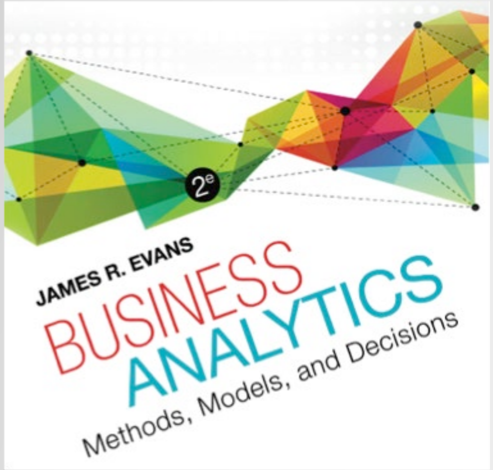


Chapter 7
Statistical Inference

Statistical Inference

- **Statistical inference** focuses on drawing conclusions about populations from samples.
  - Statistical inference includes estimation of population parameters and hypothesis testing, which involves drawing conclusions about the value of the parameters of one or more populations.

Hypothesis Testing

- **Hypothesis testing** involves drawing inferences about two contrasting propositions (each called a **hypothesis**) relating to the value of one or more population parameters.
  - H_0: **Null hypothesis**: describes an existing theory
  - H_1: **Alternative hypothesis**: the complement of H_0
- Using sample data, we either:

  \- *reject* H_0 and conclude the sample data provides sufficient evidence to support H_1, or

  \- *fail to reject* H_0 and conclude the sample data does not support H_1.

Example 7.1: A Legal Analogy for Hypothesis Testing

- In the U.S. legal system, a defendant is innocent until proven guilty.
  - H_0: Innocent
  - H_1: Guilty
- If evidence (sample data) strongly indicates the defendant is guilty, then we reject H_0.
- Note that we have not *proven* guilt or innocence!

Hypothesis Testing Procedure

Steps in conducting a hypothesis test:

1. Identify the population parameter and formulate the hypotheses to test.
2. Select a **level of significance** (the risk of drawing an incorrect conclusion).
3. Determine the decision rule on which to base a conclusion.
4. Collect data and calculate a test statistic.
5. Apply the decision rule and draw a conclusion.

One-Sample Hypothesis Tests

- Three types of one sample tests:
  1. H_0: parameter ≤ constant
     H_1: parameter > constant
  2. H_0: parameter ≥ constant
     H_1: parameter < constant
  3. H_0: parameter = constant
     H_1: parameter ≠ constant
- It is not correct to formulate a null hypothesis using >, <, or ≠.

Determining the Proper Form of Hypotheses

- Hypothesis testing always assumes that H_0 is true and uses sample data to determine whether H_1 is more likely to be true.
  - Statistically, we cannot "prove" that H_0 is true; we can only fail to reject it.
- Rejecting the null hypothesis provides strong evidence (in a statistical sense) that the null hypothesis is not true and that the alternative hypothesis is true.
- Therefore, what we wish to provide evidence for statistically should be identified as the alternative hypothesis.

Example 7.2: Formulating a One-Sample Test of Hypothesis

- CadSoft receives calls for technical support. In the past, the average response time has been at least 25 minutes. It believes the average response time can be reduced to less than 25 minutes.
  - If the new information system makes a difference, then, data should be able to confirm that the mean response time is less than 25 minutes; this defines the alternative hypothesis, H_1.

H_0: mean response time ≥ 25

H_1: mean response time < 25

	A	B	C	D	E
1	CadSoft Technical Support Response Times				
2					
3	Customer	Time (min)			
4	1	20			
5	2	12			
6	3	15			
7	4	11			
8	5	22			
9	6	6			
10	7	39			

Understanding Potential Errors in Hypothesis Testing

- Hypothesis testing can result in one of four different outcomes:

1. H_0 is true and the test correctly fails to reject H_0
2. H_0 is false and the test correctly rejects H_0
3. H_0 is true and the test incorrectly rejects H_0 (called Type I error)
4. H_0 is false and the test incorrectly fails to reject H_0 (called Type II error)

Terminology

- The probability of making a Type I error = α (level of significance) = P(rejecting H_0 | H_0 is true)
  - The value of 1 – α is called the **confidence coefficient**
    = P(not rejecting H_0 | H_0 is true),
  - The value of α can be controlled. Common values are 0.01, 0.05, or 0.10.
- The probability of making a Type II error = β = P(not rejecting H_0 | H_0 is false)
  - The value of 1 - β is called the **power of the test**
    = P(rejecting H_0 | H_0 is false).
  - The value of β cannot be specified in advance and depends on the value of the (unknown) population parameter.

Example 7.3: How β Depends on the True Population Mean

- In the CadSoft example:

  H_0: mean response time ≥ 25

  H_1: mean response time < 25

  - If the true mean is 15, then the sample mean will most likely be less than 25, leading us to reject H_0.
  - If the true mean is 24, then the sample mean may or may not be less than 25, and we would have a higher chance of failing to reject H_0.

Example 7.3: How β Depends on the True Population Mean

- In the CadSoft example:

  H_0: mean response time ≥ 25

  H_1: mean response time < 25

  - If the true mean is 15, then the sample mean will most likely be less than 25, leading us to reject H_0.
  - If the true mean is 24, then the sample mean may or may not be less than 25, and we would have a higher chance of failing to reject H_0.
- The further away the true mean is from the hypothesized value, the smaller the value of β.
- Generally, as α decreases, β increases.

