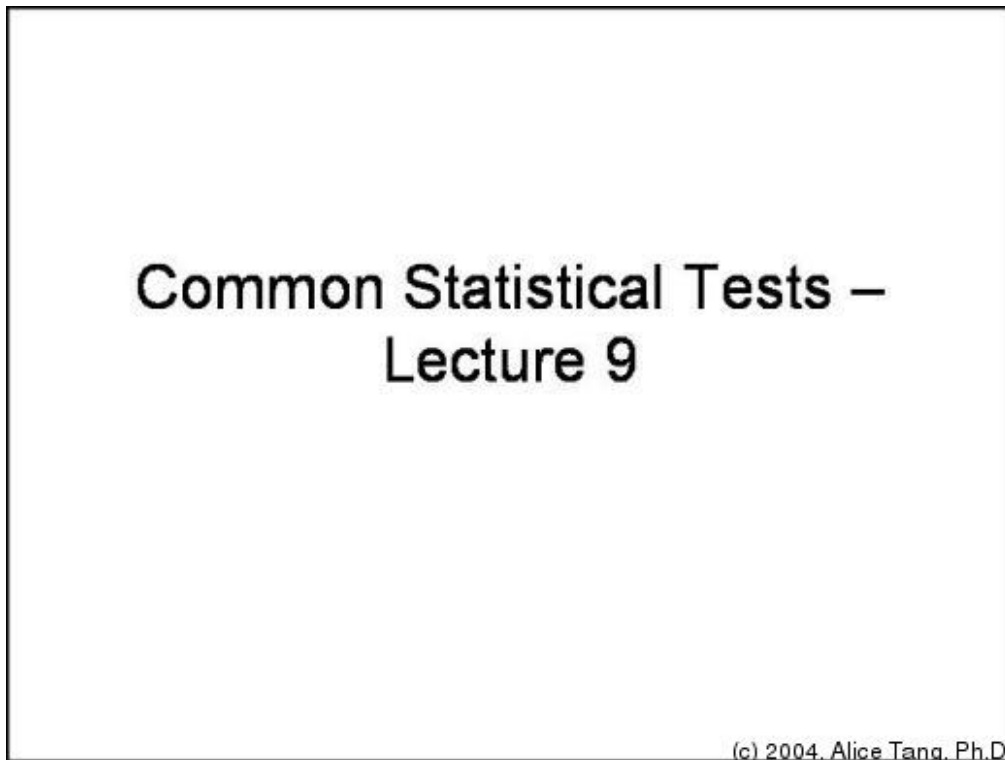
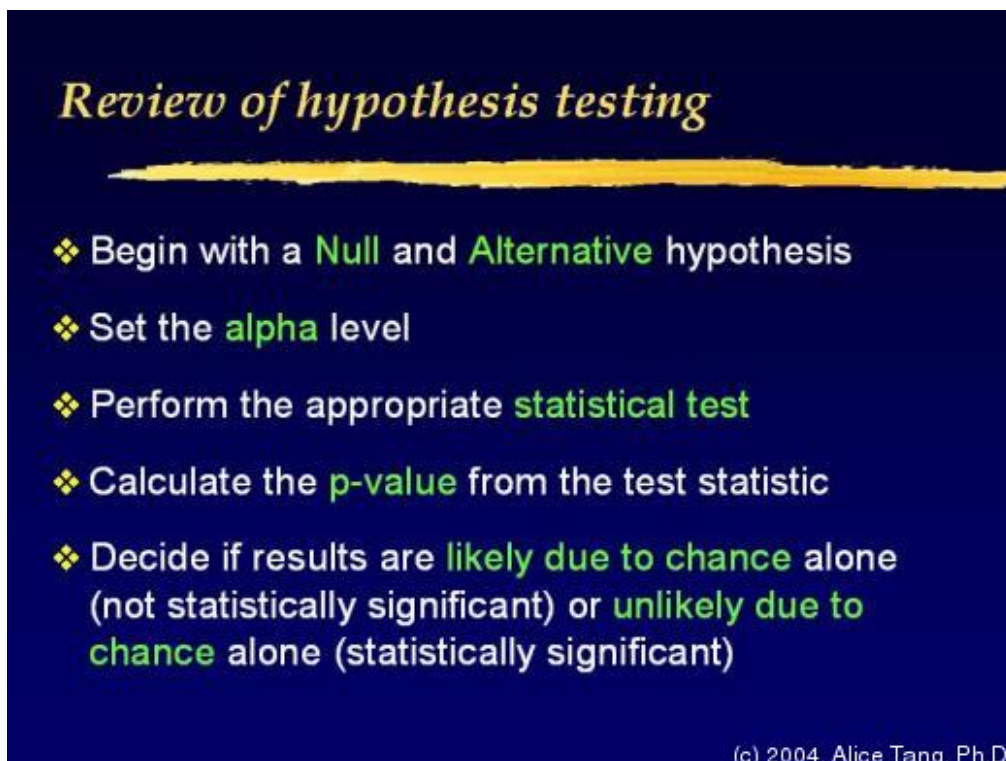


