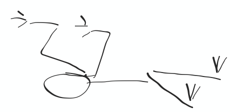
WARM UP



EXPERIMENTAL DESIGN 26

A television network conducts a weekly survey to determine the proportion of viewers who watch various programs. For the coming year, they decide to double the sample size. The main benefit of this is to

(A) reduce undercoverage bias.

(B) reduce nonresponse bias.

(C) eliminate sampling error.

(D) decrease population variability.

(E) decrease the standard deviation of the sampling distribution.

Answer: (E) The standard deviation of the sample proportions, $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ is reduced if the sample size n is increased. This reduces the margin of error but has no effect on possible bias. Sampling error, naturally present when sample statistics are used to estimate population parameters, can generally be reduced with larger sample sizes, but it cannot be eliminated.

EXPERIMENTAL DESIGN 28

Which of the following statements is *incorrect*?

(A) Voluntary response samples often underrepresent people with strong opinions.

(B) Convenience samples often lead to undercoverage bias.

(C) Questionnaires with nonneutral wording are likely to have response bias.

(D) There is no way to fix the results if a biased sampling method was employed.

(E) Nonresponse bias should be avoided because those who do not respond might have different views from those who do respond.

Answer: (A) Voluntary response samples, like radio call-in surveys, are based on individuals who offer to participate, and they typically overrepresent persons with strong opinions. Convenience samples, like shopping mall surveys, are based on choosing individuals who are easy to reach, and they typically miss a large segment of the population. Nonneutral wording can readily lead to response bias, and if surveyors want a particular result, they deliberately use certain wording. Always check carefully for bias before collecting data, because there is no recovery from a biased sampling method after the sample is collected. When a large fraction of those sampled fail to respond, the concern is whether those who do respond have different

EXPERIMENTAL DESIGN 27

A botanist is running an experiment on two fertilizers that require different amounts of watering. She has 40 test plots, half of which are in sunny locations, and half are in the shade. She randomly selects 10 sunny plots and 10 shady plots for which to use one fertilizer with its appropriate watering, while the remaining plots are for the other fertilizer with its appropriate watering. Of the following, which is the most important observation about this procedure?

(A) The variables, fertilizer and water, are confounded.

(B) The variables, fertilizer and sun, are confounded.

(C) The variables, water and sun, are confounded.

(D) No variables are confounded.

(E) There is a hidden lurking variable.

Answer: (A) When there is uncertainty with regard to which variable is causing an effect, we say the variables are *confounded*. In this experiment, it might be difficult to determine if the difference in fertilizers or the difference in watering is the real cause of observed differences in plant growth. However, with both sunny and shady plots for each fertilizer, fertilizer and sun are not confounded.

The M&M activity from our last class....

allowed us to look at a distribution of a single categorical variable (color) with 2 or more categories [brown, yellow, orange, Green, Blue, Red].

ESSENTIAL QUESTION *How do we perform significance tests for distributions of categorical variables and for relationships between categorical variables?*

The Chi-Square Test for Goodness of Fit

is a significance test that can help determine if a distribution of a claimed population [Company claimed 13% Brown, 14% Yellow, 20% orange, etc] differs from a sample distribution.

The Big Picture: Where Chapter 11 Fits

- Chapters 8–12 cover all of AP Statistics Topic Outline IV. Statistical inference: Estimating population parameters and testing hypotheses.
- About 30-40% of questions on the AP exam
- Chapters 11 focuses on inference for categorical data:
  - Section 11.1 extends on the one-sample z test for a proportion (Section 9.2) by allowing more than 2 categories for the variable.
  - Section 11.2 extends on the two-sample z test for a difference in proportions (Section 10.1) by allowing comparisons of more than 2 populations/treatments and allowing more than 2 categories for the variable.

Chapter 11: Inference for Distributions of Categorical Data

11.1 Chi-Square Tests for Goodness of Fit	2 Days
11.2 Inference for Two-way Tables	2 Days
Review, FRAPPY, and Test	2 Days

which puts the Ch. 11 test on
Tuesday, Feb 19th

Chapter 11: The Big Ideas

- Chi-square tests are about categorical variables
- Three types of chi-square tests
  - Goodness of Fit: Distribution of 1 categorical variable in 1 population
  - Homogeneity: Distribution of 1 categorical variable for 2 or more populations/treatments
  - Independence: Relationship between 2 categorical variables in 1 population

No CI's this chapter

Chi-Square tests are similar to other tests.

