Announcements:

- Final quiz begins after class today and ends on Monday at noon.
- For Monday discussion, teams will discuss results from articles in medical journals. Article summaries are posted on course website (in daily calendar).

Homework (due Mon, Nov 29):

Chapter 15: #6, #26bd, #35

Section 15.3

Chi-Square Test for Goodness-of-Fit

Research Question

For a categorical variable with *k* categories, are the population proportions (or probabilities) falling into each of the *k* categories as specified?

Examples:

- Are digits 0 to 9 equally likely to be drawn in lottery?
- In genetics, is offspring ratio 9:3:3:1, as expected by Mendel's laws?
- Is death from sudden infant death syndrome equally likely in all 4 seasons?

15.3 Testing Hypotheses about One Categorical Variable

Situation: Similar to binomial, but there can be more than two possible outcomes. Called *multinomial*.

•Measure a single categorical variable on each person or trial.

•Each person or trial falls into one of *k* mutually exclusive categories.

•Null hypothesis specifies the probabilities of falling into each of the *k* categories.

•Alternative hypothesis is that those are not all correct.

Example 15.8 *Pennsylvania Daily Number*

State lottery game: Three-digit number made by drawing a digit between 0 and 9 from each of three different containers.

Let's examine draws from the first container.
If numbers randomly selected, each value would be equally likely to occur. So, $k = 10$ and on each draw there is probability 1/10 of getting each digit (0, ..., 9)

H_0: $p = 1/10$ for each of the 10 possible digits
H_a: Not all probabilities are 1/10.

Use same 5 steps of hypothesis testing Called chi-square *goodness-of-fit* test

Step 1: *Determine the null and alternative hypotheses.*

H_0: The probabilities for k categories are $p_1, p_2, \ldots, p_k$.

H_a: Not all probabilities specified in H_0 are correct.

Note: Probabilities in the null hypothesis must sum to 1.

Pennsylvania Lottery Example:

H_0: $p_1 = p_2 = \ldots = p_{10} = 1/10$

H_a: The 10 digits are not all equally likely.