1. Lecture 9 - Introduction Slide



2. Review of hypothesis testing



3. Statistical Tests

Statistical Tests

- ❖ Statistical tests are mathematical formulas that produce p-values which allow investigators to assess the likelihood that chance accounts for the results observed in the study
- ❖ There are many different statistical tests. The choice of which test to use depends on several factors:
 - ▶ The type of data (continuous, nominal etc.)
 - ▶ The distribution of the data (normally distributed or not)
 - ▶ The type of study design (means, proportions, # of groups etc.)

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

4. Statistical Tests, cont.

Statistical Tests (cont'd)

- ❖ All test statistics follow this general format:

$$\text{Test statistic} = \frac{\text{Observed value} - \text{Expected value (H}_0\text{)}}{\text{Standard Error}}$$

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

5. Analysis of Continuous Data: 1-sample Z-test 1-sample t-test ...

Analysis of Continuous Data:

1-sample Z-test
1-sample t-test
2-sample t-test
Paired t-test

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

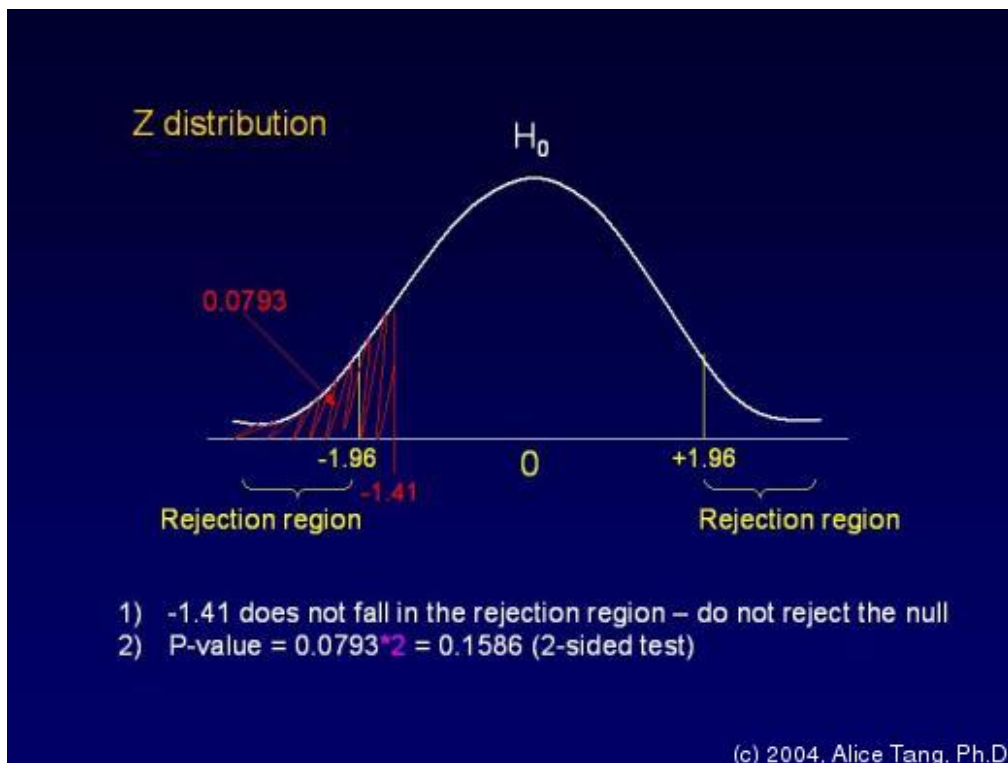
6. Example: One-sample Z-test

Example: One-sample Z-test

- ❖ Researchers are interested in whether the mean level of enzyme A in a certain population is different from 25. They measure levels of enzyme A in a sample of 10 individuals and find that the mean, $\bar{x} = 22$. Assume that the population has a known standard deviation, $\sigma = 6.7$.
- ❖ $H_0: \mu = 25$ $\alpha = 0.05$
 $H_A: \mu \neq 25$
- ❖ Calculate test statistic: $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{22 - 25}{6.7 / \sqrt{10}} = -1.41$
- ❖ Critical value (from Z-table): ± 1.96