Improving the Power of the Test

- We would like the power of the test to be high (equivalently, we would like the probability of a Type II error to be low) to allow us to make a valid conclusion.
- The power of the test is sensitive to the sample size; small sample sizes generally result in a low value of 1 - β.
- The power of the test can be increased by taking larger samples, which enable us to detect small differences between the sample statistics and population parameters with more accuracy.
- If you choose a small level of significance, you should try to compensate by having a large sample size.

Selecting the Test Statistic

- The decision to reject or fail to reject a null hypothesis is based on computing a **test statistic** from the sample data.
- The test statistic used depends on the type of hypothesis test.
  - Test statistics for one-sample hypothesis tests for means:

Type of Test	Test Statistic	
One-sample test for mean, σ known	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	(7.1)
One-sample test for mean, σ unknown	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	(7.2)

Example 7.4 Computing the Test Statistic

- In the CadSoft example, sample data for 44 customers revealed a mean response time of 21.91 minutes and a sample standard deviation of 19.49 minutes.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{21.91 - 25}{19.49/\sqrt{44}} = \frac{-3.09}{2.938} = -1.05$$

t = -1.05 indicates that the sample mean of 21.91 is 1.05 standard errors below the hypothesized mean of 25 minutes.

Drawing a Conclusion

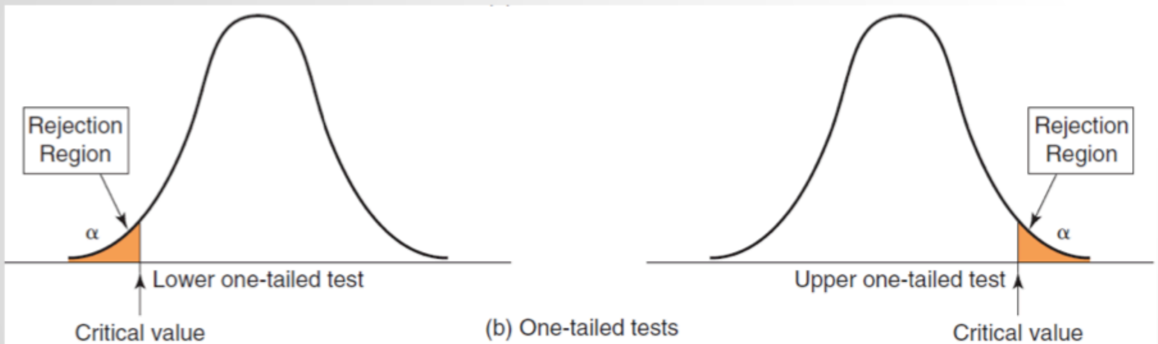
- The conclusion to reject or fail to reject H_0 is based on comparing the value of the test statistic to a "critical value" from the sampling distribution of the test statistic when the null hypothesis is true and the chosen level of significance, α.
  - The sampling distribution of the test statistic is usually the normal distribution, t-distribution, or some other well-known distribution.
- The critical value divides the sampling distribution into two parts, a rejection region and a non-rejection region. If the test statistic falls into the rejection region, we reject the null hypothesis; otherwise, we fail to reject it.

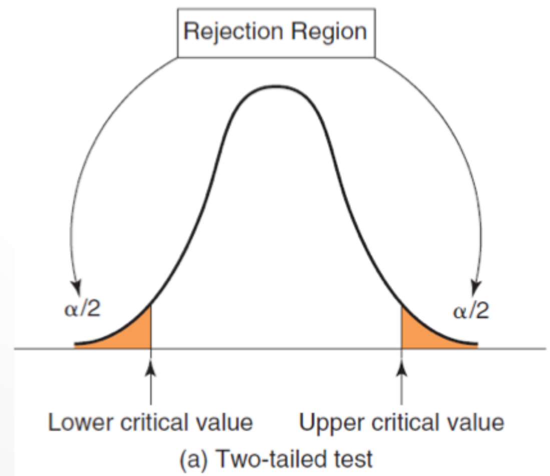
Rejection Regions

H_0: parameter ≥ constant
H_1: parameter < constant

H_0: parameter ≤ constant
H_1: parameter > constant

H_0: parameter = constant
H_1: parameter ≠ constant



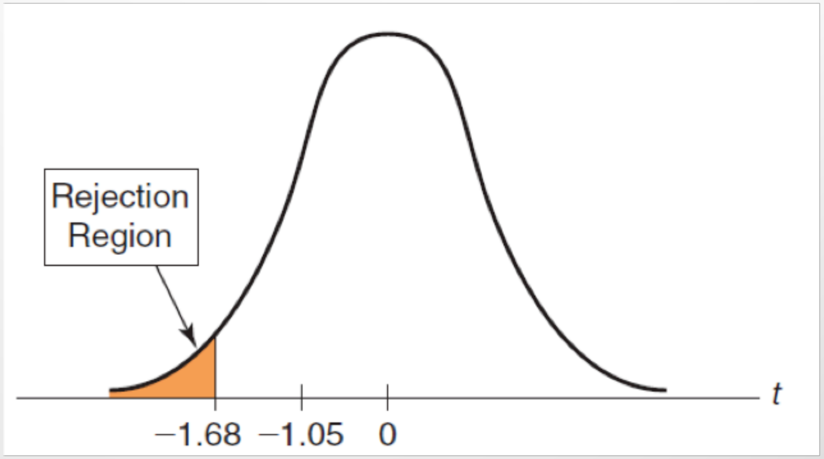


For a one-tailed test, if H_1 is stated as <, the rejection region is in the lower tail; if H_1 is stated as >, the rejection region is in the upper tail (just think of the inequality as an arrow pointing to the proper tail direction).

