Lecture 25
The Sign Test
Tests for two groups

Example 2: The Sign Test

In many states, the title of “Chess Master – X” is given to the students in grade X who rank in the States top 15 chess players in that grade. The 8th grade Chess Masters from New York and Los Angeles play in the U.S Championship. The wins (1), losses (-1), and ties (0) from 15 games appear below

Pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LA	0	1	-1	1	0	1	0	1	1	0	0	-1	1	1	1
NY	0	-1	1	-1	0	-1	0	-1	-1	0	0	1	-1	-1	-1
sign		+	-	+		+		+	+			-	+	+	+

Independent vs Dependent Samples

Often, we are interested in comparing two groups in statistical inference.

- Comparing the proportion of registered Democrats who are in favor expanding Medicare to the proportion of Republicans
- Comparing the water quality in two different rivers to assess which one has lower levels of pollution
- Most comparisons of two groups use independent samples

- **Independent samples** – when the observations in one sample are independent (have no statistical association) of the observations in the other sample – experiments that use randomization to allocate subjects to treatment groups result in independent samples!

- **Dependent samples** – when the observations in one sample are associated with the observations in another sample – this can result when the same subjects are used for each sample such as matched pair designs
  - A more technical definition is that the distribution of an observation in one sample would depend on the value of an observation in the other sample

Comparing two groups

A comparison of two groups is a type of **bivariate analysis** a statistical analysis which consists of two variables: the **response variable** and the **explanatory variable**

- **The explanatory variable** – defines two groups being compared

- **The response variable** – the variable which consists of the measured outcomes from each group.

Example: A study compares female and male college students on the proportion who say they have participated in binge drinking.

What is the response variable? What is the explanatory variable?

Heart Attacks and Aspirin

- A large-scale randomized experiment investigated the effect of regular aspirin use on myocardial infarctions (i.e., heart attacks).
- What is the response and explanatory variable?
- Are these samples independent or dependent?
- What is one question we may be interested in testing ?

	Heart Attack?		
Group	yes	no	Total
aspirin	104	10933	11037
control	189	10845	11034

Comparing two population proportions

- We can construct a test statistic which compares the population proportion between two groups

$$\hat{p}_1 - \hat{p}_2$$

What do we know about the sample distribution of $\hat{p}_1 - \hat{p}_2$?

The mean of the sampling distribution is $p_1 - p_2$

The standard deviation of the sampling distribution is $\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

assuming independent samples. Why? Because the variance of the sum or difference of two independent random variables is equal to the sum of their variances

A confidence interval for $p_1 - p_2$

- Recall the general recipe for a CI

$$\text{point estimate} \pm \text{margin of error}$$

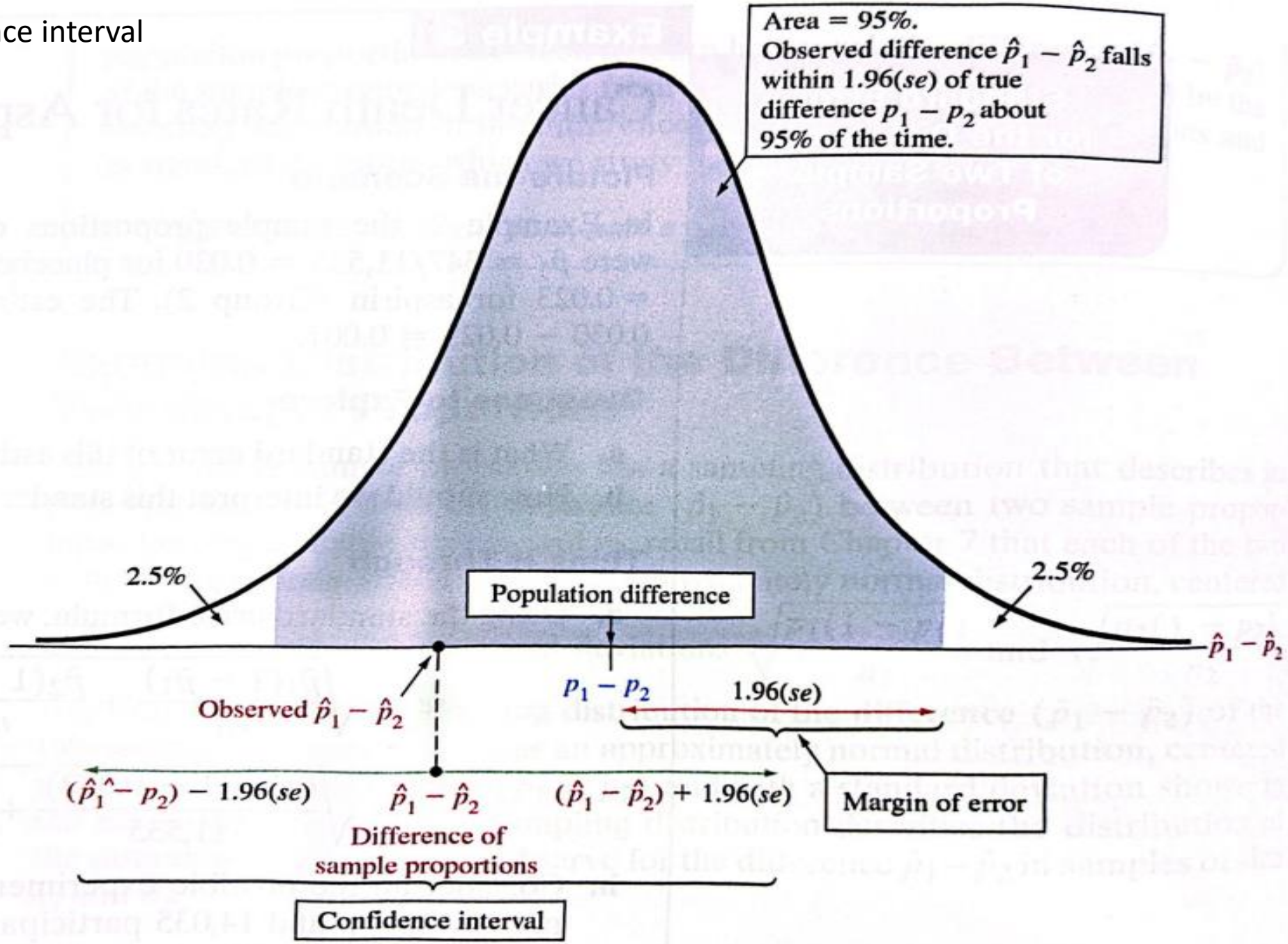
where

$$\text{margin of error} = \text{standard score} \times \text{standard error}$$

$(1-\alpha)\%$ Confidence Interval is given by:

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

95% confidence interval



A test statistic for $p_1 - p_2$

Recall the general formula for a test statistic

$$\text{Test statistic} = \frac{\text{point estimate } - \text{hypothesized value}}{\text{standard error}}$$

we can construct the following test statistic:

$$H_0: p_1 = p_2 \ (p_1 - p_2 = 0)$$

$$H_A: p_1 \neq p_2$$

$$Z_{obs} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$\hat{p}$ is called the pooled estimate of the common proportion

Example:

	Heart Attack?		
Group	**yes**	**no**	**Total**
aspirin	104	10933	11037
control	189	10845	11034