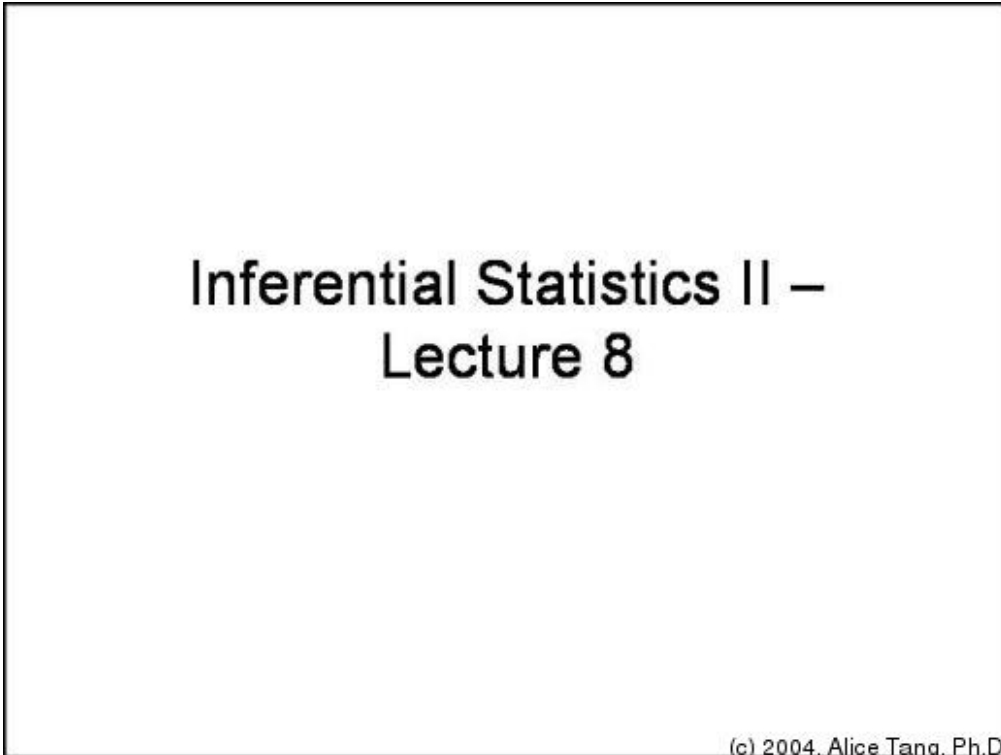


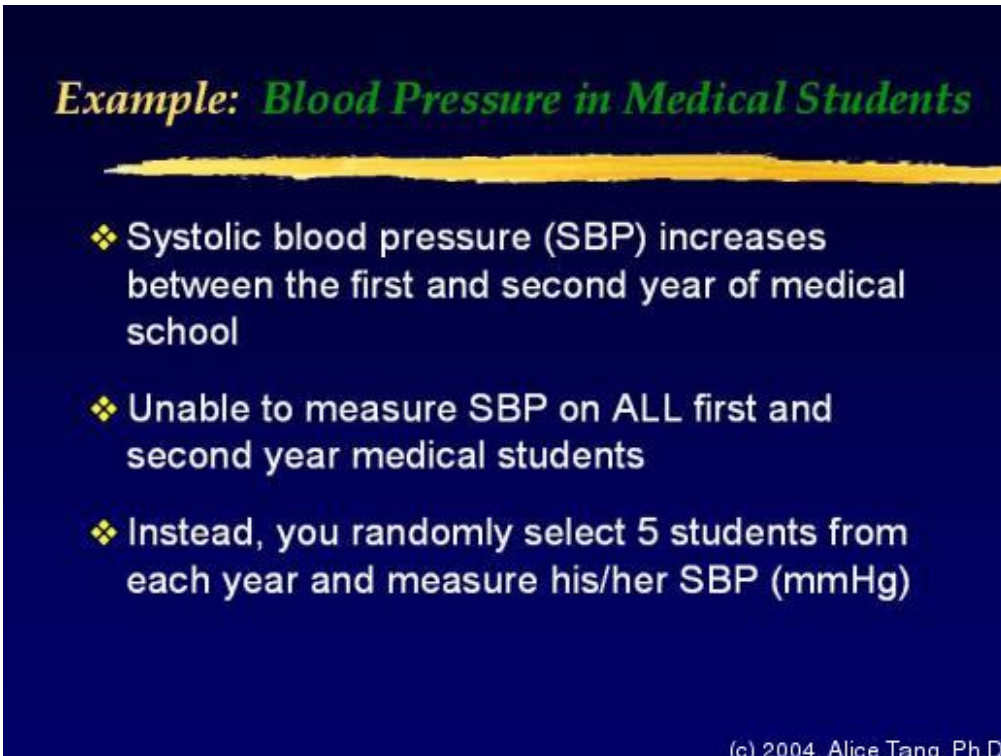
1. Lecture 8 - Introduction Slide



Inferential Statistics II –  
Lecture 8

(c) 2004, Alice Tang, Ph.D.

2. Example: Blood Pressure in Medical Students



*Example: Blood Pressure in Medical Students*

- ❖ Systolic blood pressure (SBP) increases between the first and second year of medical school
- ❖ Unable to measure SBP on ALL first and second year medical students
- ❖ Instead, you randomly select 5 students from each year and measure his/her SBP (mmHg)

(c) 2004, Alice Tang, Ph.D.

3. Example, cont.

*Example (cont'd)*

- ❖ **Data:**
  - ▶ Year 1: 120, 80, 90, 110, 95 (Mean=99)
  - ▶ Year 2: 105, 130, 145, 125, 115 (Mean=124)
- ❖ **Difference in means = +25 mm Hg**
- ❖ **Question:** Is SBP really higher in Year 2 students than Year 1 students?

(c) 2004, Alice Tang, Ph.D.

4. Concepts Illustrated by Example

*Concepts Illustrated by Example*

- ❖ Hypothesis
- ❖ Population vs. Sample
- ❖ Results (Data)
- ❖ **Conduct statistical test**
- ❖ Conclusion

(c) 2004, Alice Tang, Ph.D.

