# Biostatistics

## Chapter 9 Analysis of binary/categorical outcomes: Chi-square test

Jing Li

jing.li@sjtu.edu.cn

http://cbb.sjtu.edu.cn/~jingli/courses/2017fall/bi372/

*Dept of Bioinformatics & Biostatistics, SJTU*

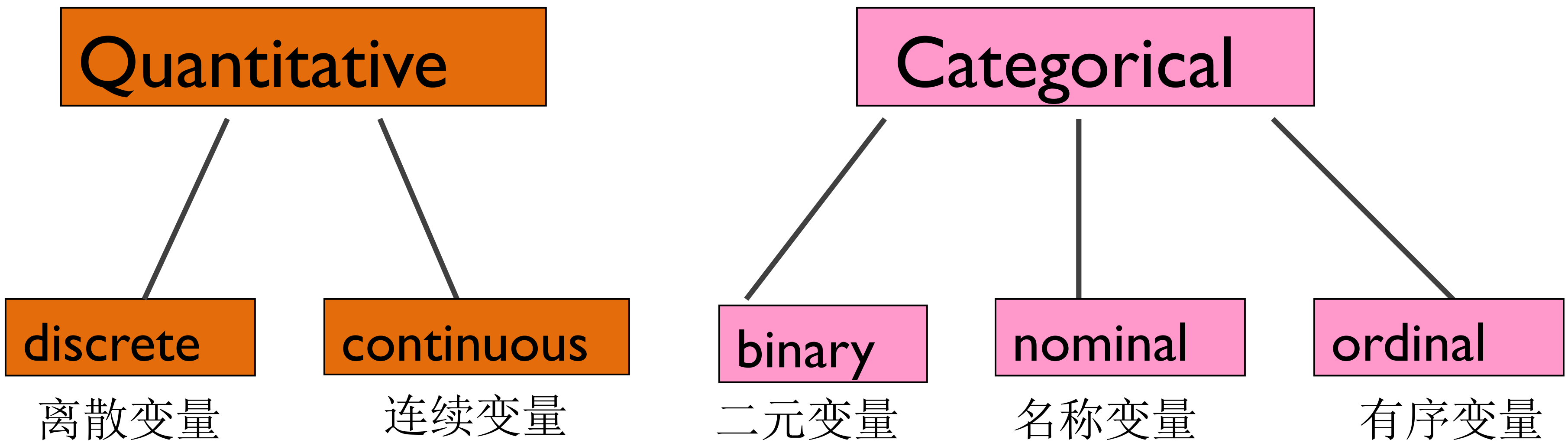
# Continuous outcome (means)

| | Parametric statistics | | Non-parametric statistics |
|---|---|---|---|
| Outcome Variable | Are the observations correlated? | | Alternatives if the normality assumption is violated (and small n): |
| | independent | correlated | |
| Continuous (e.g. blood pressure, age, pain score) | **Ttest:** compares means between two independent groups<br><br>**ANOVA:** compares means between more than two independent groups<br><br>**Pearson's correlation coefficient** (linear correlation): shows linear correlation between two continuous variables<br><br>**Linear regression:** multivariate regression technique when the outcome is continuous; gives slopes or adjusted means | **Paired ttest:** compares means between two related groups (e.g., the same subjects before and after)<br><br>**Repeated-measures ANOVA:** compares changes over time in the means of two or more groups (repeated measurements)<br><br>**Mixed models/GEE modeling:** multivariate regression techniques to compare changes over time between two or more groups | **Wilcoxon sign-rank test:** non-parametric alternative to paired ttest<br><br>**Wilcoxon sum-rank test** (=Mann-Whitney U test): non-parametric alternative to the ttest<br><br>**Kruskal-Wallis test:** non-parametric alternative to ANOVA<br><br>**Spearman rank correlation coefficient:** non-parametric alternative to Pearson's correlation coefficient |

# Review Questions (5 min)

- List the pros and  cons (优缺点) of non-parametric tests

# Types of Data

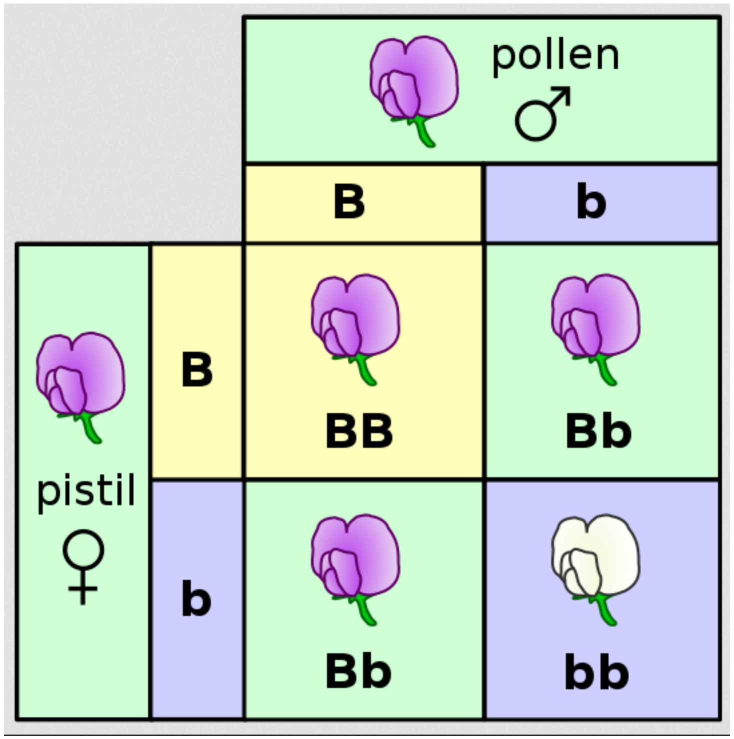


# Genetics story

A fundamental problem in genetics is determining whether the experimentally determined data fits the results expected from theory.