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

7. z distribution



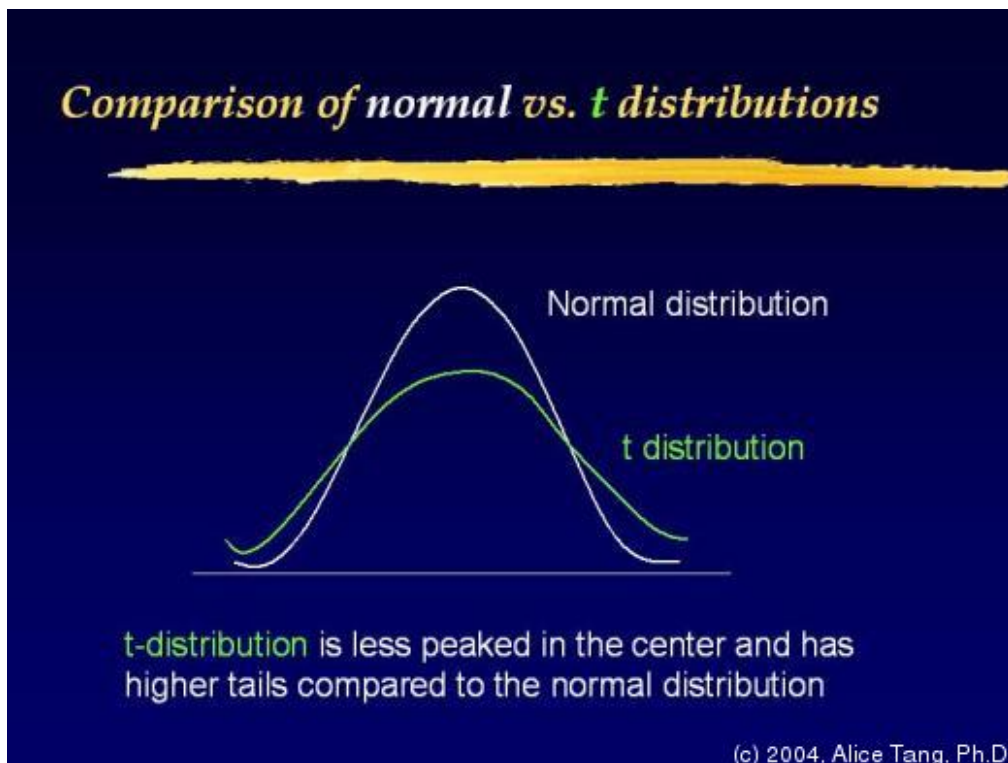
8. t-tests

t-tests

- ❖ Z-test assumes population variance (or standard deviation) is known
- ❖ When population standard deviation (σ) is unknown, but sample size is large (i.e. >30), then use sample standard deviation (s) to estimate σ and use normal distribution theory (Central Limit Theorem)
- ❖ However, when σ is unknown and sample size is small, then must use Student's *t* distribution theory

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

9. Comparison of normal vs. t distributions



10. Properties of the t-distribution

- Properties of the t-distribution*
- ❖ Family of distributions – different distribution for each sample value of $n-1$ (degrees of freedom)
 - ❖ It has a mean of 0 and is symmetrical around the mean
 - ❖ The t-distribution approaches the normal distribution as $n-1$ approaches 30.
- (c) 2004, Alice Tang, Ph.D.

11.