5. Hypothesis

## *Hypothesis*

- ❖ Formulate a **Null Hypothesis**
  - ▶ No difference between two populations, no association between exposure and disease, no effect of treatment, etc.
  - ▶  $H_0$ : Mean SBP in Year 1 - Mean SBP in Year 2 = 0
- ❖ Formulate an **Alternative hypothesis**
  - ▶ Two populations are different, there is an association between exposure and disease, there is an effect of treatment, etc.
  - ▶  $H_A$ : Mean SBP in Year 1 - Mean SBP in Year 2  $\neq$  0

(c) 2004, Alice Tang, Ph.D.

6. Parameters vs. Statistics

## *Parameters vs. Statistics*

- ❖ Review of Definitions:
  - ▶ A **parameter** is a true value in the **population** of interest
  - ▶ A **statistic** is a number you calculate from your **sample data** in order to estimate the parameter
- ❖ **Parameter**: Difference in mean SBP between all first year med students and all second year med students
- ❖ **Statistic**: Difference in mean SBP between 5 first year and 5 second year medical students chosen randomly from Tufts University

(c) 2004, Alice Tang, Ph.D.

## 7. Statistic and Parameter

What can make your sample's value  
(**statistic**) differ from the truth (**parameter**)?

How can the difference in mean SBP between 1<sup>st</sup>  
and 2<sup>nd</sup> year medical students be explained?

1. **True difference**: The two populations have different mean blood pressures
2. **Random variation**: The two populations have identical means and the observed difference is a coincidence of sampling (i.e. simply due to chance)
3. **Sampling error (bias)**: Poorly selected samples (not representative of the populations)

(c) 2004, Alice Tang, Ph.D.

## 8. Hypothesis Testing and Statistics

- ❖ Assuming the samples are appropriately selected, **hypothesis testing** and **statistics** help us decide if the difference we observe is likely due to **random variation (chance)** or due to a **real difference** between the groups.
- ❖ **Hypothesis testing** and **statistics** can **not** tell us if our results are due to **sampling error (bias)**.

(c) 2004, Alice Tang, Ph.D.

9.

Skipping ahead.....

*Skipping ahead... ..*

- ❖ **Set alpha:** usually 0.05
- ❖ **Conduct a statistical test** (covered in the next lecture)
- ❖ **Obtain a p-value**

(c) 2004, Alice Tang, Ph.D.

10.

P-value

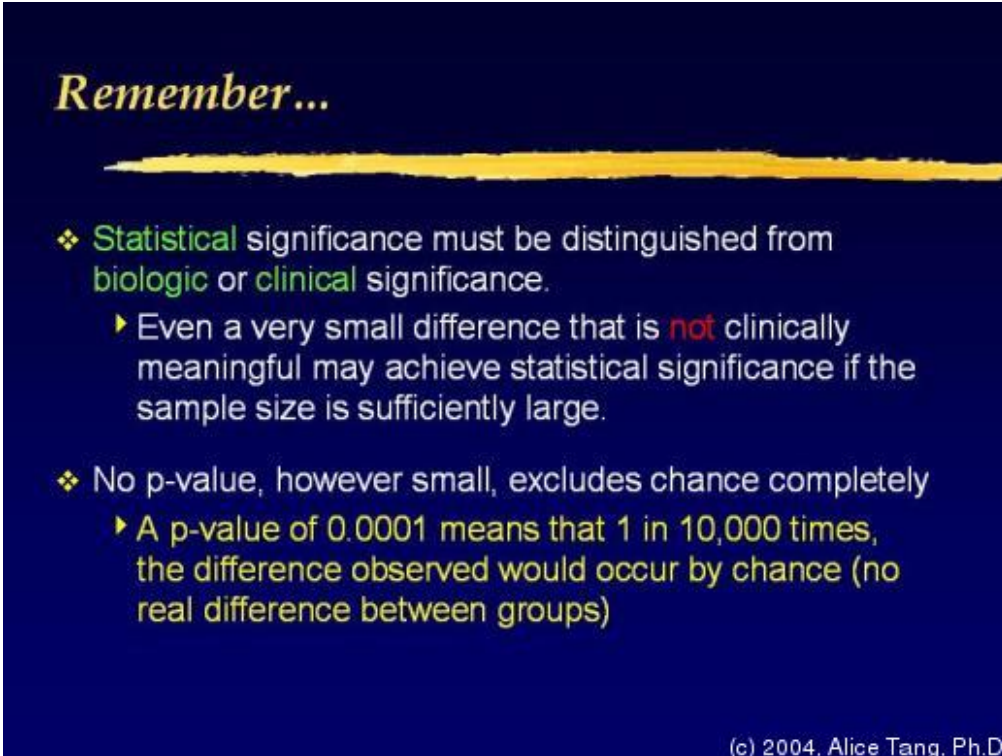
*P-value*

- ❖ You obtain a p-value of 0.03.
- ❖ **Interpretation:** Given that the null hypothesis is true (i.e. no difference in mean SBP between Year 1 and Year 2 medical students), there is a 3% chance of randomly selecting samples whose means are as far apart as (or farther than) that observed.
- ❖ In other words, the chances of observing this difference is very small, if there is no real difference between the two populations.

(c) 2004, Alice Tang, Ph.D.

11.

Remember...



