncorrected SS	17763.604	Corrected SS	768.711555
Coeff Variation	21.5136751	Std Error Mean	0.63741333
Basic Statistical Measures			
Location		Variability	
Mean	19.65318	Std Deviation	4.22812
Median	20.32000	Variance	17.87701
Mode	.	Range	15.26000
		Interquartile Range	6.53000

⁸ The VARDEF=N uses N as a divisor, while VARDEF=WDF specifies the sum of weights minus one (SAS 2005).

Tests for Location: $\mu_0=20$

Test	-Statistic-	-----p Value-----
Student's t	t -0.5441	Pr > t 0.5892
Sign	M 1	Pr >= M 0.8804
Signed Rank	S -36.5	Pr >= S 0.6752

Tests for Normality

Test	--Statistic---	-----p Value-----
Shapiro-Wilk	W 0.967845	Pr < W 0.2535
Kolmogorov-Smirnov	D 0.086184	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.063737	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.382105	Pr > A-Sq >0.2500

Quantiles (Definition 5)

Quantile	Estimate
100% Max	27.270
99%	27.270
95%	25.950
90%	25.450
75% Q3	22.815
50% Median	20.320
25% Q1	16.285

Quantiles (Definition 5)

Quantile	Estimate
10%	14.110
5%	12.120
1%	12.010
0% Min	12.010

Extreme Observations

-----Lowest----		----Highest----	
Value	Obs	Value	Obs
12.01	39	25.45	16
12.11	33	25.88	1
12.12	30	25.95	27
13.58	10	26.48	18
14.11	36	27.27	8

In the beginning of the output above, you may see the summary statistics such as N (44), mean(19.6532), variance (17.8770), skewness (-.1048), and kurtosis (-.9496). The third block

of the output entitled as “Tests for Location: $\mu_0=20$ ” reports the t statistic and its p-value. The fourth block entitled as “Tests for Normality” contains several statistics of the normality test. Since N is less than 2,000, you should read the Shapiro-Wilk W, which suggests that `lung` is normally distributed ($p < .2535$).

The MEANS procedure performs the one sample t-test using the T and PROBT options, which respectively request the two-tailed t statistic and its p-value. The CLM option produces the confidence interval (or upper and lower limits). The MEAN, STD, and STDERR respectively print the sample mean, standard deviation, and standard error.

```
PROC MEANS MEAN STD STDERR T PROBT CLM VARDEF=DF ALPHA=.01 DATA=masil.smoking;
  VAR lung;
RUN;
```

The MEANS Procedure

Analysis Variable : lung

Mean	Std Dev	Std Error	t Value	Pr > t	Lower 99% CL for Mean	Upper 99% CL for Mean
19.6531818	4.2281217	0.6374133	30.83	<.0001	17.9352878	21.3710758

The MEANS procedure does not, however, have an option to specify a hypothesized value other than zero. Thus, the null hypothesis here is that the population mean of death rate from lung cancer is zero. The t statistic is $30.83 = (19.6532 - 0) / .6374$ and the corresponding p-value is less than .0001. The large t statistic and small p-value reject the null hypothesis at the .01 significance level. The average death rate from lung cancer is not zero but much larger than zero. The confidence interval remains unchanged as long as the same significance level is used.

2.4 One Sample T-test in SPSS

SPSS has the T-TEST command for t-tests. First, open the SPSS syntax editor by clicking File→New→Syntax consecutively. The /TESTVAL subcommand specifies the value with which the sample mean is compared, whereas /VARIABLES lists the variables to be tested. Like Stata, SPSS specifies a confidence level rather than a significance level in the /CRITERIA=CI() subcommand.

```
T-TEST
  /TESTVAL = 20
  /VARIABLES = lung
  /CRITERIA = CI(.99).
```

Alternatively, you may click Analyze→Compare Means→One-Sample T Test and then provide variables to be compared.⁹ All SPSS output is skipped in this document.

⁹ Many SPSS users prefer point-and-clicking to writing commands in the SPSS syntax editor. Although convenient and easy to use, the point-and-click approach is not recommended since you cannot trace what you have done.

SPSS reports the mean difference from the hypothesized value ($-.34682=19.6532-20$) and its confidence interval. Do not be fooled by the misleading label “Sig. (2-tailed)” ; it is a simply p-value.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Deaths per 100k population from lung cancer	44	19.6532	4.22812	.63741

One-Sample Test

	Test Value = 20					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference	
					Lower	Upper
Deaths per 100k population from lung cancer	-.544	43	.589	-.34682	-2.0647	1.3711

3. Paired T-test: Dependent Samples

T-tests compare the means of two samples. Two variables may or may not be independent. When each element of a sample is matched to its corresponding element of the other sample, two samples are paired. This paired t-test examines the mean of individual differences of paired measurements and thus is appropriate for pre-post situations. Suppose we want to investigate the effectiveness of a new medicine on lung cancer by checking patients before and after they took the medicine.

The paired t-test is based on the pairwise differences in values of matched observations of two samples, $d_i = y_{1i} - y_{2i}$. The difference of matched pairs is treated as a variable; the logic of the paired t-test and one sample t-test is identical.

$$t_{\bar{d}} = \frac{\bar{d} - D_0}{s_{\bar{d}}} \sim t(n-1), \text{ where } \bar{d} = \frac{\sum d_i}{n}, s_{\bar{d}}^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}, \text{ and } s_{\bar{d}} = \frac{s_d}{\sqrt{n}}.$$

The null hypothesis is that the population mean of individual differences of paired observations is D_0 (zero unless explicitly specified), $H_0 : \mu_d = D_0$. If the null hypothesis is rejected, there must be a significant difference (effect) between two samples (pre and post outcomes).

3.1 Paired T-test in Stata

In Stata, you need to list two variables separated by an equal sign. The `level(95)` option can be omitted. Stata presents summary statistics of two variables and then conducts the paired t-test using the differences of matched pairs. The post mean 186.1944 is 10.1667 larger than 176.0278 before a treatment. But the standard deviation 32.6019 of the difference is also large.

```
. ttest pre=post0, level(95)
```

```
Paired t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
pre	36	176.0278	6.529723	39.17834	162.7717 189.2838
post0	36	186.1944	7.826777	46.96066	170.3052 202.0836
diff	36	-10.16667	5.433655	32.60193	-21.19757 .8642387

```

mean(diff) = mean(pre - post0)
Ho: mean(diff) = 0
Ha: mean(diff) < 0
Pr(T < t) = 0.0349

degrees of freedom = 35
t = -1.8711
Ha: mean(diff) != 0
Pr(|T| > |t|) = 0.0697
Ha: mean(diff) > 0
Pr(T > t) = 0.9651

```

The t statistic is $-1.8711 = (-10.1667-0) / 5.4337$. The degrees of freedom is $35 = 36 - 1$. The p-value of .0697 indicates a marginal significance but does not reject the null hypothesis at the .05 level. The treatment does not seem effective. The 95 percent confidence interval of the population mean is $-10.16667 \pm 2.03 * 5.433655$, where 2.03 is the critical value of the .05 level and 35 degrees of freedom.

Alternatively, you may first compute the difference of each pair of the two variables, and then conduct the one sample t-test. You may specify a hypothesized value other than zero in this approach.¹⁰

```
. gen d=pre-post0
. ttest d=0
(output is skipped)
```

3.2 Paired T-test in SAS

In the SAS TTEST procedure, you have to use the PAIRED statement instead of VAR. The following is the simple example of the PAIRED statement.

