Section 10.2

The Chi-Square Distribution

We now learn about the χ^2 distribution. The following is the equation for the χ^2-distribution:

$$f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2},$$

where Γ is the Gamma function and n is a positive integer.

As with the t-distribution, there are an infinite number of χ^2 distribution curves. A few of the curves are illustrated below:

You can use the χ^2cdf function on your calculator to find the area under a chi-square distribution curve.

Calculator Function: χ^2cdf(a,b,n) gives the area under the χ^2_n curve between a and b.

Remark: Unlike with normal random variables, the range of χ^2 random variables is over **nonnegative** numbers.

Exercise 1. Let $Y \sim \chi^2_7$.
(a) Find $\mathrm{P}(Y < 5)$.

This probability is

$\mathrm{P}(Y < 5) =$

(b) Find $\mathrm{P}(2 < Y < 4)$.

$$\mathrm{P}(2 < Y < 4) =$$

Class Exercise 1. Let $Y \sim \chi^2_{11}$.
(a) Find $\mathrm{P}(Y > 7.24)$. **Answer: 0.779**
(b) Find $\mathrm{P}(Y < 21.67)$. **Answer: 0.973**

Class Exercise 2. Let $Y \sim \chi^2_{34}$. Find $\mathrm{P}(4.68 < Y < 12)$. **Answer: 1.75** $\cdot 10^{-4}$

Exercise 2. (a) Suppose we roll a six-sided die 120 times. If the die is fair, how many 1's, 2's, 3's, 4's, 5's, and 6's would we expect?

If the die is fair, we would expect to get '1' of the time, '2' of the time, ..., and '6' of the time. For 120 rolls, we would expect Likewise, we would expect 2's, 3's, 4's, 5's, and 6's.

(b) When we roll the die, we get 30 1's, 10 2's, 30 3's, 10 4's, 30 5's, and 10 6's. Given part (a), how plausible is it that the die is fair?

It is

(c) Suppose we roll the die 120 times again. Let's suppose we got 17 1's, 16 2's, 15 3's, 21 4's, 22 5's, and 29 6's. How plausible is it that the die is fair in this case?

It is Some people may say it is

Part (c) in the last exercise is open for debate. To help settle this debate, we first summarize the findings in part (a) and part (c) with the following table.

Value	Observed Frequency	Expected Frequency
1	17	20
2	16	20
3	15	20
4	21	20
5	22	20
6	29	20

Exercise 3. (a) Compute the following statistic for the above table:

$\sum_{all\ values}$(observed frequency - expected frequency)^2/expected frequency.

Value	Observed Frequency	Expected Frequency	$(O\text{-}E)^2/E$
1	17	20	
2	16	20	
3	15	20	
4	21	20	
5	22	20	
6	29	20	

Notation: $(O\text{-}E)^2/E$ denotes (observed frequency - expected frequency)^2/expected frequency.

$\sum_{all\ values}$ $(O\text{-}E)^2/E$ = the sum of all values in the last column. This equals

(b) Find $\sum_{all\ values}$ $(O\text{-}E)^2/E$ for Exercise 2(b).

Value	Observed Frequency	Expected Frequency	$(O\text{-}E)^2/E$
1			
2			
3			
4			
5			
6			

$\sum_{all\ values}$ $(O\text{-}E)^2/E$ =

It turns out that

$$\sum (O - E)^2/E \sim \chi^2_k,$$

where the sum is taken over the cells in the table and k = number of categories - 1.

Goodness of Fit Test Requirement

The sample size n should be large enough that $E \geq 5$ in each category.

Exercise 4. Suppose we roll a die 120 times and get 17 1's, 16 2's, 15 3's, 21 4's, 22 5's, and 29 6's. Is there enough evidence that the die isn't fair? Carry out a hypothesis test at the 0.01 significance level.

Let p_1 = probability of getting 1.
Let p_2 = probability of getting 2.
Let p_3 = probability of getting 3.
Let p_4 = probability of getting 4.
Let p_5 = probability of getting 5.
Let p_6 = probability of getting 6.

H_0: $p_1 = \frac{1}{6}$, $p_2 = \frac{1}{6}$, $p_3 = \frac{1}{6}$, $p_4 = \frac{1}{6}$, $p_5 = \frac{1}{6}$, $p_6 = \frac{1}{6}$.

H_1: $p_1 \neq \frac{1}{6}$ or $p_2 \neq \frac{1}{6}$ or $p_3 \neq \frac{1}{6}$ or $p_4 \neq \frac{1}{6}$ or $p_5 \neq \frac{1}{6}$ or $p_6 \neq \frac{1}{6}$.

Let's assume H_o is true.

From before, we expect 20 1's, 20 2's, 20 3's, 20 4's, 20 5's, and 20 6's.

From the last exercise, $\sum_{all\ values}$ $(O\text{-}E)^2/E = 6.80$.

Using what we learned before this exercise, $\sum_{all\ values}$ $(O\text{-}E)^2/E \sim \chi^2_5$.

The p-value = $\mathrm{P}(\chi^2_5 > 6.80) = \chi^2\mathrm{cdf}(6.80,1000,5) = 0.236$.