Goodness of Fit (GOF) Test *(continued)*

Step 2: *Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.*

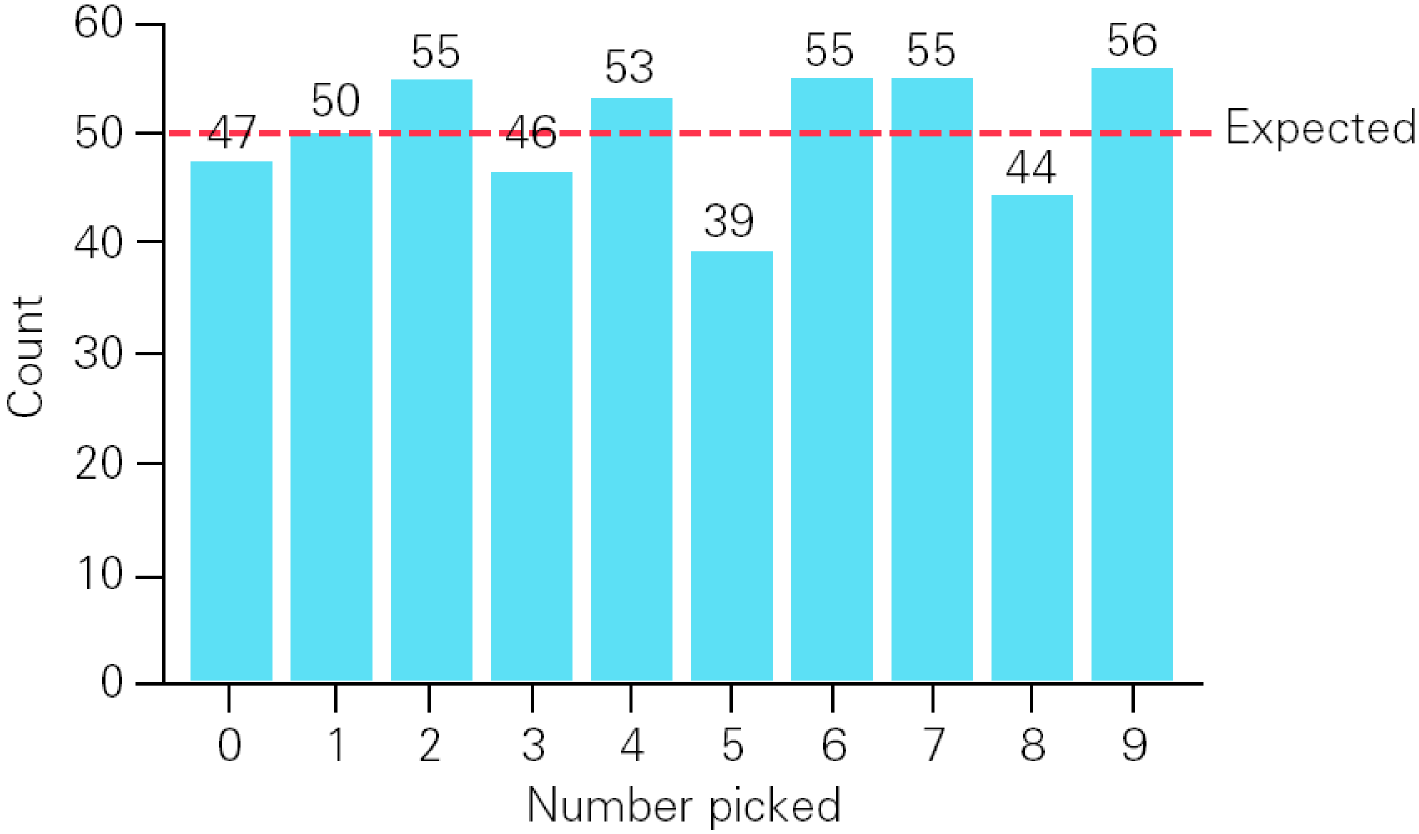
Data condition needed: At least 80% of the expected counts are greater than 5 and none are less than 1. Test statistic:

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

where the **expected count** for the i^{th} category is computed as np_i.

Example 15.8 *Pennsylvania Daily Number*

Data: n = 500 days between 7/19/99 and 11/29/00



Example, continued

Expected count = 500 (1/10) = 50 for each digit

$$\chi^2 = \sum_{\text{categories}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$= \frac{(47-50)^2}{50} + \frac{(50-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(46-50)^2}{50} + \frac{(53-50)^2}{50}$$

$$+ \frac{(39-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(55-50)^2}{50} + \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} = 6.04$$

Step 3: *p*-value of Chi-square Test

Large test statistic => evidence that values in null are not correct (observed counts don't match expected counts).

p-value = probability the chi-square test statistic could have been as large or larger if the null hypothesis were true.