- Four Step process
- Need to Check Conditions

but there are differences

today:

✓STATE appropriate hypotheses and COMPUTE the expected counts and chi-square test statistic for a chi-square test for goodness of fit.

✓STATE and CHECK the Random, 10%, and Large Counts conditions for performing a chi-square test for goodness of fit.

✓CALCULATE the degrees of freedom and *P*-value for a chi-square test for goodness of fit.

Last class we assumed that the large bag of M&M'S was the population. We investigated if the distribution of the colors differed from the claimed distribution.

Stating Hypotheses

Ho **:** The distribution of color in the large bag of M&M'S Milk Chocolate Candies **is the same as** the claimed distribution.

Ha: The distribution of color in the large bag of M&M'S Milk Chocolate Candies **is *not* the same as** the claimed distribution.

In this chapter, there will be no two-sided tests. :)

The alternative hypothesis will always be "the null hypothesis is not correct"

We can also write the hypotheses in symbols.

H_0: $p_{brown} = 0.125$, $p_{red} = 0.125$, $p_{yellow} = 0.125$, $p_{green} = 0.125$, $p_{orange} = 0.25$, $p_{blue} = 0.25$

H_a: At least two of the p_i's are incorrect

where p_{color} = the true proportion of M&M'S Milk Chocolate Candies in the large bag of that color

A third way to write hypotheses:

H_0: The company's claimed distribution of color is correct for this bag of M&M's.

H_a: The company's claimed distribution of color is incorrect for this bag of M&M's.

CAUTION:

Don't state the alternative hypothesis in a way that suggests that all the proportions in the hypothesized distribution are wrong.

Comparing Observed and Expected Counts: The Chi-Square Test Statistic

How many candies of each color should we expect to find in a sample of 60 candies? We use the claimed proportions as the truth:

Brown:	(60)(0.125) = 7.5	Green:	(60)(0.125) = 7.5
Red:	(60)(0.125) = 7.5	Orange:	(60)(0.25) = 15
Yellow:	(60)(0.125) = 7.5	Blue:	(60)(0.25) = 15

np_i

categories 1, 2, 3, ...

Calculating Expected Counts in a Chi-Square Test for Goodness of Fit

The expected count for category *i* in the distribution of a categorical variable is

$$np_i$$

where p_i is the relative frequency for category *i* specified by the null hypothesis.

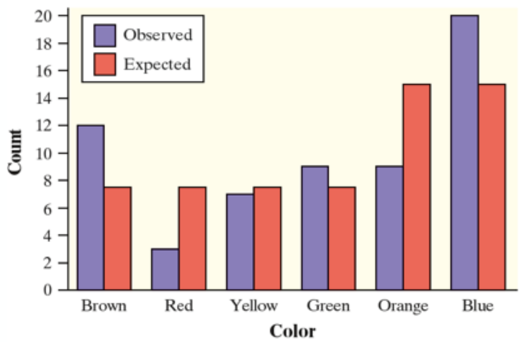
How To Pronounce ?

Walking into a room full of statisticians and referring to the "Chai" square test statistic will result in immediate loss of credibility.

["Chai" should only be used when ordering a drink at Starbucks.]

Please help your bioliogy teachers with this one!

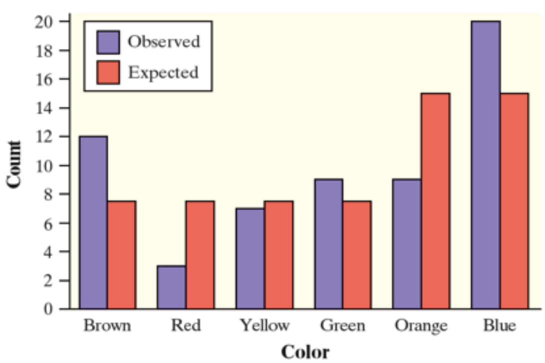
Comparing Observed and Expected Counts: The Chi-Square Test Statistic



Color	Observed	Expected
Brown	12	7.5
Red	3	7.5
Yellow	7	7.5
Green	9	7.5
Orange	9	15.0
Blue	20	15.0

