Module 17: Two-Sample t-tests, with equal variances for the two populations

This module describes one of the most utilized statistical tests, the two-sample t-test conducted under the assumption that the two populations from which the two samples were selected have the same variance.

Reviewed 11 May 05 /MODULE 17

The General Situation

Up to this point, the focus has been on a single population, for which the observations had a normal distribution with a population mean μ and standard deviation σ. From this population, a random sample of size n provided the sample statistics $\overline{X}$ and s as estimates of μ and σ, respectively.

We created confidence intervals and tested hypotheses concerning the population mean μ, using the normal distribution when we had available the value of σ and using the t distribution when we did not and thus used the estimate s from the sample. This circumstance is often described as the *one sample situation*.

Clearly, we are often faced with making judgments for circumstances that involve more than one population and sample. For the moment, we will focus on the so-called *two sample situation*. That is, we consider two populations.

	City A	City B
Mean	μ_A	μ_B
SD	σ_A	σ_B

Question:

Do you believe the two populations have the same mean?

Two Sample Hypotheses

H_0: $\mu_A = \mu_B$ versus H_1: $\mu_A \neq \mu_B$

or equivalently

H_0: $\Delta = \mu_A - \mu_B = 0$ versus H_1: $\Delta = \mu_A - \mu_B \neq 0$.

Parameters vs. Estimates

Population 1		Population 2	
Parameter	Estimate	Parameter	Estimate
Populations of individual values			
μ_1	$\bar{x}_1$	μ_2	$\bar{x}_2$
σ_1^2	s_1^2	σ_2^2	s_2^2
σ_1	s_1	σ_2	s_2
Populations of means, samples of size n_1 and n_2			
μ_1	$\bar{x}_1$	μ_2	$\bar{x}_2$
σ_1^2/n_1	s_1^2/n_1	σ_2^2/n_2	s_2^2/n_2
$\sigma_1/\sqrt{n_1}$	$s_1/\sqrt{n_1}$	$\sigma_2/\sqrt{n_2}$	$s_2/\sqrt{n_2}$

We are interested in

$$d = \bar{x}_1 - \bar{x}_2$$

$$\Delta = \mu_1 - \mu_2$$

If the samples are independent, then

$$Var(\bar{x}_1 - \bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2)$$

$$Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

When

$$\sigma_1^2 = \sigma_2^2, \quad Var(\bar{x}_1 - \bar{x}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Estimating σ^2

When $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

we have two estimates of σ^2 , one from sample 1, namely s_1^2 and one from sample 2, namely s_2^2. How can we best use these two estimates of the same thing. One obvious answer is to use the average of the two; however, it may be desirable to somehow take into account that the two samples may not the same size. If they are not the same size, then we may want the larger one to count more.

Pooled Average

Hence, we use the weighted average of the two sample variances, with the weighting done according to sample size. This weighted average is called the *pooled estimate:*

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

Estimate of Var($\bar{x}_1 - \bar{x}_2$)

To estimate Var($\bar{x}_1 - \bar{x}_2$), we can use

$$s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$$

Example 1: Blood Pressures of Children

To investigate the question of whether the children of city A and city B have the same systolic blood pressure, a random sample of n = 10 children was selected from each city and their blood pressures measured. These samples provided the following data:

Statistic	City A	City B
N	10	10
$\bar{x}$ (mmHg)	105.8	97.2
$s^2 (mmHg)^2$	78.62	22.40
s (mmHg)	8.87	4.73

We are interested in the difference:

$$\Delta = \mu_A - \mu_B$$

and we have $\bar{x}_A$ as an estimate of μ_A and $\bar{x}_B$ as an estimate of μ_B; hence it is reasonable to use:

$$d = \bar{x}_A - \bar{x}_B = 105.8 - 97.2 = 8.6 \text{ (mm Hg)}$$

as an estimate of $\Delta = \mu_A - \mu_B$.

We then can ask whether this observed difference of 8.6 mm Hg is sufficiently large for us to question whether the two population means could be the same, that is, $\mu_A = \mu_B$. Clearly, if the two population means are truly equal, that is, if $\mu_A = \mu_B$ is true, then we would expect the two sample means also to be equal, that is $\overline{x}_A = \overline{x}_B$, except for the random error that occurs as a consequence of using random samples to represent the entire populations. The question before us is whether this observed difference of 8.6 mm Hg is larger than could be reasonably attributed to this random error and thus reflects true differences between the population means.

