MA 214: Applied Statistics

Instructor: Ashis Gangopadhyay

Hypothesis Testing

Where We've Been

- Discussing inferential methods for one sample data and to "describe" the true value of a population parameter via
  - One sample confidence interval
  - One sample hypothesis testing

Where We're Going

- Two-sample inference where the objective is to "compare" between two populations parameters via
  - Confidence interval for the difference
  - Test of hypothesis for the difference

Two-Sample Inference: Confidence Interval and Hypothesis Testing

Data Types

- **Example 6:** Reaction times (in seconds) of six individuals with alcoholic stimulant and eight individuals without any stimulant.

(Two Sample Data / Independent Samples)

Reaction Times								
Without Stimulant	3.0	2.0	1.0	2.5	1.5	4.0	1.0	2.0
With Stimulant	5.0	4.0	3.0	4.5	2.0	2.5		

Two-Sample Inference: Confidence Interval and Hypothesis Testing

Data Types

- **Example 7:** Weights of eight individuals, before and after they participated in a diet program

(Paired Data)

Person	Before, X	After, Y	Difference, D=X-Y
1	175	170	5
2	168	169	-1
3	140	133	7
4	130	132	-2
5	143	150	-7
6	128	116	12
7	138	130	8
8	165	163	2

$\overline{X}_D = 3.0$

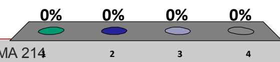
$s_D = 6.18$

Clicker Question……

- A researcher is investigating which of two newly developed automobile engine oils is better at prolonging the life of an engine.
- Since there are a variety of automobile engines, 20 different engine types were randomly selected and were tested using each of the two engine oils.
- The number of hours of continuous use before engine breakdown was recorded for each engine oil.

The data type for this experiment is……

1. One-sample data
2. Two-sample data, independent samples
3. ✔ Paired data
4. Can't say without reviewing the observations



Two-Sample Inference: Confidence Interval and Hypothesis Testing

Consider a two-sample data (independent samples)

	Population 1	Population 2
Mean	μ_1	μ_2
Standard Deviation	σ_1	σ_2
Sample Data	$X_1,\dots,X_{n1}$	$Y_1,\dots,Y_{n2}$
Sample Mean	$\bar{X}$	$\bar{Y}$
Sample Standard Deviation	s_1	s_2

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 8:** ("The Residual Cognitive Effect of Heavy Marijuana Use in College Students" by Pope and Yurgelun-Todd, JAMA, Vol. 275)

A study was conducted to measure the effect of heavy marijuana use on college students. Light and Heavy users were tested for memory recall.

Items sorted correctly by light users: n = 64, $\bar{X} = 53.3, s_1 = 3.6$

Items sorted correctly by heavy users: n = 65, $\bar{Y} = 51.3, s_2 = 4.5$

At 1% level of significance, can we conclude that the heavy users have lower mean score than light users?

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 9:** (Example 8 continued) Construct a 95% confidence interval for the difference between the two population means.

Items sorted correctly by light users: n = 64, $\bar{X} = 53.3, s_1 = 3.6$

Items sorted correctly by heavy users: n = 65, $\bar{Y} = 51.3, s_2 = 4.5$

? ?
(----------- ($\mu_1 - \mu_2$) -----------)

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Two-sample hypothesis testing for mean**

(Large sample case: $n_1 \geq 30$ and $n_2 \geq 30$)

Assumptions:

1. *Both sample sizes are large.*
2. *Samples are independent. (Two-sample data: independent samples).*

Step 1: Write down the null and the alternative hypothesis.

