Chapter 11—Inferences on Two Samples

11.1—Inference about Two Means: Dependent Samples (or Matched-Pairs)
- t hypothesis test
- t confidence interval

11.2—Inference about Two Means: Independent Samples
- t hypothesis test—unequal population variances
- t confidence interval
- t hypothesis test—equal population variances

11.3—Inference about Two Population Proportions
- Z hypothesis test
- Z confidence interval
- Sample size

Example—Comparison of two teaching methods for statistics:
- Traditional method involving class lectures
- Computer method using web-based instruction

Example—Comparison of pig gains on two diets.

Independent versus Dependent Sampling

Independent sampling—when the individuals selected for one sample do not dictate which individuals are to be in a second sample.

Dependent sampling—when the individuals in two samples are somehow related:
- husband-wife
- siblings
- similar characteristics
- same person
- same year

Example 1, p. 508, provides examples of independent versus dependent samples.

Matched-Pairs Design using the t-Distribution—Hypothesis Test Regarding the Difference of Two Means (μ_d) with σ_d Unknown

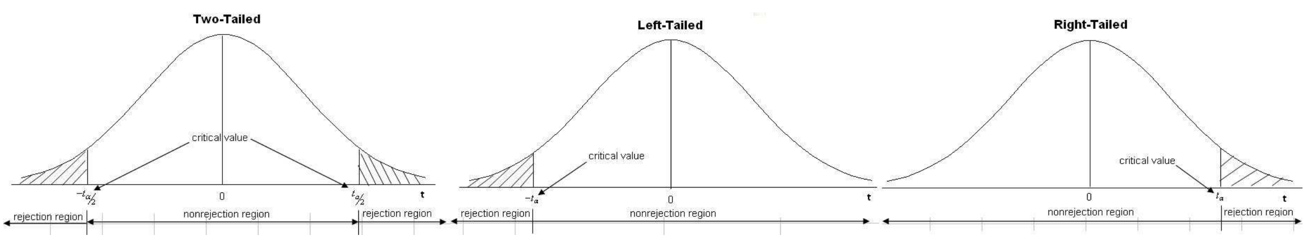
Assumptions:

- The sample is obtained using simple random sampling;
- The sample data are matched-pairs;
- The differences are normally distributed or the sample size, n, is "large" ($n \geq 30$).

Step 1: A claim is made regarding the mean difference from matched-pairs data (μ_d). The null and alternative hypotheses can be structured in three ways:

Two-Tailed	Left-Tailed	Right-Tailed
H_0: $\mu_d = 0$	H_0: $\mu_d = 0$	H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$	H_1: $\mu_d < 0$	H_1: $\mu_d > 0$
$\mu_d<0$ or $\mu_d>0$		

Step 2: Select a **level of significance**, **α**, which is generally chosen to be 0.10, 0.05, or 0.01. The significance level is used to determine the *critical value*. **Critical value** is the t-value that separates the rejection and nonrejection regions. The **rejection region** (or **critical region**) is the set of all values of the test statistic (defined in step 3) such that the null hypothesis is rejected.



Step 3: Calculate the **test statistic** or **calculated t-value**.

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{\bar{d}}{s_d / \sqrt{n}},$$

Note: Statistical inference methods on matched-pairs data use the same methods as inference on a single population mean with σ unknown, except that **differences** are analyzed.

which follows Student's t-distribution with df=n-1. The values of $\bar{d}$ and s_d are the mean and standard deviation of the differenced data. The test statistic (t) measures the number of standard deviations that the sample mean, $\bar{d}$, is from the assumed population mean, $\mu_d=0$.

Step 4: Draw a conclusion:

- Compare the calculated t-value (or test statistic) to the critical t-value and state whether or not H_0 is rejected at the specified α.