Confidence Interval for $\mu_A - \mu_B$, using s_p

$$C\left[(\bar{x}_A - \bar{x}_B) - t_{0.975} s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \le \mu_A - \mu_B \le (\bar{x}_A - \bar{x}_B) + t_{0.975} s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right] = 0.95$$

$$df = (n_A - 1) + (n_B - 1)$$

$\bar{x}_A - \bar{x}_B = 8.6$ $\qquad t_{0.975} = 2.1009$ $\qquad df = 18$

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A - 1) + (n_B - 1)} = \frac{9(78.62) + 9(22.4)}{18} = 50.51$$

$$S_p = \sqrt{50.51} = 7.11$$

$$C\left[\ 8.6 - 2.1009(7.11)\sqrt{\frac{1}{10}+\frac{1}{10}} \le \mu_A - \mu_B \le 8.6 + 2.1009(7.11)\sqrt{\frac{1}{10}+\frac{1}{10}}\ \right] = 0.95$$

$$C\left[\ 1.92 \le \mu_A - \mu_B \le 15.27\right] = 0.95$$

Example 2: *AJPH, April 1994; 84:p644*

TABLE 2—Nurse Practitioners' Responses to 10 Clinical Case Scenarios on Occupational Health Test

Case Scenario Type	Occupational Program Graduates' No. of Correct Responses, Mean ± SD $n = 31$	Nonoccupational Program Graduates' No. of Correct Responses, Mean ± SD $n = 223$
Occupational disease identification*	3.9 ± 1.0	2.2 ± 1.2
Occupational case management*	4.1 ± 1.2	3.4 ± 1.5
Both case scenario types*	8.0 ± 1.7	5.6 ± 2.4

Note. Five clinical case scenarios were included in each category.
**P* < .05 (occupational as compared with nonoccupational program graduates).

Example 2 (contd.)

	1 OCCP Prog	2 Non OCCP Prog
n	31	223
mean	4.1	3.4
SD	1.2	1.5
S^2	1.44	2.25

Example 2 (contd.)

1. **The hypothesis:** $H_0 : \mu_1 = \mu_2$ *vs.* $H_1 : \mu_1 \neq \mu_2$
2. **The assumptions:** Independent random samples from normal distributions, $\sigma_1^2 = \sigma_2^2 = \sigma^2$
3. **The α level:** $\alpha = 0.05$
4. **The test statistic:** $$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
5. **The critical region:** Reject H_0 if t is not between

$$\pm t_{0.975}(252) = 1.97$$

6. Test result:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} \qquad s_p^2 = \frac{30(1.44) + 222(2.25)}{30 + 222}$$

$$s_p^2 = \frac{542.7}{252} = 2.154 \qquad s_p = \sqrt{2.154} = 1.47$$

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{31} + \frac{1}{223}} = 0.19$$

$$t = \frac{4.1 - 3.4}{1.47(0.19)} = 2.5$$

7. The Conclusion: Reject H_0 since t = 2.5 is not between $\pm$ 1.97; $0.01 < p < 0.02$

Example 3: AJPH July 1994; 89:1068

TABLE 1—Sociodemographic Attributes, Health Status, and Use of Health Services in Major Ethnic Groups and Hispanic Subgroups of American Children

	Non-Hispanic Whites (n = 62572[a])	Non-Hispanic Blacks (n = 17324[a])	Hispanics (n = 12765)	Asians/Pacific Islanders (n = 2516)	Native Americans (n = 1067)	Mexican Americans (n = 7893)	Mainland Puerto Ricans (n = 1383)	Cuban Americans (n = 357)
Family income below poverty level	10% ± 0.3	40% ± 1	35% ± 2	18% ± 3	41% ± 4	40% ± 3	45% ± 4	13% ± 3
Parent graduated high school	93% ± 0.3	79% ± 1	61% ± 2	85% ± 3	78% ± 2	53% ± 2	65% ± 2	83% ± 3
Parents in household[b]								
Both	84% ± 0.3	42% ± 1	72% ± 1	86% ± 1	61% ± 3	75% ± 1	52% ± 2	71% ± 4
Mother only	11% ± 0.3	36% ± 1	18% ± 1	9% ± 1	21% ± 2	15% ± 1	35% ± 2	18% ± 3
Other	5% ± 0.1	22% ± 1	10% ± 0.4	6% ± 1	19% ± 2	10% ± 1	13% ± 1	10% ± 2
Parent rating of child's health								
Excellent or very good	85% ± 0.3	68% ± 1	74% ± 1	80% ± 1	66% ± 3	73% ± 1	70% ± 3	81% ± 2
Good, fair, or poor	15% ± 0.3	32% ± 1	26% ± 1	20% ± 1	34% ± 3	27% ± 1	30% ± 3	19% ± 2
Bed days for illness in past year, mean ± SE	3.1 ± 0.1	2.3 ± 0.1	2.4 ± 0.1	1.8 ± 0.2	3.2 ± 0.4	2.2 ± 0.2	3.3 ± 0.2	2.4 ± 0.7
Interval since last physician visit								
< 1 year	84% ± 0.3	78% ± 1	78% ± 1	79% ± 1	78% ± 2	74% ± 1	87% ± 2	78% ± 3
≥ 1 year or never	16% ± 0.3	22% ± 1	22% ± 1	21% ± 1	22% ± 2	26% ± 1	13% ± 2	22% ± 3
Number of physician visits in past year, mean ± SE	3.4 ± 0.04	2.4 ± 0.1	2.8 ± 0.1	2.3 ± 0.1	3.3 ± 0.4	2.4 ± 0.1	3.9 ± 0.3	2.7 ± 0.2

