Hypothesis Testing

Often, we need to choose between competing views or hypotheses concerning a population parameter(s). We usu identify the status quo as the **null hypothesis**, H_0, and the competing hypothesis, H_a, as the **alternative hypothesis**. Once we have our hypotheses, we collect and analyze sample data to determine whether it is "sufficiently" (we will be more precise on this notion soon!) consistent with the null hypothesis. If it is, we fail to reject the null hypothesis. Otherwise, we reject the null hypothesis. As we will see, the null hypothesis (like a criminal defendant assumed to be innocent until proven guilty in the US) requires lots of proof before it is rejected.

Example 1: Does a New Drug Improve Cancer Survival Rates?

Assume that the current drug used to cure pancreatic cancer results in a 10%, 5 year survival rate. Let p = fraction of pancreatic cancer patients taking a new drug that survive 5 years. Then our hypotheses are

H_0: p <=.10 H_a: p >.10.

Therefore, the new drug is assumed to not be an improvement unless we receive strong evidence to support the view that the drug improves survival.

To determine whether we should accept or reject or accept the null hypothesis we would take a sample of patients given the new drug and look at phat = fraction of patients in sample that survive at least 5 years. If phat<=.10 it is clear we should accept the null hypothesis, but what if phat = .12 or phat = .15?

Note: In this example our alternative hypothesis specifies that the population parameter is greater than the values specified in the null hypothesis. Such an alternative hypothesis is called an upper one-sided alternative hypothesis.

Example 2: Is a congressional district poorer than average?

The average US family income in 2015 was $79,263. You are interested in knowing whether your congressional district has a lower average income than the US overall. Define μ = average family income in your congressional district. Then, our hypotheses are:

H_0: μ = $79,263 or μ >= $79,263; H_a: μ < $79,263.

In this example, our alternative hypothesis specifies that the population parameter is smaller than the values specified in the null hypothesis. Such an alternative hypothesis is called a lower one-sided alternative hypothesis.

Our null hypothesis is that our district is not different than the rest of the US.

We would now take a simple random sample of families in our district, and calculate the sample mean $\bar{x}$. If $\bar{x}$ = $80,000, should we fail to reject the null hypothesis? Again, it may seem like we should, but we're dealing with sample data, which contains error. Futher, if $\bar{x} = \$75{,}000$ or $\bar{x} =$ $72,000, should we reject the null hypothesis or fail to reject it?

Example 3: Do stock and bond annual returns have equal volatility?

Often, we want to know if it is reasonable to assume that two populations have equal variance. When looking at annual investment returns, the standard deviation of annual percentage returns is referred to as volatility. In this situation, our hypotheses are:

H_0: Annual Variance Stock Returns = Annual Variance on Bond Returns.
H_a: Annual Variance Stock Returns ≠ Annual Variance on Bond Returns.

In this example, our alternative hypothesis does not specify a particular direction for the deviation of variances from equality. Therefore, the alternative hypothesis is called a two-sided alternative hypothesis.

We could now look at, say, the last 10 years of annual returns on stocks and bonds. If the sample variance of the annual percentage returns on stocks and bonds are relatively close, we would fail to reject H_0, but if the sample variance of the annual percentage returns on stocks and bonds differ greatly, we would reject the null hypothesis.

Should I Use a One-Tailed or Two-Tailed Test?

Some statisticians believe you should always use a two-tailed test because a priori you have no idea of the direction in which deviations from the null hypothesis will occur. Most statisticians feel that if a deviation from the null hypothesis in either direction is of interest, then a two-tailed alternative hypothesis should be used, while if a deviation from the null hypothesis is of interest in only one direction, then a one-tailed alternative hypothesis should be used. It is much easier to reject the null hypothesis with a one-tailed test than it is with a two-tailed test.

Types of Errors in Hypothesis Testing

To determine whether to reject or fail to reject the null hypothesis, we look at a sample statistic. We determine a set of values for the sample statistic (called the **critical region**) that result in the rejection of the null hypothesis. There are two types of errors that can be made in hypothesis testing:

Type I Error: Reject H_0 BUT H_0 true. We let α = probability of making a Type I Error. α is often called the **level of significance** of the test.

Type II Error: Fail to reject H_0 BUT H_0 not true. We define β = Probability of making a Type II Error.

In US criminal trials, the defendant is innocent until proven guilty. In this situation if we define H_0: defendant innocent and H_a: defendant guilty, then a Type I error corresponds to convicting an innocent defendant while a Type II error corresponds to allowing a guilty person to go free. Since a 12-0 vote is needed for conviction, it is clear that the US judicial system considers a Type I Error to be costlier than a Type II Error.

In a similar fashion, our approach to hypothesis testing will be to set a small probability α (usually 0.05) of making a Type I Error, and then choose a critical region that minimizes the probability of making a Type II Error.

Type I and Type II Error for Example 1

In Example 1 a Type 1 error results when we reject $p<=.10$ when in reality $p<=.10$. This corresponds to the risk of concluding the new drug is an improvement when it is not.

A Type 2 error results when we accept $p<=.10$ when actually, $p>.10$. This corresponds to the risk of concluding the new drug is not more effective when the drug is actually more effective than the old drug.

