proprep™

Workbook

Table of Contents

Chi Square Tests

Chi Square Test for Goodness of Fit

Theory

Suppose we have one categorical variable from a single population.
For example, we could take the variable of the result of a coin toss.
This variable has a hypothesized distribution – with a fair coin,
we expect equal numbers of heads and tails.

Suppose we want to test if the coin is fair. We take a random sample from the population, meaning we flip the coin some number of times. We want to test if the distribution of heads and tails in the sample matches the hypothesized distribution.

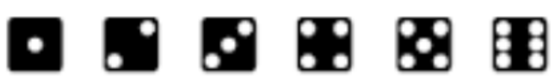
This is called a Test for Goodness of Fit, and we use a Chi-Square statistic, also known as Pearson Chi-Squared statistic.

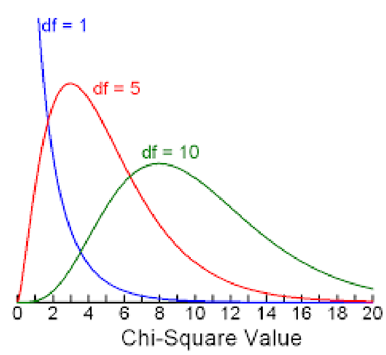
Example (solution in the recording)

Let’s use the variable of the outcome on one roll of a die.

To test whether a die is fair, we toss it 120 times, and here are the results:

17 23 20 25 18 17





Is the die fair (at 5% level of significance)?

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

df = (number of categories) – 1

Questions

1) A factory produces toys in four colors: blue, red, green, and orange. We want to check whether the proportion of blue toys is twice as large as the proportion of any other color. We sample 200 toys randomly and find: 70 blue, 50 red, 40 green, and 40 orange. What can we conclude (at a 5% level of significance)?

2) The Dept. of Ed. claims that among its employees, the ratio of those with only a high school degree, those with at least some college education, and those with a graduate degree is 1:2:1.
A random sample of 200 employees included: 56 with only a high school degree, 105 college educated, and 39 with a graduate degree.
Test the DOE's claim at a 5% level of significance.

Answer Key

1) We do not reject H_0.
We conclude that there are twice as many blues as any other color.

2) We don't reject H_0.
We conclude that the ratio is 1:2:1.

Chi Square Test of Independence

Example (solution in the recording)

Work at a factory is done in three shifts.
The following table displays the number of defective and working products made during each shift from a random sample:

	Morning Shift	Afternoon Shift	Night Shift	Total
Defective	50	60	70	180
Working	600	700	800	2100
Total	650	760	870	2280

Is there a connection between the quality of the product and the shift on which it was produced? Draw a conclusion at a $\alpha = 0.05$ level of significance.

Questions

1) Among 120 random women voters, all 120 said that they would vote for Candidate Q. Among 200 random men voters, 80 of them said they would vote for Candidate Q. Is support for Candidate Q independent of voter's gender? Test at a 5% level of significance.

2) At two specific clothing stores, the probability distribution of colors of the clothes sold on a random day was:

	Black	White	Red	Blue
Store A	15	20	15	50
Store B	60	20	10	20

a. At a 5% level of significance, test whether the probability distribution of colors at store A is in a ratio of 3:1:1:1 in favor of blue.
b. At a 2.5% level of significance, test whether there is a difference between the two stores in the probability distribution of the colors of the items sold.

Answer Key

1) We have statistical evidence to reject H_0.
There is a relationship between a voter's gender and whether or not the voter supports Candidate Q (at 0.1% level of significance).

2) a. We do not reject H_0. b. We reject H_0.