Example: One-sample t-test

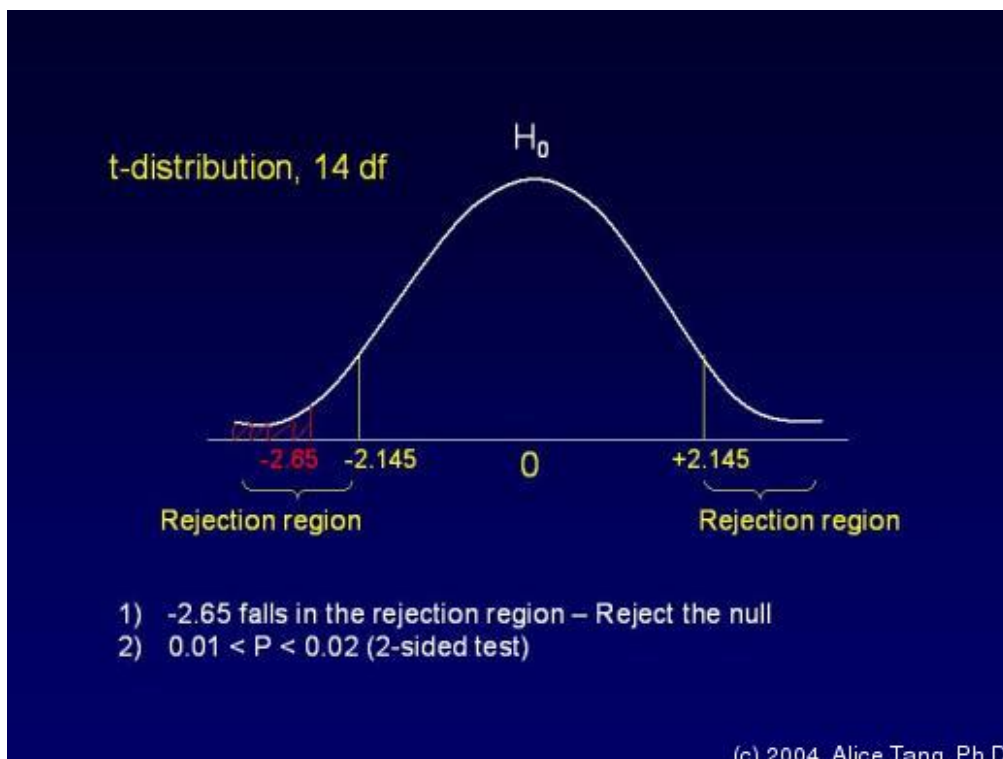
Example: One-sample t-test

- ❖ Researchers are interested in whether the mean level of enzyme B in a certain population is different from 120. They measure levels of enzyme B in a sample of 15 individuals and find that the mean, $\bar{x} = 96$ and the sample standard deviation, $s = 35$.
- ❖ $H_0: \mu = 120$ $\alpha = 0.05$
 $H_A: \mu \neq 120$
- ❖ Calculate test statistic: $t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} = \frac{96 - 120}{35 / \sqrt{15}} = -2.65$
- ❖ Critical value (t-distribution w/ 14 df) = ± 2.145

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

12.

t-distribution, 14 df



13. Two-sample t-test

Two-sample t-test

- ❖ Purpose is to compare the means of a continuous variable in two independent samples
- ❖ $H_0: \mu_1 - \mu_2 = 0$
 $H_A: \mu_1 - \mu_2 \neq 0$
- ❖ Calculate test statistic:
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\text{Sqrt} [s_p^2/n_1 + s_p^2/n_2]}$$

where,
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
- ❖ Critical Value = t-dist, with df = $n_1 + n_2 - 2$