Using the One Sample Z Test to test Hypotheses about Population Mean when n>=30 or Population is Normal and σ is known

If the sample size n is >=30, then according to the Central Limit Theorem, $\bar{x}$ will follow a normal random variable, even if the population is non-normal. In this situation, we will assume that the sample standard deviation, s, closely approximates the population standard deviation, σ. Then, the following table summarizes the **critical region** for upper one-tailed, lower one-tailed, and two-tailed hypotheses concerning μ. Because the critical regions are based on the standard normal, the tests are referred to as the **One Sample Z Tests.**

Critical Regions for One Sample Z-Test

Hypotheses	Critical Region
H_0: $\mu <= \mu_0$ H_a: $\mu > \mu_0$	$\bar{x} >= \mu_0 + \lvert z_\alpha \rvert \sigma/\sqrt{n}$
H_0: $\mu >= \mu_0$ H_a: $\mu < \mu_0$	$\bar{x} <= \mu_0 + z_\alpha \sigma/\sqrt{n}$
H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$	$\lvert \bar{x} - \mu_0 \rvert >= \lvert z_{\alpha/2} \rvert \sigma/\sqrt{n}$

In the most common case where α = 0.05, $z_{.05}$ = -1.645 and $z_{.025}$ = -1.96, and the previous formulas for the critical regions become:

Hypotheses	Critical Region
H_0: $\mu <= \mu_0$ H_a: $\mu > \mu_0$	$\bar{x} >= \mu_0 + 1.645\sigma/\sqrt{n}$
H_0: $\mu >= \mu_0$ H_a: $\mu < \mu_0$	$\bar{x} <= \mu_0 - 1.645\sigma/\sqrt{n}$
H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$	$\lvert \bar{x} - \mu_0 \rvert >= 1.96\sigma/\sqrt{n}$

Z-Test Example

Passing the HISTEP test is required for graduation in the state of Fredonia. The average state score on the test is 75. A random sample of 49 students at Carver High School have $\bar{x} = 79$ and s = 15. For α = 0.05, would you conclude that Cooley High Students perform differently than the typical state student?

There is no reason to believe that Cooley High is better or worse than the state, so we will use a two-tailed test:

H_0: μ = 75; H_a: μ ≠ 75.
$|79-75| >= 1.96*15/\sqrt{49}$.
$|79-75| = 4$.
$1.96*15/\sqrt{49} = 4.2$.

In other words, to be significant at α = 0.05, the difference must be greater than 4.2. It is not; thus we fail to reject the null hypothesis, and conclude that the average Cooley High School score does not differ from the state average.

Now, suppose we have invested resources to improve test scores at Cooley High School. Then, we might be interested in seeing if Cooley High School students performed better than the state. In this case, we would want to conduct an upper one-sided alternative hypothesis test:

H_0: μ = 75; H_a: μ > 75.
$79 >= 75 + (1.645*15)/\sqrt{49} = 78.525$
In other words, to be significant at α = 0.05, the result of the right part of the equation must be less than or equal to 79. In this case, we reject the null hypothesis, and conclude that our efforts have resulted in significant improvement.

The astute reader should realize that for a .05 level of significance our data resulted in rejection of H_o for a one-tailed test and failure to reject the H_o for a two-tailed test. This example illustrates that for the same level of significance can result in different outcomes based on the type of hypothesis test being evaluated. This is why many statisticians always recommend a two-tailed alternative, because you have made a stronger case for rejecting H_0.

Probability Values (P-Values) and Hypothesis Testing

The level of significance chosen is rather arbitrary. For that reason, most statisticians use the concept of probability values (**p-values)** to report the outcome of a hypothesis test. The p-value for a hypothesis test is the smallest value of α for which the data indicates rejection of H_0. Thus, to reject the null hypothesis the p-value must be <= α. If it is > α, we fail to reject the H_0.

The p-value may also be interpreted as the probability of observing a value of the test statistic at least as extreme as the observed value of the test statistic if H_0 is true. In other words, it's the probability of incorrectly rejecting the null hypothesis.

If we let $\boldsymbol{\bar{X}}$ represent the random variable for the sample mean under H_0 and x be the observed value of $\bar{x}$, then the p-value for the one sample Z-test is computed as follows:

P-Values for One Sample Z-Test

Hypotheses	P-Value
H_0: $\mu <= \mu_0$ H_a: $\mu > \mu_0$	Prob($\overline{\boldsymbol{X}}$ >= x)
H_0: $\mu >= \mu_0$ H_a: $\mu < \mu_0$	Prob($\overline{\boldsymbol{X}}$ <= x)
H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$	Prob($\lvert\overline{\boldsymbol{X}} - \mu_0\rvert \geq x$)

All probabilities are computed under the assumption that H_0 is true.

In our Cooley High School example, the p-value for the two-tailed test is:

2*Prob(|$\overline{\boldsymbol{X}}$-75|>=4) = 2*Prob($\overline{\boldsymbol{X}}$>=79),
which can be computed as the following in Excel:
2*(1-NORM.DIST(79,75,15/$\sqrt{49}$), True) = 2*(0.030974) = 0.061948.

Because our p-value of 0.06 > our alpha value of .05, we fail to reject H_0.

The p-value for a one-tailed test is:

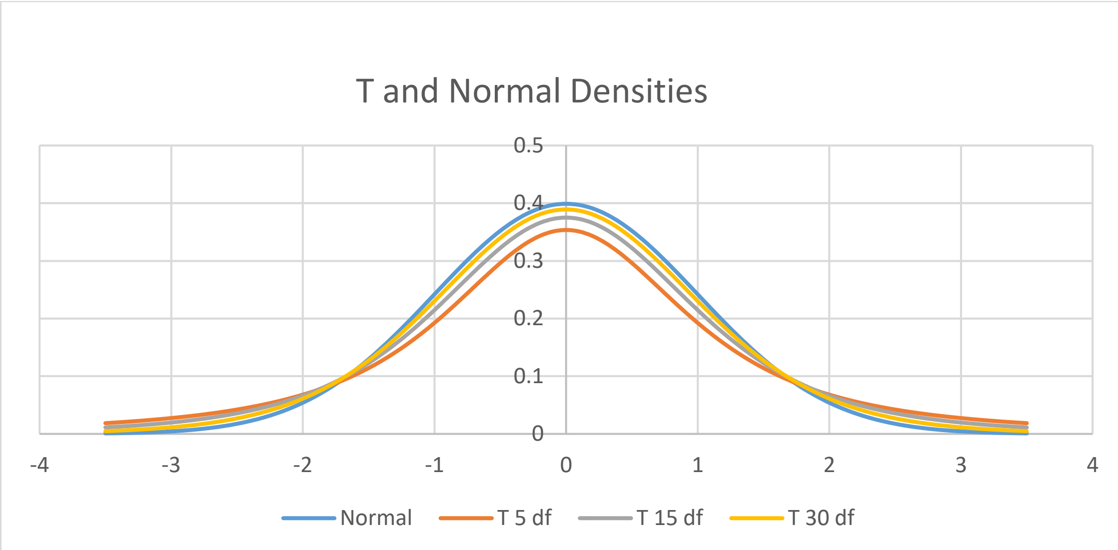
Prob($\overline{\boldsymbol{X}}$>=79) = 0.030974.