Note. Unless otherwise indicated, descriptive statistics are expressed as percentages ± standard errors (SEs).
[a]Incomplete records were omitted from the analysis.
[b]Because of rounding and small numbers of nonrespondents, the sum of a column for these characteristics may not total to 100%.

Example 3 (contd.)

	Mainland Puerto Ricans	Cuban Americans
n	1,383	357
mean	3.3	2.4
SE	0.2	0.7
$s = SE\sqrt{n}$	$(0.2)\sqrt{1383} = 7.44$	$(0.7)\sqrt{357} = 13.23$
s^2	55.35	175.03

Source: ***AJPH, July 1994; 89:1068***

1. The hypothesis: $H_0: \mu_1 = \mu_2$ *vs.* $H_1: \mu_1 \neq \mu_2$

2. The assumptions: Independent random samples from normal distributions

$\sigma_1^2 = \sigma_2^2 = \sigma^2$

3. The α level: $\alpha = 0.05$

4. The test statistic:

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

5. The critical region: Reject if t is not between

$$\pm\ t_{0.975}(1738) = 1.96$$

6. The Result:

$$s_p^2 = \frac{1382(55.35) + 356(175.03)}{1382 + 356}$$

$$s_p = \sqrt{79.86} = 8.94$$

$$\sqrt{\frac{1}{1383} + \frac{1}{357}} = 0.059$$

$$t = \frac{3.3 - 2.4}{8.94(0.059)} = \frac{0.90}{(0.527)} = 1.71$$

7. Conclusion: Accept H_0: $\mu_1 = \mu_2$, since $p > 0.05$; $0.05 < p < 0.10$

Example 4: AJPH July 1994; 89:1068

TABLE 1—Mean Values for Health Indicators and Socioeconomic Variables for Western European Countries with Social Security Systems and National Health Service Systems

	Social Security Systems (n = 7)		National Health Services (n = 10)		
	Mean	SD	Mean	SD	*P*
Gross domestic product per capita (US$)	17 799.9	1 866.2	14 281.3	3 184.3	<.01
Health care expenditures per capita (US$)	1 427.6	142.5	1 057.5	334.1	<.01
Population covered by the system, %	93.9	11.3	99.9	0.3	NS
Public health expenditures as % of total health expenditures	76.1	9.9	81.9	5.8	NS
Gini coefficient	0.3	0.04	0.3	0.04	NS
Infant mortality rate (deaths/1000 live births)	7.3	0.4	7.4	1.3	NS
PYLL, females	2 812.6	296.6	2 833.7	285.1	NS
PYLL, males	5 312.7	630.9	4 928.0	614.3	NS
Life expectancy, males, y	72.6	1.1	73.2	1.5	NS
Life expectancy, females, y	79.5	0.9	79.1	1.2	NS

Note. NS = not significant; PYLL = potential years of life lost.

1. The hypothesis: $H_0: \mu_{SSS} = \mu_{NHS}$ vs. $H_1: \mu_{SSS} \neq \mu_{NHS}$

2. The α level: $\alpha = 0.05$

3. The assumptions: Independent Samples, Normal Distribution, $\sigma^2_{SSS} = \sigma^2_{NHS}$

4. The test statistic:

$$t = \frac{\bar{X}_{SSS} - \bar{X}_{NHS}}{S_p \sqrt{\frac{1}{n_{SSS}} + \frac{1}{n_{NHS}}}}$$

5. The critical region: Reject if t is not between ± 2.1315

6. The result :

$$s_p^2 = \frac{(n_{SSS}-1)s_{SSS}^2 + (n_{NHS}-1)s_{NHS}^2}{(n_{SSS}-1)+(n_{NHS}-1)}$$

$$= \frac{6(142.5)^2 + 9(334.1)^2}{6+9} = 75,096$$

$$s_p = 274.0$$

$$t = \frac{1427.6 - 1057.5}{274.0\sqrt{\frac{1}{7}+\frac{1}{10}}} = \frac{370.1}{274.0(0.49)} = 2.75$$