Example 7.5: Finding the Critical Value and Drawing a Conclusion

- In the CadSoft example, use $\alpha = 0.05$.
  - H_0: mean response time ≥ 25
  - H_1: mean response time < 25
- $n = 44$; $df = n - 1 = 43$
- $t = -1.05$
- Critical value = $t_{\alpha/2,\, n-1}$ = T.INV(1− α , n −1) = T.INV(0.95, 43) = 1.68
- $t = -1.05$ does not fall in the rejection region.
- Fail to reject H_0.

Even though the sample mean of 21.91 is well below 25, we have too much sampling error to conclude the that the true population mean is less than 25 minutes.

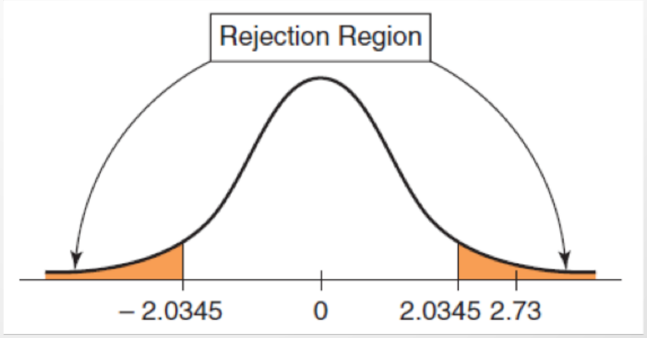


Example 7.6: Conducting a Two-Tailed Hypothesis Test for the Mean

- Excel file *Vacation Survey*
- Test whether the average age of respondents is equal to 35.
  - H_0: mean age = 35
  - H_1: mean age ≠ 35
- *n* = 34; sample mean = 38.677; sample standard deviation = 7.858.
- Test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{(38.677 - 35)}{(7.858/\sqrt{34})} = 2.73$$

- Critical value = T.INV.2T(.05, 33) = 2.0345
- *p*-value = T.DIST.2T(2.69, 33) = 0.0111
- Reject H_0.



p-Values

- A **p-value (observed significance level)** is the probability of obtaining a test statistic value equal to or more extreme than that obtained from the sample data when the null hypothesis is true.
- An alternative approach to Step 3 of a hypothesis test uses the *p*-value rather than the critical value:

Reject H_0 if the p-value $< \alpha$

Finding p-Values

- For a lower one-tailed test, the p-value is the probability to the left of the test statistic t in the t-distribution, and is found using the Excel function:
  - =T.DIST(*t, n-1, TRUE*).
- For an upper one-tailed test, the p-value is the probability to the right of the test statistic t, and is found using the Excel function:
  - 1 - T.DIST(*t, n-1, TRUE*).
- For a two-tailed test, the p-value is found using the Excel function:
  - T.DIST.2T(*t, n-1*), if t > 0
  - T.DIST.2T(*-t, n-1*), if t < 0

Example 7.7: Using p-Values

- In the CadSoft example, the *p*-value is the left tail area of the observed test statistic, $t = -1.05$.
  - *p*-value =TDIST(-1.05, 43, true) = 0.1498
  - Do not reject H_0 because the *p*-value ≥ α, i.e., 0.1498 ≥ 0.05
- For the Vacation Survey two-tailed hypothesis test in Example 7.6, the p-value for this test is
  - p-value =T.DIST.2T(2.73,33) = 0.010
  - Reject H_0 because 0.010 < 0.05

One-Sample Tests for Proportions

- Test statistic:

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \qquad (7.3)$$

- π_0 is the hypothesized value and $\hat{p}$ is the sample proportion

Example 7.8: One-Sample Test for the Proportion

- CadSoft sampled 44 customers and asked them to rate the overall quality of a software package. Sample data revealed that 35 respondents (a proportion of 35/44 = 0.795) thought the software was very good or excellent. In the past, this proportion has averaged about 75%. Is there sufficient evidence to conclude that this satisfaction measure has significantly exceeded 75% using a significance level of 0.05?

Example 7.8 Continued

- Hypotheses:
  - H_0: $\pi \leq 0.75$
  - H_1: $\pi > 0.75$
- Test statistic:

$$z = \frac{0.795 - 0.75}{\sqrt{0.75(1 - 0.75)/44}} = 0.69$$

- Critical value = NORM.S.INV(0.95) = 1.645
- *p*-value = 1 − NORM.S.DIST(0.69,TRUE) = 0.24
- Do not reject H_0.