Step 2: Calculate the test statistic:

$$Z = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Step 3:** Decision Rule

Case 1:	H_0: $\mu_1-\mu_2 = D_0$ vs. H_a: $\mu_1-\mu_2 > D_0$
	Reject H_0 if Z > z_α
Case 2:	H_0: $\mu_1-\mu_2 = D_0$ vs. H_a: $\mu_1-\mu_2 < D_0$
	Reject H_0 if Z < $-z_\alpha$
Case 3:	H_0: $\mu_1-\mu_2 = D_0$ vs. H_a: $\mu_1-\mu_2 \neq D_0$
	Reject H_0 if Z > $z_{\alpha/2}$, OR if Z < $-z_{\alpha/2}$

Note: If population standard deviations are unknown, then replace σ_1 and σ_2 by the sample standard deviations s_1 and s_2.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **A large sample (1 – α) confidence interval for ($\mu_1 - \mu_2$)**

$$\left(\bar{X} - \bar{Y}\right) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 8 (contd.):** Let μ_1 and μ_2 be the mean scores among the light and heavy users respectively.

H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 > 0$

Test Statistic:

$$Z = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{53.3 - 51.3}{\sqrt{\frac{3.6^2}{64} + \frac{4.5^2}{65}}} = \frac{2}{0.717} = 2.79$$

Decision Rule:

Reject H_0 if $Z > z_{0.01}$ ⇔ Reject H_0 if $Z > 2.33$

Thus, reject H_0 and conclude that the average score among the heavy users is lower than that of the light users.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 9 (contd.):** 95% confidence interval for ($\mu_1 - \mu_2$) is given by

$$(53.3 - 51.3) \pm 1.96\sqrt{\frac{3.6^2}{64} + \frac{4.5^2}{65}}$$

⇔

$$2 \pm 1.41 \equiv (0.59, 3.41)$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 10:**
  A study was conducted to determine the effect of alcohol.

The errors were recorded in a test of visual and motor skills for a treatment group of people who drank ethanol, and another group of people who were given a placebo.

Treatment Group: $n = 22, \bar{X} = 4.20, s_1 = 2.2$

Placebo Group: $n = 22, \bar{X} = 1.71, s_1 = 1.72$

At 5% level, test the hypothesis that there is a difference between the two population means.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 11:**
  Refer to the sample data in the last example (example 10):

Treatment Group: $n = 22, \bar{X} = 4.20, s_1 = 2.2$

Placebo Group: $n = 22, \bar{X} = 1.71, s_1 = 1.72$

Construct a 95% confidence interval for the difference in the two population means.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 1 Revisited:**
  Reaction times (in seconds) of six individuals with alcoholic stimulant and eight individuals without any stimulant.

REACTION TIMES

WITHOUT STIMULANT	3.0	2.0	1.0	2.5	1.5	4.0	1.0	2.0
WITH STIMULANT	5.0	4.0	3.0	4.5	2.0	2.5		

SAS TTEST PROCEDURE

Variable: TIME

GROUPS	N	Mean	Std Dev	Std Error	Min
Non	8	2.12500000	1.02643628	0.36290003	1.0000
Intox	6	3.50000000	1.18321596	0.48304589	2.0000

Variances	T	DF	Prob>|T|
Unequal	-2.2758	10.0	0.0462
Equal	-2.3262	12.0	0.0383

For H0: Variances are equal, F' = 1.33 DF = (5,7) Prob>F=0.7053

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Two-sample test of hypothesis about population mean difference ($\mu_1 - \mu_2$)** (Small sample case)

Assumptions:

1. *Both populations are Normally distributed.*
2. *Two samples are independent.*
3. *Two population variances are equal.*

Step 1: Write down the null and the alternative hypothesis.

Step 2: Calculate the test statistic:

$$t = \frac{\bar{X} - \bar{Y} - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{where: } s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Step 3:** Decision Rule

Case 1:	H_0: μ_1-μ_2 = D_0 vs. H_a: μ_1-μ_2 > D_0
	Reject H_0 if $t > t_\alpha$
Case 2:	H_0: μ_1-μ_2 = D_0 vs. H_a: μ_1-μ_2 < D_0
	Reject H_0 if $t < -t_\alpha$
Case 3:	H_0: μ_1-μ_2 = D_0 vs. H_a: μ_1-μ_2 ≠ D_0
	Reject H_0 if $t > t_{\alpha/2}$, OR if $t < -t_{\alpha/2}$

Note: t_α and $t_{\alpha/2}$ are based on (n_1+n_2-2) degrees of freedom.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 10 (contd.):** Let μ_1 and μ_2 be the mean scores of the treatment and placebo groups respectively.