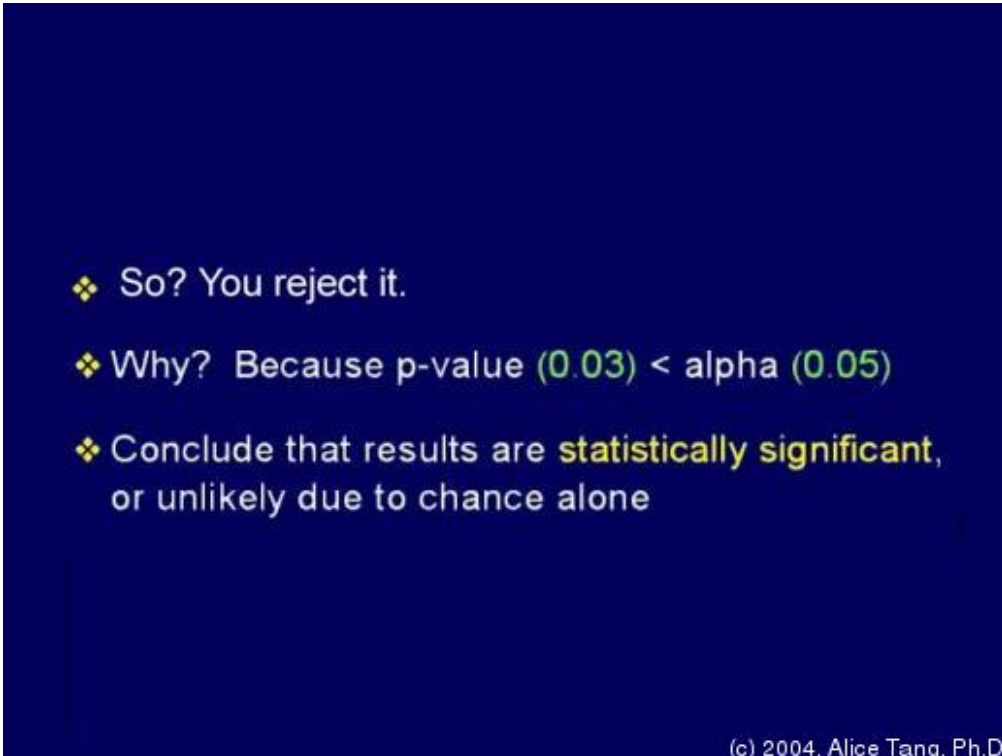
*Remember...*

- ❖ **Statistical** significance must be distinguished from **biologic** or **clinical** significance.
  - ▶ Even a very small difference that is **not** clinically meaningful may achieve statistical significance if the sample size is sufficiently large.
- ❖ No p-value, however small, excludes chance completely
  - ▶ A p-value of 0.0001 means that 1 in 10,000 times, the difference observed would occur by chance (no real difference between groups)

(c) 2004, Alice Tang, Ph.D.

12.

Results



- ❖ So? You reject it.
- ❖ Why? Because p-value (0.03) < alpha (0.05)
- ❖ Conclude that results are **statistically significant**, or unlikely due to chance alone

(c) 2004, Alice Tang, Ph.D.

13. Type I and Type II error

*Type I and Type II error*

- ❖ **Type I error (alpha error):** find that a result is **statistically significant** and reject the null hypothesis when, in fact, the null is true
- ❖ **Type II error (beta error):** find that a result is **not statistically significant** and fail to reject the null hypothesis when, in fact, the null is false.

(c) 2004, Alice Tang, Ph.D.

14. Example 2: B-vitamins for reducing risk of recurrent stroke...

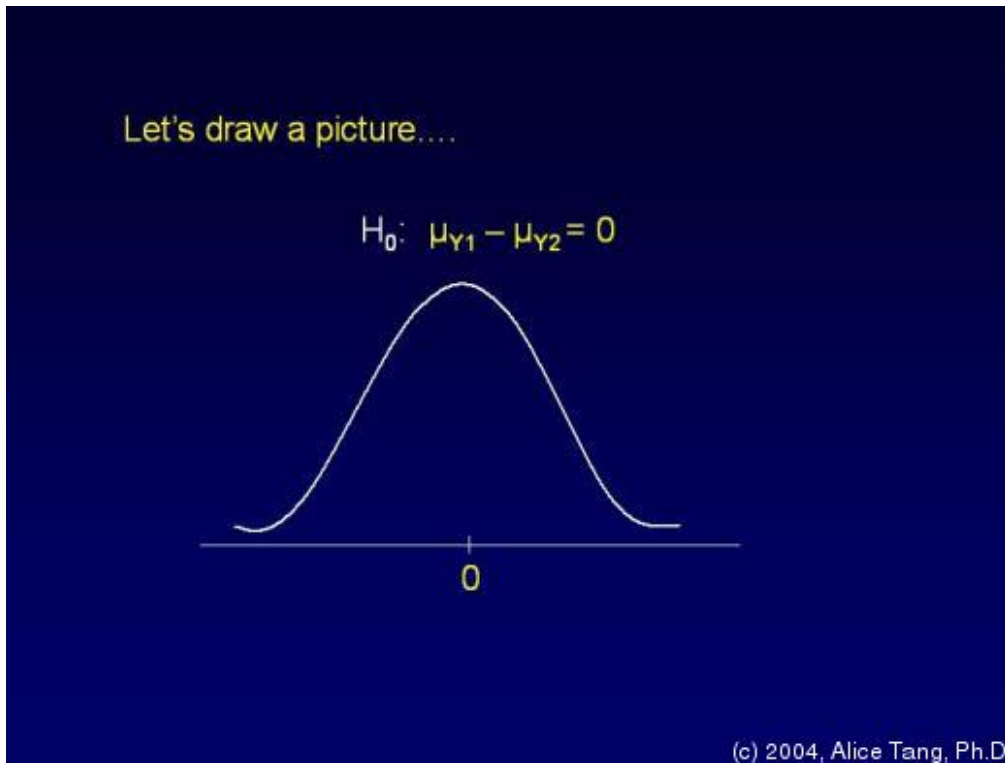
*Example 2: B-vitamins for reducing risk of recurrent stroke*

- ❖ Randomized controlled trial (RCT) to determine whether high doses of B-group vitamins reduces risk of recurrent stroke over 2 years compared to low doses of these vitamins.
  - ▶ 100 stroke patients randomly assigned to take a high dose vitamin each day
  - ▶ 100 stroke patients randomly assigned to take a low dose vitamin each day

(c) 2004, Alice Tang, Ph.D.