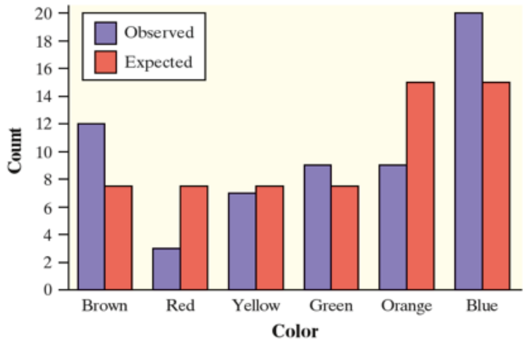
Observe Graphically
or Numerically

Comparing Observed and Expected Counts: The Chi-Square Test Statistic



Color	Observed	Expected
Brown	12	7.5
Red	3	7.5
Yellow	7	7.5
Green	9	7.5
Orange	9	15.0
Blue	20	15.0

How likely is it that differences this large or larger would occur just by chance in random samples of size 60 from the population distribution claimed by Mars, Inc.?



Color	Observed	Expected
Brown	12	7.5
Red	3	7.5
Yellow	7	7.5
Green	9	7.5
Orange	9	15.0
Blue	20	15.0

How likely is it that differences this large or larger would occur just by chance in random samples of size 60 from the population distribution claimed by Mars, Inc.?

The statistic we use to make the comparison is the **chi-square test statistic** χ^2.

The **chi-square test statistic** is a measure of how far the observed counts are from the expected counts. The formula for the statistic is

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

where the sum is over all possible values of the categorical variable.

Why divide by the expected count?

We are interested in how far away the observed count is from the expected

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

always include summation symbol.

Observed counts, not observed proportions

The **chi-square test statistic** is a measure of how far the observed counts are from the expected counts. The formula for the statistic is

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

where the sum is over all possible values of the categorical variable.

A typical calculation, by hand, look like this:

$$\chi^2 = \frac{(12-7.5)^2}{7.5} + \frac{(3-7.5)^2}{7.5} + \frac{(7-7.5)^2}{7.5} + \frac{(9-7.5)^2}{7.5} + \frac{(9-15)^2}{15} + \frac{(20-15)^2}{15}$$

category 1 2 3 3 4 . . .

The **chi-square test statistic** is a measure of how far the observed counts are from the expected counts. The formula for the statistic is

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

where the sum is over all possible values of the categorical variable.

For Jerome's data, we add six terms—one for each color category:

$$\chi^2 = \frac{(12-7.5)^2}{7.5} + \frac{(3-7.5)^2}{7.5} + \frac{(7-7.5)^2}{7.5} + \frac{(9-7.5)^2}{7.5} + \frac{(9-15)^2}{15} + \frac{(20-15)^2}{15}$$

$$= 2.7 + 2.7 + 0.03 + 0.30 + 2.4 + 1.67$$

$$= 9.8$$

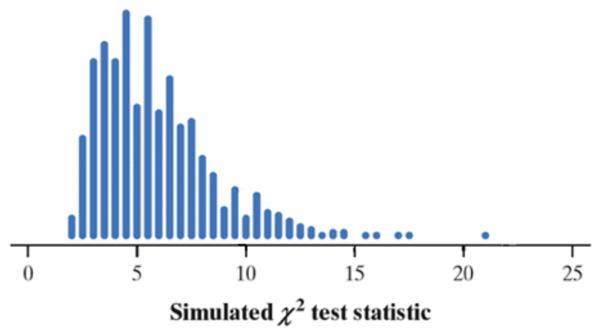
always positive!

The Chi-Square Distributions and *P*-Values

If we used software to simulate taking 1000 random samples of size 60 from the population distribution of M&M'S Milk Chocolate Candies given by Mars, Inc. Here are the values of the chi-square test statistic for these 1000 samples.

The Chi-Square Distributions and *P*-Values

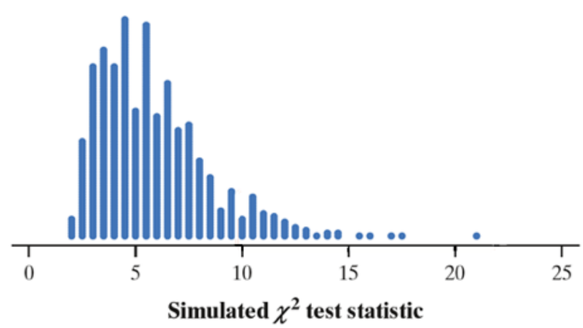
If we used software to simulate taking 1000 random samples of size 60 from the population distribution of M&M'S Milk Chocolate Candies given by Mars, Inc. Here are the values of the chi-square test statistic for these 1000 samples.