Two-Tailed	Left-Tailed	Right-Tailed
If $t < -t_{\alpha/2}$ *or* $t > t_{\alpha/2}$ *reject the null hypothesis*	*If* $t < -t_\alpha$ *reject the null hypothesis.*	*If* $t > t_\alpha$ *reject the null hypothesis.*

- Interpret the conclusion in the context of the problem.

Example 2, p. 510—Testing a Claim Regarding Matched-Pairs Data

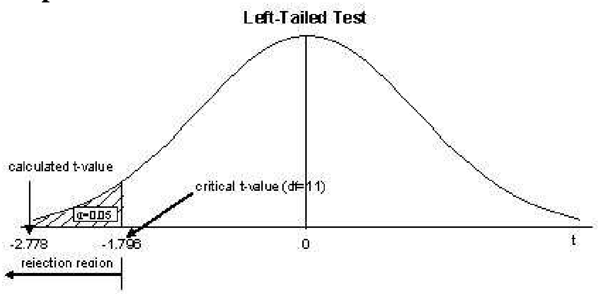
Problem: Professor Andy Neill measured the time (in seconds) required to catch a falling meter stick for 12 randomly selected students' dominant hand and nondominant hand. Professor Neill claims that the reaction time in an individual's dominant hand is less than the reaction time in their nondominant hand. Test the claim at the $\alpha=0.05$ level of significance.

Step 1: Null and Alternative Hypotheses:

H_0: $\mu_d = 0$

H_1: $\mu_d < 0$

Step 2: Select $\alpha = 0.05$ and find the critical value of t (df=12-1=11).



Step 3: Draw a random sample of n=12 students and measure their reaction times in the dominant and nondominant hands. Calculate the sample mean, $\bar{d}$, sample standard deviation, s_d, and the test statistic, *t*.

	A	B	C	D	E	F
1	Student	Dominant Hand, X1	Nondominant Hand, X2	Difference, d_i=X1-X2	$(d_i - \bar{d})$	$(d_i - \bar{d})^2$
2	1	0.177	0.179	-0.002	0.011167	0.000125
3	2	0.210	0.202	0.008	0.021167	0.000448
4	3	0.186	0.208	-0.022	-0.008833	0.000078
5	4	0.189	0.184	0.005	0.018167	0.000330
6	5	0.198	0.215	-0.017	-0.003833	0.000015
7	6	0.194	0.193	0.001	0.014167	0.000201
8	7	0.160	0.194	-0.034	-0.020833	0.000434
9	8	0.163	0.160	0.003	0.016167	0.000261
10	9	0.166	0.209	-0.043	-0.029833	0.000890
11	10	0.152	0.164	-0.012	0.001167	0.000001
12	11	0.190	0.210	-0.020	-0.006833	0.000047
13	12	0.172	0.197	-0.025	-0.011833	0.000140
14			Sum=	-0.158	0.000000	0.002970
15			$\bar{d}$ =	-0.013167	=AVERAGE(D2:D13)	
16			s_d=	0.01643	=STDEV(D2:D13)	

$$t = \frac{(\bar{d} - 0)}{\frac{s_d}{\sqrt{n}}} = \frac{(-0.013167 - 0)}{\frac{0.01643}{\sqrt{12}}} = \frac{-0.013167}{0.00474} = -2.778$$

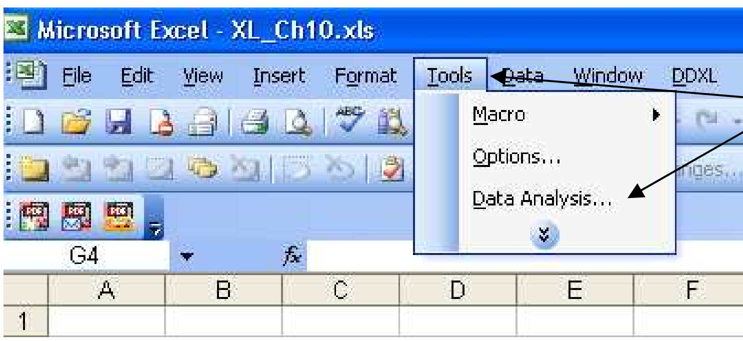
$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{0.00297}{12-1}} = 0.01643$$

Step 4: Conclusion—Because the calculated t = -2.778 is less than the critical t = -1.796 (and in the rejection region), reject H_0 at the 0.05 significance level. There is sufficient evidence at the 0.05 significance level to support Professor Neill's claim that the mean reaction time in the dominant hand is less than the mean reaction time in the nondominant hand.

