ICCPP-STATISTICS
- Chi-Square Test

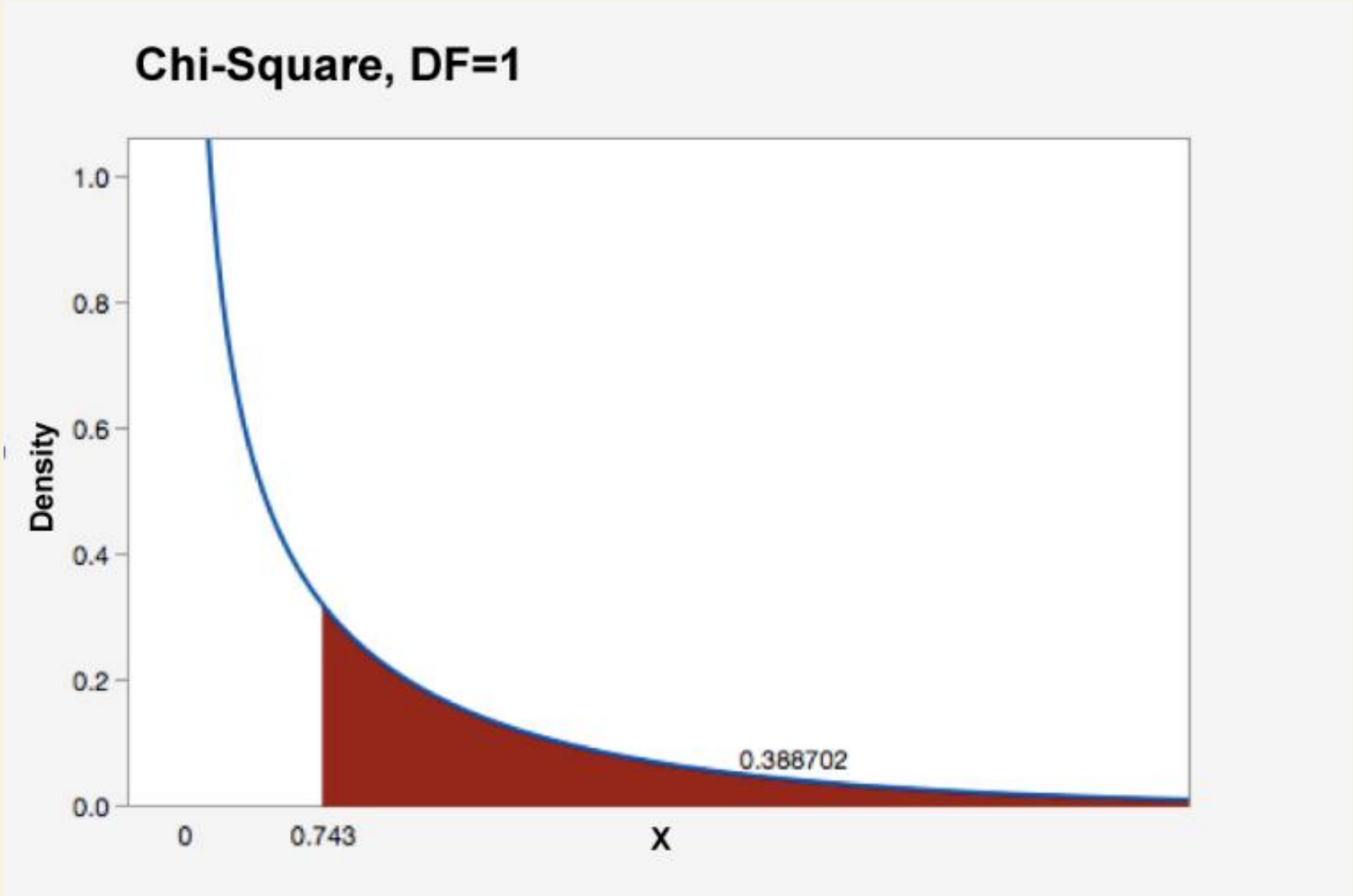
Vishal Lohchab

Scientific Assistant of
Prof. Dr. Hans-Werner Gessmann
Director ICCPP International

Karl Pearson (1900)

Chi-Square-Test

Definition



Definition

A Chi-square test is a hypothesis testing method.