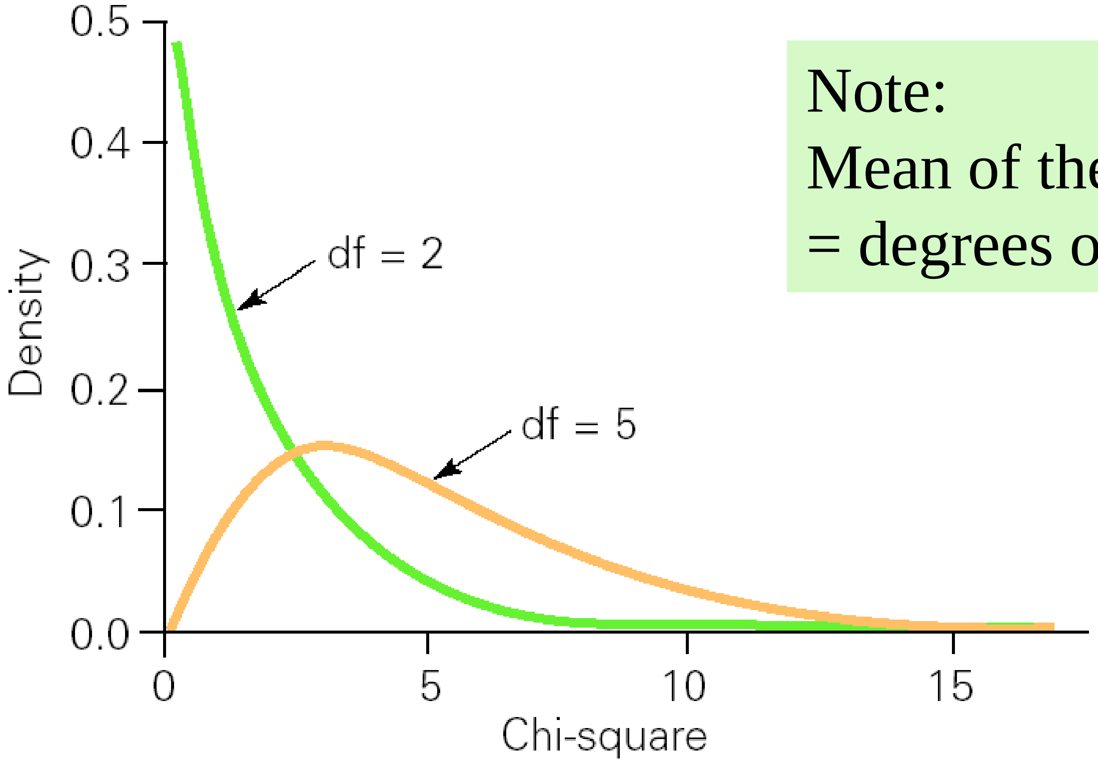
<u>Chi-square probability distribution</u> used to find *p*-value.

Degrees of freedom: **df = k – 1**

This is because we are free to specify k – 1 totals, then the last one is determined.

Chi-square Distributions

- Skewed to the right distributions.
- Minimum value is 0.
- Indexed by the *degrees of freedom*.



Note:
Mean of the distribution
= degrees of freedom.

Finding the *p*-value from Table A.5, p. 732:

Look in the corresponding **"df" row** of Table A.5. Scan across until you find where the statistic falls.

- If value of statistic falls between two table entries, *p*-value is between values of *p* (column headings) for these two entries.
- If value of statistic is larger than entry in rightmost column (labeled $p = 0.001$), *p*-value is less than 0.001 (written as $p < 0.001$).
- If value of statistic is smaller than entry in leftmost column (labeled $p = 0.50$), *p*-value is greater than 0.50 (written as $p > 0.50$).

Step 4: Making a Decision

Large test statistic => small *p*-value
=> evidence that the proportions are *not* as specified.

Two equivalent rules:

Reject H_0 when …

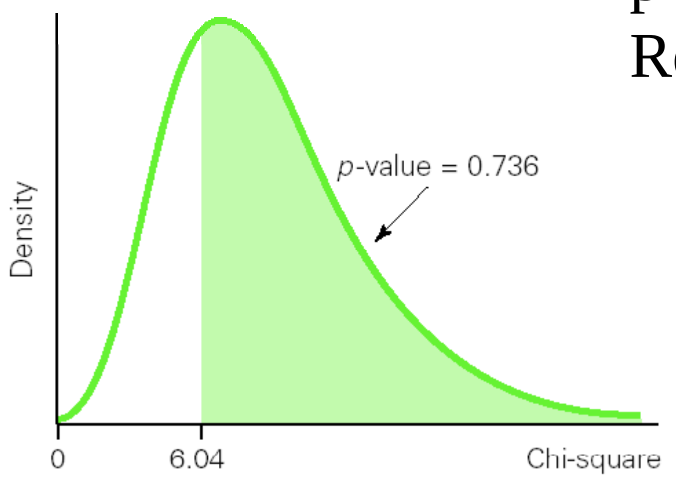
- *p*-value ≤ 0.05
- Chi-square statistic is greater than the entry in the 0.05 column of Table A.5 (the critical value). That defines the *rejection region.*

Example 15.8 *Daily Number (cont)*

Chi-square goodness of fit statistic:

From Table A.5 (page 732) gives areas to the *right* of the chi-square value, because that's the *p*-value in this situation.:

Example: Chi-square value = 6.04.
df = k – 1 = 10 – 1 = 9
p-value > 0.50 (note it is 0.736)
Rejection region: Above 16.92.