**Gregor Mendel** (1822-1884)

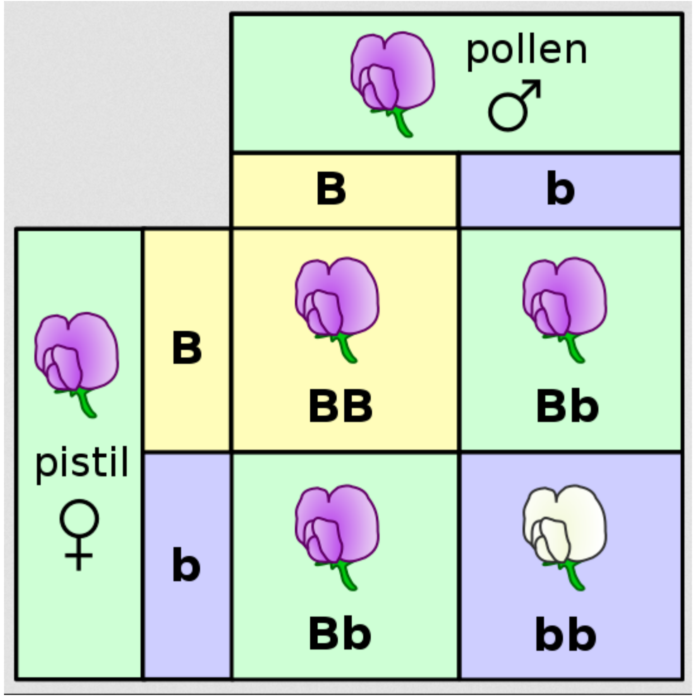
"Father of Modern Genetics"




**(the Punnett square 旁氏表)**

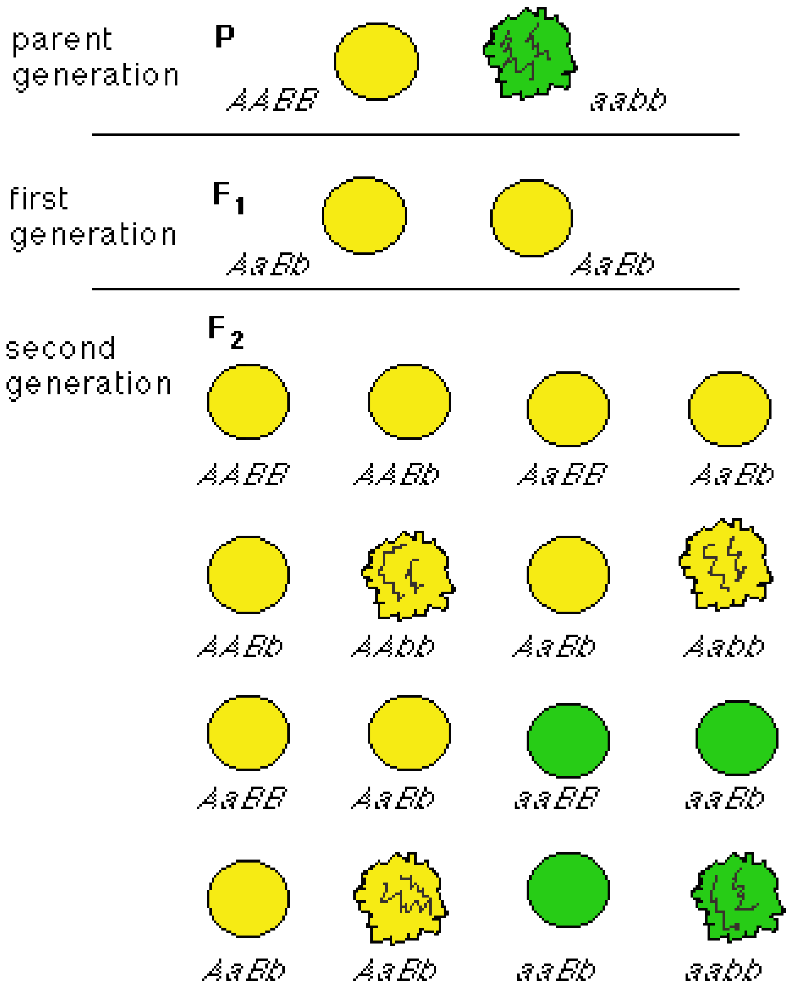
# Chi-Square Test for Goodness of Fit

- For example, do a cross and see 290 purple flowers and 110 white flowers in the offspring.
- This is pretty close to a **3/4 : 1/4** ratio, but how do you formally define "pretty close"?
- What about 250:150?




[Law of Segregation]

# Mendel Pea Plant Experiment



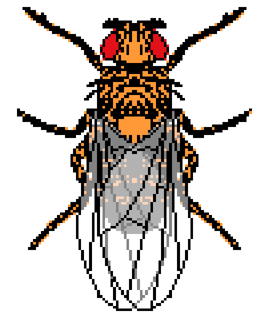

315 105 101 35

9 3 3 1 ?

# Another Example from Mendel

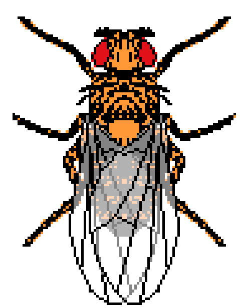
## A fruit fly cross

x



Black body, eyeless

wild

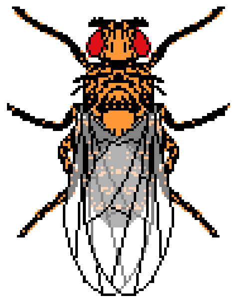


F1: all wild

# F1 x F1

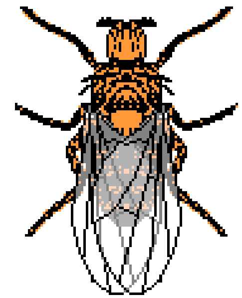
## F2

5610



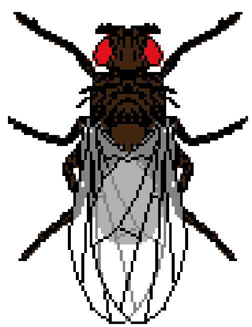
wild

1881



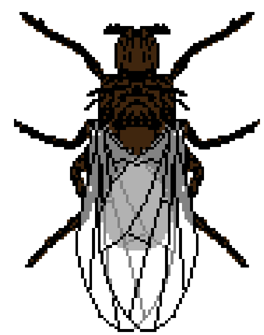
Normal body, eyeless

1896



Black body, wide eyes

622



Black body, eyeless

# Analysis of the results

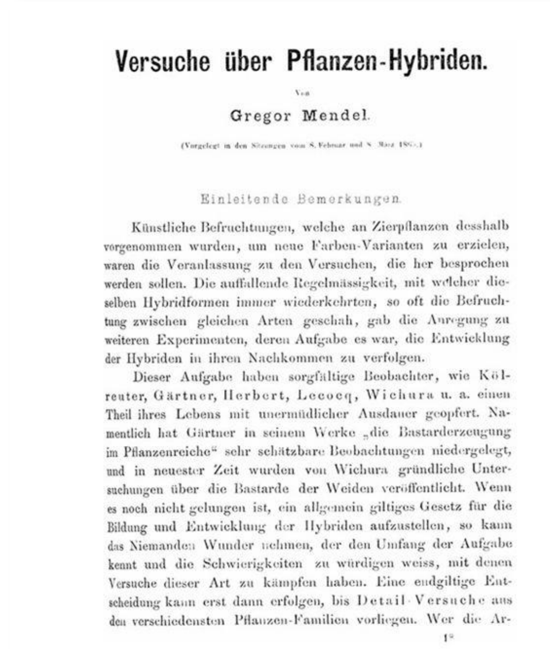
- What is the expected outcome of this cross?