7. The conclusion: Reject H_0: $\mu_{SSS} = \mu_{NHS}$; $0.01 < p < 0.02$

Independent Random Samples from Two Populations of Serum Uric Acid Values

	Sample 1	Sample 2
	1.2	1.7
	0.8	1.5
	1.1	2.0
	0.7	2.1
	0.9	1.1
	1.1	0.9
	1.5	2.2
	0.8	1.8
	1.6	1.3
	0.9	1.5
Sum	10.6	16.1
Mean	1.06	1.61

Serum Acid Worksheet

	Sample 1		Sample 2	
	x	x^2	x	x^2
	1.2	1.44	1.7	2.89
	0.8	0.64	1.5	2.25
	1.1	1.21	2.0	4.00
	0.7	0.49	2.1	4.41
	0.9	0.81	1.1	1.21
	1.1	1.21	0.9	0.81
	1.5	2.25	2.2	4.84
	0.8	0.64	1.8	3.24
	1.6	2.56	1.3	1.69
	0.9	0.81	1.5	2.25
Sum	10.6	12.06	16.1	27.59
Mean	1.06		1.61	
Sum^2/n	1.236		25.921	
SS	0.824		1.669	
Variance	0.092		0.185	
SD	0.303		0.431	

$$s_1^2 = 0.09, \quad s_2^2 = 0.19$$

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

$$s_p^2 = \frac{(9)(0.09) + (9)(0.19)}{9 + 9}$$

$$s_p^2 = \frac{0.81 + 1.71}{18} = 0.14, \quad s_p = 0.37$$

Testing the Hypothesis That The Two Serum Uric Acid Populations Have The Same Mean

1. The hypothesis: $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$

2. The α-level: $\alpha = 0.05$

3. The assumptions: Independent Random Samples
Normal Distribution, $\sigma_1^2 = \sigma_2^2$

4. The test statistic:

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

5. The reject region: Reject H_0: $\mu_1 = \mu_2$ if t is not between $\pm\, t_{0.975}(18) = 2.1009$

6. The result:

$$t = \frac{1.06 - 1.61}{0.37\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{-0.55}{0.37(0.45)} = -3.3$$

7. The conclusion: Reject H_0: $\mu_1 = \mu_2$, since t is not between $\pm$ 2.1009

Serum Uric Acid Values Before And After a Special Meal

	Before	After
	1.2	1.7
	0.8	1.5
	1.1	2.0
	0.7	2.1
	0.9	1.1
	1.1	0.9
	1.5	2.2
	0.8	1.8
	1.6	1.3
	0.9	1.5
Sum	10.6	6.1
Mean	1.06	1.61

Serum Uric Acid Values Before And After A Special Meal

Worksheet

Person	Before	After	d^a	d^2
1	1.2	1.7	0.5	0.25
2	0.8	1.5	0.7	0.49
3	1.1	2.0	0.9	0.81
4	0.7	2.1	1.4	1.96
5	0.9	1.1	0.2	0.04
6	1.1	0.9	-0.2	0.04
7	1.5	2.2	0.7	0.49
8	0.8	1.8	1.0	1.00
9	1.6	1.3	-0.3	0.09
10	0.9	1.5	0.6	0.36
Sum	10.6	16.1	5.5	5.53
Mean	1.06	1.61	0.55	
Sum^2/n			3.025	
SS			2.505	
Variance			0.278	
SD			0.528	

d^a = After - Before

Testing the Hypothesis That The Serum Uric Acid Levels Before and After A Special Meal Are The Same

1. The hypothesis: $H_0: \Delta = 0$ vs. $H_1: \Delta \neq 0$,
where $\Delta = \mu_{After} - \mu_{Before}$

2. The α-level: $\alpha = 0.05$

3. The assumptions: Random Sample of Differences,
Normal Distribution

4. The test statistic:

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{\bar{x}_{After} - \bar{x}_{Before}}{s_d \big/ \sqrt{n}}$$

5. The rejection region: Reject H_0: $\Delta = 0$, if t is not between $\pm\, t_{0.975}(9) = 2.26$

6. The result:

$$t = \frac{0.55}{0.528/\sqrt{10}} = \frac{0.55}{0.528/(3.16)} = 3.29$$

7. The conclusion: Reject H_0: $\Delta = 0$ since t is not between $\pm$ 2.26