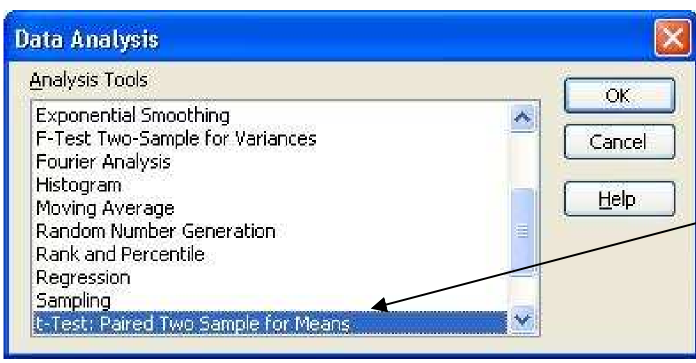
Example 2, p. 510—Using Excel to Test a Claim Regarding Matched-Pairs Data

Problem (see previous page): This problem can be solved using Excel. One advantage is that Excel provides a calculated P-value as well as a calculated t-value. The P-value serves as a measure of the strength of evidence against H_0. A **small P-value** means that the null hypothesis is **strongly rejected** or the result is **highly statistically significant**.

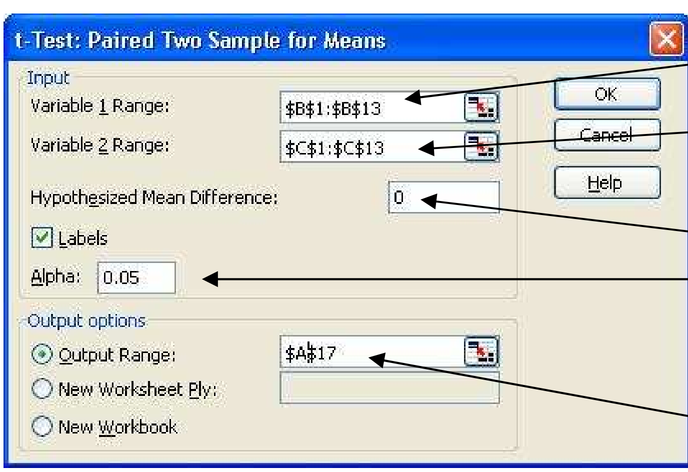
Excel: Two-sample t-tests, Dependent Sampling (or Matched-Pairs):
Step 1: Enter raw data in columns B and C (the Excel worksheet page is shown on the next page).
Step 2: Select the **Tools** menu, highlight **Data Analysis**.
Step 3: Select "t-test: Paired Two-Sample for Means." With the cursor in the "Variable 1 Range" box, highlight the data in column B. With the cursor in the "Variable 2 Range" box, highlight the data in column C. Enter the hypothesized mean difference (usually 0) and a value for alpha (e.g., α=0.05). In the "Output Range" box, specify a cell for the output (upper left corner of the output range). Click **OK**.



From **Tools** menu select **Data Analysis.**



Select an Analysis Tool: t-Test: Paired Two Sample for Means



Highlight the data for Variable 1 and Variable 2

Enter the hypothesized mean difference of 0, and enter α=0.05

Specify a cell for the output.