Law of Independent Assortment

- 9/16 wild type: 3/16 normal body eyeless: 3/16 black body wild eyes: 1/16 black body eyeless.

# Goodness of Fit

- Mendel has no way of solving this problem （ 1865 ）.
- Shortly after the rediscovery of his work in 1900, Karl Pearson and R.A. Fisher developed the "chi-square" test for this purpose.
- The chi-square test is a "goodness of fit" test: it answers the question of how well do experimental data fit expectations.



Versuche über Pflanzen-Hybriden.

Von

Gregor Mendel.

(Vorgelegt in den Sitzungen vom 8. Februar und 8. März 1865.)

Einleitende Bemerkungen.

Künstliche Befruchtungen, welche an Zierpflanzen desshalb vorgenommen wurden, um neue Farben-Varianten zu erzielen, waren die Veranlassung zu den Versuchen, die her besprochen werden sollen. Die auffallende Regelmässigkeit, mit welcher dieselben Hybridformen immer wiederkehrten, so oft die Befruchtung zwischen gleichen Arten geschah, gab die Anregung zu weiteren Experimenten, deren Aufgabe es war, die Entwicklung der Hybriden in ihren Nachkommen zu verfolgen.

Dieser Aufgabe haben sorgfältige Beobachter, wie Kölreuter, Gärtner, Herbert, Lecocq, Wichura u. a. einen Theil ihres Lebens mit unermüdlicher Ausdauer geopfert. Namentlich hat Gärtner in seinem Werke „die Bastarderzeugung im Pflanzenreiche" sehr schätzbare Beobachtungen niedergelegt, und in neuester Zeit wurden von Wichura gründliche Untersuchungen über die Bastarde der Weiden veröffentlicht. Wenn es noch nicht gelungen ist, ein allgemein giltiges Gesetz für die Bildung und Entwicklung der Hybriden aufzustellen, so kann das Niemanden Wunder nehmen, der den Umfang der Aufgabe kennt und die Schwierigkeiten zu würdigen weiss, mit denen Versuche dieser Art zu kämpfen haben. Eine endgiltige Entscheidung kann erst dann erfolgen, bis Detail-Versuche aus den verschiedensten Pflanzen-Familien vorliegen. Wer die Ar-

1*

Karl Pearson

R.A. Fisher

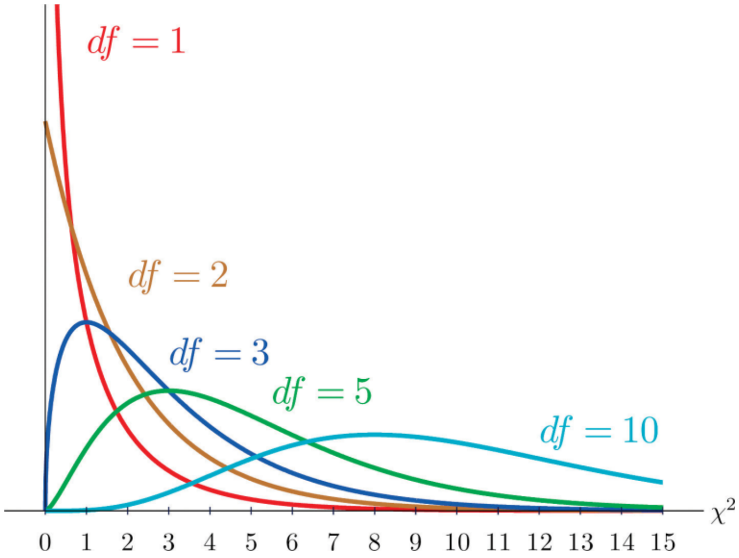
# Chi-Square Distribution

In probability theory and statistics, the **chi-square distribution** ( $\chi^2$**-distribution**) with *k* degrees of freedom is the distribution of a sum of the squares of *k* independent standard normal random variables. It is one of the most widely used probability distributions in inferential statistics.

# Chi-Square Distribution

## Characteristics of the Chi-Square Distribution

1. It is not symmetric.
2. The shape of the chi-square distribution depends upon the degrees of freedom, just like Student's *t*-distribution.
3. As the number of degrees of freedom increases, the chi-square distribution becomes more symmetric.
4. The values are non-negative. That is, the values of are greater than or equal to 0.



# Chi-Square Test

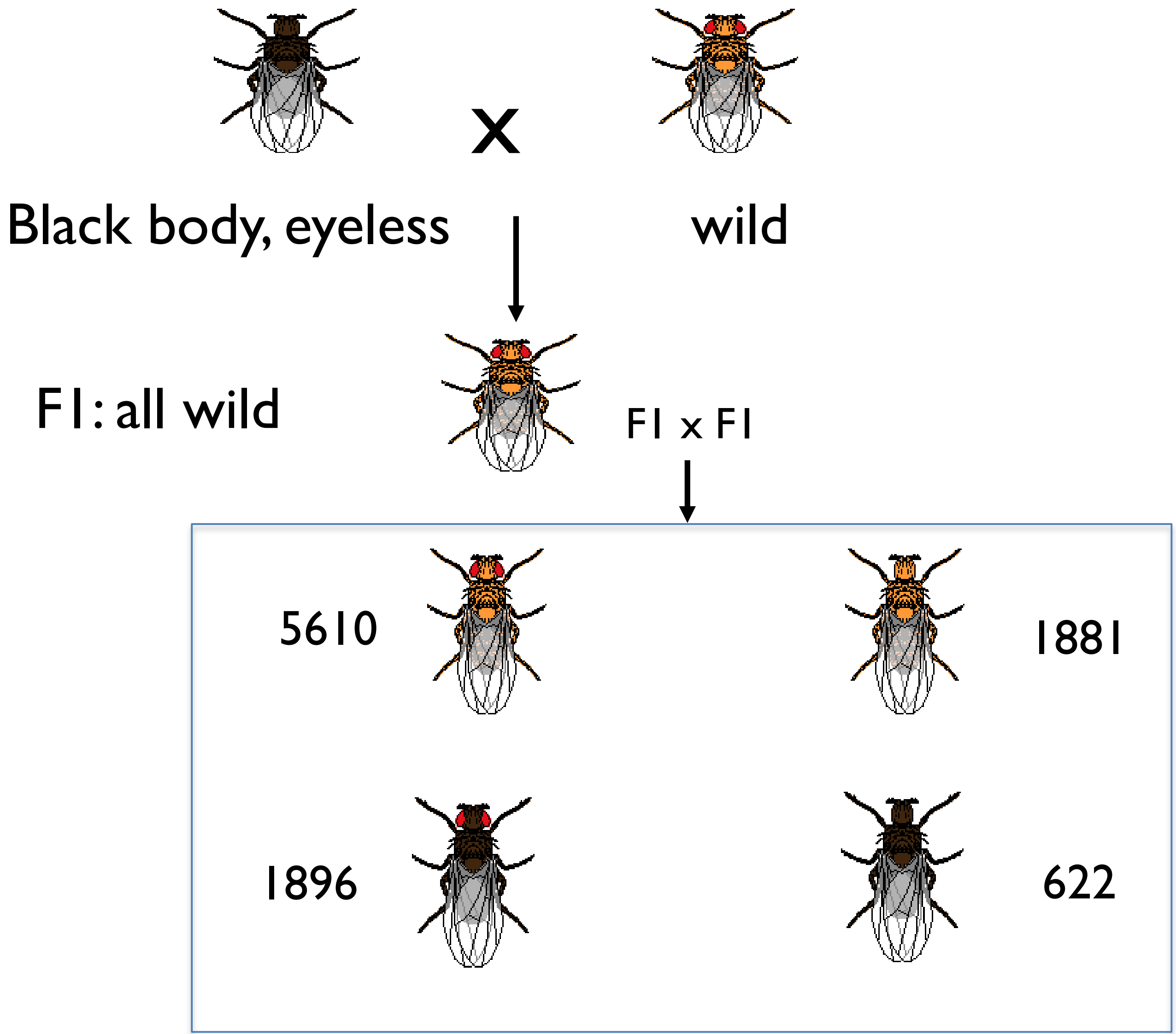
The chi-square distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one, the independence of two criteria of classification of qualitative data