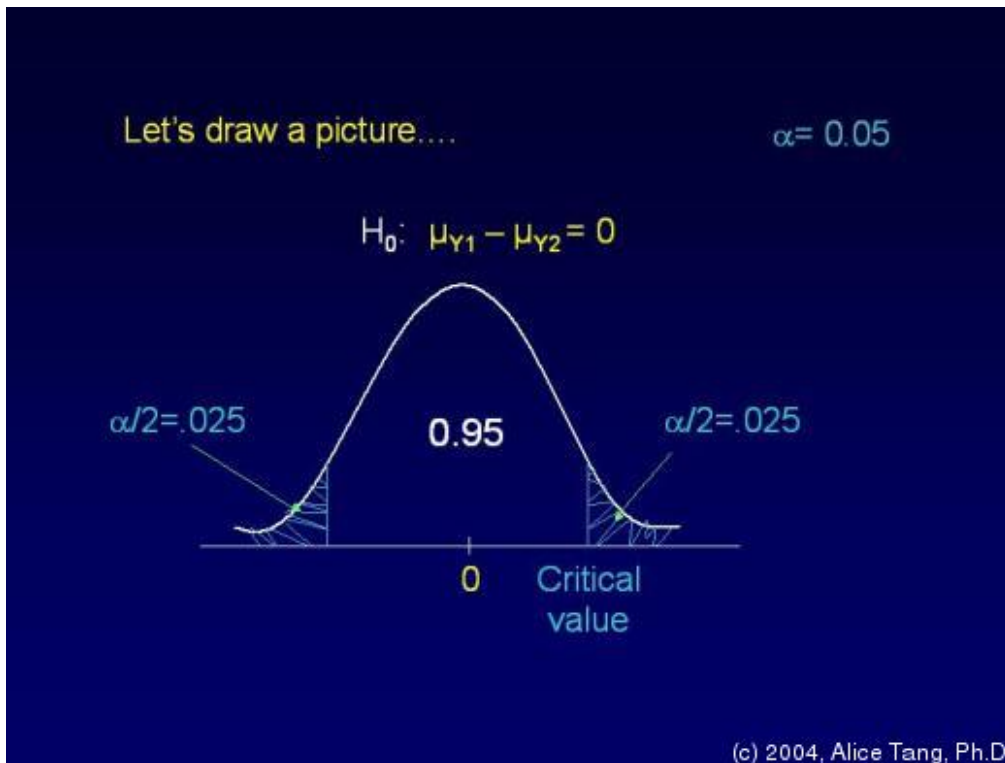
15.

### Bell Curve 1



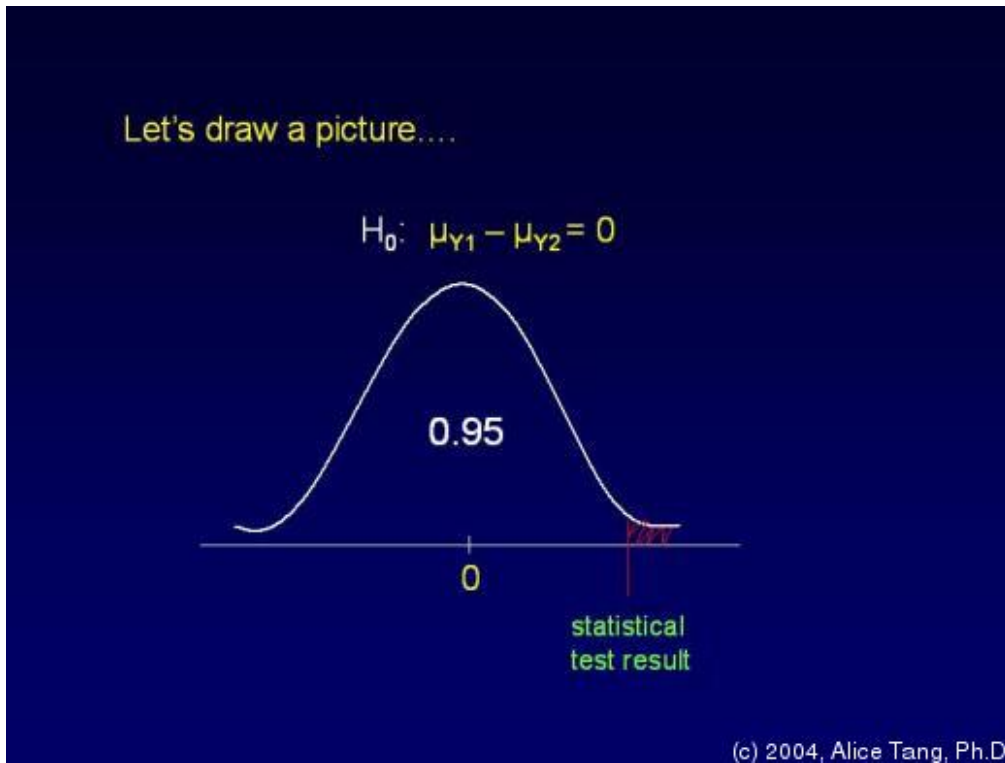
16.