The Chi-Square Distributions and *P*-Values

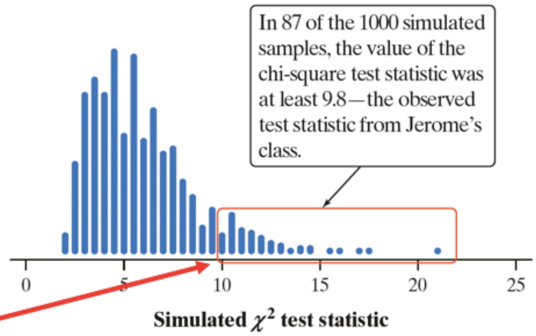
If we used software to simulate taking 1000 random samples of size 60 from the population distribution of M&M'S Milk Chocolate Candies given by Mars, Inc. Here are the values of the chi-square test statistic for these 1000 samples.

Larger values of χ^2 give more convincing evidence against H_0 and in favor of H_a.



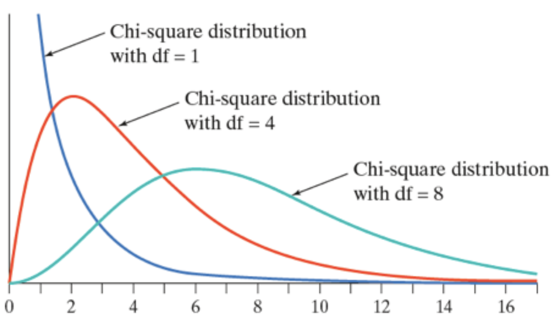
We used software to simulate taking 1000 random samples of size 60 from the population distribution of M&M'S Milk Chocolate Candies given by Mars, Inc. Here are the values of the chi-square test statistic for these 1000 samples.

Larger values of χ^2 give more convincing evidence against H_0 and in favor of H_a.



Our estimated *P*-value is 87/1000 = 0.087

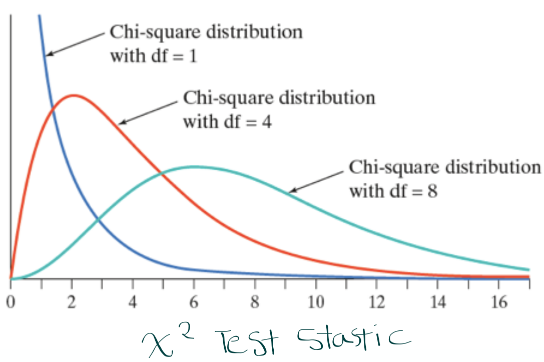
A **chi-square distribution** is defined by a density curve that takes only nonnegative values and is skewed to the right. A particular chi-square distribution is specified by its degrees of freedom.



df = No. of Categories − 1

Pick Up
the
Hand Out

A **chi-square distribution** is defined by a density curve that takes only nonnegative values and is skewed to the right. A particular chi-square distribution is specified by its degrees of freedom.



- As the degrees of freedom (df) increase, the density curves become less skewed.
- The mean of a particular chi-square distribution is equal to its degrees of freedom.
- For df > 2, the mode (peak) of the chi-square density curve is at df – 2.

The χ^2 distributions are the third and final type of distribution that you will use regularly.

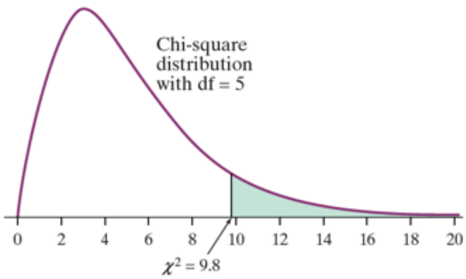
[Normal distrib., t-distributions, χ^2 distrib.]

find table C

There are 6 color categories for M&M's® Milk Chocolate Candies, so df = 6 – 1 = 5.