1. Chi-square test for goodness of fit
2. Chi-square test for independence

# Chi-square goodness-of-fit test (拟合优度检验)

A goodness-of-fit test is an inferential procedure used to determine whether a frequency distribution follows a claimed distribution.

# Let's look at a fruit fly cross



# Chi-Square Analysis

- The chi-square analysis allows you to use statistics to determine if your data "good" or not.
- We are using laws of probability to determine possible outcomes for genetic crosses.
- How will we know if our flower or fruit fly data is "good"?
- Do observed and expected differ more than expected due to chance?

# Test for Goodness-of-Fit



Let $O_i$ represent the observed counts of category $i$, $E_i$ represent the expected counts of an category $i$, **$k$** represent the number of categories, and $n$ represent the number of independent trials of an experiment. Then,

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \qquad i = 1, 2, \ldots, k$$

approximately follows the chi-square distribution with **$k - 1$** degrees of freedom provided. $E_i = np_i$ for $i = 1, 2, \ldots, k$.

**NOTE:**

1. All expected frequencies are greater than or equal to 1 (all $E_i \geq 1$)
2. No more than 20% of the expected frequencies are less than 5.

# Test for Gooness-of-fit

If your hypothesis is supported by data

- you are claiming that mating is random and so is segregation and independent assortment （分离与自由组合定律）.

If your hypothesis is not supported by data

- you are seeing that the deviation between observed and expected is very far apart…something non-random must be occurring….

# The Chi-Square Goodness-of-Fit Test

**Step 1:** A claim is made regarding a distribution. The claim is used to determine the null and alternative hypothesis.

$H_o$: the random variable follows the claimed distribution

$H_A$: the random variable does not follow the claimed distribution

**Step 2**: Calculate the expected frequencies for each of the $k$ categories. The expected frequencies are $np_i$ for $i = 1, 2, \ldots, k$ where $n$ is the number of trials and $p_i$ is the probability of the $i$th category assuming the null hypothesis is true.

**Step 3:** Verify the requirements fort he goodness-of-fit test are satisfied.

(1) all expected frequencies are greater than or equal to 1 (all $E_i \geq 1$)

(2) no more than 20% of the expected frequencies are less than 5.

# Now Conduct the Analysis:

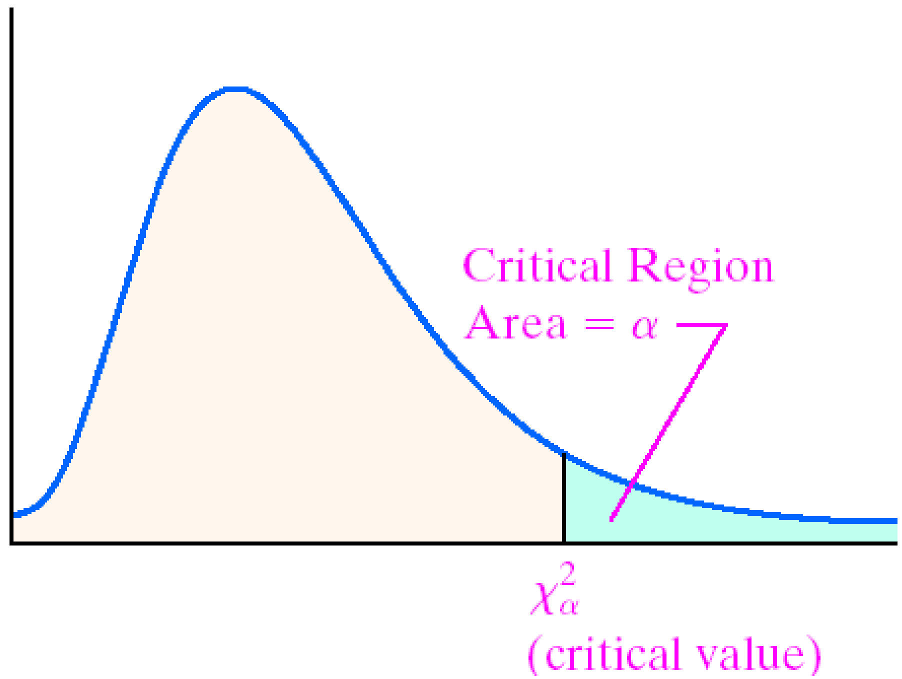
| Phenotype | Observed | | Hypothesis |
|---|---|---|---|
| Wild | 5610 | 9/16 | |
| Eyeless | 1881 | 3/16 | |
| Black body | 1896 | 3/16 | |
| Eyeless, black body | 622 | 1/16 | 626 |
| Total | 10009 | | |

To compute the hypothesis value take 10009/16 = 626

# Now Conduct the Analysis:

| Phenotype | Observed | Hypothesis |
|---|---|---|
| Wild | 5610 | 5634 |
| Eyeless | 1881 | 1878 |
| Black body | 1896 | 1878 |
| Eyeless, black body | 622 | 626 |
| Total | 10009 | |

**Step 4:** Select a level of significance α based upon the seriousness of making a Type I error. The level of significance is used to determine the critical value. All chi-square goodness-of-fit tests are right tests, the shaded region represents the critical region



# Step 5: Compute the **test statistic**

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

| Phenotype | Observed | Hypothesis |
|---|---|---|
| Wild | 5610 | 5634 |
| Eyeless | 1881 | 1878 |
| Black body | 1896 | 1878 |
| Eyeless, black body | 622 | 626 |
| Total | 10009 | |

- $(5610 - 5630)^2 / 5630 = .07$
- $(1881 - 1877)^2 / 1877 = .01$
- $(1896 - 1877)^2 / 1877 = .20$
- $(622 - 626)^2 / 626 = .02$
- $\chi^2 = .30$
- How many degrees of freedom? **4-1=3**

STEP 6: Compare the critical value to the test statistic.
If $\chi^2 > \chi^2_\alpha$ reject the null hypothesis

STEP 6: Compare the critical value to the test statistic.