### Bell Curve 2



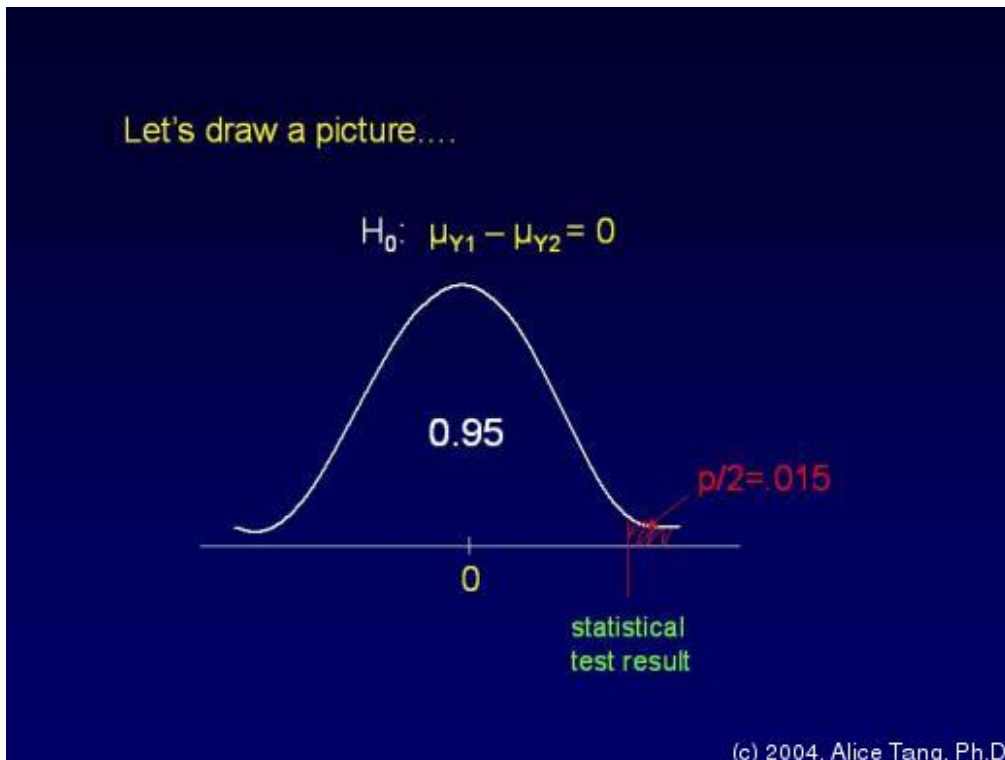
17.

### Bell Curve 3



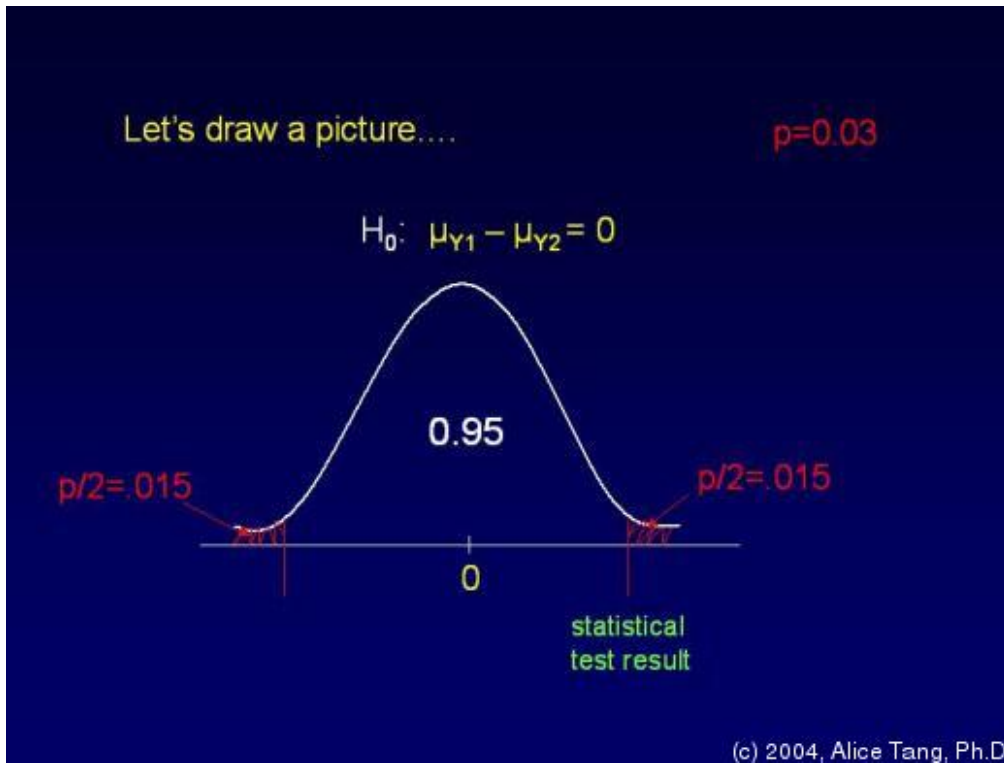
18.

### Bell Curve 4



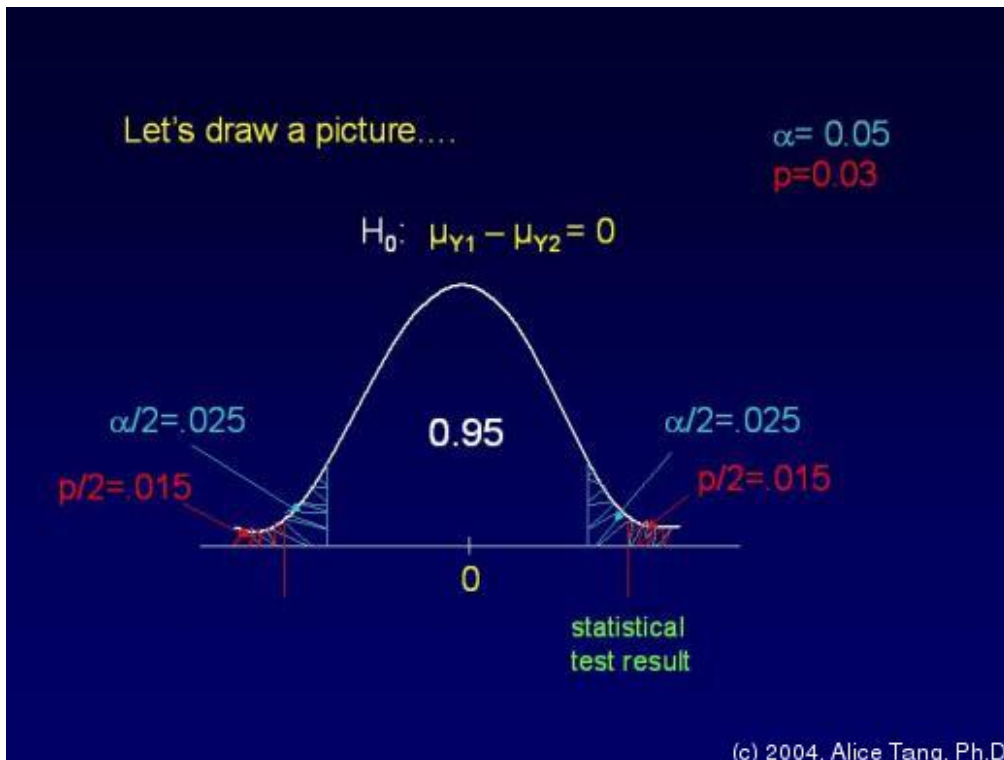
19.