```
PROC TTEST DATA=masil.drug;
  PAIRED pre*post0;
RUN;
```

The PAIRED statement provides various ways of comparing variables using asterisk (*) and colon (:) operators. The asterisk requests comparisons between each variable on the left with each variable on the right. The colon requests comparisons between the first variable on the left and the first on the right, the second on the left and the second on the right, and so forth. Consider the following examples.

```
PROC TTEST DATA=masil.drug;
  PAIRED pro:post0;
  PAIRED (a b)*(c d); /* Equivalent to PAIRED a*c a*d b*c b*d; */
  PAIRED (a b):(c d); /* Equivalent to PAIRED a*c b*c; */
  PAIRED (a1-a10)*(b1-b10);
RUN;
```

The first PAIRED statement is the same as PAIRED pre*post0. The second and the third PAIRED statements contrast differences between asterisk and colon operators.¹¹ The hyphen (-) operator in the last statement indicates a1 through a10 and b1 through b10. Take a look at the following example of a paired t-test in SAS.

```
PROC TTEST H0=0 DATA=masil.drug;
  PAIRED (pre)*(post0-post1);
RUN;
```

The TTEST Procedure								
Statistics								
Difference	N	Lower CL Mean	Upper CL Mean	Lower CL Std Dev	Upper CL Std Dev	Lower CL Std Dev	Upper CL Std Dev	Std Err
pre - post0	36	-21.2	-10.17	0.8642	26.443	32.602	42.527	5.4337
pre - post1	36	-30.43	-20.39	-10.34	24.077	29.685	38.723	4.9475

¹⁰ In Stata, the hypothesized value is fixed to zero in paired and independent sample t-tests.

¹¹ Any text between /* and */ is considered a comment and thus ignored when SAS executes that part.

T-Tests			
Difference	DF	t Value	Pr > t
pre - post0	35	-1.87	0.0697
pre - post1	35	-4.12	0.0002

The TTEST procedure presents descriptive statistics of the differences of paired observations. The small t statistic of -1.87 does not reject the null hypothesis of no difference between pre and post0 at the .05 significance level. However, -4.12 ($= -20.39 / 4.9475$) for the difference between pre and post1 rejects the null hypothesis and suggests a significant effect of the treatment ($p < .0002$).

In order to use the UNIVARIATE and MEANS procedures, the differences of matched pairs of two variables should be computed in advance.

```
DATA masil.drug2;
  SET masil.drug;
  d1 = pre - post0;
  d2 = pre - post1;
RUN;

PROC UNIVARIATE MU0=0 VARDEF=DF NORMAL; VAR d1 d2; RUN;
PROC MEANS MEAN STD STDERR T PROBT CLM; VAR d1 d2; RUN;
PROC TTEST ALPHA=.05; VAR d1 d2; RUN;
(output is skipped)
```

3.3 Paired T-test in SPSS

In SPSS, you need to list paired variables in the PAIRS subcommand. You may click Analyze→ Compare Means→ Paired-Samples T Test and then provide paired variables to be compared.

```
T-TEST
  /PAIRS = pre post0
  /CRITERIA = CI(.95).
```

4. Comparing Independent Samples with Equal Variances

This section discusses the most typical form of t-test that compares the means of two independent random samples y_1 and y_2 . They are independent in the sense that they are drawn from different populations and each element of one sample is not paired (linked to) with its corresponding element of the other sample.

An example is the death rate from lung cancer between heavy cigarette consuming states and light consuming states. Since each state is either a heavy or light consumer, observations of two groups are not linked. The typical null hypothesis of the independent sample t-test is that the mean difference of the two groups is zero, $H_0 : \mu_1 - \mu_2 = 0$.

4.1 F test for Equal Variances

T-tests assume that samples are randomly drawn from normally distributed populations with unknown parameters. In addition to these random sampling and normality assumptions, you should check the equal variance assumption when examining the mean difference of two independent samples. The population variances of the two groups σ_1^2 and σ_2^2 need to be equal in order to use the pooled variance.¹² Otherwise, the t-test is not reliable due to the incorrect variance and degrees of freedom used.

In practice, unequal variances of two independent samples are less problematic when two samples have the same number of observations (balanced data) (Hildebrand et al. 2005: 362). The problem will be critical if one sample has a larger variance and a much smaller sample size compared to the other (362).

The folded form F-test is commonly used to examine whether two populations have the same variance, $H_0 : \sigma_1^2 = \sigma_2^2$. The F statistic is

$$\frac{s_L^2}{s_S^2} \sim F(n_L - 1, n_S - 1)$$

where L and S respectively indicate groups with larger and smaller sample variances.

The SAS TTEST procedure and SPSS T-TEST command conduct the F-test for equal variances. SAS reports the folded form F statistic, whereas SPSS computes Levene's weighted F statistic. In Stata, the `.oneway` command performs the Bartlett's chi-squared test for equal variances.

The following is an example of the F-test using the `.oneway` command. The chi-squared statistic of .1216 ($p < .727$) does not reject the null hypothesis of equal variances at the .05 significance level. Two samples appear to have the equal variance. The F statistic of 28.85 in the ANOVA table indicates a significant mean difference in the death rate from lung cancer

¹² $E(\bar{y}_1 - \bar{y}_2) = \mu_1 - \mu_2$, $Var(\bar{y}_1 - \bar{y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$ when $\sigma_1^2 = \sigma_2^2 = \sigma^2$

between heavy cigarette consuming states and light consumers ($p < .0000$). Do not be confused with the two different F statistics.

. oneway lung smoke

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	313.031127	1	313.031127	28.85	0.0000
Within groups	455.680427	42	10.849534		
Total	768.711555	43	17.8770129		

Bartlett's test for equal variances: $\text{chi2}(1) = 0.1216$ Prob>chi2 = 0.727

4.2 Overview of the Independent Sample T-test

If the null hypothesis of equal variances is not rejected, the pooled variance s_{pool}^2 can be used. The pooled variance consists of individual sample variances weighted by the number of observations of the two groups. The null hypothesis of the independent sample t-test is $H_0: \mu_1 - \mu_2 = D_0$ and the degrees of freedom are $n_1 + n_2 - 2 = (n_1 - 1) + (n_2 - 1)$.¹³ The t statistic is computed as follows.

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s_{pool} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$\text{where } s_{pool}^2 = \frac{\sum (y_{1i} - \bar{y}_1)^2 + \sum (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

When the equal variance assumption is violated, the t-test needs to use individual variances in the approximate t and the degrees of freedom. This test may be called the unequal variance t-test (Hildebrand et al. 2005: 363). Notice that the approximation below is based both on the number of observations and variances of two independent samples. The approximate t is

$$t' = \frac{\bar{y}_1 - \bar{y}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(df_{\text{Satterthwaite}})$$

$$\text{where } df_{\text{Satterthwaite}} = \frac{(n_1 - 1)(n_2 - 1)}{(n_1 - 1)(1 - c)^2 + (n_2 - 1)c^2} \text{ and } c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}$$

In case of unequal variances, the t-test requires the approximation of the degrees of freedom (Satterthwaite 1946; Welch 1947; Cochran and Cox 1992; SAS 2005; Stata 2007). Among several approximation methods, Satterthwaite's approximation is commonly used. Note that the approximation is a real number, not necessarily an integer. SAS, Stata, and SPSS all compute

¹³ In SAS, you may set a specific value of D_0 to other than zero but may not in Stata and SPSS, which do not have such an option in the `.ttest` and `T-TEST` commands.

rate from lung cancer than the heavy counterparts. Notice that the F statistic of 28.85 in section 4.1 is $t^2 = (-5.3714)^2$.

If only aggregated data of the two variables are available, use the `.ttesti` command, the immediate form of `.ttest`, and list the number of observations, mean, and standard deviation of the two variables.