H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 \neq 0$

Pooled standard deviation:

$$s_p = \sqrt{\frac{21 \times 2.2^2 + 21 \times 1.72^2}{22 + 22 - 2}} = 1.975$$

Test Statistic:

$$t = \frac{4.20 - 1.71}{1.975\sqrt{\frac{1}{22} + \frac{1}{22}}} = 4.18$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 10 (contd.):** Let μ_1 and μ_2 be the mean scores of the treatment and placebo groups respectively.

H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 \neq 0$

Test Statistic: 4.18 (as calculated on previous page)

Decision Rule:

Reject H_0 if $t > t_{0.025}$ OR if $t < -t_{0.025}$ (based on 22+22-2) = 42 df

⇔ Reject H_0 if $t > 2.021$ OR if $t < -2.021$

We conclude that the mean error scores of the treatment and control groups are statistically different.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Small sample confidence interval for ($\mu_1 - \mu_2$)**

Assumptions: *Same as the two sample hypothesis testing (small sample case).*

Confidence interval with confidence coefficient (1-α) is given by:

$$(\bar{X} - \bar{Y}) \pm t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 11 (Contd.)**

We wish to compute a 95% CI for ($\mu_1 - \mu_2$).

$1 - \alpha = 0.95$, $\alpha = 0.05$, $t_{\alpha/2} = t_{0.025} = 2.021$ (based on 42 df)

Thus, a 95% CI for ($\mu_1 - \mu_2$) is given by

$$(4.20 - 1.71) \pm 2.021(1.975)\sqrt{\frac{1}{22} + \frac{1}{22}}$$

$$\equiv 2.49 \pm 1.20$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 13**
  A medical research was conducted to determine whether there is a difference between effectiveness of two pain relief medicines used for headaches.

  Over a six month period, a sample of individuals used one of the medicines and a second sample of individuals used the other.

  Data collected during the study showed the time required to receive the pain relief.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 13**

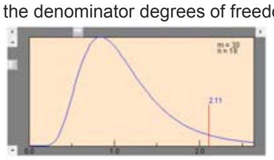
 

Individuals using Medicine 1	Individuals using Medicine 2
$n_1 = 8$	$n_2 = 12$
$\overline{X} = 24.8 \text{ min}$	$\overline{Y} = 26.1 \text{ min}$
$s_1 = 3.3 \text{ min}$	$s_2 = 4.2 \text{ min}$

Obtain a 95% confidence interval for the difference in the mean effects of the two medicines.

F-Distribution

- A continuous distribution that sits on the positive side of the real line, and shaped skewed to the right.
- F distribution is determined by two parameters (v_1,v_2)
- v_1 is called the numerator degrees of freedom
- v_2 is called the denominator degrees of freedom



http://www.econtools.com/jevons/java/Graphics2D/FDist.html

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Two-sample test of hypothesis about population variance σ²**

Assumptions:

1. *Both populations are Normally distributed.*
2. *Two samples are independent.*

We consider three cases separately.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Case 1:**

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_a : \sigma_1^2 > \sigma_2^2$$

- **Test Statistic:**

$$F = \frac{s_1^2}{s_2^2}$$

- **Decision Rule:**

  Reject H_0 if $F > F_\alpha$ where F_α is based on (n_1-1, n_2-1) df.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Case 2:**

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_a : \sigma_1^2 < \sigma_2^2$$

- **Test Statistic:**

$$F = \frac{s_2^2}{s_1^2}$$

- **Decision Rule:**

  Reject H_0 if $F > F_\alpha$ where F_α is based on (n_2-1, n_1-1) df.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Case 3:**

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_a : \sigma_1^2 \neq \sigma_2^2$$

- **Test Statistic:**

$$F = \frac{\max\left(s_1^2, s_2^2\right)}{\min\left(s_1^2, s_2^2\right)}$$

- **Decision Rule:**

Reject H_0 if $F > F_{\alpha/2}$ where the df is given by case 1 or case 2 depending on the form of the F-statistic.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 10 (Contd.):**

We wish to verify if the equal variance condition is satisfied at 5% level of significance.