Because our p-value of 0.03 < than our alpha value of .05, we reject H_0.

One Sample Hypothesis for Mean: Small Sample, Normal Population, Variance Unknown

Suppose we are interested in testing a hypothesis about the mean of a normal population where the population variance is unknown, and the sample size n is <30. Then, it can be shown that ($\bar{x}$ - μ)/(s/√n) follows a t-distribution with n-1 degrees of freedom. Here s = sample standard deviation.

Like the standard normal distribution, the t-distribution has a density symmetric around 0. As shown below, the t-distribution has fatter tails than the standard normal density, and as n increases the t-distribution approaches the standard normal density.



As shown in the **Hypothesis Testing.xlsx** spreadsheet, T random variable worksheet shows the percentiles of the t-distribution and how they can be computed using the T.INV function. We let $t_{(\alpha,n-1)}$ represent the α percentile of a t-distribution with n-1 degrees of freedom. Below we find, for example, $t_{(.025,28)}$ = -2.04841.

	D	E	F	G	H
6	**2.5 %ile 28 df**	**-2.04841**	**=T.INV(0.025,28)**		
7	**97.5%ile 28 df**	**2.048407**	**=T.INV(0.975,28)**		
8	**0.5% ile 13 df**	**-3.01228**	**=T.INV(0.005,13)**		
9	**99.5%ile 13 df**	**3.012276**	**=T.INV(0.995,13)**		

Basically, one sample t-tests look just like one sample z-tests with s replacing σ and the t percentiles replacing the z percentiles.

Critical Region for One Sample t-tests

Hypotheses	Critical Region
H_0: $\mu <= \mu_0$ H_a: $\mu > \mu_0$	$\bar{x} >= \mu_0 + \lvert t_{(\alpha,n-1)} \rvert s/\sqrt{n}$
H_0: $\mu >= \mu_0$ H_a: $\mu < \mu_0$	$\bar{x} <= \mu_0 + t_{(\alpha,n-1)} s/\sqrt{n}$
H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$	$\lvert \bar{x} - \mu_0 \rvert >= \lvert t_{(\alpha/2,n-1} \rvert s/\sqrt{n}$

If we let $\mathbf{T_{n-1}}$ stand for a t-distribution with n-1 df and t represent the observed value of $(\bar{x} - \mu)/(s/\sqrt{n})$, then the p-values for a one sample t-test may be computed as follows:

P-Values for One Sample t-Test

Hypotheses	P-Value
H_0: $\mu <= \mu_0$ H_a: $\mu > \mu_0$	Prob($\mathbf{T}_{n-1} >= t$)
H_0: $\mu >= \mu_0$ H_a: $\mu < \mu_0$	Prob($\mathbf{T}_{n-1} <= t$)
H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$	2*Prob($\mathbf{T}_{n-1} >= \|t\|$)

As shown below, in a fashion similar to the NORM.DIST function, the T.DIST function in Excel can be used to compute probabilities for the t-distribution.

	D	E	F
12	**Prob T_{10}>=2**	**0.036694**	**=1-T.DIST(2,10,TRUE)**
13	**Prob T_{10}<=-2**	**0.036694**	**=T.DIST(-2,10,TRUE)**

Example 5

Passing the HISTEP test is required for graduation in the state of Fredonia. The average state score on the test is 75. A random sample of 25 students at Carver High School has $\bar{x} = 81$ and s = 15. For α= 0.05, would you conclude that Cooley High School students perform differently than the typical state student?

We use a two-tailed test because we have no a priori view about whether Cooley High School students will perform better or worse than the typical state student.

H_0: μ= 75, H_a: $\mu \neq 75$.
Using the function T.INV(0.025,24) We find $t_{(..025,24)}$ = -2.06.
We reject H_0 if |81-75| >= 2.06*15/$\sqrt{25}$ = 6.18.
Because 6 is not >= 6.18, we fail to reject H_0.

The p-value for this test is 2*Prob($\mathbf{T}_{24}$ >= (81-75)/(15/$\sqrt{25}$))
= 2*Prob($\mathbf{T}_{14}$ >= 2).
Prob($\mathbf{T}_{14}$ >= 2) may be computed with the formula = 1-T.DIST(2,14,TRUE) in Excel, which returns 0.033. Therefore, the p-value for this test is 2*0.033 = 0.066. Because .066 is > 0.05, we fail to reject H_0.

Single Sample Test about Population Proportion

Often a population proportion is unknown. The "One Sample Proportion" worksheet in **Hypothesis Testing.xlsx** spreadsheet uses the BINOM.DIST.RANGE function to calculate the p-value for both one-tailed and two-tailed alternative hypotheses.

	A	B	C	D	E	F	G	H	I	J	K
1	**Testing**	**a Proportion**				**Player makes 300 of 400 Free Throws?**					
2						**Has she improved from being a 70% foul shooter?**					
3	**trials**	**400**		**Ho: p<=p0**							
4	**successes**	**300**		**Ha: p>p0**		**$H_{0:}$ p<=0.70 $H_{a:}$ P>0.70**					
5	**Pzero**	**0.7**									
6				**Righttailedpvalue**	**0.0155**	**=BINOM.DIST.RANGE(trials,Pzero,successes,trials)**					
7											
8											
9				**Ho: p>=p0**							
10				**Ha: p<p0**							
11											
12				**Lefttailedpvalue**	**0.9884**	**=BINOM.DIST.RANGE(trials,Pzero,0,successes)**					
13											
14				**Ho: p=p0**							
15				**Ha: p≠p0**							
16											
17				**Twotailedpvalue**	**0.0311**	**=2*MIN(Lefttailedpvalue,Righttailedpvalue)**					
18											
19			**Reject H0 if pvalue<=α**								

To use this worksheet, simply enter the number of trials in cell B3, the number of successes in B4, and in cell B5 enter the value of the population proportion (p0) assumed in the null hypothesis. Then, the p-value for a right-tailed test can be found in cell E6, the p-value for a left-tailed test can be found in cell E12, and the p-value for a two-tailed test can be found in cell E17.