### Bell Curve 5



20.

### Bell Curve 6



21. Four Possible Outcomes to a Hypothesis Test

*Four possible outcomes to a hypothesis test:*

		Truth	
		$H_0$ is true	$H_0$ is false
Decision based on sample	Reject $H_0$	Type I error ( $\alpha$ )	Correct decision
	Do not reject $H_0$	Correct decision	Type II error ( $\beta$ )

(c) 2004, Alice Tang, Ph.D.

22. Hypothesis

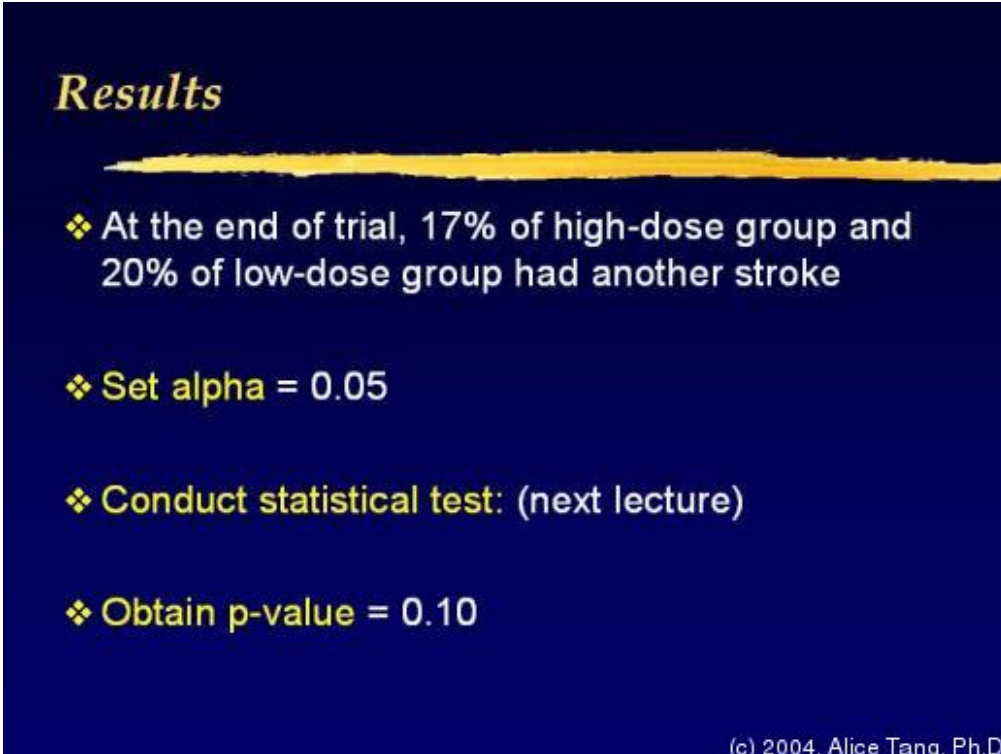
## Hypothesis

- ❖ **Null hypothesis**  
 $H_0$ : 2 year incidence of stroke in high dose group – 2 year incidence of stroke in low dose group = 0
- ❖ **Alternative hypothesis:**  
 $H_A$ : 2 year incidence of stroke in high dose group – 2 year incidence of stroke in low dose group  $\neq 0$

(c) 2004, Alice Tang, Ph.D.

23.

## Results



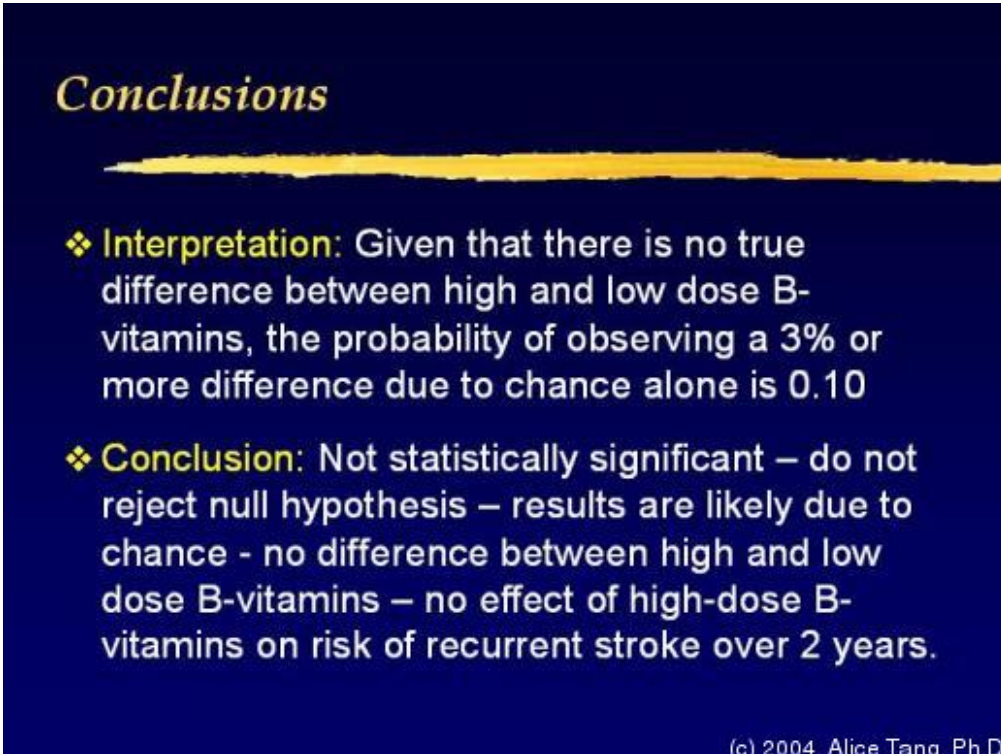
*Results*

- ❖ At the end of trial, 17% of high-dose group and 20% of low-dose group had another stroke
- ❖ Set  $\alpha = 0.05$
- ❖ Conduct statistical test: (next lecture)
- ❖ Obtain  $p\text{-value} = 0.10$

(c) 2004, Alice Tang, Ph.D.