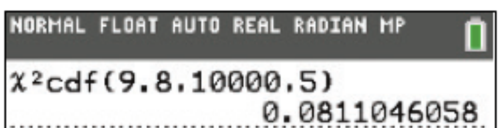
The *P*-value is the probability of getting a χ^2 value of as large as or larger than 9.8 when H_0 is true.

TEST STATISTIC



	P		
df	**.15**	**.10**	**.05**
4	6.74	7.78	9.49
5	8.12	9.24	11.07
6	9.45	10.64	12.59

The *P*-value for a test based on Jerome's data is between 0.05 and 0.10.



As usual... specify lower, upper df

We now have 3 calculator functions that can help us find areas for different distributions

normalcdf

tcdf

χ^2cdf

Conditions for Performing a Chi-Square Test for Goodness of Fit

Random: The data come from a random sample from the population of interest.

10%: When sampling without replacement, $n < 0.10N$.

Large Counts: All *expected* counts are at least 5.

Why Check

Random: So.... we can generalize to the appropriate population

10%: So.... sampling without replacement is ok.

Large Counts: So.... so the sampling distribution is approximately a chi-square distribution and we can use χ^2 to find a P-Value

The Chi-Square Test for Goodness of Fit

Suppose the conditions are met. To perform a test of H_0: The stated distribution of a categorical variable in the population of interest is correct compute the chi-square test statistic:

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

where the sum is over the k different categories. The P-value is the area to the right of χ^2 under the chi-square density curve with $k - 1$ degrees of freedom.

A chi-square density curve is a good model for the sampling distribution of the chi-square statistic but only when the sample size is large enough that the expected counts are all at least 5

[similar to Large Counts Condition for $\hat{p}$]

ie

Outcome of roll	1	2	3	4	5	6	Total
Observed count	12	28	12	13	10	15	90
Expected count	15	15	15	15	15	15	90

The color of Reese's Pieces®

The Hershey Company makes Reese's Pieces and claims the distribution of colors is as follows: 50% orange, 25% brown, and 25% yellow. Skeptical of this claim, Trey purchases a very large bag of Reese's Pieces and selects a random sample of 80 pieces. Here are the results:

Color	Orange	Brown	Yellow
Count	31	22	27

(a) State the hypotheses that Trey should test.

The color of Reese's Pieces®

The Hershey Company makes Reese's Pieces and claims the distribution of colors is as follows: 50% orange, 25% brown, and 25% yellow. Skeptical of this claim, Trey purchases a very large bag of Reese's Pieces and selects a random sample of 80 pieces. Here are the results:

Color	Orange	Brown	Yellow
Count	31	22	27

(a) State the hypotheses that Trey should test.

H_0: Hershey's claimed distribution of color for Reese's pieces is correct for this bag.

H_a: Hershey's claimed distribution of color for Reese's Pieces is incorrect...

(b) Calculate the expected count for each of the possible outcomes.

(c) Calculate the value of the chi-square test statistic.

(b) Calculate the expected count for each of the possible outcomes.

If H_0 is true:

$80(.5) = 40$ for orange

$80(.25) = 20$ for brown

$80(.25) = 20$ for yellow

(c) Calculate the value of the chi-square test statistic.

31	22	27
40	20	20

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

$$\chi^2 = \frac{(31-40)^2}{40} + \frac{(22-20)^2}{20} + \frac{(27-20)^2}{20} = 4.675$$

AP Exam Tip

Don't round the expected counts!
(just because you think they should be integers)

- it's the average number of observations in a given category in many many random samples.

(d) Find the P-value using Table C, then calculate a more precise value using technology.

$df = 3 - 1 = 2$ → Table C

P-Value is between 0.05 and 0.10

(d) Find the P-value using Table C, then calculate a more precise value using technology.

$df = 3 - 1 = 2$ → Table C

P-Value is between 0.05 and 0.10

$\chi^2\text{cdf}[\text{lower} \quad \text{upper} \quad \text{df}]$

$\chi^2\text{cdf}[\underset{\text{lower},}{4.675} \quad \underset{\text{upper},}{10000} \quad \underset{\text{df}}{2}] = .0966$

See your
Ch. 10 Test

11.1:....1, 3, 5, 7, 22

Study pp.709-716

FR E 1
P $\frac{1}{2}$
I 0

13 a
b
c
d