Output from Excel—t-Test: Paired Two Sample for Means

	A	B	C	D	E	F
1	Student	Dominant Hand, X1	Nondominant Hand, X2	Difference, d_i=X1-X2	$(d_i - \bar{d})$	$(d_i - \bar{d})^2$
2	1	0.177	0.179	-0.002	0.011167	0.000125
3	2	0.210	0.202	0.008	0.021167	0.000448
4	3	0.186	0.208	-0.022	-0.008833	0.000078
5	4	0.189	0.184	0.005	0.018167	0.000330
6	5	0.198	0.215	-0.017	-0.003833	0.000015
7	6	0.194	0.193	0.001	0.014167	0.000201
8	7	0.160	0.194	-0.034	-0.020833	0.000434
9	8	0.163	0.160	0.003	0.016167	0.000261
10	9	0.166	0.209	-0.043	-0.029833	0.000890
11	10	0.152	0.164	-0.012	0.001167	0.000001
12	11	0.190	0.210	-0.020	-0.006833	0.000047
13	12	0.172	0.197	-0.025	-0.011833	0.000140
14			Sum=	-0.158	0.000000	0.002970
15			$\bar{d}$ =	-0.013167		
16						
17	t-Test: Paired Two Sample for Means					
18			*Dominant Hand, X1*	*Nondominant Hand, X2*		
19	Mean		0.179750	0.192917		
20	Variance		0.000307	0.000324		
21	Observations		12	12		
22	Pearson Correlation		0.572			
23	Hypothesized Mean Difference		0			
24	df		11			
25	t Stat		-2.776			
26	P(T<=t) one-tail		0.00902			
27	t Critical one-tail		1.796			
28	P(T<=t) two-tail		0.018			
29	t Critical two-tail		2.201			

Confidence Interval for a Matched-Pairs Test:

(1-α)100% CI: point estimate ± margin of error

$$\bar{d} \pm t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$$

where $t_{\alpha/2}$ is computed using df=n-1.

Independent-Samples Design using the t-Distribution—Hypothesis Test Regarding the Difference of Two Means with Unknown Population Standard Deviations (σ_1 and σ_2)

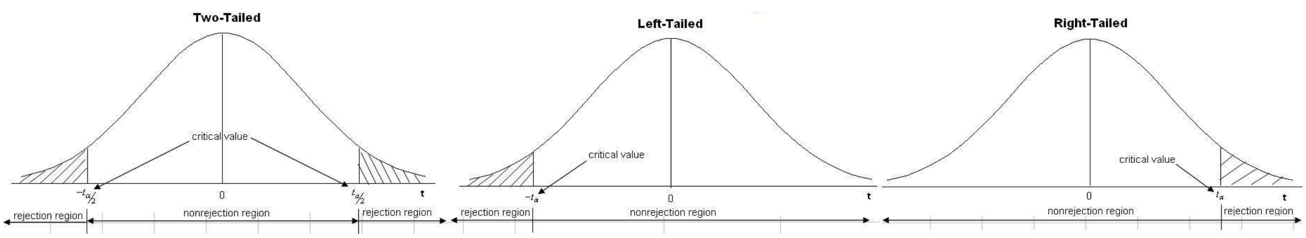
Assumptions:

- The samples are obtained using simple random sampling;
- The samples are independent;
- The populations from which the samples are drawn are normally distributed or the sample sizes are "large" ($n_1 \geq 30$ and $n_2 \geq 30$).

Step 1: A claim is made regarding the two population means (μ_1 and μ_2). The null and alternative hypotheses can be structured in three ways:

Two-Tailed	Left-Tailed	Right-Tailed
H_0: $\mu_1 = \mu_2$	H_0: $\mu_1 = \mu_2$	H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$	H_1: $\mu_1 < \mu_2$	H_1: $\mu_1 > \mu_2$
$\mu_1<\mu_2$ or $\mu_1>\mu_2$		

Step 2: Select a **level of significance**, **α**, which is generally chosen to be 0.10, 0.05, or 0.01. The critical value is determined using the smaller of n_1-1 or n_2-1 degrees of freedom. **Critical value** is the t-value that separates the rejection and nonrejection regions. The **rejection region** (or **critical region**) is the set of all values of the test statistic (defined in step 3) such that the null hypothesis is rejected.