Example 6

An NBA player has made 70% of his foul shots in the past. The owner has hired a free throw coach to improve his free throw shooting. So far this year, the player has shot 400 free throws and made 300. At the .05 level of significance, can you conclude that the coach has succeeded in improving the player's free throw shooting?

Because we are only interested in improvement we use an upper one-sided test.

H_0: p = 0.70; H_a: p > .0.70. Here p = the probability that the player makes a free throw.

If we enter these values into the "One Sample Proportion" worksheet, we see the p-value is 0.016, which is <= 0.05, so we reject the null hypothesis and conclude the coach has improved the player's free throw shooting. In this situation, the p-value is simply the probability of a result as extreme as making 300 of 400 free throws, or the probability that a 70% free throw shooter will make >= 300 free throws in 400 attempts.

Testing Hypotheses about Equality of Variances

Often, we want to test if two populations have identical variances. The "T Test Equal Variance" worksheet in the Hypothesis Testing.xlsx spreadsheet contains a template to test for α= 0.05 equality of variance between two populations. The test assumes the two populations are normal random variables.

Simply enter the sample size and sample variance for the two populations in D3:D6. D11 and D12 give a 95% confidence interval for the ratio of the population variances.

If the 95% confidence interval contains 1, you fail to reject the null hypothesis that the populations have equal variances; if the 95% confidence interval does not contain 1 reject the hypothesis of equal variances.

In this worksheet, we have the grades on the final exam of 14 statistics students who took a hybrid statistics class (mostly online) and the final exam grades of 18 students who took the same final exam but took the course in person with the same instructor as the hybrid group. In D3:D6, we entered the sample size and sample variance for each population. We are 95% sure that the ratio of the population variances is between 0.68 and 5.66. Since this interval includes 1, we conclude the variance of the scores for the two classes are identical.

	A	B	C	D
1	**TESTING IF VARIANCES OF TWO POPULATIONS ARE EQUAL**			
2				
3			**SAMPLEVAR1**	**33.91758242**
4			**SAMPLESIZE1**	**14**
5			**SAMPLEVAR2**	**18.01633987**
6			**SAMPLESIZE2**	**18**
7			**SVAR1OVERSVAR2**	**1.882601164**
8			**LOWERCI**	**0.358900271**
9			**UPPERCI**	**3.003895725**
10				
11			**LOWERLIMIT**	**0.675666068**
12			**UPPERLIMIT**	**5.65513759**
13				
14			**Variances Equal**	

Testing the Difference Between Two Population Means

There are four important hypothesis tests that can be used to evaluate the difference in two population means. Within Excel, the "Data Analysis Add-In" makes ii simple to perform each of these tests:

Situation	Name of Test
Large sample size (n >= 30) from each population and samples are independent	z-Test Two Sample for Means
Small sample size (n < 30) for at least one population, populations are normal, variances unknown but equal, and the samples are independent	t-Test Two Sample Assuming Equal Variances
Small sample size (n < 30) for at least one population, populations are normal, variances unknown but unequal, and the samples are independent	t-Test Two Sample Assuming Unequal Variances
The two populations are normal and the observations from the two populations can be paired in a natural fashion	t-Test Paired Two Sample for Means

z-test Two Samples for Means

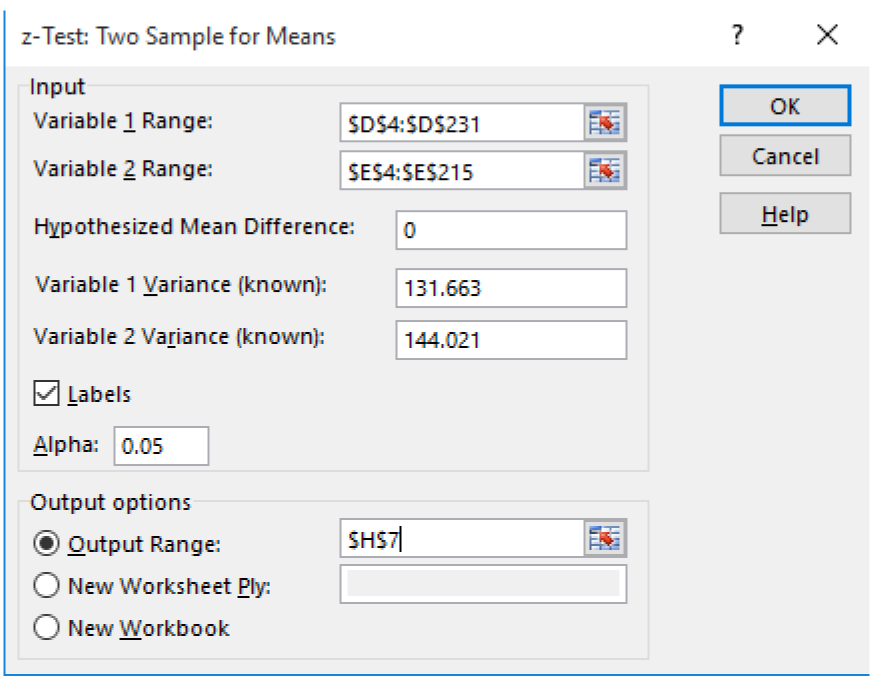
Suppose we have a large sample size of at least 30 from two populations, and the samples from the two populations are independent (that is, the values in the sample from the first population have no effect on the values in the sample from the second population). Let μ_i = unknown mean for population i. Then, the z-Test Two Sample for Means test from the Data Analysis Add-In can be used to test:

H_0: $\mu_1 = \mu_2$ against a one tailed or two-tailed alternative.