If $\chi^2 > \chi_\alpha^2$ reject the null hypothesis

# CHI-SQUARE DISTRIBUTION TABLE

| Accept Hypothesis | | | | | | | | | Reject Hypothesis | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Probability (p) | | | | | | | | | | | |
| Degrees of Freedom | 0.95 | 0.90 | 0.80 | 0.70 | 0.50 | 0.30 | 0.20 | 0.10 | **0.05** | 0.01 | 0.001 |
| 1 | 0.004 | 0.02 | 0.06 | 0.15 | 0.46 | 1.07 | 1.64 | 2.71 | **3.84** | 6.64 | 10.83 |
| 2 | 0.10 | 0.21 | 0.45 | 0.71 | 1.39 | 2.41 | 3.22 | 4.60 | **5.99** | 9.21 | 13.82 |
| 3 | 0.35 | 0.58 | 1.01 | 1.42 | 2.37 | 3.66 | 4.64 | 6.25 | **7.82** | 11.34 | 16.27 |
| 4 | 0.71 | 1.06 | 1.65 | 2.20 | 3.36 | 4.88 | 5.99 | 7.78 | **9.49** | 13.38 | 18.47 |
| 5 | 1.14 | 1.61 | 2.34 | 3.00 | 4.35 | 6.06 | 7.29 | 9.24 | **11.07** | 15.09 | 20.52 |
| 6 | 1.63 | 2.20 | 3.07 | 3.83 | 5.35 | 7.23 | 8.56 | 10.64 | **12.59** | 16.81 | 22.46 |
| 7 | 2.17 | 2.83 | 3.82 | 4.67 | 6.35 | 8.38 | 9.80 | 12.02 | **14.07** | 18.48 | 24.32 |
| 8 | 2.73 | 3.49 | 4.59 | 5.53 | 7.34 | 9.52 | 11.03 | 13.36 | **15.51** | 20.09 | 26.12 |
| 9 | 3.32 | 4.17 | 5.38 | 6.39 | 8.34 | 10.66 | 12.24 | 14.68 | **16.92** | 21.67 | 27.88 |
| 10 | 3.94 | 4.86 | 6.18 | 7.27 | 9.34 | 11.78 | 13.44 | 15.99 | **18.31** | 23.21 | 29.59 |

# Conclusions

- the P value read off the table places our chi square number of .30 close to .95 or 95%
- This means that 95% of the time when our observed data is this close to our expected data, this deviation is due to random chance.
- We therefore fail to reject our null hypothesis. We accept Mendel's conclusion that the observed results for a 9/16 : 3/16 : 3/16 : 1/16 ratio.

# Practice

| phenotype | observed | Expected proportion | Expected number |
|---|---|---|---|
| round yellow | 315 | | |
| round green | 101 | | |
| wrinkled yellow | 108 | | |
| wrinkled green | 32 | | |
| total | 556 | | |

# Practice

| phenotype | observed | Expected proportion | Expected number |
| --- | --- | --- | --- |
| round yellow | 315 | 9/16 | 312.75 |
| round green | 101 | 3/16 | 104.25 |
| wrinkled yellow | 108 | 3/16 | 104.25 |
| wrinkled green | 32 | 1/16 | 34.75 |
| total | 556 | 1 | 556 |

$X^2$ =0.47, df=3

# Chi-Square Test

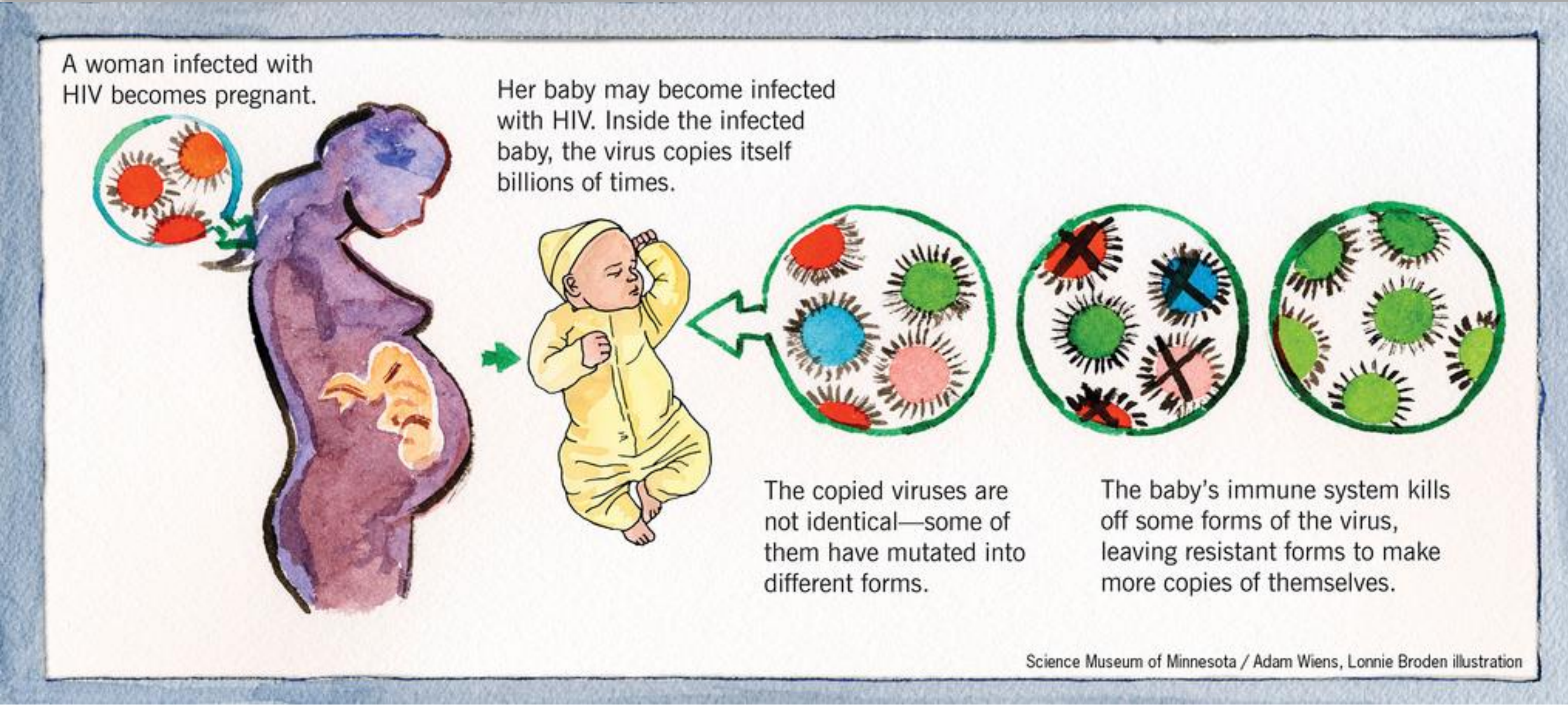
The chi-square distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one, the independence of two criteria of classification of qualitative data