```
. ttesti 22 16.98591 3.164698 22 22.32045 3.418151, level(95)
```

Suppose the data set is arranged in the second type of Figure 3 so that one variable `high_lung` has data for heavy cigarette consuming states and the other `low_lung` for light consuming states. You have to use the `unpaired` option to indicate that two variables are not paired. A grouping variable here is not necessary.

```
. ttest high_lung=low_lung, unpaired
```

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
high_lung	22	22.32045	.7287523	3.418151	20.80493	23.83598
low_lung	22	16.98591	.6747158	3.164698	15.58276	18.38906
combined	44	19.65318	.6374133	4.228122	18.36772	20.93865
diff		5.334545	.9931371		3.330313	7.338777
diff = mean(high_lung) - mean(low_lung)				t =	5.3714	
Ho: diff = 0				degrees of freedom =	42	
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 1.0000		Pr(T > t) = 0.0000		Pr(T > t) = 0.0000		

Since the variable order is reversed here, the summary statistics of heavy cigarette consumers are displayed first and t statistic has the opposite sign. However, this outcome leads to the same conclusion. The large test statistic of 5.3714 rejects the null hypothesis at the .05 level; heavy cigarette consuming states on average have a higher average death rate from lung cancer than light consumers.

The `unpaired` option is very useful since it enables you to conduct a t-test without additional data manipulation. You need to use the `unpaired` option to compare two variables, say leukemia and kidney, as independent samples in Stata.¹⁵

```
. ttest leukemia=kidney, unpaired
```

Two-sample t test with equal variances

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
leukemia	44	6.829773	.0962211	.6382589	6.635724	7.023821
kidney	44	2.794545	.0782542	.5190799	2.636731	2.95236

¹⁵ In SAS and SPSS, however, you have to stack up two variables and generate a grouping variable before performing the independent sample t-test.

```

-----+-----
combined |      88      4.812159      .2249261      2.109994      4.365094      5.259224
-----+-----
diff      |      4.035227      .1240251      3.788673      4.281781
-----+-----
diff = mean(leukemia) - mean(kidney)          t = 32.5356
Ho: diff = 0                                degrees of freedom = 86

Ha: diff < 0                                Ha: diff != 0                                Ha: diff > 0
Pr(T < t) = 1.0000                          Pr(|T| > |t|) = 0.0000                          Pr(T > t) = 0.0000

```

The average death rate from leukemia cancer is 6.8298, which is about 4 higher than the average rate from kidney cancer. But we want to know if there is any statistically significant difference in the population means of two death rates. The $F 1.5119 = (.6382589^2) / (.5190799^2)$ and its p-value ($=.1794$) do not reject the null hypothesis of equal variances. The large t statistic 32.5356 rejects the null hypothesis that average death rates from leukemia and kidney cancers have the same mean at the .05 level; the average death rate from leukemia cancer is higher than that from kidney cancer.

4.4 Independent Sample T-test in SAS

The TTEST procedure by default examines the hypothesis of equal variances, and then provides t statistics for both cases. The procedure by default reports Satterthwaite's approximation for the degrees of freedom. SAS requires that a data set is arranged in the first type of Figure 3 for the independent sample t-test; a variable to be tested is classified by a grouping variable, which should be specified in the CLASS statement. You may specify a hypothesized value other than zero using the H0 option.¹⁶

```

PROC TTEST H0=0 ALPHA=.05 DATA=masil.smoking;
  CLASS smoke;
  VAR lung;
RUN;

```

TTEST displays summary statistics of two samples and then reports the results of the t-test and F-test. First, look at the last block of the output entitled as "Equality of Variances." The labels "Num DF" and "Den DF" are respectively numerator's and denominator's degrees of freedom.

The TTEST Procedure								
Statistics								
Variable	smoke	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev
lung	0	22	15.583	16.986	18.389	2.4348	3.1647	4.5226
lung	1	22	20.805	22.32	23.836	2.6298	3.4182	4.8848

¹⁶ For example, $H_0 = -5$ influences the t statistic $-3368 = [(16.98591 - 22.32045) - (-5)] / \sqrt{10.8497 * (1/22 + 1/22)}$ and its p-value of .7379. Other statistics remain unchanged. Therefore, you do not reject the null hypothesis of 5 difference in average death rates and conclude that heavy cigarette consuming states have the 5 larger average death rate than light consumers. This conclusion is consistent with the t-test with $H_0 = 0$.

lung	Diff (1-2)	-7.339	-5.335	-3.33	2.7159	3.2939	4.1865
------	------------	--------	--------	-------	--------	--------	--------

Statistics

Variable	smoke		Std Err	Minimum	Maximum
lung		0	0.6747	12.01	25.45
lung		1	0.7288	12.11	27.27
lung	Diff (1-2)		0.9931		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > t
lung	Pooled	Equal	42	-5.37	<.0001
lung	Satterthwaite	Unequal	41.8	-5.37	<.0001

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
lung	Folded F	21	21	1.17	0.7273

The small F statistic 1.17 and its large p-value do not reject the null hypothesis of equal variances ($p < .7273$). The t-test uses the pooled variance, thus you need to read the line labeled as “Pooled”; otherwise, read the lines of “Satterthwaite” and “Cochran.” The large t -5.37 and its p-value .0001 reject the null hypothesis. Heavy cigarette consuming states have a higher average death rate from lung cancer than light consuming states.

If you have a summary data set with the values of variables `lung` and their frequency count, specify the count variable in the `FREQ` statement.

```
PROC TTEST DATA=masil.smoking;
  CLASS smoke;
  VAR lung;
  FREQ count;
RUN;
```

Now, let us compare the death rates from leukemia and kidney by stacking up two variables into one and generating a grouping variable. The following `DATA` step reshapes the data set from the second type of Figure 3 to the first type. The new variable `rate` contains death rates and `leu_kid` identifies groups.

```
DATA masil.smoking2;
  SET masil.smoking;
  rate = leukemia; leu_kid = 'Leukemia'; OUTPUT;
  rate = kidney; leu_kid = 'Kidney'; OUTPUT;
  KEEP leu_kid rate;
RUN;

PROC TTEST COCHRAN DATA=masil.smoking2;
```

```

CLASS leu_kid;
VAR rate;
RUN;

```

TTEST presents summary statistics and then conducts the t-test and F-test. The F 1.51 does not reject the null hypothesis of equal variances at the .05 level ($p < .1794$). Accordingly, you have to read the t statistic and p-value in the line labeled as "Pooled." The t -32.54 is large sufficiently to reject the null hypothesis ($p < .0001$). In this example, the violation of the equal variance assumption does not make any difference in t statistic and p-value. Note that Cochran approximation for the degrees of freedom is generated by the COCHRAN option in the TTEST statement.

The TTEST Procedure

Statistics

Variable	leu_kid	N	Lower CL		Upper CL		Lower CL Std Dev	Upper CL Std Dev	Std Err
			Mean	Mean	Mean	Mean			
rate	Kidney	44	2.6367	2.7945	2.9524	0.4289	0.5191	0.0783	
rate	Leukemia	44	6.6357	6.8298	7.0238	0.5273	0.6383	0.0962	
rate	Diff (1-2)		-4.282	-4.035	-3.789	0.5063	0.5817	0.124	

T-Tests

Variable	Method	Variances	DF	t Value	Pr > t
rate	Pooled	Equal	86	-32.54	<.0001
rate	Satterthwaite	Unequal	82.6	-32.54	<.0001
rate	Cochran	Unequal	43	-32.54	<.0001

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
rate	Folded F	43	43	1.51	0.1794

4.5 Independent Sample T-test in SPSS

In the T-TEST command, you need to provide a grouping variable in the /GROUP subcommand. The Levene's F statistic of .0000 does not reject the null hypothesis of equal variances ($p < .995$). You may click Analyze → Compare Means → Independent-Samples T Test and then provide variables to be compared and a grouping variable with specific values specified.

