Module 1. Introduction to Statistics (April 2023)

Question 1. There is an assumption that there is no significant difference between boys and girls with respect to intelligence. Tests are conducted on two groups and the following are the observations

Mean Standard Deviation Size

Girls 89 4 50

Boys 82 9 120

Validate the claim with 5% LoS (Level of Significance).

Answer 1.

H0 (Null Hypothesis): No difference between boys and girls in terms of intelligence. ($\mu 2 = \mu 2$)

H1 (Alternate Hypothesis): Boys and girls are different in terms of intelligence ($\mu 2 \neq \mu 2$) => two tailed tests

x1bar = 75 (boys sample mean)

x2bar = 73 (girls sample mean)

LoS (α) = 5%

In question, we have two sample mean. Boys sample mean and girls sample mean. Hence this can be solved with two mean problem.

Next both samples size n1 = 60 and n2 = 100 are greater than 30 hence will use z-test.

Step 1:

calculate z value from the two mean z test formula as below:

$z = [(x1bar - x2bar) - (\mu 2 - \mu 2)]/\sqrt{(s12/n1 + s22/n2)}$

$\mu 2 - \mu 2 = 0$ assuming null hypothesis is true

$z = (75-73)/\sqrt{(82/60 + 102/100)}$

$z = 1.39$

step 2:

calculate z critical value for $\alpha = 5\%$ from z-table.

So, from z-table Z critical value = -1.96, +1.96 (will get two values due two tailed test)

step 3:

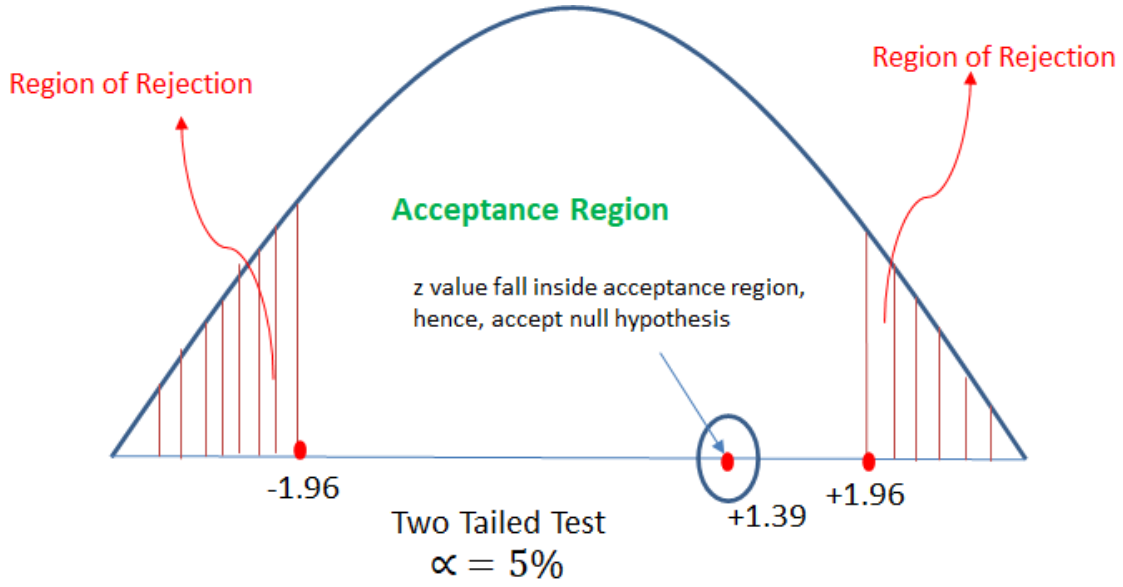
check if calculated z value is in between z critical value then accept the null hypothesis if z calculated is outside z critical then reject the null hypothesis.

Here, z calculated value is in between the z critical values. -1.96 < 1.39 < 1.96

Hence will accept the null hypothesis.

Conclusion:

with given data it is significantly proven that there is no significant difference between the intelligence of boys and girls.



Question 2. Analyse the below data and tell whether you can conclude that smoking causes cancer or not?

Category Diagnosed as Cancer Without Cancer Total

Smokers 220 230 550

Non-Smokers 350 640 990

Total 680 910 1590

Answer 2.

Step 1:

H0 (Null Hypothesis): Cancer is dependent on smoking

H1 (Alternate Hypothesis): cancer is not dependent on smoking

Step 2:

Calculate the expected value for each cell of the table (when null hypothesis is true)

The expected values specify what the values of each cell of the table would

be if there is no association between the two variables.

The formula for computing the expected values requires the sample size, the

row totals, and the column totals.

expected value (e) = (row total * column total)/table total

Now let's create another table with observed and expected values both:

Category	Diagnosed as Cancer	Without Cancer	Total
Smokers	o = 400, e = 700*700/1500 = 326	o = 300, e = 700*800/1500 = 373	700
Non-Smokers	o = 300, e = 800*700/1500 = 373	o = 500, e = 800*800/1500 = 426	800
Total	700	800	1500

Step 3:

calculate the chi square value:

$\chi 2 = \Sigma[(o-e)2]/e$

$\chi 2 = (400-326)2/326 + (300-373)2/373 + (300-373)2/373 + (500-426)2/426$

$\chi 2 = 16.79 + 14.28 + 14.28 + 12.85$

$\chi 2 = 58.2$

Step 4:

Decide if $\chi 2$ is statistically significant.

The final step of the chi-square test of significance is to determine if the value

of the chi-square test statistic is large enough to reject the null hypothesis.

Now will check χ2 table for the critical value with α = 5%

So, from table we got χ2 (critical value at α = 5%) = 3.841

The chi-square value of 58.2 is much larger than the critical value of 3.84, so the null hypothesis can be rejected.

Which means with given data, it can be significantly concluded that cancer is not dependent on smoking.