Two common Chi-square tests involve checking if observed frequencies in one or more categories match expected frequencies.

Definition

You use a Chi-square test for hypothesis tests about whether your data is as expected.

The basic idea behind the test is to compare the observed values in your data to the expected values that you would see if the null hypothesis is true.

The formula for Chi-Square

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

χ^2 = chi squared

O_i = observed value

E_i = expected value

Types of Chi-square tests

- There are two commonly used Chi-square tests:

  Chi-square goodness of fit test and the Chi-square test of independence.

  Both tests involve variables that divide your data into categories. Difference between both tests are as follows:

+

Chi-square goodness of fit test

Number of variables = One

Purpose of test =

Decide if one variable is likely to come from a given distribution or not.

Chi-square goodness of fit test

Example = Decide if bags of candy have the same number of pieces of each flavor or not

Hypotheses in example =

H_o: proportion of flavors of candy are the same

H_a: proportions of flavors are not the same

+

Chi-square goodness of fit test

Number of variables = Two

Purpose of test =

Decide if two variables might be related or not

How to perform a Chi-square test

For both the Chi-square goodness of fit test and the Chi-square test of independence, perform these same analysis steps:

- Define your null and alternative hypotheses before collecting your data.
- Decide on the alpha value. This involves deciding the risk you are willing to take of drawing the wrong conclusion.

How to perform a Chi-square test

For both the Chi-square goodness of fit test and the Chi-square test of independence, perform these same analysis steps:

- Check the data for errors.
- Check the assumptions for the test.
- Perform the test and draw your conclusion.

Solution

The chi-square formula is a difficult formula to deal with.
That's mostly because you're expected to add a large amount of numbers.
The easiest way to solve the formula is by making a table.

Solution

Step 1

Make a table with columns for "Categories," "Observed," "Expected," "Residual (Obs-Exp)", "(Obs-Exp)^2" and "Component (Obs-Exp)^2 / Exp."

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp

+

Solution

Step 2

Fill in your categories. Categories should be given to you in the question. There are 12 zodiac signs, so:

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries					
Taurus					
Gemini					
Cancer					
Leo					
Virgo					
Libra					
Scorpio					
Sagittarius					
Capricorn					
Aquarius					
Pisces					

+

Solution

Step 3

Write your counts. Counts are the number of each items in each category in column 2. You're given the counts in the question:

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries	29				
Taurus	24				
Gemini	22				
Cancer	19				
Leo	21				
Virgo	18				
Libra	19				
Scorpio	20				
Sagittarius	23				
Capricorn	18				
Aquarius	20				
Pisces	23				

+

Solution

Step 4

Calculate your expected value for column 3. In this question, we would expect the 12 zodiac signs to be evenly distributed for all 256 people, so 256/12 = 21.333. Write this in column 3.

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries	29	21.333			
Taurus	24	21.333			
Gemini	22	21.333			
Cancer	19	21.333			
Leo	21	21.333			
Virgo	18	21.333			
Libra	19	21.333			
Scorpio	20	21.333			
Sagittarius	23	21.333			
Capricorn	18	21.333			
Aquarius	20	21.333			
Pisces	23	21.333			

Solution

Step 5

Subtract the expected value (Step 4) from the Observed value (Step 3) and place the result in the "Residual" column. For example, the first row is Aries: 29-21.333 = 7.667.