```

T-TEST
  /GROUPS = smoke(0 1)
  /VARIABLES = lung
  /CRITERIA = CI(.95).

```

The second table below has two parts: one for equal variance test and the other for t-test. - 5.33455 is the difference between two group means (16.9859-22.3205). Again read p-values from columns labeled “Sig.” and “Sig. (2-tailed).”

Group Statistics

	1 for those whose cigarette consumption is larger than the median	N	Mean	Std. Deviation	Std. Error Mean
Deaths per 100k population from lung cancer	0	22	16.9859	3.16470	.67472
	1	22	22.3205	3.41815	.72875

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Upper	Lower
Equal variances assumed	.000	.995	-5.371	42	.000	-5.33455	.99314	-7.3388	-3.33031
Equal variances not assumed			-5.371	41.75	.000	-5.33455	.99314	-7.3391	-3.32996

Table 3 compares the independent sample t-test under the equal variance assumption using three statistical software packages. For testing the assumption of equal variances, Stata, SAS, and SPSS respectively report the Bartlett’s test (chi-squared), the folded form F test, and Levene’s F test. Despite different test statistics in Stata and SAS, their p-values are almost the same. Stata and SAS respectively produce Welch’s and Cochran-Cox’s approximations of the degrees of freedom in addition to the Satterthwaite’s approximation. Three software packages report the same result. It is not surprising that t and approximate t (t') are the same because two population variances are equal.

Table 3. T-test for Independent Samples with Equal Variances (Lung Cancer)

	Stata 10	SAS 9.1	SPSS 15
T	-5.3714 (<.0000)	-5.37 (<.0001)	-5.371
Degrees of Freedom (DF)	42	42	42
Equal Variance Test (F-Test)	.1216 (<.727)	1.17 (<.7273)	.0000 (<.995)
T' *	-5.3714	-5.37	-5.371
Satterthwaite DF*	41.7532 (<.0000)	41.8 (<.0001)	41.753
Cochran-Cox/Welch DF*	43.7296 (<.0000)	21 (<.0001)	N/A

* These statistics for unequal variances are provided only for comparison.

5. Comparing Independent Samples with Unequal Variances

If the assumption of equal variances is violated, you have to compute the approximate t using individual sample variances instead of the pooled variance. It is also necessary to obtain the Satterthwaite's, Cochran-Cox's (SAS), or Welch's (Stata) approximation of the degrees of freedom. However, the null hypothesis remains unchanged. This section compares mean death rates from kidney cancer between the west/south areas and the east/north counterparts.

5.1 Independent Sample T-test in Stata

Let us check equality of variances using the `.oneway` command. The `tabulate` option produces a table of summary statistics of the groups. The west and south areas (`west=1`) have a lower average death rate of 2.6183 but a larger standard deviation of .5984 than the east and north counterparts.

```
. oneway kidney west, tabulate
```

Summary of kidney					
west	Mean	Std. Dev.	Freq.		
0	3.006	.3001298	20		
1	2.6183333	.59837219	24		
Total	2.7945455	.51907993	44		

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	1.63947758	1	1.63947758	6.92	0.0118
Within groups	9.94661333	42	.236824127		
Total	11.5860909	43	.269443975		

```
Bartlett's test for equal variances: chi2(1) = 8.6506 Prob>chi2 = 0.003
```

The chi-squared 8.6506 of the Bartlett's test rejects the null hypothesis of equal variances ($p < .003$). The folded F statistic is $3.9749 = .59837219^2 / .3001298^2$ with 23 and 19 degrees of freedom, which also rejects the null hypothesis ($p < .0034$).¹⁷ Therefore, the approximate t is computed using individual variances.

$$t' = \frac{3.006 - 2.6183}{\sqrt{\frac{.3001^2}{20} + \frac{.5984^2}{24}}} = 2.78187,$$

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2} = \frac{.3001^2/20}{.3001^2/20 + .5984^2/24} = .2318, \text{ and}$$

$$df_{\text{Satterthwaite}} = \frac{(n_1 - 1)(n_2 - 1)}{(n_1 - 1)(1 - c)^2 + (n_2 - 1)c^2} = \frac{(20 - 1)(24 - 1)}{(20 - 1)(1 - .2318)^2 + (24 - 1).2318^2} = 35.1071$$

¹⁷ In Stata, run `.di Ftail(23,19, 3.974885)*2` to get the p-value.

The approximate t of 2.7819 rejects the null hypothesis of no difference in mean death rates between the two areas ($p < .0086$).¹⁸ The c is .2318 and Satterthwaite's approximation of the degrees of freedom is 35.1071.

In Stata, you have to use the `unequal` option in the `.ttest` command to perform the t -test for two independent samples with unequal variances.

```
. ttest kidney, by(west) unequal level(95)
```

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	20	3.006	.0671111	.3001298	2.865535	3.146465
1	24	2.618333	.1221422	.5983722	2.365663	2.871004
combined	44	2.794545	.0782542	.5190799	2.636731	2.95236
diff		.3876667	.139365		.1047722	.6705611

diff = mean(0) - mean(1) t = 2.7817
 Ho: diff = 0 Satterthwaite's degrees of freedom = 35.1098

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.9957 Pr(|T| > |t|) = 0.0086 Pr(T > t) = 0.0043

The approximate t 2.7817 and Satterthwaite's approximation 35.1098, due to the rounding error, are slight different from what is manually computed above. Notice that the 2.7817 is the square root of the F statistic 6.92 of the `.oneway` output above. If you want to get Welch's approximation, use the `welch` as well as `unequal` options; without the `unequal` option, `welch` is ignored.

```
. ttest kidney, by(west) unequal welch
```

Two-sample t test with unequal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	20	3.006	.0671111	.3001298	2.865535	3.146465
1	24	2.618333	.1221422	.5983722	2.365663	2.871004
combined	44	2.794545	.0782542	.5190799	2.636731	2.95236
diff		.3876667	.139365		.1050824	.6702509

diff = mean(0) - mean(1) t = 2.7817
 Ho: diff = 0 Welch's degrees of freedom = 36.2258

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.9957 Pr(|T| > |t|) = 0.0085 Pr(T > t) = 0.0043

Satterthwaite's approximation is slightly smaller than Welch's 36.2258. Again, these approximations are not integers, but real numbers. The approximate t 2.7817 remains unchanged, but the p -value becomes slightly smaller due to the different approximation used.

¹⁸ In Stata, run `. di (1-ttail(35.1071, -2.78187))*2` to get the p -value.


```

PROC TTEST COCHRAN DATA=masil.smoking;
  CLASS west;
  VAR kidney;
RUN;

```

The TTEST Procedure

		Statistics						
Variable	s_west	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev
kidney	0	20	2.8655	3.006	3.1465	0.2282	0.3001	0.4384
kidney	1	24	2.3657	2.6183	2.871	0.4651	0.5984	0.8394
kidney	Diff (1-2)		0.0903	0.3877	0.685	0.4013	0.4866	0.6185

Statistics

Variable	west	Std Err	Minimum	Maximum
kidney	0	0.0671	2.34	3.62
kidney	1	0.1221	1.59	4.32
kidney	Diff (1-2)	0.1473		

T-Tests

Variable	Method	Variances	DF	t Value	Pr > t
kidney	Pooled	Equal	42	2.63	0.0118
kidney	Satterthwaite	Unequal	35.1	2.78	0.0086
kidney	Cochran	Unequal	.	2.78	0.0109

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
kidney	Folded F	23	19	3.97	0.0034

The F 3.97 rejects the null hypothesis of equal variances ($p < .0034$). Thus, individual sample variances are used to compute the approximate t and the approximation of the degrees of freedom needs to be used in statistical inferences. In the above output, Cochran-Cox's approximation is not defined since sample sizes are not equal. The t' 2.78 is large enough to reject the null hypothesis of the same population death rate ($p < .0086$). If you mistakenly read the smaller t statistic 2.63 for the pooled variance ($p < .0118$), you might not reject the null hypothesis at the .01 level.

Now, let us compare the average death rates from bladder and kidney cancers that are arranged in the second type of Figure 3. We need to reshape the data set first to make it into the first data arrangement type.