Since $0.236 > 0.01$, we cannot reject H_o. There is not enough evidence that the die isn't fair.

Remark: For **all** problems in this section, p-values are of the form

$$p\text{-value} = \mathrm{P}(\chi^2_k > a),$$

where k is a positive integer and a is a positive real number.

Exercise 5. The manufacturer of M&M's candies claims that the number of different-colored candies in bags of dark chocolate M&M's is uniformly distributed. To test this claim, you randomly select a bag that contains 504 dark chocolate M&M's. The results are shown in the table below:

Color	Frequency, f
Brown	81
Yellow	96
Red	89
Blue	84
Orange	76
Green	78

Using $\alpha = 0.10$, perform a chi-square goodness-of-fit test to test the claimed or expected distribution. What can you conclude?

Let p_{br} = proportion of M & M's that are brown.
Let p_y = proportion of M & M's that are yellow.
Let p_r = proportion of M & M's that are red.
Let p_{bl} = proportion of M & M's that are blue.
Let p_o = proportion of M & M's that are orange.
Let p_g = proportion of M & M's that are green.

H_0:

H_1:

Let's assume H_o is true.

The expected number of brown M&M's (in a bag of 504 M&M's) =

The expected number of yellow M&M's =

The expected number of red M&M's =

The expected number of blue M&M's =

The expected number of orange M&M's =

The expected number of green M&M's =

Color	Observed Frequency	Expected Frequency	$(O\text{-}E)^2/E$
Brown			
Yellow			
Red			
Blue			
Orange			
Green			

$\sum_{all\ values} (O\text{-}E)^2/E$ = sum of values in last column =

We know that $\sum_{all\ values} (O\text{-}E)^2/E \sim \chi^2_5$.

Therefore, the p-value =

Since 0.10, we reject H_o. There is evidence from the bag that the M&M's are not evenly distributed.

Exercise 6. Results from a survey two years ago asking college students what their top motivations are for using a credit card are shown below:

Motivation	Percent of People
Reward Program	28%
Low Interest Rate	24%
Cash back	22%
Special Store Discounts	8%
Improve Credit Rating	18%

To determine whether this distribution has changed, you randomly selected 425 college students and ask each what the top motivation is for using a credit card. The results are shown in the table below:

Response	Frequency, f
Reward Program	112
Low Interest Rate	98
Cash back	107
Special Store Discounts	46
Improve Credit Rating	62

Can you conclude that there has been a change in the claimed or expected distribution? Use $\alpha = 0.05$.

Let p_{rp} = proportion of people whose motivation is the reward program.
Let p_{lir} = proportion of people whose motivation is the low interest rate.
Let p_{cb} = proportion of people whose motivation is cash back.
Let p_{ssd} = proportion of people whose motivation is special store discounts.
Let p_{icr} = proportion of people whose motivation is to improve credit rating.

H_0:

H_1:

Let's assume H_o is true.

The expected number of reward program people (in a sample of 425) =

The expected number of low interest rate people =

The expected number of cash back people =

The expected number of special store discounts people =

The expected number of improve credit rating people =

Response	Observed Frequency	Expected Frequency	$(O\text{-}E)^2/E$
Reward Program			
Low Interest Rate			
Cash back			
Special Store Discounts			
Improve Credit Rating			

$\sum_{all\ values} (O\text{-}E)^2/E$ = sum of values in last column =

We know that $\sum_{all\ values} (O\text{-}E)^2/E \sim \chi^2_4$.

Therefore, the p-value =

Since 0.05, we reject H_o. There is that there has been a change in the distribution of reasons.

Class Exercise 3. A personnel director believes that the distribution of the reasons workers leave their jobs is different from the one shown in the table below:

Response	Percent
Limited advancement potential	41
Lack of recognition	25
Low salary/benefits	15
Unhappy with management	10
Bored/don't know	9

The director randomly selects 200 workers who recently left their jobs and asks each his or her reason for doing so. The results are in the table below:

Response	Frequency, f
Limited advancement potential	78
Lack of recognition	52
Low salary/benefits	30
Unhappy with management	25
Bored/don't know	15

At $\alpha = 0.01$, are the distributions different? **Answer: p-value = 0.731**

Class Exercise 4. A bicycle safety claims that fatal bicycle accidents are uniformly distributed throughout the week. The table below shows the day of the week for which 782 randomly selected fatal bicycle accidents occurred.

Day	Frequency, f
Sunday	108
Monday	112
Tuesday	105
Wednesday	111
Thursday	123
Friday	105
Saturday	118

At $\alpha = 0.10$, can you reject the claim that the distribution is uniform? **Answer: p-value = 0.876**

Class Exercise 5. A golf instructor wants to test the following claim: of all golf students in the United States, 65% need the most help with short game shots, 22% need the most help with approach and swing, 9% need the most help with driver shots, and 4% need the most help with putting. A random sample of golf students finds that 276 need the most help with short-game shots, 99 need the most help with approach and swing, 42 need the most help with driver shots, and 18 need the most help with putting. Test the claim at $\alpha = 0.10$. **Answer: p-value = 0.918**

Homework

C Problems

5-17 ODD

B Problems

1, 3

A Problems

None