Treatment Group: $n = 22, \bar{X} = 4.20, s_1 = 2.2$

Placebo Group: $n = 22, \bar{X} = 1.71, s_1 = 1.72$

Test Statistic: $F = \frac{(2.2)^2}{(1.7)^2} = 1.67$

Decision Rule: Reject H_0 if $F > F_{0.025}$ ⇔ Reject H_0 if $F > 2.4247$

Thus, there is not enough evidence in the data to conclude that the two variances are different.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Unequal variance two-sample t-test for independent samples**

If the population variances are unequal, then the pooled variance based t-test is not applicable. Instead, one can carry out the Welch-Satterthwaite approximate t-test. The test statistic is given by:

Test Statistic: $$t = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Unequal variance two-sample t-test for independent samples**

The decision follows the same rule as in the equal variance t-test, with one important difference. The df of the t-distribution is not (n_1+n_2-2) anymore. The df is now given by the formula

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Unequal variance two-sample t-test for independent samples**

Thus, the df is calculated from the sample data. It is possible that v may not be a whole number. In that case, we look up the t-table for the closest df listed in the table.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 14**

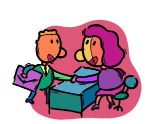
In a wage discrimination case involving male and female employees, independent samples of male and female employees with 5 years or more experience provided the following hourly wage results:

Male

$n_1 = 21$

$\bar{X} = \$9.25$

$s_1 = \$1.4$



Female

$n_2 = 16$

$\bar{Y} = \$8.70$

$s_2 = \$0.80$

Is there evidence that the female employees are paid less than their male counterpart? Use α = 0.01.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Solution:**
  First note that the F-statistics for testing the equality of the variance is given by

$$F = \frac{(1.4)^2}{(0.8)^2} = 3.06$$

The tabulated value of $F_{0.025}$ = 2.7559 for (20,15) df. Thus, we reject the H_0 at 5% level of significance and conclude that there is enough evidence to conclude that the variances are unequal.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Solution (Contd.):**
  Now let μ_1 and μ_2 be the mean hourly wages of the male and female employees respectively. We wish to test

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_a : \mu_1 - \mu_2 > 0$$

Since the variances are unequal, we will carry out the Welch-Satterthwaite approximate t-test. The test statistic is given by

$$t = \frac{9.75 - 8.70}{\sqrt{\frac{(1.4)^2}{21} + \frac{(0.8)^2}{16}}} = 2.87$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

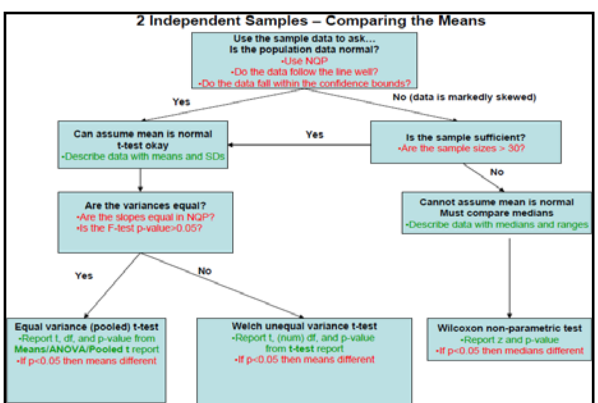
- **Solution (Contd.):**
  Thus the rejection rule is to reject H_0 if t > $t_{0.01}$. However, the df of the t-distribution is given by

$$v = \frac{\left(\frac{(1.4)^2}{21} + \frac{(0.8)^2}{16}\right)^2}{\frac{(1.4)^4}{21^2(21-1)} + \frac{(0.8)^4}{16^2(16-1)}} = 32.78$$

Thus, we look up the $t_{0.01}$ value from the t-table with 33 df, which is 2.46 (based on 30 df). That is, reject H_0 if t > 2.46.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Solution (Contd.):**
  Since the value of the t-statistic is 2.87 and that is greater than 2.46, we reject H_0. There is enough evidence in the data to conclude that the mean hourly wage of the male employees is significantly higher than the female employees.



Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 16**
  A study was conducted to determine if a new method to help fourth grade students with their deficiencies in basic math skills is more effective than the standard method.

  Students were paired off according to their deficiency levels using a baseline measure, and one student from each pair was randomly assigned to one of the two methods, and the other student in the pair was assigned to the other method.

  After each pair received a certain amount of training, their math proficiency levels was measured.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 16**

Pair	New Method Score (X)	Standard Method Score (Y)	Difference D = X-Y
1	77	72	5
2	74	68	6
3	82	76	6
4	73	68	5
5	87	84	3
6	69	68	1
7	66	61	5
8	80	76	4

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 16**
  From the data gathered, we can calculate:

Mean: $\bar{X} = 76.0, \bar{Y} = 71.6$

SD: $s_X = 6.93, s_Y = 7.0$

F: 1.02 (do not reject the null hypothesis of equal variance)

- Additionally,

Pooled standard deviation: 6.96

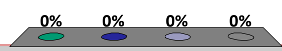
Two-sample t-test: t=1.26, p-value =0.1149

Is that the right conclusion?

Conclusion: There is NO significant difference between the two methods!

Is this the right conclusion?

1. Yes, as long as the computation is correct.
2. No, we didn't check normality of the populations.
3. No, the method used is not applicable.
4. Not enough information to judge.



Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Paired t-test**

Assumptions:

1. *The differenced data ($D_i = X_i - Y_i$) must be a sample from a normal population.*

Step 1: Write down the null and the alternative hypothesis.

Step 2: Calculate the test statistic:

$$t = \frac{\bar{D} - D_0}{s_D / \sqrt{n}}$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Step 3:** Decision Rule ($\mu_D = \mu_1 - \mu_2$)

Case 1:	$H_0: \mu_D = D_0$ vs. $H_a: \mu_D > D_0$
	Reject H_0 if $t > t_\alpha$
Case 2:	$H_0: \mu_D = D_0$ vs. $H_a: \mu_D < D_0$
	Reject H_0 if $t < -t_\alpha$
Case 3:	$H_0: \mu_D = D_0$ vs. $H_a: \mu_D \neq D_0$
	Reject H_0 if $t > t_{\alpha/2}$, OR if $t < -t_{\alpha/2}$

Note: t_α and $t_{\alpha/2}$ are based on (n – 1) degrees of freedom.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 16:**

$$\sum D_i = 35, \sum D_i^2 = 173$$

$$\bar{D} = \frac{35}{8} = 4.375,$$

$$s_p = \sqrt{\frac{1}{n-1}\left(\sum D_i^2 - \frac{\left(\sum D_i\right)^2}{n}\right)} = \sqrt{\frac{1}{7}\left(173 - \frac{(35)^2}{8}\right)} = 1.685$$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 16:**

We will use a paired t-test for the following hypothesis:

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_a: \mu_1 - \mu_2 > 0$$

Test Statistic:

$$t = \frac{\overline{D} - D_0}{\frac{s_D}{\sqrt{n}}} = \frac{4.375 - 0}{\frac{1.685}{\sqrt{8}}} = 7.34$$

Reject H_0 if t > $t_{0.05}$ (at 7 df); or p-value = 0.

Reject H_0 and conclude that there is overwhelming evidence that the new method is better than the standard method.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 17**

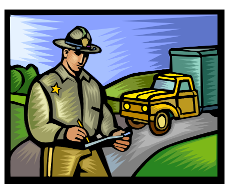
Records of a hospital show that 52 men in a sample of 1000 men vs. 23 women in a sample of 1000 women were admitted because of a heart disease.