```

DATA masil.smoking3;
  SET masil.smoking;
  rate = bladder; bla_kid = 'Bladder'; OUTPUT;
  rate = kidney; bla_kid = 'Kidney'; OUTPUT;

```

```
KEEP bla_kid rate;
RUN;
```

The CLASS statement specifies a categorical variable (grouping variable) `bla_kid` and the new variable `rate` contains both death rates from bladder and kidney cancers. Unlike the Stata `ttest` command, the SAS TTEST procedure allows you to specify a hypothesized value other than zero.

```
PROC TTEST COCHRAN H0=0 DATA=masil.smoking3;
  CLASS bla_kid;
  VAR rate;
RUN;
```

The TTEST Procedure

Statistics

Variable	bla_kid	N	Lower CL		Upper CL		Lower CL Std Dev	Upper CL Std Dev	Std Err
			Mean	Mean	Mean	Mean			
rate	Bladder	44	3.8278	4.1211	4.4145	0.7972	0.9649	1.2226	0.1455
rate	Kidney	44	2.6367	2.7945	2.9524	0.4289	0.5191	0.6577	0.0783
rate	Diff (1-2)		0.9982	1.3266	1.655	0.6743	0.7748	0.9107	0.1652

T-Tests

Variable	Method	Variances	DF	t Value	Pr > t
rate	Pooled	Equal	86	8.03	<.0001
rate	Satterthwaite	Unequal	66	8.03	<.0001
rate	Cochran	Unequal	43	8.03	<.0001

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
rate	Folded F	43	43	3.46	<.0001

The 3.46 of the F test rejects the null hypothesis of equal variances ($p < .0001$). The t statistic 8.03 for the case of unequal variances is large and rejects the null hypothesis of the same population mean at the .01 level. The population means of the two death rates are different. Interestingly, these t statistics in this case happen to be same regardless whether the variances are equal or not. However, the degrees of freedom vary across approximation methods. Remember the Welch's approximation in Stata is 67.0324.

5.3 Independent Sample T-test in SPSS

Like SAS, SPSS reports t statistics for both cases of equal and unequal variances. Levene's F 5.466 rejects the null hypothesis of equal variances at the .05 level ($p < .024$). Other statistics are the same as those of SAS.

T-TEST

```

/GROUPS = west(0 1)
/VARIABLES = kidney
/CRITERIA = CI(.95).

```

Table 4 summarizes the independent sample t-test in Stata, SAS, and SPSS. These software packages respectively report the Bartlett's test, the folded form F test, and Levene's F test. Despite different test statistics in Stata and SAS, their p-values are almost same. Stata and SAS additionally produce Welch's and Cochran-Cox's approximations of the degrees of freedom, respectively. The significant difference in population variances produces different t and t' and requires approximations of the degrees of freedom.

While the t' 2.7817 rejects the null hypothesis at the .01 level, the t statistic of 2.6311 do not. Heavy cigarette consuming states have a higher average death rate from kidney cancer than the light counterparts. If you mistakenly read the latter, you may reach a misleading conclusion at the .01 level.

Table 4. T-test for Independent Samples with Unequal Variances (Kidney Cancer)

	Stata 10	SAS 9.1	SPSS 15
<i>T statistic</i> *	2.6311 (<.0118)	2.63 (<.0118)	2.631 (<.012)
<i>Degrees of Freedom (DF)</i> *	42	42	42
Equal Variance Test (F-Test)	8.6506 (<.003)	3.97 (<.0034)	5.466 (<.024)
Approximate T (T')	2.7817	2.78	2.782
Satterthwaite DF	35.1098 (<.0086)	35.1 (<.0086)	35.110 (<.009)
Cochran-Cox/Welch DF	36.2258 (<.0085)	. (<.0109)	N/A

* These statistics for equal variances are provided only for comparison.

Corrected Total	43	768.7115545			
	R-Square	Coeff Var	Root MSE	lung Mean	
	0.407215	16.75995	3.293863	19.65318	
Source	DF	Anova SS	Mean Square	F Value	Pr > F
smoke	1	313.0311273	313.0311273	28.85	<.0001

The degrees of freedom in the t-test are the same as the degrees of freedom of error in the ANOVA table. The F statistic 28.85 with 1 degree of freedom is t squared, $(-5.3714)^2$. Accordingly, their p-values are identical. Stata `.anova` and `.oneway` commands produce the same result.

```
. anova lung smoke
```

	Number of obs =	44	R-squared =	0.4072	
	Root MSE =	3.29386	Adj R-squared =	0.3931	
Source	Partial SS	df	MS	F	Prob > F
Model	313.031127	1	313.031127	28.85	0.0000
smoke	313.031127	1	313.031127	28.85	0.0000
Residual	455.680427	42	10.849534		
Total	768.711555	43	17.8770129		

In SPSS, the ONEWAY command conducts the same one-way ANOVA.

```
ONEWAY lung BY smoke.
```

6.2 Generalized Linear Model (GLM)

GLM can handle both balanced and unbalanced data. Like ANOVA, GLM produces the same result of the one-way ANOVA. The SAS GLM and MIXED procedures report the F statistic for the one-way ANOVA. However, Stata's `.glm` command does not perform the one-way ANOVA.

```
PROC GLM DATA=masil.smoking;
  CLASS smoke;
  MODEL lung=smoke /SS3;
RUN;
```

The GLM Procedure

Dependent Variable: lung

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	313.0311273	313.0311273	28.85	<.0001
Error	42	455.6804273	10.8495340		
Corrected Total	43	768.7115545			

	R-Square	Coeff Var	Root MSE	lung Mean
	0.407215	16.75995	3.293863	19.65318

Source	DF	Type III SS	Mean Square	F Value	Pr > F
smoke	1	313.0311273	313.0311273	28.85	<.0001

The MIXED procedure has the similar usage as the GLM procedure.

```
PROC MIXED; CLASS smoke; MODEL lung=smoke; RUN;
```

In SPSS, the GLM (or UNIANOVA) commands perform the one-way ANOVA. You may not exclude the intercept in these commands. The MIXED command also fits this model.

```
GLM lung BY smoke
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /DESIGN = smoke.
```

```
MIXED lung BY smoke
  /FIXED = smoke |SSTYPE(3)
  /METHOD = REML.
```

6.3 Linear Regression (Ordinary Least Squares)

The linear regression or OLS give you the same answer that the t-test and the one-way ANOVA produced. In OLS, only a dummy variable (grouping variable) is included in the right-hand side. Both OLS and ANOVA use the same covariance structure in analysis. Hence, these three methods present the same result in their own ways.

The SAS REG procedure, Stata `.regress` command, and SPSS REGRESSION command produce the same output of a linear regression model with only a dummy for the grouping variable `smoke`.