For example, in the "Two Sample Z test" worksheet of the **Hypothesis Testing.xlsx** workbook, we are given the starting salaries (in thousands of dollars) for 227 marketing and 211 finance graduates of a leading MBA program. We want to conduct a two-tailed test to determine if the average starting salaries for marketing and finance majors are equal.

	B	C	D	E	F	G	H	I	J	K
1										
2		**variance**	**131.6628591**	**144.021**						
3										
4			**Marketing**	**Finance**						
5			**118**	**105**						
6			**110**	**90**						
7			**106**	**101**			**z-Test: Two Sample for Means**			
8			**94**	**130**						
9			**91**	**124**				***Marketing***	***Finance***	
10			**102**	**104**			**Mean**	**98.64758**	**109.1896**	
11			**96**	**129**			**Known Variance**	**131.66**	**144.02**	
12			**116**	**110**			**Observations**	**227**	**211**	
13			**106**	**110**			**Hypothesized Mean Differe**	**0**		
14			**117**	**126**			**z**	**-9.38203**		
15			**90**	**116**			**P(Z<=z) one-tail**	**0**		
16			**113**	**97**			**z Critical one-tail**	**1.644854**		
17			**112**	**123**			**P(Z<=z) two-tail**	**0**		
18			**109**	**124**			**z Critical two-tail**	**1.959964**		
19			**106**	**115**						
20			**114**	**100**			**H_0: Mean Marketing=Mean Finance**			
21			**92**	**130**			**H_a: Mean Marketing ≠Mean Finance**			
22			**99**	**93**						
23			**105**	**97**			**P-Value =0 so reject null hypothesis**			
24			**81**	**93**			**and conclude significant difference**			
25			**82**	**110**			**between Average salary of marketing and finance majors**			

After selecting the Data Analysis Add-in from the Data ribbon, we select the “z-Test Two Samples for Means.” We assume α = 0.05. After computing the sample variance (with the VAR function) for each population (in cells D2 and E2) fill in the dialog box as shown below:



We find a p-value of 0, so for any alpha, we would reject the hypothesis that average salaries for marketing and finance majors are equal. Clearly, finance majors have significantly larger salaries.

t-Test Two Sample Assuming Equal Variances

Suppose we want to compare the means of two normal populations which have unknown but equal variances. If we take samples from the two populations that are independent and at least one of the sample sizes is <30, then we can use the t-Test Two Sample Assuming Equal Variances to compare the population means. First, of course, we should test the hypothesis that the two populations have Equal Variances.

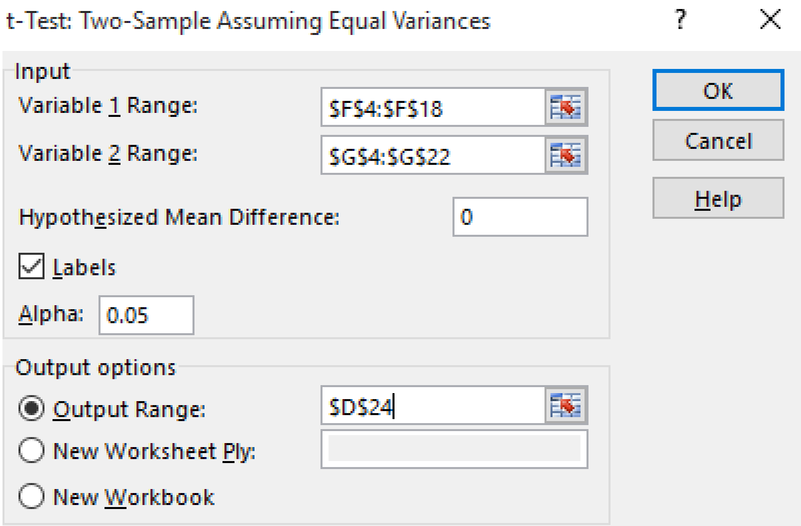
In the "Test Equal Variance" worksheet of **Hypothesis Testing.xlsx** workbook, we are given the final exam grades for 14 students who took statistics in a hybrid (mostly online) format and final exam grades of 18 students who took the course in a traditional classroom format. Does performance of the students in the two classes differ significantly? Our hypotheses are:

H_0: $\mu_{Hybrid} = \mu_{Inclass}$; H_a: $\mu_{Hybrid} \neq \mu_{Inclass}$

Our test requires both populations be normal. A quick eyeball test for normality is to compute the skewness and kurtosis of a sample. If both the skewness (computed with SKEW function) and kurtosis (computed with KURT function) of a sample are between -1 and +1 the assumption of normality is almost surely justified. From cells F2:G3 the assumption of normal populations appears reasonable.

	A	B	C	D	E	F	G
1	**TESTING IF VARIANCES OF TWO POPULATIONS ARE EQUAL**						
2					**SKEW**	**-0.50544**	**0.144793**
3			**SAMPLEVAR1**	**33.91758242**	**KURT**	**0.103766**	**-0.61404**
4			**SAMPLESIZE1**	**14**		Hybrid	In Person
5			**SAMPLEVAR2**	**18.01633987**		87	88
6			**SAMPLESIZE2**	**18**		94	96
7			**SVAR1OVERSVAR2**	**1.882601164**		86	84
8			**LOWERCI**	**0.358900271**		89	82
9			**UPPERCI**	**3.003895725**		74	81
10						84	85
11			**LOWERLIMIT**	**0.675666068**		85	90
12			**UPPERLIMIT**	**5.65513759**		85	90
13						92	89
14			**Variances Equal**			90	95
15						77	88
16			**Test H_0: MeanHybrid=Mean In Person**			82	89
17			**Test Ha: MeanHybrid≠Mean In Person**			94	93
18						84	85
19			**Test H_0: VarianceHybrid=Variance In Person**				87
20			**Test Ha: VarianceHybrid≠Variance In Person**				83
21			**Accept H_0**				88
22							84
23			**Now test mean difference using equal variance t test**				
24				t-Test: Two-Sample Assuming Equal Variances			

