Department of biology

Mathmatics

1 stage

chapter_4

By

JAAFAR ALMAMOORI

Department of biology

fourth chapter

Inferential statistics

First: testing hypotheses

Testing statistical hypotheses has become one of the most important characteristics that distinguishes field and experimental research in the field of education, psychology, and human sciences in general. The primary goal of testing hypotheses is to infer the characteristics of society or some of them from observing the sample from which it was taken, with the aim of generalizing the results we reach in our study of the sample. Small of the population represented by that sample.

Hypotheses are usually of two types. The first is the null hypothesis (H0), which is tested statistically. The second is the alternative hypotheses (H1), which is usually the opposite of the null hypothesis .

Second: Steps for testing hypotheses

1. Formulate the null and alternative hypotheses.

2. Determine the appropriate significance level (α). There is almost agreement that the significance levels (0.0.), (,.,.) and (...,.) are the criteria for rejecting the null hypothesis and accepting the alternative hypothesis.

.3Determine the appropriate type of test. This depends on the nature of the statistical data and whether testing the hypothesis requires fulfilling certain assumptions about the population from which the sample is drawn or not**.**

Chi-square

The first person to describe the chi-square was Carl Pearson around the year 1900 AD. The chi-square, symbolized by (X2), is also a random variable and is used to test hypotheses.

Chi-square tests are used to test hypotheses and significance of nominal data. They are types, including:

1 - Significance test for one variable, Chi-square, for goodness of fit
2 - Testing the significance of more than one variable, Chi-square for independence

We use chi-square tests to test the importance of the significant difference between observed data and expected data, or between the factors and forces that the researcher wants to study and analyze. Chi-square tests depend on the null hypothesis, which always claims that there is no significant difference between the real or observed data and the expected data, or between the factors that the researcher wants to examine. When conducting the test, we obtain a result called the calculated value, and this value must be compared with tabular values on degrees of freedom (such as It is -2-3-1) and at confidence levels (00% - 00%). If the calculated value is smaller or equal to the tabulated value at a certain degree of freedom and a significance level of 0.05 or 0.01, that is, the confidence level is 00% and 00%, respectively, then there is no significant difference, and therefore we accept the null hypothesis and reject the alternative hypothesis of the research, while if the calculated value is greater From the table value, there is a statistically significant difference, and therefore we reject the null hypothesis and accept the alternative hypothesis.

First: Testing the significance of one variable (Chi-square - for goodness of fit)

The chi-square goodness of fit test is used to test whether the observed results differ from the expected results or not.

Steps for the Kai goodness-of-match test:

1 - Formulation of the null hypothesis and alternative hypothesis:

- There is no difference between the observed results and the expected results: 0 H
- There is a difference between the observed results and the expected results: 1 H

2 - Extracting the calculated chi-square value after creating a table that helps us calculate it as follows:

Levels of change or category	Viewer repetition O	Expected frequency E	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$

$$=X^2 = \Sigma\frac{(O-E)^2}{E}$$

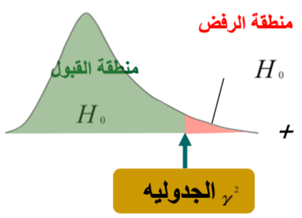
1. Chi-square tabular value:

We determine the level of significance α and the degree of freedom from (number of categories - 1)

We extract the tabular chi-square value

2. Make decision

We make the decision based on the Chi-square value (we determine the rejection region and the acceptance region on the following diagram:(



If the value of the test statistic falls in the rejection region, we reject the null hypothesis H0 and accept the alternative hypothesis H1, but if the value of the test statistic falls in the acceptance region, we accept the null hypothesis H0.

Example 1: In previous studies on Corona patients, they were asked about their educational level, and the results were as follows:

%5 at the university level - 15% at the secondary level - 30% at the intermediate level - 50% at the primary level

But currently the results for 60 patients were as follows:

6 patients in the university stage - 20 patients in the secondary stage - 10 patients in the middle stage - 24 patients in the primary stage

Can we decide that the results of this year's current study differ from the results of previous studies? Use α=0.05

the solution:

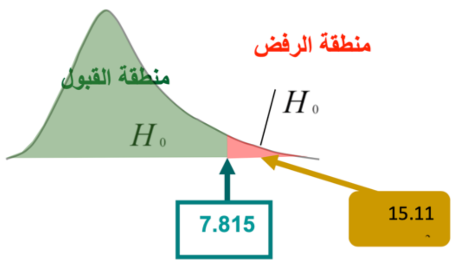
Hypotheses are formulated

1 - There is no difference between the results of the previous study and the current study, H0
2 - There is a difference between the results of the previous study and Solution H1

Categories	Duplicates	ratio	Expected frequencies	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
University	6	5%	0.05*60=3	3	9	**3**
High School	20	15%	0.15*60=11	9	81	**7.36**
Medium	10	30%	0.30*60=18	-8	64	**3.55**
Primary	24	50%	0.05*60=30	-6	36	**1.2**
total	60					**15.11**

We determine the tabular values of x by the following law

)Number of rows -1) (Number of columns -1) = 3 degrees of freedom, meaning the tabular X values are 7.81



Make decision:

Since the value of Kai fell in the rejection region, we reject the null hypothesis and accept the alternative hypothesis, meaning that there is a difference between the previous and current results.

Example 2: Two cities in Babylon. It was found that the number of Coronavirus infections currently in the city of Al-M

ahawil is 36, and the number of infections in the city of Al-Hilla is 40, while previously the infections were equal. What is required is to test the importance of the moral difference between the two cities in terms of previous and current results from the Coronavirus?

