AP Statistics
Mr. Murphy
12.1 Chi-Square Tests for Univariate Data
HW Pg. 662 #12.1, 12.2, 12.7, 12.8, 12.9, 12.12

Goals: 1. Run a Chi-squared Test for Goodness of Fit.
2. Have I mentioned we're almost done!?

We have been looking at quantitative data for the past few chapters…now we will switch over to **categorical data.**

Ex1 For 1000 shoppers donating blood at a mall, the frequencies of blood types were as shown in the table below. Consider this an SRS of all mall shoppers.

Blood Type	Frequency
O	465
A	294
B	196
AB	45
Total	1000

In the general population, the blood type distribution is as follows:

Type O = 45%, Type A = 40%,
Type B = 10%, Type AB = 5%,.

Do these data provide evidence that the blood type proportions of mall shoppers differ from the blood type proportions of the general public? Test the appropriate hypotheses using $\alpha = 0.01$.

We will test the following hypotheses:

H_0 : Mall shoppers have the same blood type proportions as the general public
H_a : Mall shoppers DO NOT have the same blood type proportions as the general public

Data was given for n = 1000 mall shoppers. If H_0 is true, what are the expected counts?

Note: A different look for H_0 could be

$$H_0 : p_O = 0.45,\ p_A = 0.40,\ p_B = 0.10,\ p_{AB} = 0.05$$

	Blood Type			
	O	A	B	AB
Observed Count	465	294	196	45
Expected Count	450	400	100	50

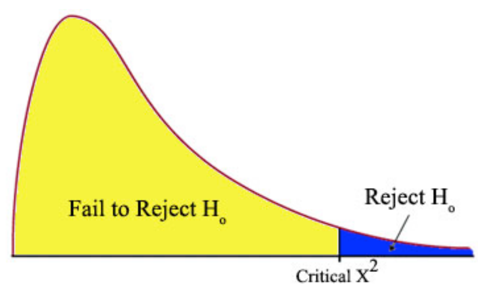
We need a test statistic. We call it the χ^2 (Chi-square) test statistic. Here's how to calculate it.

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed cell count - expected cell count})^2}{\text{expected cell count}}$$

Let's calculate our χ^2 test statistic.

Now how do we get our *P*–value?

χ^2 Goodness-Of-Fit Procedure



1. Hypotheses:

$H_0 : p_1 =$ hypothesized proportion for Category 1

$p_2 =$ hypothesized proportion for Category 2

.

.

.

$p_k =$ hypothesized proportion for Category k

2. Test Statistic: $\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed cell count - expected cell count})^2}{\text{expected cell count}}$

3. P – values: The P – value associated with the computed test statistic value is the area to the **right** of χ^2 under the $df = k - 1$ chi-square curve.

4. <u>*Assumptions*</u>: 1. Observed cell counts are based on a *random sample.*
2. The *sample size is large*. The sample size is large enough for the chi-square test to be appropriate as long as every expected cell count is at least 5.

Ex2 Five partners in a law firm brought in the following numbers of new clients during the past year.

Partner	King	Wong	Patel	Allen	Pickens
Number of New Clients	35	42	22	41	30

Is there sufficient evidence at the 5% or 10% level of significance that the partners do not bring in equal numbers of new clients?

Ex3 A genetic model for offspring of two Labrador retrievers states the ratios for dog colors to be 5:4:1 for black:yellow:chocolate. Two labrador retrievers are bred and a litter consisting of 3 black lab puppies, 5 yellow lab puppies, and 2 chocolate lab puppies is produced. For a goodness of fit test, the χ^2 statistic would be:

(a) 1.79
(b) 2.05
(c) 2.92
(d) 4.94
(e) 7.08

Ex4 A die was rolled 24 times with the following results:

Number of Dots	1	2	3	4	5	6
Frequency	2	8	2	1	3	8

A goodness-of-fit chi-squared test is to be used to test the null hypothesis that the die is fair. At a significance level of $\alpha = 0.01$, the value of the chi-square test statistic and the decision reached is

(a) 12.5; fail to reject the null hypothesis
(b) 12.5; reject the null hypothesis
(c) 25.0; fail to reject the null hypothesis
(d) 25.0; reject the null hypothesis
(e) 75.5; reject the null hypothesis

Checkpoint:
Multiple Choice

1. A highway superintendent states that five bridges into a city are used in the ratio 2:3:3:4:6 during the morning rush hour. A highway study of a simple random sample of 6000 cars indicates that 720, 970, 1013, 1380, and 1917 cars use the five bridges, respectively. Can the superintendent's claim be rejected at the 2.5% or 5% level of significance?

(a) There is sufficient evidence to reject the claim at either of these two levels.
(b) There is sufficient evidence to reject the claim at the 2.5% but not at the 5% level.
(c) There is sufficient evidence to reject the claim at the 5% but not at the 2.5% level.
(d) There is not sufficient evidence to reject the claim at either of these two levels.
(e) There is not sufficient information to answer this question.

2. In a study to compare movie preferences among different age groups, a χ^2 statistic was used. If a small value of the test statistic is obtained, it suggests that

(a) the null hypothesis may not be rejected, since the differences between the observed and expected values are relatively large.
(b) the null hypothesis may be rejected, since the differences between the observed and expected values are relatively large.
(c) the null hypothesis may not be rejected, since the differences between the observed and expected values are relatively small.
(d) the null hypothesis may be rejected, since the differences between the observed and expected values are relatively small.
(e) the null hypothesis may not be rejected, since the differences between the observed and expected values are the same.

3. A reporter believed that police officers were required to write a specific quota of traffic tickets during a month. In order to meet the alleged quota, he believed officers would need to write more tickets during the last week of the month. To investigate the claim, the reporter collected the number of tickets written by the local police force in a month and organized them by weeks as shown in the table below.

Week	First Week	Second Week	Third Week	Fourth Week	Total
Tickets Written	133	112	154	165	564

A chi-square analysis was performed to test the claim that there is a relationship between the week of the month and the number of tickets written. What is the P-value of the test?

(a) $0.0005 < P < 0.001$
(b) $0.0025 < P < 0.005$
(c) $0.005 < P < 0.01$
(d) $0.01 < P < 0.02$
(e) $0.025 < P < 0.05$