Two-Sample Hypothesis Tests

- <u>Lower-tailed test</u>
  - H_0: population parameter (1) - population parameter (2) ≥ D_0
  - H1: population parameter (1) - population parameter (2) < D_0
- This test seeks evidence that the difference between population parameter (1) and population parameter (2) is less than some value, D_0.
- When $D_0 = 0$, the test simply seeks to conclude whether population parameter (1) is smaller than population parameter (2).

Two-Sample Hypothesis Tests

- <u>Upper-tailed test</u>
  - H_0: population parameter (1) - population parameter (2) ≤ D_0
  - H1: population parameter (1) - population parameter (2) > D_0
- This test seeks evidence that the difference between population parameter (1) and population parameter (2) is greater than some value, D_0.
- When $D_0 = 0$, the test simply seeks to conclude whether population parameter (1) is larger than population parameter (2).

Two-Sample Hypothesis Tests

- <u>Two-tailed test</u>
  - H_0: population parameter (1) - population parameter (2) = D_0
  - H1: population parameter (1) - population parameter (2) ≠ D_0
- This test seeks evidence that the difference between the population parameters is equal to D_0.
- When $D_0 = 0$, we are seeking evidence that population parameter (1) differs from population parameter (2).
  - In most applications, $D_0 = 0$, and we are simply seeking to compare the population parameters.

Excel Analysis Toolpak Procedures for Two-Sample Hypothesis Tests

Type of Test	Excel Procedure
Two-sample test for means, σ^2 known	Excel *z*-test: Two-sample for means
Two-sample test for means, σ^2 unknown, assumed unequal	Excel *t*-test: Two-sample assuming unequal variances
Two-sample test for means, σ^2 unknown, assumed equal	Excel *t*-test: Two-sample assuming equal variances
Paired two-sample test for means	Excel *t*-test: Paired two-sample for means
Two-sample test for equality of variances	Excel *F*-test Two-sample for variances

Two-Sample Tests for Difference in Means

- Forms of the hypothesis test:

$$H_0: \mu_1 - \mu_2 \{\geq, \leq, \text{or} =\} 0$$

$$H_1: \mu_1 - \mu_2 \{<, >, \text{or} \neq\} 0 \quad \textbf{(7.4)}$$

Example 7.9: Comparing Supplier Performance

- *Purchase Orders* database

	A	B	C	D	E	F	G	H	I	J	K
1	Purchase Orders										
2											
3	Supplier	Order No.	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Months)	Order Date	Arrival Date	Lead Time
4	Hulkey Fasteners	Aug11001	1122	Airframe fasteners	$ 4.25	19,500	$ 82,875.00	30	08/05/11	08/13/11	8
5	Alum Sheeting	Aug11002	1243	Airframe fasteners	$ 4.25	10,000	$ 42,500.00	30	08/08/11	08/14/11	6
6	Fast-Tie Aerospace	Aug11003	5462	Shielded Cable/ft.	$ 1.05	23,000	$ 24,150.00	30	08/10/11	08/15/11	5
7	Fast-Tie Aerospace	Aug11004	5462	Shielded Cable/ft.	$ 1.05	21,500	$ 22,575.00	30	08/15/11	08/22/11	7
8	Steelpin Inc.	Aug11005	5319	Shielded Cable/ft.	$ 1.10	17,500	$ 19,250.00	30	08/20/11	08/31/11	11
9	Fast-Tie Aerospace	Aug11006	5462	Shielded Cable/ft.	$ 1.05	22,500	$ 23,625.00	30	08/20/11	08/26/11	6
10	Steelpin Inc.	Aug11007	4312	Bolt-nut package	$ 3.75	4,250	$ 15,937.50	30	08/25/11	09/01/11	7

- Determine if the mean lead time for Alum Sheeting (μ_1) is greater than the mean lead time for Durrable Products (μ_2).

	A	B
1		
2		
3	Row Labels	Average of Lead Time
4	Alum Sheeting	7.00
5	Durrable Products	4.92
6	Fast-Tie Aerospace	8.47
7	Hulkey Fasteners	6.47
8	Manley Valve	6.45
9	Pylon Accessories	8.00
10	Spacetime Technologies	15.25
11	Steelpin Inc.	10.20
12	Grand Total	8.41