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

14. Example: two-sample t-test

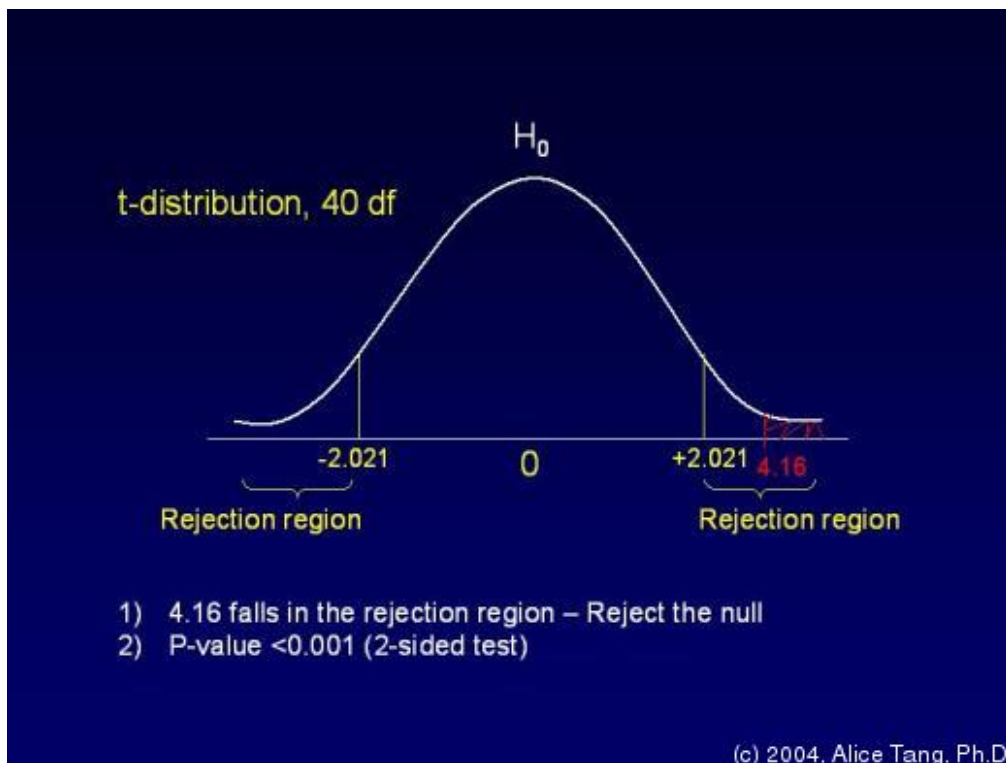
Example: two-sample t-test

- ❖ Researchers are interested in knowing whether people with diabetes have the same SBP as people without diabetes.
- ❖ $H_0: \mu_1 - \mu_2 = 0$ $\alpha = 0.05$
 $H_A: \mu_1 - \mu_2 \neq 0$
- ❖ Data:
 - ▶ Diabetics: $n_1=20, \bar{x}_1=135, s_1=10$ mm Hg
 - ▶ Nondiabetics: $n_2=25, \bar{x}_2=125, s_2=6$ mm Hg
- ❖ Test statistic: $t=(135-125)/2.4 = 4.16$
- ❖ Critical value (t-distribution w/ 43 df): ± 2.021
(use 40 df from t-table)

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

15.

t-distribution, 40 df



16.

Paired t-test

Paired t-test

- ❖ **Purpose** is to compare means of two non-independent samples. For example, measurements on the same individuals before and after a treatment
- ❖ Analysis performed on **differences** between individual pairs of observations
- ❖ $H_0: \mu_d = 0$
 $H_A: \mu_d \neq 0$
- ❖ **Calculate test statistic:** $t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$
- ❖ **Critical Value** = t-dist. with n-1 df

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

17. Example: Paired t-test

Example: Paired t-test

- ❖ 12 subjects participated in a study on the effectiveness of a certain diet on serum cholesterol levels.
- ❖ Data:

Subject	Before	After	Difference
1	201	200	-1
2	231	236	+5
3	221	216	-5
4	260	233	-27
5	228	224	-4
etc.	etc.	etc.	etc.

$\mu_d = -20.17, s_d = 23.13$

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

18. Example: paired t-test (cont'd)

Example: paired t-test (cont'd)

- ❖ Calculate test statistic:

$$t = \frac{-20.17 - 0}{23.13 / \sqrt{12}} = \frac{-20.17}{6.68} = -3.02$$

- ❖ Critical value (t-dist w/ 11 df) = ± 2.201
- ❖ -3.02 falls within the rejection region – Reject the null
- ❖ $0.01 < p < 0.02$
- ❖ Conclude that diet was effective

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

19. What if your data are not normally distributed and your samp...

What if your data are not normally distributed and your sample size is relatively small?

- ❖ Mathematically transform the data into a normal distribution (take the log or square root of all values)
- ❖ Use a different class of tests called '**non-parametric**' tests. These tests are based on the ranking or ordering of data rather than their numerical values.