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries	29	21.333	7.667		
Taurus	24	21.333	2.667		
Gemini	22	21.333	0.667		
Cancer	19	21.333	-2.333		
Leo	21	21.333	-0.333		
Virgo	18	21.333	-3.333		
Libra	19	21.333	-2.333		
Scorpio	20	21.333	-1.333		
Sagittarius	23	21.333	1.667		
Capricorn	18	21.333	-3.333		
Aquarius	20	21.333	-1.333		
Pisces	23	21.333	1.667		

Solution

Step 6

Square your results from Step 5 and place the amounts in the $(\text{Obs-Exp})^2$ column

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries	29	21.333	7.667	58.782889	
Taurus	24	21.333	2.667	7.112889	
Gemini	22	21.333	0.667	0.44889	
Cancer	19	21.333	-2.333	5.442889	
Leo	21	21.333	-0.333	0.110889	
Virgo	18	21.333	-3.333	11.108889	
Libra	19	21.333	-2.333	5.442889	
Scorpio	20	21.333	-1.333	1.776889	
Sagittarius	23	21.333	1.667	2.778889	
Capricorn	18	21.333	-3.333	11.108889	
Aquarius	20	21.333	-1.333	1.776889	
Pisces	23	21.333	1.667	2.778889	

Solution

Step 7

Divide the amounts in Step 6 by the expected value (Step 4) and place those results in the final column.

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries	29	21.333	7.667	58.782889	2.755490976
Taurus	24	21.333	2.667	7.112889	0.333421882
Gemini	22	21.333	0.667	0.44889	0.021042048
Cancer	19	21.333	-2.333	5.442889	0.255139408
Leo	21	21.333	-0.333	0.110889	0.005198003
Virgo	18	21.333	-3.333	11.108889	0.520737308
Libra	19	21.333	-2.333	5.442889	0.255139408
Scorpio	20	21.333	-1.333	1.776889	0.083292973
Sagittarius	23	21.333	1.667	2.778889	0.130262457
Capricorn	18	21.333	-3.333	11.108889	0.520737308
Aquarius	20	21.333	-1.333	1.776889	0.083292973
Pisces	23	21.333	1.667	2.778889	0.130262457

+

Solution

Step 8

Add up (sum) all the values in the last column.

Category	Observed	Expected	Residual= (Obs-Exp)	(Obs-Exp)^2	Component = (Obs-Exp)^2 / Exp
Aries	29	21.333	7.667	58.782889	2.755490976
Taurus	24	21.333	2.667	7.112889	0.333421882
Gemini	22	21.333	0.667	0.44889	0.021042048
Cancer	19	21.333	-2.333	5.442889	0.255139408
Leo	21	21.333	-0.333	0.110889	0.005198003
Virgo	18	21.333	-3.333	11.108889	0.520737308
Libra	19	21.333	-2.333	5.442889	0.255139408
Scorpio	20	21.333	-1.333	1.776889	0.083292973
Sagittarius	23	21.333	1.667	2.778889	0.130262457
Capricorn	18	21.333	-3.333	11.108889	0.520737308
Aquarius	20	21.333	-1.333	1.776889	0.083292973
Pisces	23	21.333	1.667	2.778889	0.130262457
					5.094017203

This is the chi-square: 5.094

Test a Chi Square Hypothesis

(df) Degrees of freedom equals the number of categories minus 1.

(p-value) A p-value is a measure of the probability that an observed difference could have occurred just by random chance.

+

Test a Chi Square Hypothesis

Step 1

Take the chi-square statistic. Find the p-value in the chi-square table. The closest value for df=11 and 5.094 is between .900 and .950.

Note

The chi square table doesn't offer exact values for every single possibility. If you use a calculator, you can get an exact value. The exact p value is 0.9265.

Test a Chi Square Hypothesis

Step 2

Use the p-value and decide whether to support or reject the null hypothesis.

In general, small p-values (1% to 5%) would cause you to reject the null hypothesis.

This very large p-value (92.65%) means that the null hypothesis should not be rejected.

+
References

1. Kenney J F; Keeping E S: Mathematics of Statistics, Pt. 2, 2nd ed. Princeton, NJ: Van Nostrand, 1951.
2. Stephanie G: "Welcome to Statistics How To!" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/