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_1: \mu_1 - \mu_2 > 0$$

Selecting the Proper Excel Procedure

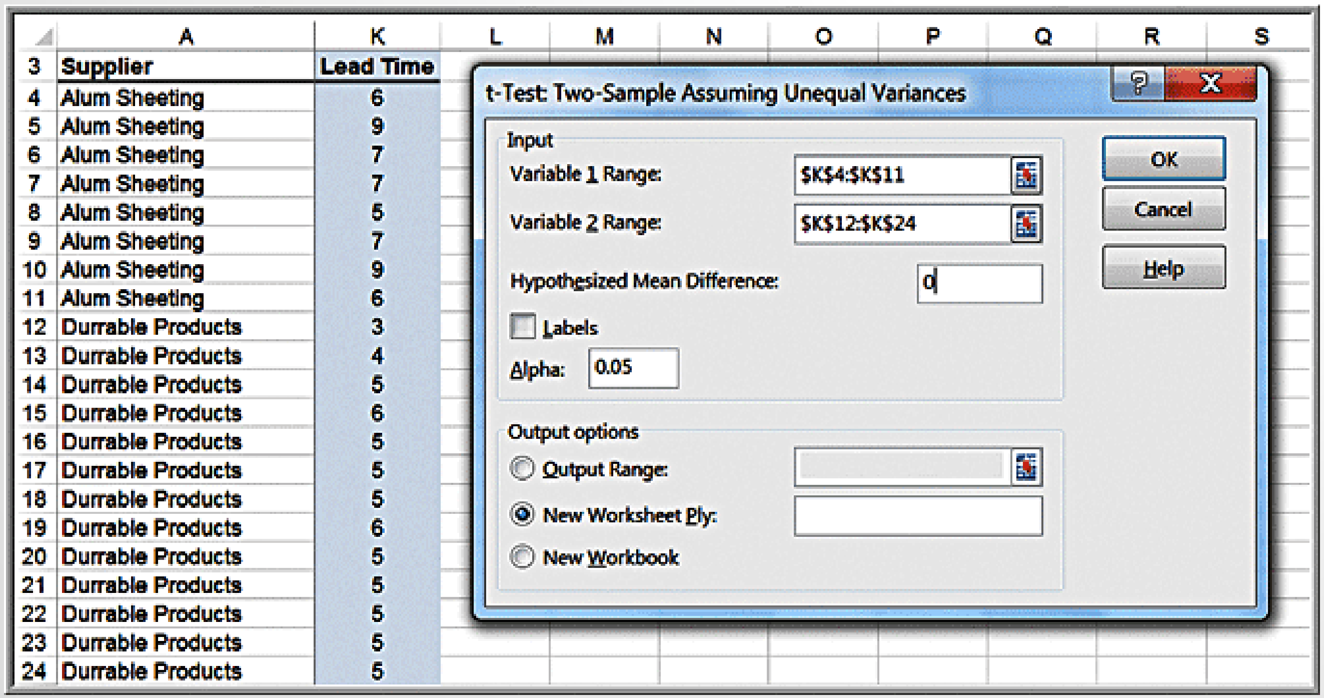
- Population variances are known:
  - *z-Test: Two-Sample for Means*
- Population variances are unknown and assumed unequal:
  - *t-Test: Two-Sample Assuming Unequal Variances*
- Population variances are unknown but assumed equal:
  - *t-Test: Two-Sample Assuming Equal Variances*
- These tools calculate the test statistic, the p-value for both a one-tail and two-tail test, and the critical values for one-tail and two-tail tests.

Intepreting Excel Output

- If the test statistic is negative, the one-tailed *p*-value is the correct *p*-value for a lower-tail test; however, for an upper-tail test, you must subtract this number from 1.0 to get the correct *p*-value.
- If the test statistic is nonnegative (positive or zero), then the *p*-value in the output is the correct *p*-value for an upper-tail test; but for a lower-tail test, you must subtract this number from 1.0 to get the correct *p*-value.
- For a lower-tail test, you must change the sign of the one-tailed critical value.

Example 7.10: Testing the Hypotheses for Supplier Lead-Time Performance

- *t-Test: Two-Sample Assuming Unequal Variances*
  - *Variable 1 Range:* Alum Sheeting data
  - *Variable 2 Range:* Durrable Products data



	A	K
3	Supplier	Lead Time
4	Alum Sheeting	6
5	Alum Sheeting	9
6	Alum Sheeting	7
7	Alum Sheeting	7
8	Alum Sheeting	5
9	Alum Sheeting	7
10	Alum Sheeting	9
11	Alum Sheeting	6
12	Durrable Products	3
13	Durrable Products	4
14	Durrable Products	5
15	Durrable Products	6
16	Durrable Products	5
17	Durrable Products	5
18	Durrable Products	5
19	Durrable Products	6
20	Durrable Products	5
21	Durrable Products	5
22	Durrable Products	5
23	Durrable Products	5
24	Durrable Products	5

Example 7.10 Continued

- Results
  - Rule 2: If the test statistic is nonnegative (positive or zero), then the *p*-value in the output is the correct *p*-value for an upper-tail test.

$t = 3.83$
Critical value = 1.81
p-value = 0.00166

Reject H_0.

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2		Alum Sheeting	Durrable Products
3		*Variable 1*	*Variable 2*
4	Mean	7	4.923076923
5	Variance	2	0.576923077
6	Observations	8	13
7	Hypothesized Mean Difference	0	
8	df	10	
9	t Stat	3.827958507	
10	P(T<=t) one-tail	0.001664976	
11	t Critical one-tail	1.812461123	
12	P(T<=t) two-tail	0.003329952	
13	t Critical two-tail	2.228138852	