24.

## Conclusions



*Conclusions*

- ❖ **Interpretation:** Given that there is no true difference between high and low dose B-vitamins, the probability of observing a 3% or more difference due to chance alone is 0.10
- ❖ **Conclusion:** Not statistically significant – do not reject null hypothesis – results are likely due to chance - no difference between high and low dose B-vitamins – no effect of high-dose B-vitamins on risk of recurrent stroke over 2 years.

(c) 2004, Alice Tang, Ph.D.

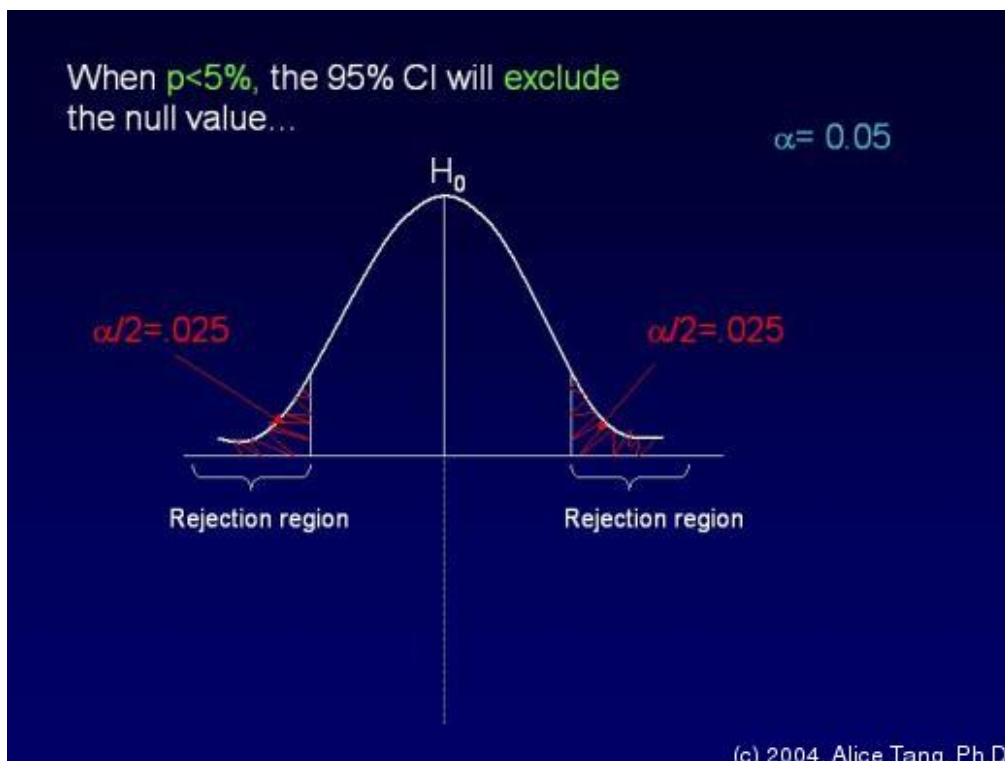
25. More on confidence intervals

### More on confidence intervals

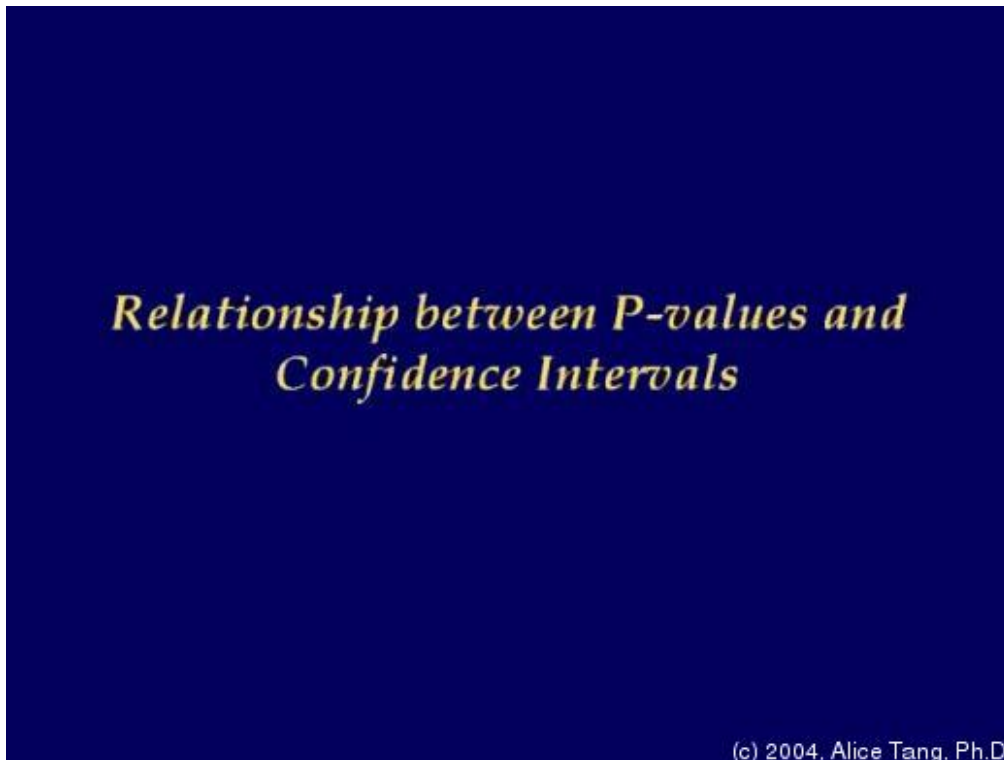
- ❖ There is a direct relationship between the CI and the p-value
- ❖ When the p-value is **less** than  $\alpha$  (.05), then the 95% CI will **exclude** the **null** value (statistically significant)
- ❖ When the p-value is **greater** than  $\alpha$  (0.05), then the 95% CI will **include** the **null** value (not statistically significant)

(c) 2004, Alice Tang, Ph.D.