In cells B3 and B5, we use the VAR function to determine the sample variance for each data set. After entering the sample sizes in B4 and B6, we find from D11 and D12 that we are 95% sure the ratio of the population variances is between .68 and 5.65. This interval includes one, so the assumption of equal variances is justified. Then, from the Data ribbon, we choose Data Analysis and select the “t-Test Two Sample Assuming Equal Variances” and fill in the dialog box as follows:



The results of the hypothesis test are:

	D	E	F	G
24	t-Test: Two-Sample Assuming Equal Variances			
25				
26		*Hybrid*	*In Person*	
27	Mean	85.92857	87.61111	
28	Variance	33.91758	18.01634	
29	Observations	14	18	
30	Pooled Variance	24.90688		
31	Hypothesized Mea	0		
32	df	30		
33	t Stat	-0.94609		
34	P(T<=t) one-tail	0.175831		
35	t Critical one-tail	1.697261		
36	P(T<=t) two-tail	0.351663		
37	t Critical two-tail	2.042272		

The p-value for a two-tailed test is .35. Because 0.35 >= .0.05, we fail to reject the null hypothesis that the mean performance of students is equivalent regardless of class delivery.

t-Test Two Sample Assuming Unequal Variances

Suppose we want to compare the means of two normal populations which have unknown but unequal variances. If we take samples from the two populations that are independent and at least one of the sample sizes is < 30, then we might need to use the t-Test Two Sample

Assuming Unequal Variances to compare the population means. First, of course, we should test the hypothesis that the two populations have unequal variances.

In the "T Test Unequal Variance" worksheet of Hypothesis Testing.xlsx spreadsheet, we illustrate how to use the t-Test Two Sample Assuming Equal Variances analysis in Excel.

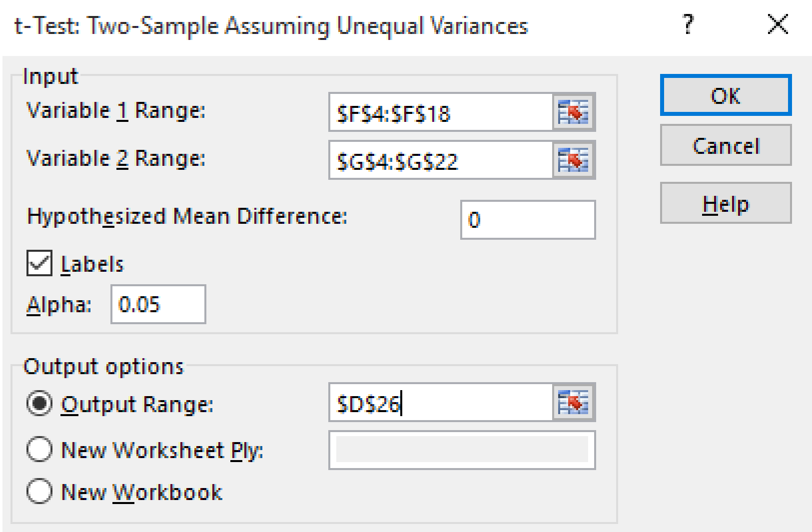
	C	D	E	F	G
1	**TESTING IF VARIANCES OF TWO POPULATIONS ARE EQUAL**				
2			**SKEW**	**-0.5968**	**0.484351**
3	**SAMPLEVAR1**	**11.0289756**	**KURT**	**-0.21044**	**-0.22528**
4	**SAMPLESIZE1**	**14**		Placebo	Drug
5	**SAMPLEVAR2**	**1.119337276**		2.907128	10.41907
6	**SAMPLESIZE2**	**18**		4.40015	7.99349
7	**SVAR1OVERSVAR2**	**9.853129914**		5.492934	8.61935
8	**LOWERCI**	**0.358900271**		4.55906	9.152468
9	**UPPERCI**	**3.003895725**		7.550103	8.599242
10				-2.77164	8.922706
11	**LOWERLIMIT**	**3.536290994**		-3.85383	9.382386
12	**UPPERLIMIT**	**29.59777483**		1.075804	9.259688
13				3.053792	9.446886
14	**Variances Not Equal**			3.256407	10.71418
15				2.996793	11.12724
16	**Test H_0: MeanPlacebo=Mean Drug**			-2.98564	9.476701
17	**Test Ha: MeanPlacebo<MeanDrug**			1.396865	9.680459
18				2.843022	8.138261
19	**Test H_0: VariancePlacebo=VarianceDrug**				10.3785
20	**Test Ha: VariancePlacebo≠VarianceDrug**				8.954938
21	**Reject H_0**				11.99119
22					10.30647
23	**Now test mean difference using Unequal variance t test**				

You are conducting a study to determine if a new anti-cholesterol drug is more effective than the placebo in reducing cholesterol. Fourteen patients were given a placebo, and eighteen patients were given a new anti-cholesterol drug. Your data indicates the change (reduction) in cholesterol for each patient from month 1 to month 2. We are trying to determine if the mean change cholesterol for the patients receiving the drug is bigger than that of those receiving the placebo.

First, we confirm the assumption of normality using SKEW and KURT functions. As you can see by the results in F2:G3, the assumption of normal populations appears reasonable.

Next, we determine if the variances are equal. In cells D3 and D5, we compute the sample variances with the VAR function, and in cells D4 and D6, we note the sample sizes. From cells D11 and D12, we are 95% sure the ratio of the population variances is between 3.54 and 29.56. This interval does not include one, so we conclude the population variances are unequal.