```
PROC REG DATA=masil.smoking;
  MODEL lung=smoke;
RUN;
```

```

                The REG Procedure
                Model: MODEL1
                Dependent Variable: lung

                Number of Observations Read      44
                Number of Observations Used      44
```

Analysis of Variance

```

                Sum of          Mean
```

Source	DF	Squares	Square	F Value	Pr > F
Model	1	313.03113	313.03113	28.85	<.0001
Error	42	455.68043	10.84953		
Corrected Total	43	768.71155			

Root MSE	3.29386	R-Square	0.4072
Dependent Mean	19.65318	Adj R-Sq	0.3931
Coeff Var	16.75995		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	16.98591	0.70225	24.19	<.0001
smoke	1	5.33455	0.99314	5.37	<.0001

The coefficient of the intercept 16.9859 is the mean of the first group ($smoke=0$). The group coded as zero is the baseline of comparison. The coefficient of `smoke` is the mean difference between two groups ($5.33455=22.3205-16.9859$). The average death rate of heavy cigarette consuming states is about 5.3346 higher than that of the light counterparts.

The standard error of the coefficient of `smoke` is the denominator of the independent sample t-

test, $.9931 = s_{pool} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 3.2939 \sqrt{\frac{1}{22} + \frac{1}{22}}$, where the pooled variance estimate is

$10.8497 = 3.2939^2$ (see 4.3). Thus, the t 5.37 for the dummy coefficient in OLS is identical to the t statistic for the independent samples with equal variances.

The following is an example of the Stata `.regress` command. A dependent variable precedes a list of independent variables. `_cons` in the output means the intercept term.

```
. regress lung smoke
```

Source	SS	df	MS	Number of obs = 44		
Model	313.031127	1	313.031127	F(1, 42) =	28.85	
Residual	455.680427	42	10.849534	Prob > F =	0.0000	
Total	768.711555	43	17.8770129	R-squared =	0.4072	
				Adj R-squared =	0.3931	
				Root MSE =	3.2939	

lung	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
smoke	5.334545	.9931371	5.37	0.000	3.330314	7.338777
_cons	16.98591	.702254	24.19	0.000	15.5687	18.40311

The SPSS REGRESSION command looks complicated compared to the SAS REG procedure and Stata `.regress` command. You need to specify a dependent variable in the `/DEPENDENT` subcommand and a list of independent variables in `/METHOD`.

```

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT lung
/METHOD=ENTER smoke.

```

Table 5 compares three methods when comparing two independent samples. Although they presents different statistics but reach the same conclusion. ANOVA, GLM, and OLS report the same $F(1, 42)$ of 28.85, which is equivalent to $t(42)$ of -5.3714 since $28.85 = (-5.3714)^2$. In OLS, the intercept is sample mean of the baseline group coded as zero, while the dummy coefficient is the mean difference of two samples. Hence, the t-test, ANOVA, and OLS are the same test when comparing the means of two independent samples. However, the t-test is recommended for comparing group means to take advantage of simple computation and easy interpretation.

Table 5. Comparison of the T-test, One-way ANOVA/GLM, and Linear Regression (OLS)

	T-Test	ANOVA/GLM	OLS*
Effect Size	Mean difference		b (dummy coeff.)
Denominator of T Statistic	$s_{pooled} \sqrt{1/n_1 + 1/n_2}$		Standard error of b
Variance Partition		ANOVA Table	ANOVA Table
Test Statistic	T	$F = T^2 = (T_b)^2$	T_b
P-value	p-value	p-value	p-value
Degree of Freedom (of Error)	$n_1 + n_2 - 2$	$n_1 + n_2 - 2$	$n_1 + n_2 - 2$
Baseline for Comparison			a (coded as 0)

* The a and b of OLS are the intercept and the dummy coefficient, respectively.

7. Comparing the Proportions of Binary Variables

When a left-hand side variable to be compared is binary, the mean of the binary variable is the same as the proportion or the percentage of success. The data generation process (DGP) looks like the binomial distribution with two outcomes (i.e., 0 and 1) and an independent probability p of success.

When $np \geq 5$ and $n(1-p) \geq 5$ in general, the binomial distribution is approximated to the normal probability distribution. Therefore, we can compare proportions using the properties of these probability distributions. However, if a probability is extremely small or large, the z test becomes less accurate (Hildebrand et al. 2005: 332, 388).

7.1 Comparing a Proportion with a Hypothesized Proportion

Suppose n measurements of a binary variable y were taken from the binomial distribution with a probability of success p . If you want to compare the proportion with the hypothesized proportion, compute the z score for the standard normal distribution as follows (Hildebrand et al. 2005; Stata 2007; Bluman 2008).

$$z = \frac{y - \mu}{\sigma} = \frac{n\hat{p} - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

where \hat{p} is the proportion of success in a sample, p_0 is a hypothesized proportion, and n is the number of observations. The null hypothesis is that the population proportion and the hypothesized proportion are equal, $H_0 : p = p_0$.

The $(1-\alpha)100$ percent confidence interval of p_0 is based on the sample proportion \hat{p} .

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $z_{\alpha/2}$ is the critical value of the two-tailed test at the α significance level.

Let us compare the proportions of y_1 and y_2 using SAS and Stata. Their proportions (means) are .6667 (=20/30) and .3333 (10/30), respectively.

. sum

Variable	Obs	Mean	Std. Dev.	Min	Max
y1	30	.6666667	.4794633	0	1
y2	30	.3333333	.4794633	0	1

We are interested in whether the population proportion of y_1 is .5. The test statistic z is $1.8257 = (2/3 - .5) / \sqrt{.5 * .5 / 30}$ and the corresponding p -value is .0679.²⁰ Therefore, the null hypothesis is rejected at the .05 level.

²⁰ In Stata, run `.di (1-normal(1.8257419))*2` to get the two-tailed p -value.

In Stata, use the `.prtest` command followed by a binary variable and the hypothesized proportion separated by an equal sign.

```
. prtest y1=.5
```

```
One-sample test of proportion                y1: Number of obs =      30
-----
Variable |      Mean   Std. Err.      [95% Conf. Interval]
-----+-----
      y1 |   .6666667   .0860663      .4979798      .8353535
-----+-----
      p = proportion(y1)                    z =      1.8257
Ho: p = 0.5

      Ha: p < 0.5                Ha: p != 0.5                Ha: p > 0.5
Pr(Z < z) = 0.9661            Pr(|Z| > |z|) = 0.0679            Pr(Z > z) = 0.0339
```

The z score of 1.8257 above is not large sufficiently to reject the null hypothesis ($p < .0679$). The population proportion of y_1 appears to be .5. The 95 percent confidence interval is $.6667 \pm 1.96 * \text{sqrt}((.6667*(1-.6667))/30)$.

If you have aggregated information only, try to use the `.prtesti` command, which is the immediate form of `.prtest`. This command is followed by the number of observations, the proportion of a sample, and hypothesized proportion.

```
. prtesti 30 .3333 .5
(output is skipped)
```

In SAS, you need to take the point-and-click approach to compare proportions since SAS does not have a procedure to do this task in an easy manner. Click Solution→Analyst→Statistics→Hypothesis Tests→One-Sample Test for a Proportion. You are asked to choose the category of success or the level of interest (0 or 1).

One Sample Test of a Proportion

Sample Statistics

y1	Frequency
0	10
1	20
Total	30

Hypothesis Test

Null Hypothesis: Proportion = 0.5
 Alternative: Proportion ^= 0.5

y1	Proportion	Z Statistic	Pr > Z
1	0.6667	1.83	0.0679

In SAS and Stata, the test is based on the large-sample theory. If you have a small sample, you need to conduct the binomial probability test using the `.bitest` (or `.bitesti`) command in Stata (Stata 2007). The p-value .0987 below is slightly larger than .0679 above.