Two-Sample Test for Means with Paired Samples

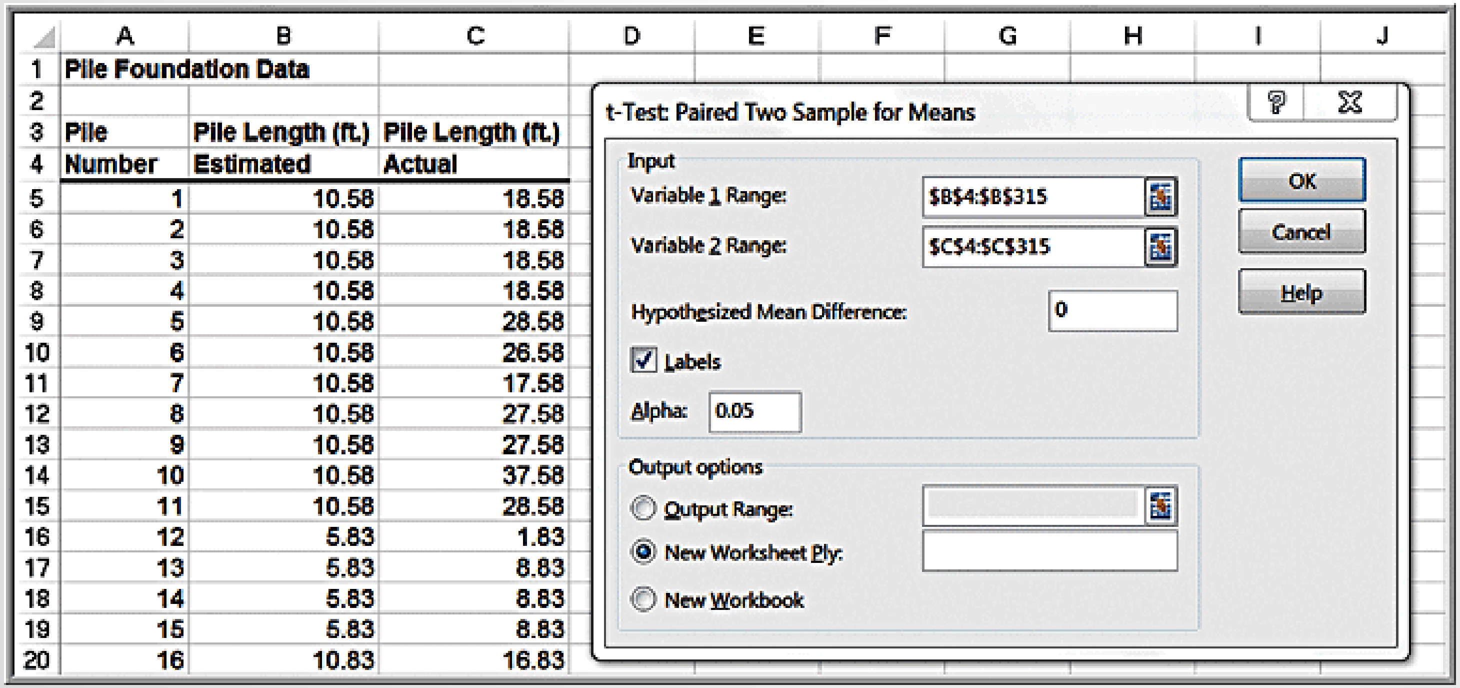
- In many situations, data from two samples are naturally paired or matched.
- When paired samples are used, a paired t-test is more accurate than assuming that the data come from independent populations.
- Hypotheses (μ_D is the mean difference between the paired samples):

$$H_0: \mu_D \{\geq, \leq, \text{or} =\} 0$$
$$H_1: \mu_D \{<, >, \text{or} \neq\} 0.$$

- Excel *Data Analysis* tool: *t-Test: Paired Two-Sample for Means*

Example 7.11 Using the Paired Two-Sample Test for Means

- Excel file *Pile Foundation*
  - Test for a difference in the means of the estimated and actual pile lengths (two-tailed test).



Example 7.11 Continued

- Results:
- $t = -10.91$
- t is smaller than the lower critical value
- p-value ≈ 0
- Reject the null hypothesis

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		*Estimated*	*Actual*
4	Mean	28.17755627	34.55623794
5	Variance	255.8100385	267.0113061
6	Observations	311	311
7	Pearson Correlation	0.79692836	
8	Hypothesized Mean Difference	0	
9	df	310	
10	t Stat	-10.91225025	
11	P(T<=t) one-tail	5.59435E-24	
12	t Critical one-tail	1.649783823	
13	P(T<=t) two-tail	1.11887E-23	
14	t Critical two-tail	1.967645929	

Test for Equality of Variances

- Test for equality of variances between two samples using a new type of test, the *F*-test.
  - To use this test, we must assume that both samples are drawn from normal populations.
- Hypotheses:

$$H_0: \sigma_1^2 - \sigma_2^2 = 0$$
$$H_1: \sigma_1^2 - \sigma_2^2 \neq 0 \quad \textbf{(7.5)}$$

- F-test statistic:

$$F = \frac{s_1^2}{s_2^2} \quad \textbf{(7.6)}$$

- Excel tool: *F-test for Equality of Variances*

F-Distribution

- The *F*-distribution has two degrees of freedom, one associated with the numerator of the *F*-statistic, $n_1 - 1$, and one associated with the denominator of the *F*-statistic, $n_2 - 1$.
- Table 4, Appendix A provides only upper-tail critical values, and the distribution is not symmetric.

Upper critical values of the *F* distribution for numerator degrees of freedom ν_1 and denominator degrees of freedom ν_2, 5% significance level

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.882	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978

Conducting the *F*-Test

- Although the hypothesis test is really a two-tailed test, we will simplify it as an upper-tailed, one-tailed test to make it easy to use tables of the *F*-distribution and interpret the results of the Excel tool.
  - We do this by ensuring that when we compute F, we take the ratio of the larger sample variance to the smaller sample variance.
- Find the critical value $F_{\alpha/2,df1,df2}$ of the *F*-distribution, and then we reject the null hypothesis if the *F*-test statistic exceeds the critical value.
- Note that we are using $\alpha/2$ to find the critical value, not α. This is because we are using only the upper tail information on which to base our conclusion.