Hypotheses are formulated

3 - There is no difference between the results of the previous study and the current study, H0

4 - There is a difference between the results of the previous study and Solution H1

variable	Current Results	Ratio	Expected results	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Mahaweel	36	50%	38	-2	4	0.105
Hilla city	40	50%	38	2	4	0.105
total	76					0.21

Degree of freedom (2-1) (2-1) =1

From the chi-square table, the tabular value of X = 3.84

Calculated X value = 0.21

Make decision:

There is no significant difference because the calculated value is smaller than the tabulated value and has a degree of freedom (1) and a confidence level of 95%. Therefore, we accept the null hypothesis and reject the alternative hypothesis.

Second: Moral test for more than one variable (chi-square - independence)

In many cases, we need to recognize whether there is a relationship between two characteristics of a society. For example, we may need to know whether there is a relationship between income level and educational level? Or is there a relationship between eye color and hair color in a society? Or is there a relationship between achievement level and household income?

The chi-square test of independence is used to answer such questions (is there a relationship between two nominal variables) and is based on comparing the observed values with the expected values. Therefore, we must select a random sample from the population under study and then classify the observations of this sample according to the levels of each of the two characteristics and put them in a table called the compatibility table.

Steps to test the chi-square for independence:

1- Formulation of the null hypothesis and the alternative hypothesis: there is no relationship between the two qualities or there is no correlation between the two qualities: H0

There is a relationship between the two qualities or there is a correlation between the two qualities: 1 H

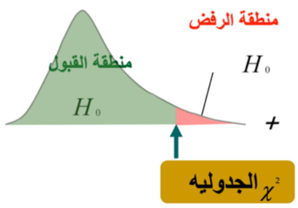
2- The calculated chi-square value can be calculated from the following application:

	B1	B2
A1	a	b
A2	c	d

$$X^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

3-Chi-square tabular value: Extracted from the chi-square statistical table based on a level based on the level of morality and degree of freedom.

4- **Decision:**



Example:1

In research to study the relationship between tea drinking and gender or sex (males, females), a sample of 88 residents in the city of Hilla was selected and through the questionnaire they were classified in the table as follows. Do these data indicate a relationship between tea drinking and gender?

Use a morale level of 0.05=α

	male	female	total
Drink tea	40	33	73
Don't Drink tea	3	12	15
Total	43	45	88

sol:

H 0: There is no relationship between tea drinking and gender.

H 1: There is a relationship between tea drinking and gender.

$$X^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$X^2 = \frac{88(480 - 99)^2}{(73)(15)(43)(45)} = 6.029$$

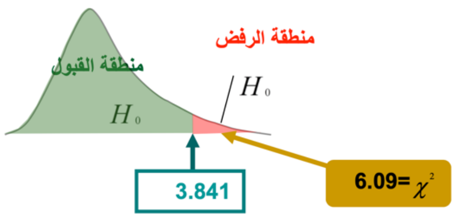
We get the critical or tabular value from the chi-square distribution table,

DF=(R-1) (C-1)

DF= (2-1) (2-1) =1

DF= 1

We get the critical or tabular value from the chi-square distribution table, and we find it: **3.841**



The calculated chi-square value is greater than the tabular value, i.e. it is in the rejection zone and therefore we reject H0 and accept H1, which is that there is a relationship between tea drinking and sex.

First exercise:

Social research was conducted to study the relationship between sex (male, female) and the desire to marry relatives, a sample of 57 individuals was taken, and the results were as follows:

	Male	Female	Total
agree	10	15	25
Not agree	20	12	32
Total	30	27	57

Is there a link or relationship between the sexes and the tendency to marry relatives or are the two characteristics independent of each other, i.e. no relationship between the sexes and the tendency to marry relatives with a moral level of 0.05?

Second exercise:

An experiment was conducted on two groups, the first group includes 95 people who played romantic music and the second group also includes 95 people who played neutral or regular music, and the aim was to urge them to donate, so does the effect of the type of music have an impact on donation in the two groups at the level of a sample of 0.05, noting that the results were as follows:

	Yes	No	Total
Romantic	27	23	50
Neutral	15	35	50
Total	42	58	100

df	0.05	0.025	0.01	0.005
1	3.84	5.02	6.63	7.88
2	5.99	7.38	9.21	10.60
3	7.82	9.35	11.35	12.84
4	9.49	11.14	13.28	14.86
5	11.07	12.83	15.09	16.75
6	12.59	14.45	16.81	18.55
7	14.07	16.01	18.48	20.28
8	15.51	17.54	20.09	21.96
9	16.92	19.02	21.66	23.59
10	18.31	20.48	23.21	25.19
11	19.68	21.92	24.72	26.75
12	21.03	23.34	26.21	28.30
13	22.36	24.74	27.69	29.82
14	23.69	26.12	29.14	31.31
15	25.00	27.49	30.58	32.80
16	26.30	28.85	32.00	34.27
17	27.59	30.19	33.41	35.72
18	28.87	31.53	34.81	37.15
19	30.14	32.85	36.19	38.58
20	31.41	34.17	37.56	40.00
21	32.67	35.48	38.93	41.40
22	33.93	36.78	40.29	42.80
23	35.17	38.08	41.64	44.18
24	36.42	39.37	42.98	45.56
25	37.65	40.65	44.32	46.93
26	38.89	41.92	45.64	48.29
27	40.11	43.20	46.96	49.64
28	41.34	44.46	48.28	50.99
29	42.56	45.72	49.59	52.34
30	43.77	46.98	50.89	53.67
40	55.75	59.34	63.71	66.80
50	67.50	71.42	76.17	79.52
100	124.34	129.56	135.82	140.19