To test H_0: $\mu_{Placebo}=\mu_{Drug}$ and H_a: $\mu_{Drug}>\mu_{Placebo}$, we select Data Analysis from the Data ribbon, and choose "t-Test Two Sample Assuming Unequal Variances." Then, we fill in the dialog box as shown below:



The results of the hypothesis test are:

	D	E	F	G
26	t-Test: Two-Sample Assuming Unequal Variances			
27				
28		*Placebo*	*Drug*	
29	Mean	2.137211	9.586846	
30	Variance	11.02898	1.119337	
31	Observations	14	18	
32	Hypothesized Mea	0		
33	df	15		
34	t Stat	-8.08041		
35	P(T<=t) one-tail	3.81E-07		
36	t Critical one-tail	1.75305		
37	P(T<=t) two-tail	7.61E-07		
38	t Critical two-tail	2.13145		

The p-value for a one tailed test is 4 in 10 million, a number that is much less than 0.05, so we reject the null hypothesis that the placebo and drug are equally effective at reducing cholesterol, and conclude that the new drug is more effective than the placebo at reducing cholesterol.

t-Test Paired Two Sample for Means

Often observations from two populations can be paired in a meaningful way. In such situations, the t-Test Paired Two Sample for Means (often called Matched Pairs) can be used. Both populations need to be normal random variables. For example:

Goal	Design
To test if a drug reduces cholesterol	Pick ten pairs of two people who are matched on age, weight and cholesterol. We flip a coin to randomly choose which member of each pair receives the drug and which receives the placebo.
To test if a new type of insulation reduces heating bills	Pick ten pairs of two houses that had the same heating bill last winter. Flip a coin to choose house gets the new type of insulation; the other retains its old insulation.
To test if cross training (not just swimming) improves a swimmer's time	Pick 15 pairs of swimmers who had identical best times in their event. Flip a coin to determine which in the pair starts cross training.

In each of these situations, we are **blocking** the effect of a variable on the response and focusing on the difference between the **treatment** variable. Blocking the effect of non-treatment variables makes it easier to isolate the effect of the treatment variable.

Blocking Variable	Treatment Variable
Physical characteristics of patients	Difference between drug and placebo
Size and design of home	Difference between new and old insulation
Swimmer's ability	Difference between cross training and just in water training

In the "Matched Pairs" worksheet, we use the t-Test Paired Two Sample for Means to determine whether a new type of insulation reduces heating bills.

The homes in each row had identical winter heating bills in 2016. Before the winter of 2017, a coin is flipped for each pair of homes. If the coin is heads the first home in the pair is given a new type of insulation while the other home keeps its old insulation. If the coin is tails the first home keeps its current insulation, and the second home gets the new insulation. This randomization blocks out random differences in homes and allows us to be more confident that any differences in heating bills that we observe are based on the type of insulation in the home rather than something else.

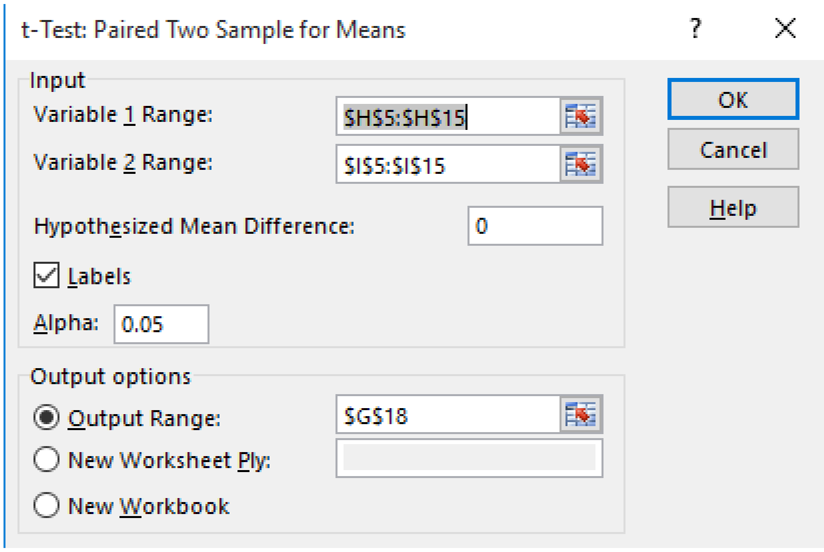
The numerical data in rows 6-15 of this worksheet indicate the change in the monthly winter heating bill during the 2017 winter. For example, the first home given new insulation saw their heating bill increase by an average of $23 per month, and the first home that kept their current insulation saw a $34 per month reduction in their heating bill insulation. We wish to test H_o: $\mu_{New} = \mu_{Old}$ against H_a: $\mu_{New} < \mu_{Old}$.

Here μ_{New} = mean reduction in monthly winter 2017 heating bill for homes with new insulation and μ_{old} = mean reduction in monthly winter 2017 heating bill for homes with old insulation.

As with the previous tests, the first step is to check of the skewness and kurtosis for each sample to ensure they are consistent with normality. Based on the values in H3:I4, we can be confident that the normality assumption is not violated.

	G	H	I
3	**skewness**	**0.43421808**	**-0.197214711**
4	**kurtosis**	**-0.057517024**	**-2.107287845**
5	**Observation**	**Old Insulation**	**New Insulation**
6	**1**	**-34**	**23**
7	**2**	**6**	**16**
8	**3**	**31**	**-28**
9	**4**	**10**	**29**
10	**5**	**-2**	**30**
11	**6**	**-12**	**-72**
12	**7**	**49**	**-46**
13	**8**	**-15**	**-55**
14	**9**	**-45**	**21**
15	**10**	**-17**	**-61**
16			
17			
18	**t-Test: Paired Two Sample for Means**		
19			
20		***Old Insulation***	***New Insulation***
21	**Mean**	**-2.9**	**-14.3**
22	**Variance**	**806.3222222**	**1750.233333**
23	**Observations**	**10**	**10**
24	**Pearson Correlation**	**-0.206862333**	
25	**Hypothesized Mean**	**0**	
26	**df**	**9**	
27	**t Stat**	**0.652971466**	
28	**P(T<=t) one-tail**	**0.265049676**	
29	**t Critical one-tail**	**1.833112933**	
30	**P(T<=t) two-tail**	**0.530099352**	
31	**t Critical two-tail**	**2.262157163**	

To do this analysis, from the Data ribbon, we select Data Analysis, choose “t-Test Paired Two Sample for Means” and fill in the dialog box as shown below:



We obtain a p-value for a one-tailed test of 0.265, so for $\alpha = 0.05$, we would fail to reject H_0, and conclude the new insulation is ineffective.