1. Chi-square test for goodness of fit
2. Chi-square test for independence

# Chi-square test of Independence

Chi-square test allows you to compare **proportions** between 2 or more groups (ANOVA for means; chi-square for proportions).

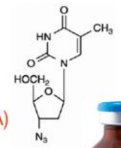
# Infant HIV/ AZT Study




Zidovudine

analogue of thymidine

AZT (azidothymidine – USA)




•Retrovir®

caps. 100 mg


*葛兰素威康制药公司生产*

# Infant HIV/ AZT Study

We will motivate by using data from the Pediatric AIDS Clinical Trial Group (ACTG) Protocol 076 Study Group

- Study design

  - "We conducted a randomized, double-blinded, placebo- controlled trial of the efficacy and safety of Zidovudine (AZT) in reducing the risk of maternal-infant HIV transmission"

  - 363 HIV infected pregnant women were randomized to AZT or placebo

# Note

- Random assignment of treatment
  - ✓Helps insure two groups are comparable
  - ✓Patient and physician could not request a particular treatment
- Double-blind
  - ✓Patient and physician did not know treatment assignment

# Observed HIV Transmission

- Results
  - Of the 180 women randomized to AZT group, 13 gave birth to children who tested positive for HIV within 18 months of birth
  - Of the 183 women randomized to the placebo group, 40 gave birth to children who tested positive for HIV within 18 months of birth

| | HIV+ | HIV- | Total |
|---|---|---|---|
| AZT | 13 | | 180 |
| Placebo | 40 | | 183 |
| Total | | | |

# Observed HIV Transmission Proportions

- AZT

$$\hat{p}_{AZT} = \frac{13}{180} = 0.07 = 7\%$$

- Placebo

$$\hat{p}_{PLAC} = \frac{40}{183} = 0.22 = 22\%$$

Is the difference significant, or can it be explained by chance?

Two sample Z-test (Compare Two Proportions)

# Hypothesis Test to Compare Two Proportions

- Two sample z-test

- Are the proportions of infants contracting HIV within 18 months-of-birth equivalent at the population level for those whose mothers are treated with AZT versus untreated (placebo)?
  - $H_o$: $p_1 = p_2$
  - $H_A$: $p_1 \neq p_2$

- In other words, is the expected difference in proportions zero?
  - $H_o$: $p_1 - p_2 = 0$
  - $H_A$: $p_1 - p_2 \neq 0$

# Hypothesis Test to Compare Two Proportions

- Recall, general "recipe" for hypothesis testing . . .
  1. Start by assuming $H_o$ true
  2. Measure distance of sample result from $\mu_o$ (here again its 0)
  3. Compare test statistic (distance) to appropriate distribution to get p-value

$$z = \frac{(observed\ diff) - (null\ diff)}{SE\ of\ observed\ differenc}$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{SE}(\hat{p}_1 - \hat{p}_2)} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1 \times (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \times (1 - \hat{p}_2)}{n_2}}}$$

# Infant HIV/ AZT Study

- In the infant HIV/AZT study, recall:

$$\hat{p}_1 - \hat{p}_2 = -0.15$$

$$S\hat{E}(\hat{p}_1 - \hat{p}_2) = 0.036$$

- So in this study:

$$z = \frac{-0.15}{0.36} \approx -4.2$$

  - So this study result was 4.2 standard errors below the null mean of 0 (i.e., 4.2 standard errors from the difference in the proportion of HIV+ infants between the AZT and placebo groups expected if null was true)

# Infant HIV/ AZT Study

- If we were to look this up on a normal table, we would find a very low p-value ($p < .001$)

# Infant HIV/ AZT Study

- If we were to look this up on a normal table, we would find a very low p-value ($p < .001$)
- We can also using **chi-square test** in this study

  - (Pearson's) Chi-Square Test ($x^2$)
    - Calculation is easy (can be done by hand)
  - Works well for "big" sample sizes
  - Gives (essentially) the same p-value as z-test for comparing two proportions
  - Unlike z-test, can be extended to compare proportions between more than two independent groups in one test

# Recall: HIV/AZT Example

| | | Drug Group | | |
|---|---|---|---|---|
| | | AZT | Placebo | |
| HIV Transmission | Yes | 13 | 40 | 53 |
| | No | 167 | 143 | 310 |
| | | 180 | 183 | 363 |

# The procedure to test for the independence:

1. State a hypotheses based on the fit of the data

2. Make a table of the observed and expected values. You will most likely be given the observed values.

3. Calculate the chi-squared test statistic, this is $\chi^2_{test} = \sum \frac{(O_i - E_i)^2}{E_i}$

4. Look up the chi-squared critical value from your chi-squared tables in the information booklet.

5. Compare your test statistic with your critical value and make a conclusion.

If the test statistic lies in the critical region then reject $H_0$ in favour of $H_1$.
Otherwise do not reject $H_0$.

# Expected values

$$\chi^2_{test} = \sum \frac{(O_i - E_i)^2}{E_i}$$

| | | Drug Group | | |
|---|---|---|---|---|
| | | AZT | Placebo | |
| HIV Transmission | Yes | 13 | 40 | 53 |
| | No | 167 | 143 | 310 |
| | | 180 | 183 | 363 |

For the goodness of fit test, the expected frequency for each category is obtained by

expected frequency = p*n

(p is the proportion from the null hypothesis and n is the size of the sample)

# Expected values

$$\chi^2_{test} = \sum \frac{(O_i - E_i)^2}{E_i}$$

| | | Drug Group | | |
|---|---|---|---|---|
| | | AZT | Placebo | |
| HIV Transmission | Yes | 13 | 40 | 53 |
| | No | 167 | 143 | 310 |
| | | 180 | 183 | 363 |