Step 3: Calculate the **test statistic** or **calculated t-value**.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

which approximately follows Student's t-distribution. The test statistic (t) measures the number of standard deviations that the difference in sample means, ($\bar{x}_1 - \bar{x}_2$), is from the assumed difference in the population means, (μ_1-μ_2)=0.

Step 4: Draw a conclusion:

- Compare the calculated t-value (or test statistic) to the critical t-value and state whether or not H_0 is rejected at the specified α.

Two-Tailed	Left-Tailed	Right-Tailed
If $t < -t_{\alpha/2}$ *or* $t > t_{\alpha/2}$ *reject the null hypothesis*	*If* $t < -t_\alpha$ *reject the null hypothesis.*	*If* $t > t_\alpha$ *reject the null hypothesis.*

- Interpret the conclusion in the context of the problem.

Independent-Samples Design using the t-Distribution—Hypothesis Test Regarding the Difference of Two Means with Unknown Population Standard Deviations (σ_1 and σ_2)

Problem (p. 523): In the Spacelab Life Sciences 2 payload, 14 male rats were sent to space. Upon their return, the red blood cell mass (in milliliters) of the rats was determined. A *control* group of 14 male rats was held under the same conditions (except for space flight) as the space rats and their red blood cell mass was also determined when the space rats returned. Test the claim that the flight rats have a different mean red blood cell mass from the control rats at the α = 0.05 significance level.

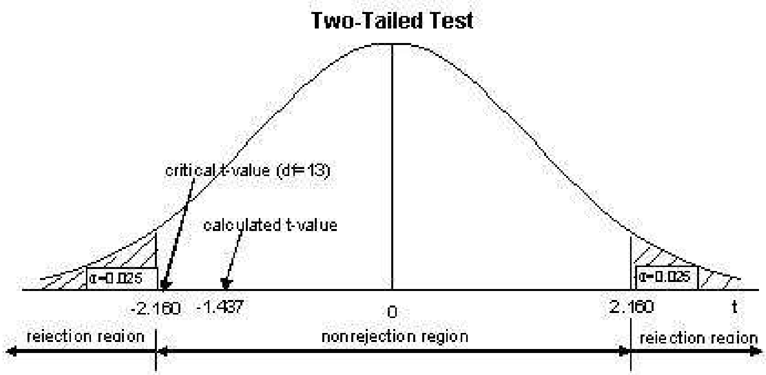
Step 1: Null and Alternative Hypotheses:

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

H_1: $\mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0$

$\mu_1 < \mu_2 \text{ or } \mu_1 > \mu_2$

Step 2: Select α = 0.05 and find the critical value of t (df=14-1=13).



Step 3: Draw a random sample of n=28 rats and randomly assign them to the flight group or the control group. Calculate the sample means, $\bar{x}_1 \text{ and } \bar{x}_2$, standard deviations, s_1 and s_2, and test statistic, *t*.

	A	B
1	Flight, X1	Control, X2
2	8.59	8.65
3	8.64	6.99
4	7.43	8.40
5	7.21	9.66
6	6.87	7.62
7	7.89	7.44
8	9.79	8.55
9	6.85	8.70
10	7.00	7.33
11	8.80	8.58
12	9.30	9.88
13	8.03	9.94
14	6.39	7.14
15	7.54	9.14
16	$\bar{x}_1 = 7.881$	$\bar{x}_2 = 8.430$
17	s_1=1.017	=STDEV(A2:A15)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(7.881 - 8.430) - 0}{\sqrt{\frac{1.017^2}{14} + \frac{1.005^2}{14}}} = \frac{-0.549}{0.3821288115} = -1.437$$

$$s_1 = \sqrt{\frac{\sum (x_{1i} - \bar{x}_1)^2}{n_1 - 1}} = \sqrt{\frac{13.4458}{14 - 1}} = 1.017$$

Step 4: Conclusion—Because the calculated t = -1.437 is greater than the left critical t = -2.160 (and in the nonrejection region), do not reject H_0 at the 0.05 significance level. There is not a significant difference in the mean red blood cell mass of the flight rats and the control rats at the 0.05 level.