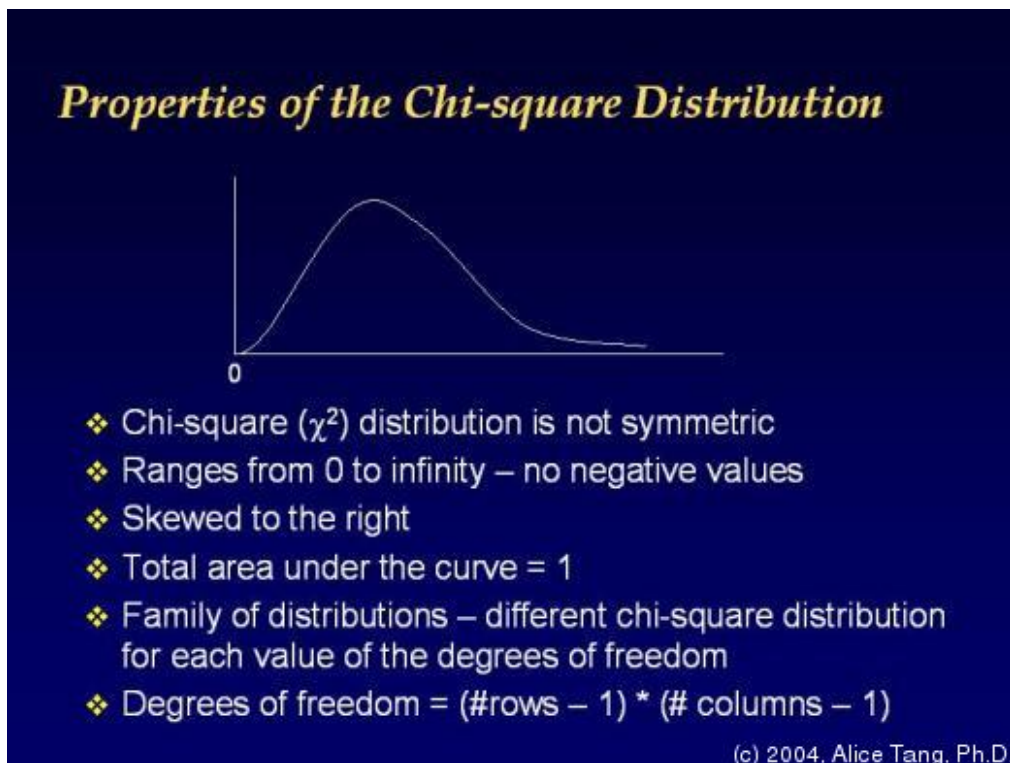
(c) 2004, Alice Tang, Ph.D.

20. Analysis of Nominal Data: Chi-square test

*Analysis of Nominal Data:
Chi-square test*

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

21. Properties of the Chi-square Distribution



22. Chi-square test

Chi-square test

- ❖ Used when both exposure and outcome are categorical (nominal)... can be 2 x 2 or larger tables
- ❖ **Null hypothesis:** no difference in proportions between groups
- ❖ **Test statistic:**
$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \quad (\text{in each cell})$$

where, **Expected** = (row total x column total) / TOTAL population
- ❖ **Critical Value:** χ^2 with (r-1)(c-1) degrees of freedom

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

23. Example: Chi-square test

Example: Chi-square test

- ❖ In an air pollution study, a random sample of 200 households were selected from each of 2 communities. A respondent in each household was asked whether or not anyone in the household was bothered by air pollution
- ❖ Data:

Bothered by air pollution?	Community		Total
	A	B	
Yes	43	81	124
No	157	119	276
Total	200	200	400

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

24. Hypotheses for Chi-square example

Hypotheses for Chi-square example

- ❖ H_0 : proportion bothered by air pollution in Community A = proportion bothered by air pollution in Community B
- ❖ H_A : proportion bothered by air pollution in Community A \neq proportion bothered by air pollution in Community B
- ❖ Alpha = 0.05

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

25. Example: Chi-square test, cont. 1

Example: Chi-square test (cont'd)

Bothered by air pollution?	Community		Total
	A	B	
Yes	43	81	124
No	157	119	276
Total	200	200	400

Expected = (row total x column total) / total pop.
 $(124 \times 200) / 400 = 62$ OR $(124 / 400) \times 200 = 62$

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

26. Example: Chi-square test, cont. 2

Example: Chi-square test (cont'd)

Bothered by air pollution?	Community		Total
	A	B	
Yes	43	81	124
No	157	119	276
Total	200	200	400

$$\chi^2 = \sum [(Obs - Exp)^2 / Exp]$$

$$= (43-62)^2 / 62 + (81-62)^2 / 62 + (157-138)^2 / 138 + (119-138)^2 / 138 = 16.88$$

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

27.

Example: Chi-square test, cont. 3