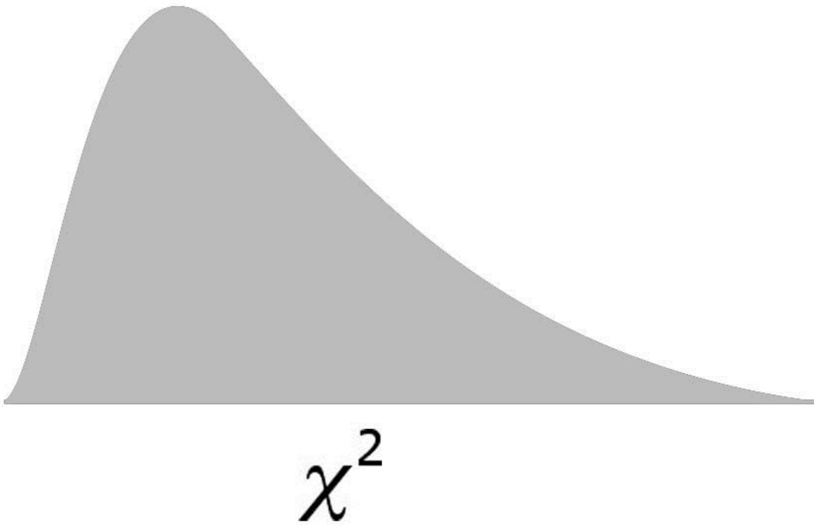
For the test for independence, the expected frequency for each cell in the matrix is obtained by

expected frequency = (row total)(column total) / n

# Degrees of freedom, *v*.

The degrees of freedom will be defined as:

$v$=(number of rows-1)(number of columns-1)



# Expected values

$$\chi^2_{test} = \sum \frac{(O_i - E_i)^2}{E_i}$$

| | | Drug Group | | |
|---|---|---|---|---|
| | | AZT | Placebo | |
| HIV Transmission | Yes | 13 | 40 | 53 |
| | No | 167 | 143 | 310 |
| | | 180 | 183 | 363 |

$$E = \frac{RC}{N} = \frac{53(180)}{363} = 26.3$$

# Expected values

- We could do the same for the other three cells:

| | | Drug Group | | |
|---|---|---|---|---|
| | | AZT | Placebo | |
| HIV Transmission | Yes | 26.3 | 26.7 | 53 |
| | No | 153.7 | 156.3 | 310 |
| | | 180 | 183 | 363 |

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

# Expected values

- We could do the same for the other three cells:

| | | Drug Group | | |
|---|---|---|---|---|
| | | AZT | Placebo | |
| HIV Transmission | Yes | 26.3 | 26.7 | 53 |
| | No | 153.7 | 156.3 | 310 |
| | | 180 | 183 | 363 |

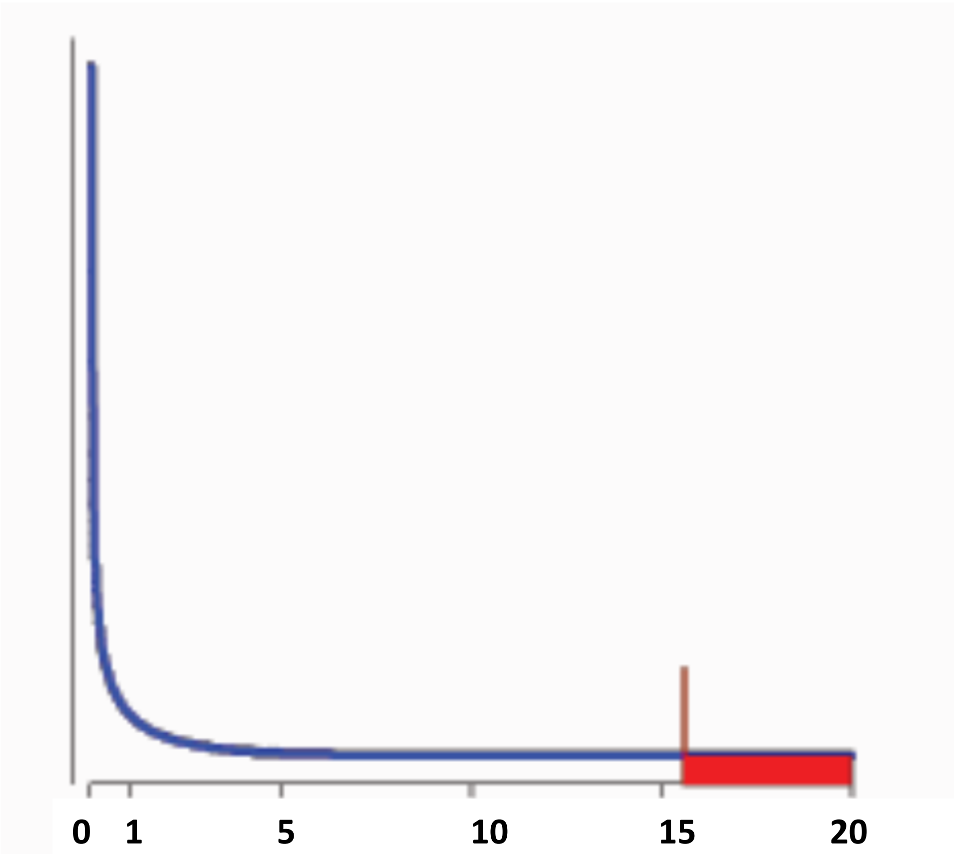
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(13-26.3)^2}{26.3} + \frac{(40-26.7)^2}{26.7} + \frac{(167-153.7)^2}{153.7} + \frac{(143-156.3)^2}{156.3}$$

$$= 15.6$$

# P value

- P=0.0001



# More example



## Vaccine& Influenza

# Example 2 × 2 table

| Influenza | | | |
|---|---|---|---|
| | yes | No | Total |
| Vaccine | 20 | 220 | 240 |
| Placebo | 80 | 140 | 220 |
| Total | 100 | 360 | 460 |

# Example 2×2 table

| Influenza | | | |
|---|---|---|---|
| | yes | No | Total |
| Vaccine | 20 | 220 | 240 |
| Placebo | 80 | 140 | 220 |
| Total | 100 | 360 | 460 |

**Expected numbers**

| Influenza | | | |
|---|---|---|---|
| | yes | No | Total |
| Vaccine | 52.2 | 187.8 | 240 |
| Placebo | 47.8 | 172.2 | 220 |
| Total | 100 | 360 | 460 |

# Example 2×2 table

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 53.09$$

df=1 for 2×2 table

**Yates's continuity correction (连续性矫正)**

$$\chi^2 = \sum \frac{(|O - E| - 0.5)^2}{E}, d.f. = 1$$

**The corrected value is 51.46**

# Yates's continuity correction (连续性矫正)

**Yates's continuity correction is needed, when**

df=1, and total sample size (total of numbers in the table) is less than 40

or, df=1, and the smallest expected number is less than 5

Cochran (1954) recommended

It reduces the size of the chi-square value and so reduces the chance of finding a statistically significant difference, so that correction for continuity makes the test more conservative.