The red blood cell mass problem for flight and control rats was calculated in Excel using a "t-Test: Two-Sample Assuming Unequal Variances."—the Excel output is shown at the bottom of this page. Notice that df=26 in the Excel output versus df=13 in the hand calculations above. Your textbook (p. 525) explains that a more accurate way to calculate df is using the following df formula, which is the formula used in Excel:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{1.017^2}{14} + \frac{1.005^2}{14}\right)^2}{\frac{\left(\frac{1.017^2}{14}\right)}{14-1} + \frac{\left(\frac{1.005^2}{14}\right)}{14-1}} = 25.996 \; or \; 26$$

With the higher df from Excel, the critical t is smaller (2.056 with df=26 versus 2.160 with df=13).

- The smaller critical t decreases the number of standard deviations the difference in the sample means must be from the hypothesized mean difference before the null hypothesis is rejected.
- In using the larger df, we need less evidence to reject the null hypothesis.

Excel Output for "t-Test: Two-Sample Assuming Unequal Variances"

	Flight, X1	*Control, X2*
Mean	7.880714286	8.43
Variance	1.035207143	1.010969231
Observations	14	14
Hypothesized Mean Difference	0	
df	26	
t Stat	-1.436781704	
P(T<=t) one-tail	0.081352709	
t Critical one-tail	1.705617901	
P(T<=t) two-tail	0.162705419	
t Critical two-tail	2.055529418	

Confidence Interval for (μ_1-μ_2) with Independent Samples (Unequal Variances):

(1-α)100% CI: point estimate ± margin of error

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $t_{\alpha/2}$ is computed using the smaller of n_1-1 or n_2-1 df.

Independent-Samples Design using the t-Distribution with a Pooled Variance, s_p^2.

Assumption: population variances are homogeneous, i.e., $\sigma_1^2=\sigma_2^2$.

Pooled variance: $$s_p^2 = \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{(n_1-1)+(n_2-1)}$$

Calculated t-value: $$t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s_p\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

The red blood cell mass problem for flight and control rats was calculated in Excel using a "t-Test: Two-Sample Assuming **Equal Variances**."—the Excel output is shown at the bottom of this page. Notice that the result is the same as that for "**Unequal Variances**" shown above. This is because the two sample sizes are the same (i.e., $n_1=n_2$). In situations where the sample sizes are different, the results would be different.

Excel Output for "t-Test: Two-Sample Assuming Equal Variances"

	Flight, X1	*Control, X2*
Mean	7.880714286	8.43
Variance	1.035207143	1.010969231
Observations	14	14
Pooled Variance	1.023088187	
Hypothesized Mean Difference	0	
df	26	
t Stat	-1.436781704	
P(T<=t) one-tail	0.081352709	
t Critical one-tail	1.705617901	
P(T<=t) two-tail	0.162705419	
t Critical two-tail	2.055529418	

Using a pooled variance has a statistical advantage if the two population variances are in fact equal. However, in practice, it is difficult to verify that two sample variances might be equal, so your textbook advises to use Welch's t (which assumes unequal population variances) when comparing two means from independent samples.

Confidence Interval for (μ_1-μ_2) with Independent Samples (Pooled Variance):

$(1-\alpha)100\%$ CI: point estimate ± margin of error

$$(\bar{x}_1-\bar{x}_2) \pm t_{\alpha/2}\cdot s_p\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

where $t_{\alpha/2}$ is computed using $df=(n_1+n_2-2)$.