CHI SQUARE TEST

The chi square Test for Goodness of fit test claims about population proportions.

It is a non parametric test that is performed on categorical [ordinal and nominal] data.

There is a population of Male who likes different color bikes

	Theory	Sample
Yellow Bike	1/3	22
Red Bike	1/3	17
Orange Bike	1/3	59

↓

Theory categorical distribution

→ Observed categorical distribution

(*) Goodness of fit test

Goodness Of fit test

In a science class of 75 students, 11 are left handed. Does this class fit the theory that 12% of people are left handed.

$$\frac{3}{\cancel{100}\ \cancel{25}_{1}} \times \overset{3}{\cancel{75}} = 9$$

Ans)

	O	E
left handed	11	9
Right handed	64	66
	75	75

CHI SQUARE For Goodness of Fit

In 2010 Census of the city, the weight of the individuals in a small city were found to be the following

<50kg	50 - 75	>75
20%	30%	50%

In 2020, weight of n=500 individuals were sampled. Below are the results

<50	50-75	>75
140	160	200

In 2020, weight of n=500 individuals were sampled. Below are the results

<50	50-75	>75
140	160	200

Using $\alpha = 0.05$, would you conclude the population diffcuences of weights has changed in the last 10 years?

Ans)

Ans)

2010

Expected

<50kg	50 - 75	>75
20%	30%	50%

2020

n=500

Observed

<50	50-75	>75
140	160	200

Expand

<50	50-75	>75
0.2 x 500 = 100	0.3 x 500 = 150	0.5 x 500 = 250

① Null Hypothesis : H_0 : The data meets the expectation

Alternate Hyp : H_1 : The data does not meet the expectation

② $\alpha = 0.05$ CI: 95%

③ Degree of freedom

$df = K - 1 = 3 - 1 = 2$

③ Degree of freedom

$$df = K - 1 = 3 - 1 = 2$$

④ Decision Boundary

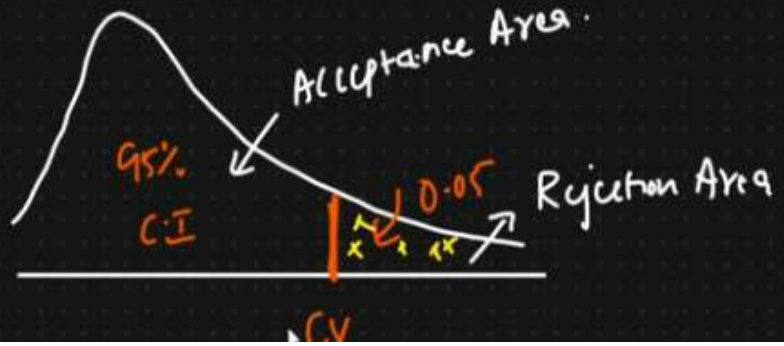
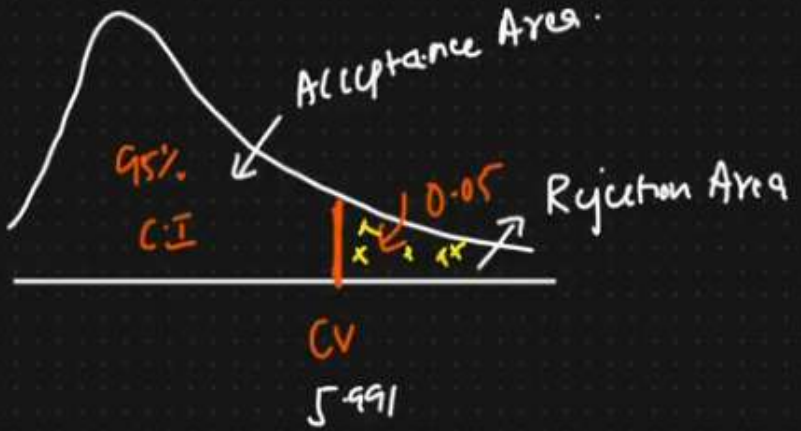


Table: Chi-Square Probabilities

The areas given across the top are the areas to the right of the critical value. To look up an area on the left, subtract it from one, the right)

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997

(4) Decision Boundary



If χ^2 is greater than 5.99, Reject Ho

etc

We fail to reject the Null Hypothesis

(5) Calculate Chi Square Test Statistic

2020

n = 500

Observed

<50	50-75	>75
140	160	200

Expected

<50	50-75	>75
0.2 x 500 = 100	0.3 x 500 = 150	0.5 x 500 = 250

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

$$= \frac{(140-100)^2}{100} + \frac{(160-150)^2}{150} + \frac{(200-250)^2}{250}$$

$$= \frac{1600}{100} + \frac{100}{150} + \frac{2500}{250}$$

$= 16 + 0.66 + 10$

$= 26.66$

$\chi^2 = 26.66$

If χ^2 is greater than 5.99, Reject Ho,

else

We fail to reject the Null Hypothesis

If χ^2 is greater than 5.99, Reject H0

else

We fail to reject the Null Hypothesis

$26.66 > 5.99$, Reject H0

Answer

The weights of 2020 population are different than those expected in the 2010 population