```
. bitest y1=.5
(output is skipped)

. bitesti 30 .3333 .5
```

N	Observed k	Expected k	Assumed p	Observed p
30	10	15	0.50000	0.33333

```
Pr(k >= 10)          = 0.978613 (one-sided test)
Pr(k <= 10)          = 0.049369 (one-sided test)
Pr(k <= 10 or k >= 20) = 0.098737 (two-sided test)
```

7.2 Comparing Two Proportions

If you wish to compare two proportions, apply the following formula.²¹ The pooled (weighted) proportion is used under the null hypothesis of the equal proportion, $p_1 = p_2$.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{pooled}(1 - \hat{p}_{pooled})(1/n_1 + 1/n_2)}}$$

where $\hat{p}_{pooled} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{y_1 + y_2}{n_1 + n_2}$, \hat{p}_1 and \hat{p}_2 are the proportions of two samples, y_1 and y_2 are the numbers of success of two samples, and n_1 and n_2 are the numbers of observations. The $(1-\alpha)100$ percent confidence interval of $p_1 - p_2$ is based on the sample proportions \hat{p}_1 and \hat{p}_2 .

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

In Stata, the `.prtest` command enables you to use both types of data arrangement illustrated in Figure 3. If you have a data set arranged in the second type, list the two variables separated by an equal sign.

```
. prtest y1=y2
```

Two-sample test of proportion

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
y1	.6666667	.0860663			.4979798 .8353535
y2	.3333333	.0860663			.1646465 .5020202
diff	.3333333	.1217161			.0947741 .5718926
	under Ho:	.1290994	2.58	0.010	

```
diff = prop(y1) - prop(y2)          z = 2.5820
Ho: diff = 0
```

²¹ For other types of formulae, see <http://www.tufts.edu/~gdallal/p.htm>

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(Z < z) = 0.9951	Pr(Z < z) = 0.0098	Pr(Z > z) = 0.0049

The pooled proportion \hat{p}_{pooled} is $.5 = (20+10)/(30+30)$. The z score is $2.5820 = (2/3-1/3) / \sqrt{.5*.5*(1/30+1/30)}$ and its p -value is $.0098$. You may reject the null hypothesis of the equal proportion at the $.05$ level; two population proportions are different. The 95 percent confidence interval is $.3333 \pm 1.96 * \sqrt{.6667*(1-.6667)/30 + .3333*(1-.3333)/30}$.

If the data set is arranged in the first type, run the following command.

```
. prtest y1, by(group)
(output is skipped)
```

Alternatively, you may use the following formula (Hildebrand et al. 2005: 386-388). Note that its denominator is used to construct the confidence interval of $p_1 - p_2$.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}}$$

This formula returns $2.7386 = (2/3-1/3) / \sqrt{.6667*(1-.6667)/30 + .3333*(1-.3333)/30}$, which is slightly larger than 2.5820 above. We can reject the null hypothesis of equal proportion ($p < .0062$).

SAS produces the same output as the above. You need to select Solution→Analyst→Statistics→Hypothesis Tests→Two-Sample Test for a Proportion.

Two Sample Test of Equality of Proportions

Sample Statistics

Value	- Frequencies of -	
	v1	v2
0	10	20
1	20	10

Hypothesis Test

Null hypothesis: Proportion of v1 - Proportion of v2 = 0
 Alternative: Proportion of v1 - Proportion of v2 ^= 0

Value	- Proportions of -			
	v1	v2	Z	Prob > Z
1	0.6667	0.3333	2.58	0.0098

If you have aggregated information only, use the `.prtesti` command with the number observations and the proportion of success of two samples consecutively.

```
. prtesti 30 .6667 30 .3333
(output is skipped)
```

7.3 Comparing Means versus Comparing Proportions

Now, you may ask yourself: “What if I conduct the t-test to compare means of two binary variables?” or “What is the advantage of comparing proportions over comparing means (t-test)?” The simple answer is no big difference in case of a large sample size. Only difference between comparing means and proportions comes from the computation of denominators in the formula. The difference becomes smaller as sample size increases. If N is sufficiently large, the t probability distribution and the binomial distribution are approximated to the normal distribution.

Let us perform the independent sample t-test on the same data and check the difference. The unpaired option indicates that the two samples are not paired but independent of each other.

```
. ttest y1=y2, unpaired
```

```
Two-sample t test with equal variances
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
y1	30	.6666667	.0875376	.4794633	.4876321	.8457012
y2	30	.3333333	.0875376	.4794633	.1542988	.5123679
combined	60	.5	.0650945	.5042195	.3697463	.6302537
diff		.3333333	.1237969		.0855269	.5811397
diff = mean(y1) - mean(y2)					t =	2.6926
Ho: diff = 0					degrees of freedom =	58
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0		
Pr(T < t) = 0.9954		Pr(T > t) = 0.0093		Pr(T > t) = 0.0046		

The t of 2.6926 is similar to the z score of 2.5820. Their p -values are respectively .0093 and .0098; the null hypothesis is rejected in both tests.

Table 6 suggests that the difference between comparing means (t-test) and proportions (z-test) become negligible as N becomes larger. The random variable a was drawn from $\text{RAND}('BERNOULLI', .50)$ in SAS, which is the random number generator for the Bernoulli distribution with a probability of .50. Similarly, the variable b is generated from $\text{RAND}('BERNOULLI', .55)$. Roughly speaking, the p -values of t and z become almost same if sample size exceeds 30.

Table 6. Comparing Means and Proportions with Different Sample Sizes

	N=10	N=20	N=30	N=50	N=100	N=200	N=500
$a: n_a(p_a)$	5 (.40)	10 (.40)	15 (.33)	25 (.40)	50 (.44)	100(.46)	250(.47)
$b: n_b(p_b)$	5 (1.00)	10 (.90)	15 (.73)	25 (.68)	50 (.70)	100(.62)	250(.60)
Means (t)	-2.4495 (.0400)	-2.6112 (.0177)	-2.3155 (.0281)	-2.0278 (.0482)	-2.6940 (.0083)	-2.2883 (.0232)	-2.8877 (.0040)
Proportions (z)	-2.0702 (.0384)	-2.3440 (.0191)	-2.1958 (.0281)	-1.9863 (.0470)	-2.6259 (.0086)	-2.2700 (.0232)	-2.8696 (.0041)

* P-values in parentheses.

8. Conclusion

The t-tests is a basic statistical method for examining the mean difference between two groups. The one-way ANOVA can compare means of more than two groups. T-tests also compare the proportions of binary variables when sample size is large. Whether data are balanced or unbalanced does not matter in t-tests and one-way ANOVA. The one-way ANOVA, GLM, and linear regression model all use the variance-covariance structure in their analysis, but present equivalent results in different ways.

Here are key checklists for researchers who want to conduct t-tests. First, a variable to be tested should be interval or ratio-scaled so that its mean is substantively meaningful. Do not, for example, compare the means of skin colors (white=0, yellow=1, black=2) of children between two cities. In case of binary variables, the t-test compares the proportions of success of the variables. If you have a latent variable measured by several Likert-scaled manifest variables, first run a factor analysis to construct that latent variable before t-test.

Second, the data generation process (sampling and data collection) should be carefully explored to ensure that the samples were randomly drawn. If each observation is not independent of other observations and selection bias is involved in the sampling process, sophisticated methods need to be employed to deal with the non-randomness. In case of self-selection, for example, the propensity score matching appears to be a good candidate.