Chi Square Test for Independence

The **Chi-Square test for Independence** is used to determine if there is a significant relationship between two categorical variables. For example, the distribution of eye color by gender in an Iowa State statistics class is summarized in the following **contingency table:**

	Eye Color			
Gender	Blue	Brown	Green	Hazel
Female	370	352	198	187
Male	359	290	110	169
Total	729	642	308	356

The question is whether eye color is independent of or dependent on gender. As shown in the "Chi-Square" worksheet, we find the percentage of eye color by gender is:

	K	L	M	N	O	P
11		Eye Color				
12	Gender	Blue	Brown	Green	Hazel	Total
13	Female	33.42%	31.80%	17.89%	16.89%	100.00%
14	Male	38.69%	31.25%	11.85%	18.21%	100.00%
15	Total	35.82%	31.55%	15.14%	17.49%	100.00%

If eye color was completely independent of gender, then eye color would not depend on gender, and we would find that:

- 35.82% of each gender with blue eyes.
- 31.55% of each gender with brown eyes
- 15.14% of each gender with green eyes.
- 17.49% of each gender with hazel eyes.

(These percentages are the marginal probabilities for each eye color in the table.)

Are the observed discrepancies from the expected percentages simply occurring by chance? If so, gender and eye color are independent. In this situation, our hypotheses are:

H_0: Gender and Eye Color are Independent
H_a: Eye Color depends on Gender

If a contingency table has R rows and C columns, then the relevant test statistic follows a χ^2 random variable with (R-1) * (C-1) degrees of freedom. In this example, R = 2 and C = 4, giving us 3 degrees of freedom.

To compute the relevant test statistic, r, we define:

- N = Total number of observations
- O_{ij} = Observed number of observations in row i and column j of the contingency table.
- E_{ij} =N * (proportion of observations in row i)*(proportion of observations in column j); or the row marginal probability multiplied by the column marginal probability multiplied by the total number of observations. E_{ij} is simply the expected number of observations in the row i column j cell if the row and column categories are independent.

Then, for each cell, compute $(O_{ij} – E_{ij})^2/E_{ij}$. In other words, observed value – expected value squared divided by the expected value.

Summing up this quantity for each cell yields the χ^2 statistic. If there is perfect independence between the row and column categories, then for each cell E_{ij} =O_{ij}, (the number of expected observations equals the number observed), and the observed χ^2 statistic would equal 0. Therefore, a large χ^2 statistic would result in rejection of the null hypothesis.

If the level of significance is α, then we reject the null hypothesis when the test statistic is >= the (1-α) %ile of the Chi Square random variable with the appropriate degrees of freedom. This value for the Chi-Square test can be found using the CHISQ.INV(1-α,degrees of freedom) function in Excell. The cutoffs for 1,2,3, and 4 degrees of freedom for α=0.05 are as follows:

	D	E	F
14			
15	**df**	**Cutoff**	
16	**1**	**3.841459**	**=CHISQ.INV(0.95,D16)**
17	**2**	**5.991465**	**=CHISQ.INV(0.95,D17)**
18	**3**	**7.814728**	**=CHISQ.INV(0.95,D18)**
19	**4**	**9.487729**	**=CHISQ.INV(0.95,D19)**

In our example, we have R = 2 and C = 4, so we have (2-1)*(4-1) = 3 degrees of freedom and the cutoff point for rejecting the null hypothesis of independence for α= 0.05 is 7.81.

The total number of observations (2035) is computed in cell I17.

To calculate the Chi Square, you must first calculate the expected values. Obtain these values, E_{ij}, by copying the formula from L19 to L19:O22. The formula in L19 = ($P7/Total)*(L$9/Total)*Total, which is the observed marginal value for the associated row divided by the total multiplied by the marginal value for the associated column divided by the total; this product is then multiplied by the total. For example, the expected value for females with blue eyes is calculated like this:

Expected Number of Greened Eyed Females = (1107/2035)*(729*2035)*2035 = 396.56.

Note the row and column totals (L21:O21 and P19:P20) of each E_{ij} match the observed data.

	K	L	M	N	O	P
17		Eye Color				
18	Gender	Blue	Brown	Green	Hazel	Total
19	Female	396.56	349.24	167.55	193.66	1107.00
20	Male	332.44	292.76	140.45	162.34	928.00
21	Total	729.00	642.00	308.00	356.00	

Copying the formula ((L7-L19)^2/L19) from L25 to L25:O26 computes $(O_{ij} - E_{ij})^2/E_{ij}$ for each cell.

	H	I	J	K	L	M	N	O
22	**Test Statistic**							
23					Eye Color			
24				Gender	Blue	Brown	Green	Hazel
25				Female	1.78	0.02	5.54	0.23
26				Male	2.12	0.03	6.60	0.27
27	Chi Square Total							
28	16.59							

Summing those values, the Chi Square statistic is 16.59 which is >= cutoff of 7.81, so we reject the hypothesis that eye color and gender are independent. We can find the p-value for the test using the CHISQ.DIST.RT(test statistic value, degrees of freedom) in cell H29. RT in this formula stands for "right tailed probability" which is the probability that we evaluate in Chi Square analysis.

	H	I	J	K
29				
30	0.000858	=CHISQ.DIST.RT(H28,3)		

In this case, the p-value of 0.000858 is <= 0.05, which is consistent with our rejection of H_0.