Example 7.12: Applying the F-Test for Equality of Variances

- Determine whether the variance of lead times is the same for Alum Sheeting and Durrable Products in the *Purchase Orders* data.
  - The variance of the lead times for Alum Sheeting is larger than the variance for Durable Products, so this is assigned to *Variable 1*.

F = 3.47
Critical value = 2.91
P-value = 0.029
Reject H_0

	A	B	C
1	F-Test Two-Sample for Variances		
2		Alum Sheeting	Durrable Products
3		*Variable 1*	*Variable 2*
4	Mean	7	4.923076923
5	Variance	2	0.576923077
6	Observations	8	13
7	df	7	12
8	F	3.466666667	
9	P(F<=f) one-tail	0.028595441	
10	F Critical one-tail	3.606514642	

Analysis of Variance (ANOVA)

- Used to compare the means of two or more population groups.

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_m$$

H_1: at least one mean is different from the others

- ANOVA derives its name from the fact that we are analyzing variances in the data.
- ANOVA measures variation between groups relative to variation within groups.
- Each of the population groups is assumed to come from a normally distributed population.

Example 7.13: Difference in *Insurance Survey* Data

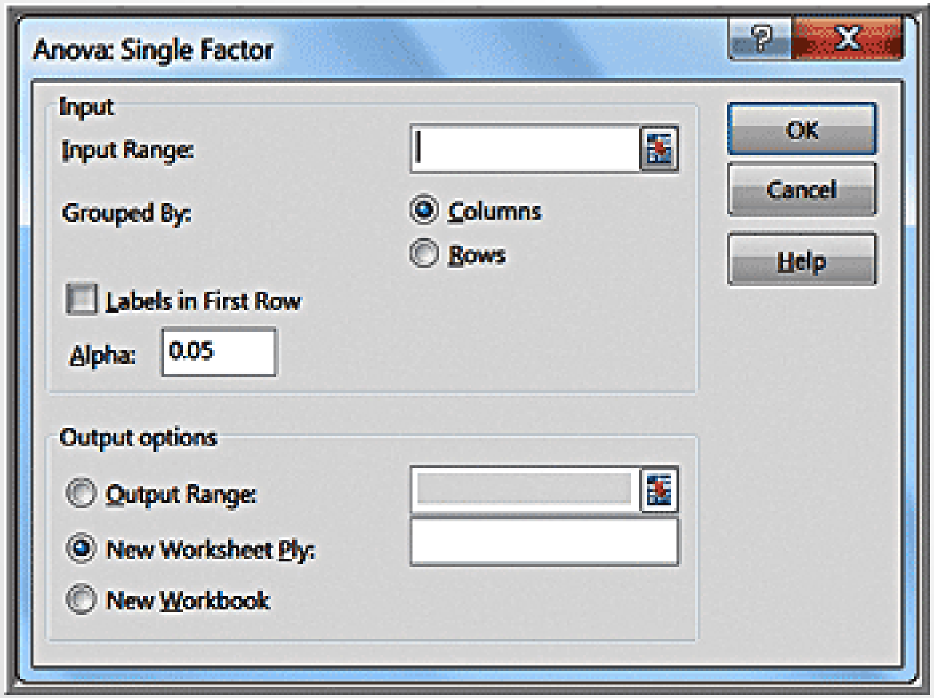
- Determine whether any significant differences exist in satisfaction among individuals with different levels of education.

	College Graduate	Graduate Degree	Some College
	5	3	4
	3	4	1
	5	5	4
	3	5	2
	3	5	3
	3	4	4
	3	5	4
	4	5	
	2		
Average	3.444	4.500	3.143
Count	9	8	7

- The variable of interest is called a **factor**. In this example, the factor is the educational level, and we have three categorical levels of this factor, college graduate, graduate degree, and some college.

Example 7.14: Applying the Excel ANOVA Tool

- *Data Analysis* tool: *ANOVA: Single Factor*
  - The input range of the data must be in contiguous columns



Example 7.14: Continued

- Results
  - $F = 3.92$
  - $F_{crit} = 3.46$
  - $F > F_{crit}$
  - p-value = 0.0356
  - Reject $H_{0.}$

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	College graduate	9	31	3.444444444	1.027777778		
6	Graduate degree	8	36	4.5	0.571428571		
7	Some college	7	22	3.142857143	1.476190476		
8							
9							
10	ANOVA						
11	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
12	Between Groups	7.878968254	2	3.939484127	3.924651732	0.035635398	3.466800112
13	Within Groups	21.07936508	21	1.003779289			
14							
15	Total	28.95833333	23				

Assumptions of ANOVA

- The *m* groups or factor levels being studied represent populations whose outcome measures
  1. are randomly and independently obtained,
  2. are normally distributed, and
  3. have equal variances.
- If these assumptions are violated, then the level of significance and the power of the test can be affected.