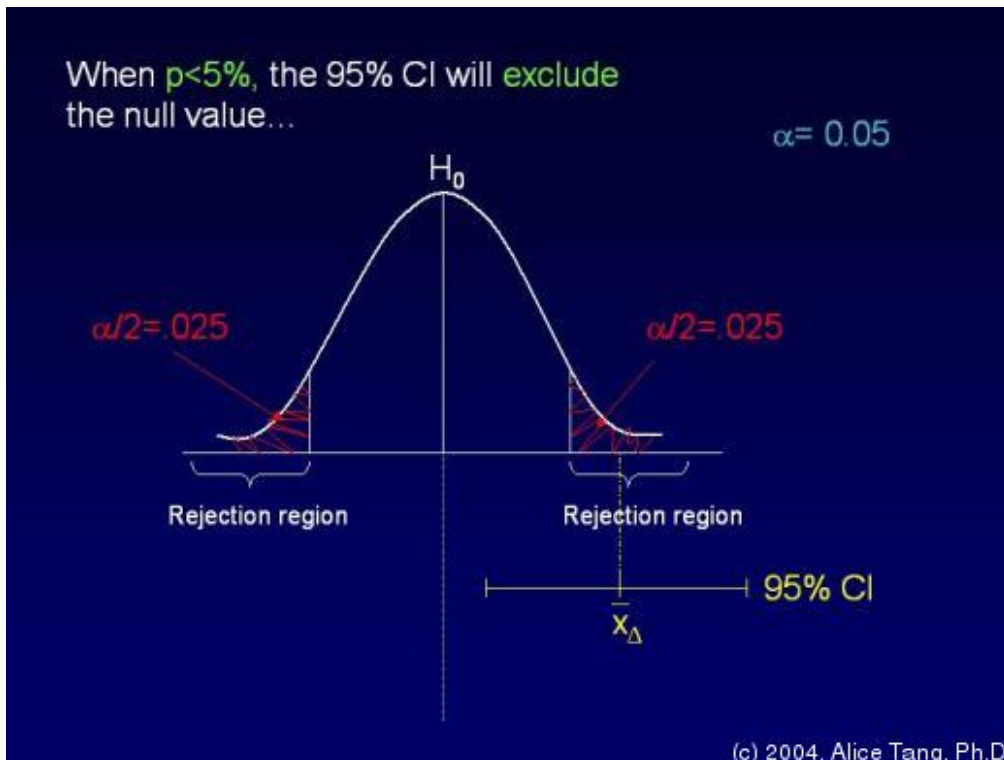
26. Rejection Region 1



27. Relationship between P value and Confidence Intervals

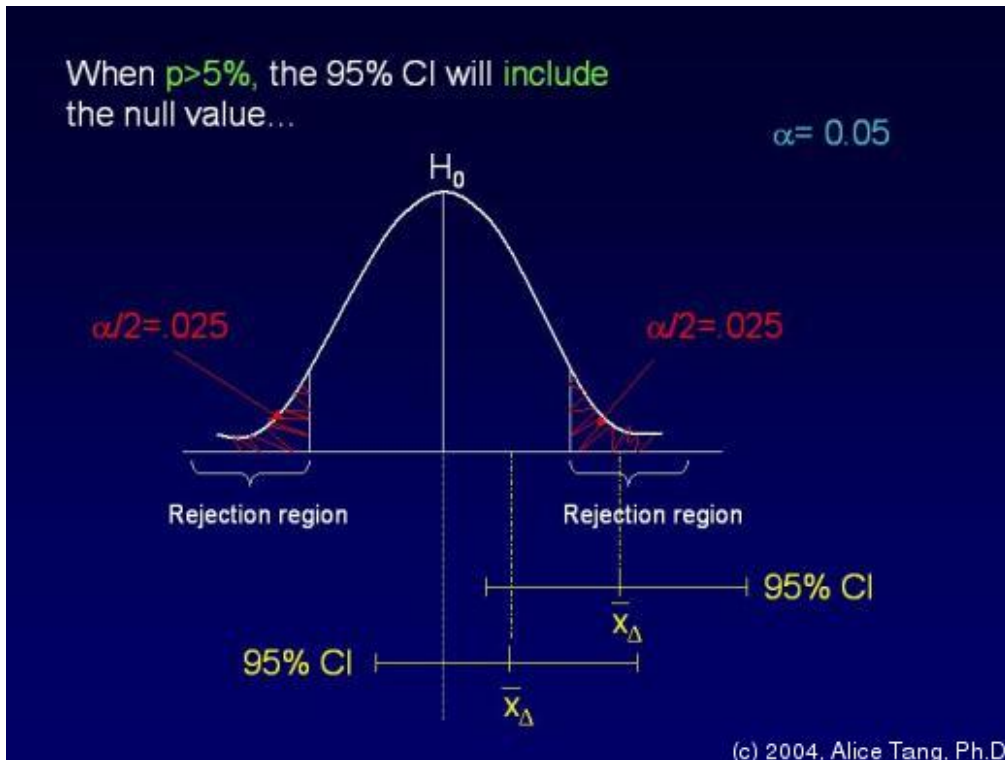


28. Rejection Region 2



29.

### Rejection Region 3



30.

### Null values

## *Null values*

- ❖ Null values depend on your null hypothesis:
  - ▶ Null value for RR and OR is 1.0 (no difference between exposed and unexposed, or diseased and non-diseased)
  - ▶ Null value for comparison of two means is 0 (no difference between groups)

(c) 2004, Alice Tang, Ph.D.

31.

## Take-home points

### *Take-home points*

- ❖ Know the steps in conducting a hypothesis test
- ❖ Know how to interpret a p-value
- ❖ Know the four possible outcomes to a hypothesis test
- ❖ Know the relationship between p-values and confidence intervals.

(c) 2004, Alice Tang, Ph.D.