Do these data present sufficient evidence that indicate a higher rate of heart disease among men admitted to the hospital?

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 18**

A sample of driving records over a two year period show that 16 of 400 adult drivers had received traffic citations, whereas 24 of 300 teenage drivers had received traffic citations.



Use the data to construct a 95% confidence interval of the difference between the rates of traffic citations among adult and teenage drivers.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Test for the equality of two proportions**

Test Statistic:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where: $\hat{p}_1 = \frac{x}{n_1}, \hat{p}_2 = \frac{y}{n_2}, \text{and } \hat{p} = \frac{x+y}{n_1+n_2}$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Step 3:** Decision Rule

Case 1:	H_0: p_1-p_2 = 0	vs. H_a: p_1-p_2 > 0
	Reject H_0 if Z > z_α	
Case 2:	H_0: p_1-p_2 = 0	vs. H_a: p_1-p_2 < 0
	Reject H_0 if Z < -z_α	
Case 3:	H_0: p_1-p_2 = 0	vs. H_a: p_1-p_2 ≠ 0
	Reject H_0 if Z > $z_{\alpha/2}$, OR if Z < -$z_{\alpha/2}$	

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 17 (continued)**

p_1 = rate of heart attack among men admitted to the hospital

p_2 = rate of heart attack among women admitted to the hospital

$$H_0: p_1 - p_2 = 0 \quad \text{vs.} \quad H_a: p_1 - p_2 > 0$$

$$\hat{p}_1 = \frac{52}{1000} = 0.052, \quad \hat{p}_2 = \frac{23}{1000} = 0.023$$

And the pooled proportion: $\hat{p} = \frac{52+23}{1000+1000} = 0.0375$

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 17 (continued)**

Thus, the test statistic:

$$Z = \frac{0.052 - 0.023}{\sqrt{0.0375(1-0.0375)\left(\frac{1}{1000}+\frac{1}{1000}\right)}} = 3.41$$

Note that the p-value of the test, $P(Z > 3.41) \approx 0.$

Thus, we reject H_0, and conclude that the rate of heart attack among men admitted to the hospital is greater than that among women.

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Confidence Interval for the difference between two proportions ($p_1 - p_2$):**

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

Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 18 (continued)**

Let p_1 be the rate among adult drivers who received citation and p_2 is the rate among teenage drivers.

The sample proportions:

$$\hat{p}_1 = \frac{16}{400} = 0.04 \text{ and } \hat{p}_2 = \frac{24}{300} = 0.08$$

Thus 95% CI for (p_1-p_2) is given by:

$$(0.04-0.08) \pm 1.96\sqrt{\frac{0.04\times 0.96}{400}+\frac{0.08\times 0.92}{300}} \equiv (-0.076, -0.004)$$

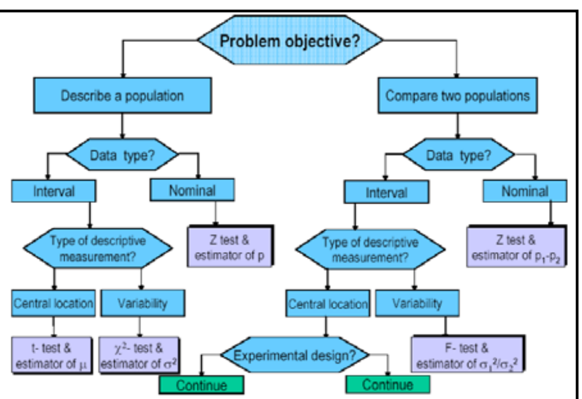
Two-Sample Inference: Confidence Interval and Hypothesis Testing

- **Example 18 (continued)**

Thus 95% CI for (p_1-p_2) is given by:

$$(-0.076, -0.004)$$

Thus with 95% confidence we can say that the difference between the two rates belongs to the interval (-0.076,-0.004).



Two-Sample Inference: Confidence Interval and Hypothesis Testing