Researchers should also examine the normality assumption especially when N is small. It is awkward to compare the means of random variables that are not normally distributed. If N is not large and normality is questionable, conduct the Shapiro-Wilk W, Shapiro-Francia W, Kolmogorov-Smirnov D, or Jarque-Bera test. If the normality assumption is violated, try nonparametric methods such as the Kolmogorov-Smirnov Test and Wilcoxon Rank-Sum Test.

Table 7. Comparison of T-test Features of Stata, SAS and SPSS

	Stata 10	SAS 9	SPSS 15
Test for equal variances	Bartlett's chi-squared (.oneway)	Folded form F (TTEST)	Levene's weighted F (T-TEST)
Comparing means (T-test)	.ttest; .ttesti	TTEST	T-TEST
Comparing proportions	.prtest; .prtesti	(point-and-click)	
Approximation of the degrees of freedom (DF)	Satterthwaite's DF Welch's DF	Satterthwaite's DF Cochran-Cox's DF	Satterthwaite's DF
Hypothesized value anything other than 0 in H_0	Only in one sample t-test	H0	Only in one sample t-test
Data arrangement for the independent samples	Type I and II	Type I	Type I
Aggregated Data	.ttesti; .prtesti	FREQ statement	N/A

There are four types of t-tests. If you have a variable whose mean is compared to a hypothesized value, this is the case of the one sample t-test. If you have two variables and they are paired, that is, if each element of one sample is linked to its corresponding element of the other sample, conduct the paired t-test. You are checking if differences of individual pairs have

mean 0 (no effect). If two independent variables are compared, run the independent sample t-test.

When comparing two independent samples, first check if the variances of two variables are equal. Try to conduct the folded form F test. If their variances are not significantly different, you may use the pooled variance to compute standard error. If the equal variance assumption is violated, use individual variances and approximate degrees of freedom.

If you need to compare means of more than two groups, conduct the one-way ANOVA. See Figure 3 for summary of t-tests and one-way ANOVA.

Next, consider types of t-tests, data arrangement, and software issues to determine the best strategy for data analysis (Table 7). The first data arrangement in Figure 4 is commonly used for the independent sample t-test, whereas the second arrangement is appropriate for the paired sample t-test. If independent samples are arranged in the second type in SAS and SPSS, you should reshape data into the first type.

Finally, report and interpret results clearly. You need to report mean difference (or group means), standard error, degrees of freedom, and N. Table 8 is a simple example of reporting t-test results. And then interpret results substantively rather than simply report numbers. For instance, “The average death rate from lung cancer of heavy cigarette consuming states is on average 5.335 higher than that of other states.”

Table 8. Reporting T-Test Results

	Pre versus Post1	Lung Cancer/Smoking	Kidney Cancer/Region
Mean difference	-20.39**	-5.335**	.3877**
(standard error)	(4.9475)	(.9931)	(.1472)
Degrees of freedom	35	42	35.1

Statistical significance: * <.05 and ** <.01

SAS has several procedures (e.g., TTEST, MEANS, and UNIVARIATE) and useful options. For example, the H0 option allows you to specify a hypothesized value other than zero. The Stata `.ttest` and `.ttesti` commands provide very flexible ways of handling different data arrangements and aggregated data. Table 9 summarizes options used in these commands.

Table 9. Options of the .ttest Command and its Immediate Form of .ttesti

	Usage	by()	unequal	welch	unpaired
One sample	var=c				
Paired (dependent) sample	var1=var2				
Equal variance (1 variable)	var	X			
Equal variance (2 variables)	var1=var2				X
Unequal variance (1 variable)	var	X	X	X	
Unequal variance (2 variables)	var1=var2		X	X	X

* The `.ttesti` command does not allow the `unpaired` option.

** The “var1=var2” assumes the second data arrangement in Figure 4.

Appendix: Data Set

The data set used here is adopted from J. F. Fraumeni's (1968) study on cigarette smoking and cancer. The data are per capita numbers of cigarettes sold (smoked) by 43 states and the District of Columbia in 1960 together with death rates per hundred thousand people from various forms of cancers. Two variables were added to categorize states into two groups.

Fraumeni, J. F. 1968. "Cigarette Smoking and Cancers of the Urinary Tract: Geographic Variations in the United States," *Journal of the National Cancer Institute*, 41(5): 1205-1211. Downloaded from <http://lib.stat.cmu.edu>.

cigar = number of cigarettes smoked (hds per capita)

bladder = deaths per 100k people from bladder cancer

lung = deaths per 100k people from lung cancer

kidney = deaths per 100k people from kidney cancer

leukemia = deaths per 100k people from leukemia

smoke = 1 for those whose cigarette consumption is larger than the median and 0 otherwise.

west = 1 for states in the South or West and 0 for those in the North, East or Midwest.

The followings are summary statistics and normality tests of these variables.

. sum cigar-leukemia

Variable	Obs	Mean	Std. Dev.	Min	Max
cigar	44	24.91409	5.573286	14	42.4
bladder	44	4.121136	.9649249	2.86	6.54
lung	44	19.65318	4.228122	12.01	27.27
kidney	44	2.794545	.5190799	1.59	4.32
leukemia	44	6.829773	.6382589	4.9	8.28

. sfrancia cigar-leukemia

Variable	Shapiro-Francia W' test for normal data				
	Obs	W'	V'	z	Prob>z
cigar	44	0.93061	3.258	2.203	0.01381
bladder	44	0.94512	2.577	1.776	0.03789
lung	44	0.97809	1.029	0.055	0.47823
kidney	44	0.97732	1.065	0.120	0.45217
leukemia	44	0.97269	1.282	0.474	0.31759

. tab west smoke

west	smoke		Total
	0	1	
0	7	13	20
1	15	9	24
Total	22	22	44

References

- Bluman, Allan G. 2008. *Elementary Statistics: A Step by Step Approach, A Brief Version*, 4th ed. New York: McGraw Hill.
- Cochran, William G., and Gertrude M. Cox. 1992. *Experimental Designs*, 2nd ed. New York: John Wiley & Sons.
- Fraumeni, J. F. 1968. "Cigarette Smoking and Cancers of the Urinary Tract: Geographic Variations in the United States," *Journal of the National Cancer Institute*, 41(5): 1205-1211.
- Gosset William S. 1909. "The Probable Error of a Mean," *Biometrika*, 6(1):1-25.
- Hildebrand, David K., R. Lyman Ott, and J. Brian Gray. 2005. *Basic Statistical Ideas for Managers*, 2nd ed. Belmont, CA: Thomson Brooks/Cole.
- SAS Institute. 2005. *SAS/STAT User's Guide, Version 9.1*. Cary, NC: SAS Institute.
- Satterthwaite, F.W. 1946. "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2:110-114.
- SPSS. 2007. *SPSS 15.0 Syntax Reference Guide*. Chicago, IL: SPSS Inc.
- Stata Press. 2007. *Stata Reference Manual Release 10*. College Station, TX: Stata Press.
- Walker, Glenn A. 2002. *Common Statistical Methods for Clinical Research with SAS Examples*. Cary, NC: SAS Institute.
- Welch, B. L. 1947. "The Generalization of Student's Problem When Several Different Population Variances Are Involved," *Biometrika*, 34: 28-35.

Acknowledgements

I am grateful to Jeremy Albright, Takuya Noguchi, Kevin Wilhite, and Kaigang Li at the UITS Center for Statistical and Mathematical Computing, Indiana University, who provided valuable comments and suggestions.

Revision History

- 2003. First draft
- 2004. Second revision (nonparametric methods excluded).
- 2005. Third revision (data arrangements and conclusion added).
- 2007. Fourth revision (comparing proportions added).
- 2008. Fifth revision (introduction and conclusion revised; SPSS output added).