Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

Comparison of Means
Equal Population Variances

We select two independent random samples:
from population 1 of size n_1 with mean $\bar{x}_1$ and variance s_1^2
from population 2 of size n_2 with mean $\bar{x}_2$ and variance s_2^2

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{s_P^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Estimate of Standard Error

Pooled Estimate of Variance

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

$$df = n_1 + n_2 - 2$$

For large n's, use z

Comparison of Means
Equal Population Variances
Example

Compare a new method of teaching reading to "slow learners" to the current standard method.

You decide to base this comparison on the results of a reading test given at the end of a learning period of 6 months.

Of a random sample of 22 slow learners, 10 are taught by the new method and 12 are taught by the standard method. Qualified instructors under similar conditions teach all 22 children for a 6-month period. The results of the reading test at the end of 6 months are as follows:

New Method: $\bar{x}_1 = 76.4$; $s_1^2 = 34.04$

Standard Method: $\bar{x}_2 = 72.33$; $s_2^2 = 40.24$

Are the reading scores of children using the new method greater than those of children using the standard method with alpha=.05?

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

Comparison of Means Equal Population Variances Solution

H0: $\mu_1 - \mu_2 \leq 0$

Ha: $\mu_1 - \mu_2 > 0$

a = .05

df = $n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ df

Critical Value(s):

n<30, so use t. $t_? = 1.725$

Decision Rule:

$t > t_?$

Test Statistic:

Decision:

Conclusion:

Small-Sample t Test Solution

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(m_1 - m_2\right)}{\sqrt{S_P^{\,2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\left(76.4 - 72.33\right) - \left(0\right)}{\sqrt{37.45 \cdot \left(\frac{1}{10} + \frac{1}{12}\right)}} = 1.55$$

$$S_P^{\,2} = \frac{\left(n_1 - 1\right) \cdot S_1^{\,2} + \left(n_2 - 1\right) \cdot S_2^{\,2}}{n_1 + n_2 - 2}$$

$$= \frac{\left(10 - 1\right) \cdot \left(34.04\right) + \left(12 - 1\right) \cdot \left(40.24\right)}{10 + 12 - 2} = 37.45$$

Comparison of Means Equal Population Variances Solution

H0: $\mu_1 - \mu_2 \leq 0$

Ha: $\mu_1 - \mu_2 > 0$

a = .05

df = $n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ df

Critical Value(s):

n<30, so use t. t? =1.725

Decision Rule:

t>t?

Test Statistic:

$$t = \frac{(76.4 - 72.33) - (0)}{\sqrt{37.45 \cdot \left(\frac{1}{10} + \frac{1}{12}\right)}} = 1.55$$

Decision:

t is less than t?. We cannot reject the null hypothesis.

Conclusion:

We do not have enough evidence to reject the null hypothesis and conclude that the new method of testing does not improve reading scores.

Confidence Interval for the Difference (equal population variances)

Does one method of teaching reading produce better results than another?

- How much better are the results of the method?

Construct a 95% confidence interval for the difference between the two means and interpret it.

$$\left(\overline{x}_1 - \overline{x}_2\right) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = (76.4 - 72.33) \pm 2.086\sqrt{37.45\left(\frac{1}{10}+\frac{1}{12}\right)}$$

$$= 4.07 \pm (2.086)(2.62) = 4.07 \pm 5.47 = [-1.4, 9.54]$$

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

Paired-Sample t Test Test Statistic

$$z \text{ or } t = \frac{\overline{x}_D - D_0}{\frac{s_D}{\sqrt{n_D}}}$$

$$df = n_D - 1$$

When n>30,use z
When n<30 use t(n-1) df

D_0=0 when testing whether there is any difference or not.

Sample Mean

$$\overline{x}_D = \frac{\sum_{i=1}^{n} D_i}{n_D}$$

Sample Standard Deviation

$$s_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \overline{x}_D)^2}{n_D - 1}}$$

Paired-Sample t Test Example

Do energy conservation efforts really reduce consumption over time? By how much?

A study is undertaken to determine how consumers react to energy conservation efforts. A random group of 60 families is chosen. Each family's rate of consumption of electricity is monitored in equal length time periods before and after they are offered financial incentives to reduce their energy consumption rate. The difference in electric consumption between the periods is recorded for each family. The average reduction in consumption is 0.2 kW and the standard deviation of the differences s_D=1.0 kW. At α=0.01, is there evidence to conclude that the incentives reduce consumption?