Chi-Square Test for Independence

- Test for independence of two categorical variables.
  - H_0: two categorical variables are independent
  - H_1: two categorical variables are dependent

Example 7.15: Independence and Marketing Strategy

- *Energy Drink Survey* data. A key marketing question is whether the proportion of males who prefer a particular brand is no different from the proportion of females.
  - If gender and brand preference are indeed independent, we would expect that about the same proportion of the sample of female students would also prefer brand 1.
  - If they are not independent, then advertising should be targeted differently to males and females, whereas if they are independent, it would not matter.

	A	B	C	D	E	F	G	H	I
1	**Energy Drink Survey**								
2									
3	**Respondent**	**Gender**	**Brand Preference**						
4	1	Male	Brand 3		**Count of Respondent**	**Column Labels**			
5	2	Female	Brand 3		**Row Labels**	**Brand 1**	**Brand 2**	**Brand 3**	**Grand Total**
6	3	Male	Brand 3		Female	9	6	22	37
7	4	Male	Brand 1		Male	25	17	21	63
8	5	Male	Brand 1		**Grand Total**	**34**	**23**	**43**	**100**
9	6	Female	Brand 2						
10	7	Male	Brand 2						

Chi-Square Test Calculations

- Step 1: Using a cross-tabulation of the data, compute the expected frequency if the two variables are independent.

$$\text{expected frequency in row } i \text{ and column } j = \frac{(\text{grand total row } i)(\text{grand total column } j)}{\text{total number of observations}} \quad \textbf{(7.7)}$$

Example 7.16: Computing Expected Frequencies

$$\text{expected frequency in row } i \text{ and column } j = \frac{(\text{grand total row } i)(\text{grand total column } j)}{\text{total number of observations}} \quad (7.7)$$

	E	F	G	H	I	J	K
1	Chi-Square Test						
2							
3	Count of Respondent	Column Labels					
4	Row Labels	Brand 1	Brand 2	Brand 3	Grand Total		
5	Female	9	6	22	37		
6	Male	25	17	21	63		
7	Grand Total	34	23	43	100		
8							
9							
10	Expected Frequency	Brand 1	Brand 2	Brand 3	Grand Total		
11	Female	12.58	8.51	15.91	37		
12	Male	21.42	14.49	27.09	63		
13	Grand Total	34	23	43	100		

Expected frequency of Female and Brand 1 = 37*34/100

Chi-Square Test Calculations

- Step 2: Compute a test statistic, called a **chi-square statistic**, which is the sum of the squares of the differences between observed frequency, f_o, and expected frequency, f_e, divided by the expected frequency in each cell:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \qquad \textbf{(7.8)}$$

Chi-Square Distribution

- The sampling distribution of X^2 is a special distribution called the **chi-square distribution**.
  - The chi-square distribution is characterized by degrees of freedom.
  - Table 3 in Appendix A

Degrees of Freedom	Upper Tail Areas (α)											
	.995	.99	.975	.95	.90	.75	.25	.10	.05	.025	.01	.005
1			0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	12.549	15.987	18.307	20.483	23.209	25.188

Chi-Square Test Calculations (continued)

- *Step 3*: Compare the chi-square statistic for the level of significance α to the critical value from a chi-square distribution with (r – 1)(c – 1) degrees of freedom, where r and c are the number of rows and columns in the cross-tabulation table, respectively.
  - The Excel function CHISQ.INV.RT(*probability, deg_ freedom*) returns the value of X^2 that has a right-tail area equal to probability for a specified degree of freedom.
  - By setting *probability* equal to the level of significance, we can obtain the critical value for the hypothesis test.
  - The Excel function CHISQ.TEST(*actual_range, expected_range*) computes the p-value for the chi-square test.

Example 7.17: Conducting the Chi-Square Test

- Test statistic = 6.49
- d.f. = (2 – 1)(3 – 1) = 2
- Critical value = CHISQ.INV.RT(0.05,2) = 5.99
- p-value = CHISQ.TEST(F6:H7,F12:H13) = 0.0389
- Reject H_0

	E	F	G	H	I
1	Chi-Square Test				
2					
3	Count of Respondent	Column Labels			
4	Row Labels	Brand 1	Brand 2	Brand 3	Grand Total
5	Female	9	6	22	37
6	Male	25	17	21	63
7	Grand Total	34	23	43	100
8					
9					
10	Expected Frequency	Brand 1	Brand 2	Brand 3	Grand Total
11	Female	12.58	8.51	15.91	37
12	Male	21.42	14.49	27.09	63
13	Grand Total	34	23	43	100
14					
15					
16	Chi Square Statistic	Brand 1	Brand 2	Brand 3	Grand Total
17	Female	1.02	0.74	2.33	4.09
18	Male	0.60	0.43	1.37	2.40
19	Grand Total	1.62	1.18	3.70	6.49
20					
21		Chi-square critical value			5.99146455
22		p-value			0.03892134

Test statistic

Google Analytics Project 3 Email me (albert.kalim@asbury.edu) your answers by Sunday, 6/19,11:59 p.m. ET (10 points total)

Log in to your Google Analytics dashboard here and click Sign in to Analytics.

After you logged in, focus on the Audience tools on the left-hand side. Pick two channels with data (active users, user explorer, geo, behavior, etc.) and **list two business recommendations per channel on how to increase the audience and elaborate on them**. You will be graded based on your familiarity with the tools and how you read/interpret the data.