Result is ***not statistically significant***; the ***null hypothesis is not rejected***.

Step 5: Report the Conclusion in Context

Conclusion: Pennsylvania lottery digits drawn *are not statistically different* from what's expected by chance.

New Example: Is Sudden Infant Death Syndrome (SIDS) Seasonal?

Data from King County, Washington

Define p_1, p_2, p_3, p_4 to be the proportion of deaths from SIDS that happen in the winter, spring, summer and fall. They are defined so that the seasons have about equal days.

Step 1: *Determine the null and alternative hypotheses.*

H_0: $p_1 = ¼$, $p_2 = ¼$, $p_3 = ¼$, $p_4 = ¼$

H_a: Not all probabilities specified in H_0 are correct.

Note: Probabilities in the null hypothesis must sum to 1.

Step 2: *Verify necessary data conditions, and if met, summarize the data into an appropriate test statistic.*

Data condition needed: At least 80% of the expected counts are greater than 5 and none are less than 1. Test statistic:

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

where the **expected count** for the i^{th} category is computed as np_i.

Example: Counts for the 4 seasons were 78, 71, 87, 86

Are these different enough to conclude a difference exists in the population?

Example, continued

Total n = 78 + 71 + 87 + 86 = 322

Expected count = 322 (1/4) = 80.5 for each season.

$$\chi^2 = \frac{(78-80.5)^2}{80.5} + \frac{(71-80.5)^2}{80.5} + \frac{(87-80.5)^2}{80.5} + \frac{(86-80.5)^2}{80.5}$$

= 2.10

Step 3: Finding the *p*-value

Degrees of freedom = 4 – 1 = 3.

From Table A.5, smallest entry is 2.37, the value with .50 below it. So, for our test statistic of 2.10 all we can say is *p*-value > .50.

<u>Rejection region approach:</u>

For df = 3, reject the null hypothesis if the test statistic is greater than 7.81. (Ours is not.)

Step 4: Making a Decision

Large test statistic => small *p*-value
=> evidence that the proportions are *not* as specified.

Two equivalent rules:
Reject H_0 when …

- *p*-value ≤ 0.05; in our example it is.
- Chi-square statistic is greater than the entry in the 0.05 column of Table A.5 (the critical value). That defines the *rejection region.* In our example, the test statistic is *not* in the rejection region.
- **So we do not reject the null hypothesis.**

Step 5: Report the Conclusion in Context

Conclusion: Sudden infant death syndrome proportions across seasons *are not statistically different* from what's expected by chance (i.e. all seasons being equal).

Use of chi-square test in genetics

- Based on Mendel's laws, expect certain ratios of phenotypes. Can be tested using chi-square goodness-of-fit tests.
- Example: In a dihybrid cross (AaBb x AaBb), the expected proportions of 4 phenotypes are 9:3:3:1.
- Data from classic experiment with Starchy/sugary and Green/white seedlings, progeny of 3839 self-fertilized heterozygotes (Starchy/green, Starchy/white, Sugary/green, Sugary/white):

  1997, 906, 904, 32.
- Null hypothesis probabilities are 9/16, 3/16, 3/16, 1/16

Results from Minitab for this example

Chi-Square Goodness-of-Fit Test

Category	Observed	Test Proportion	Expected	Contribution to Chi-Sq
1	1997	0.5625	2159.44	12.219
2	906	0.1875	719.81	48.159
3	904	0.1875	719.81	47.130
4	32	0.0625	239.94	180.205

N	DF	Chi-Sq	P-Value
3839	3	287.714	0.000

Because the *p*-value is 0.000, reject the null hypothesis. Conclude that in this case, the genetics did not work out to be the 9:3:3:1 ratio expected.
Note that the largest contribution to the large test statistic is (4)Sugary/white. Observed = 32, expected = 239.94.