Paired-Sample t Test Solution

H0: $\mu_D = 0$

Ha: $\mu_D < 0$

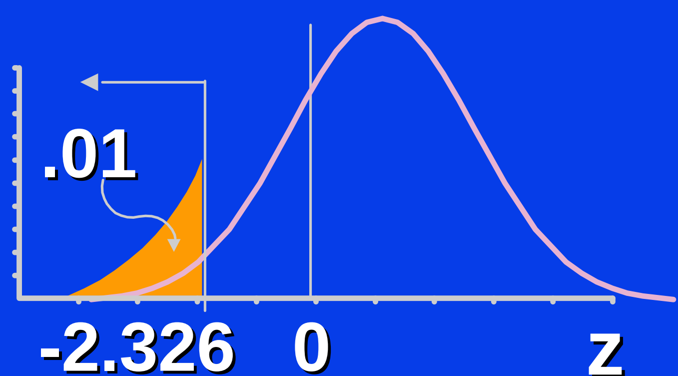
a = .01

Decision Rule:

n>30, so use z;

z<z?

Critical Value(s):



Test Statistic:

$$t = \frac{\bar{x}_D - D_0}{\frac{s_D}{\sqrt{n_D}}} = \frac{-0.2 - 0}{\frac{1.0}{\sqrt{60}}} = -1.55$$

Decision: z is not less than z?; we cannot reject the null hypothesis.

Conclusion:

We do not have enough evidence to say that incentives reduce consumption.

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

Confidence Interval for Paired Observations

Do energy conservation efforts really reduce consumption over time? By how much?

A (1-α)100% confidence interval for the mean difference is constructed using the t distribution for small sample sizes and z distribution for large sample sizes.

$$\bar{x}_D \pm z_{a/2} \frac{s_D}{\sqrt{n}}$$

When n>30,use z
When n<30 use t(n-1) df

$$\bar{x}_D \pm z_{a/2} \frac{s_D}{\sqrt{n}} = -0.2 \pm 2.576 \frac{1.0}{\sqrt{60}} = [-0.5, 0.1]$$

Z Test for Difference in Two Proportions

Is the proportion of sub-prime mortgages to low-income households greater than than for moderate-income households? How much greater?

In 1998, a sample of mortgages was taken from the over 1 million mortgages disclosed nationally under HMDA. Here is a decription of the sample:

Income Group	Percent Subprime	n
Low-income	26%	400
Moderate-income	11%	600

Is there sufficient evidence to claim that the proportion of sub-prime mortgages to low-income households exceeds that among moderate income households? Test using ?=.01

Portions of these notes are adapted from Statistics 7e © 1997 Prentice-Hall, Inc.

Z Test for Two Proportions Solution

H_0: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 > 0$

a = .01

***n_1* =** 400 ***n_2* =** 600

Decision Rule:

$z > z_?$

Critical Value(s):

$z_? = 2.326$

Test Statistic:

Decision:

Conclusion:

Z Test for Two Proportions Solution

$n_1 \cdot \hat{p}_1 = 400 \cdot .26 = 104$ $n_2 \cdot \hat{p}_2 = 600 \cdot .11 = 66$

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{104 + 66}{400 + 600} = .17$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(.26 - .11) - (0)}{\sqrt{(.17) \cdot (1 - .17) \cdot \left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$$= 6.19$$

Z Test for Two Proportions Solution

H_0: $p_1 - p_2 = 0$

H_a: $p_1 - p_2 > 0$

a = .01

n_1 = 400 **n_2 =** 600

Decision Rule:

$z > z_?$

Critical Value(s):

$z_? = 2.326$

Test Statistic:

Z=6.19

Decision:

z is greater than $z_?$. There is evidence to reject the null hypothesis at the 1% level.

Conclusion:

We have evidence that higher proportions of low-income households receive subprime mortgages than moderate income households.

Confidence Interval for the Difference in Proportions

Is the proportion of sub-prime mortgages to low-income households greater than than for moderate-income households? How much greater?

A large sample (1-α)100% confidence interval for the difference between two population proportions:

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

How much greater is the proportion of subprime mortgages to low-income buyers compared to moderate-income buyers?

$$(.26 - .11) \pm 2.575\sqrt{\frac{.26(1-.26)}{400} + \frac{.11(1-.11)}{600}} = .15 \pm 2.575(0.02538) = [.12, .18]$$