# Yates's continuity correction (连续性矫正)

**When df>1, NO Yates's continuity correction !**

**When df>1, the chi-squared test is valid when less 20% of the expected numbers are under 5 and none is less than 1; data is randomly selected.**

Cochran (1954) recommended

# Summary: the Chi-Square Test

- Both chi-square tests use the same statistic. The calculation of the chi-square statistic requires two steps:

1. The null hypothesis is used to construct an idealized sample distribution of **expected frequencies** that describes how the sample would look if the data were in perfect agreement with the null hypothesis.

# Expected Frequency in The Chi-Square Test

For the goodness of fit test, the expected frequency for each category is obtained by

expected frequency = p*n

(p is the proportion from the null hypothesis and n is the size of the sample)

For the test for independence, the expected frequency for each cell in the matrix is obtained by

expected frequency = (row total)(column total) / n

# Caveat

**When the sample size is less than 40 or the expected frequency is very small in any cell (<1), **Fisher's exact test** is used as an alternative to the chi-square test.

# What do we do if the expected values in any of the cells in a 2x2 table is below 5?

For example, a sample of teenagers might be divided into male and female on the one hand, and those that are and are not currently dieting on the other. We hypothesize, perhaps, that the proportion of dieting individuals is higher among the women than among the men, and we want to test whether any difference of proportions that we observe is significant. The data might look like this:

# What do we do if the expected values in any of the cells in a 2x2 table is below 5?

For example, a sample of teenagers might be divided into male and female on the one hand, and those that are and are not currently dieting on the other. We hypothesize, perhaps, that the proportion of dieting individuals is higher among the women than among the men, and we want to test whether any difference of proportions that we observe is significant. The data might look like this:

| | men | women | total |
|---|---|---|---|
| dieting | 1 | 9 | 10 |
| not dieting | 11 | 3 | 14 |
| totals | 12 | 12 | 24 |

# What do we do if the expected values in any of the cells in a 2x2 table is below 5?

For example, a sample of teenagers might be divided into male and female on the one hand, and those that are and are not currently dieting on the other. We hypothesize, perhaps, that the proportion of dieting individuals is higher among the women than among the men, and we want to test whether any difference of proportions that we observe is significant. The data might look like this:

**The expected values:**

| | men | women | total |
|---|---|---|---|
| dieting | 5 | 5 | 10 |
| not dieting | 7 | 7 | 14 |
| totals | 12 | 12 | 24 |

# Binary or categorical outcomes (proportions)

| Outcome Variable | Are the observations correlated? | | Alternative to the chi-square test if sparse cells: |
|---|---|---|---|
| | independent | correlated | |
| Binary or categorical (e.g. fracture, yes/no) | **Chi-square test:** compares proportions between more than two groups<br><br>**Relative risks:** odds ratios or risk ratios (for 2x2 tables)<br><br>**Logistic regression:** multivariate technique used when outcome is binary; gives multivariate-adjusted odds ratios | **McNemar's chi-square test:** compares binary outcome between correlated groups (e.g., before and after)<br><br>**Conditional logistic regression:** multivariate regression technique for a binary outcome when groups are correlated (e.g., matched data)<br><br>**GEE modeling:** multivariate regression technique for a binary outcome when groups are correlated (e.g., repeated measures) | **Fisher's exact test:** compares proportions between independent groups when there are sparse data (some cells <5).<br><br>**McNemar's exact test:** compares proportions between correlated groups when there are sparse data (some cells <5). |

# Fisher's Exact Test

**The exact test is recommended for a 2×2 table, when**

1. the overall total of the table is less than 20, or
2. the overall total is between 10 and 40 and the smallest of the four expected numbers is less than 5

**The chi-squared test is valid when the overall total is more than 40**

Cochran (1954) recommended

The question we ask about these data is: knowing that 10 of these 24 teenagers are dieters, what is the probability that these 10 dieters would be so unevenly distributed between the girls and the boys? If we were to choose 10 of the teenagers at random, what is the probability that 9 of them would be among the 12 girls, and only 1 from among the 12 boys?

**--Hypergeometric distribution!**

a discrete probability distribution that describes the probability of ***k*** successes in ***n*** draws from a finite population of size ***N*** <u>**without**</u> replacement.

--<u>**Fisher's exact test**</u> uses hypergeometric distribution to calculate the "exact" probability of obtaining such set of the values.

# Fisher's exact test

Before we proceed with the Fisher test, we first introduce some notation. We represent the cells by the letters *a*, *b*, *c* and *d*, call the totals across rows and columns *marginal totals*, and represent the grand total by *n*. So the table now looks like this:

| | men | women | total |
|---|---|---|---|
| dieting | $a$ | $b$ | $a + b$ |
| not dieting | $c$ | $d$ | $c + d$ |
| totals | $a + c$ | $b + d$ | $n$ |

| | men | women | total |
|---|---|---|---|
| dieting | $a$ | $b$ | $a+b$ |
| not dieting | $c$ | $d$ | $c+d$ |
| totals | $a+c$ | $b+d$ | $n$ |

Fisher showed that the probability of obtaining any such set of values was given by the hypergeometric distribution:

$$p = \binom{a+b}{a}\binom{c+d}{c} \Big/ \binom{n}{a+c}$$

$$= \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

# In our example

As extreme as observed

More extreme than observed

| | men | women | total |
|---|---|---|---|
| dieting | 1 | 9 | 10 |
| not dieting | 11 | 3 | 14 |
| totals | 12 | 12 | 24 |

| | men | women | total |
|---|---|---|---|
| dieting | 0 | 10 | 10 |
| not dieting | 12 | 2 | 14 |
| totals | 12 | 12 | 24 |

$$p = \frac{10!14!12!12!}{24!1!9!11!3!} = 0.00134$$

$$p = \frac{10!14!12!12!}{24!0!10!12!2!} = 0.00003$$

Recall that p-value is the probability of observing data as extreme or more extreme if the null hypothesis is true. So the p-value is this problem is 0.00137.

# The fisher Exact Probability Test

- Used when one or more of the expected counts in a contingency table is small.
- Fisher's Exact Test is based on exact probabilities from a specific distribution (the hypergeometric distribution).
- There's really no lower bound on the amount of data that is needed for Fisher's Exact Test. You can use Fisher's Exact Test when one of the cells in your table has a zero in it. Fisher's Exact Test is also very useful for highly imbalanced tables. If one or two of the cells in a two by two table have numbers in the thousands and one or two of the other cells has numbers less than 5, you can still use Fisher's Exact Test.
- Fisher's Exact Test has no formal test statistic and no critical value, and it only gives you a p-value.

END