Introduction

The Kolmogorov Smirnov Goodness of fit test compares your data with a known distribution and lets you know if they have the same distribution.

Commonly used as a test for normality to see if your data is normally distributed

Assumptions

- The sample is a random sample.
- The scale of measurement is at least ordinal.
- The test is only exact for continuous variables.

Procedure

- Hypothesis:

H_0: The data follow a specified distribution.

H_1: The data do not follow the specified distribution.

- Level of Significance:

$$\alpha=0.05$$

$$\alpha= 0.01$$

$$\alpha=0.1$$

- Test Statistic:
- Arrange the data in ascending order then compute

D= max (D^+ , D^-)

$$D^+ = \max\left[\frac{i}{N} - R_i\right]$$

$$D^- = \max\left[R_i - \left(\frac{i-1}{N}\right)\right]$$

N= no .sample observations

Calculation:

- Arrange in ascending order
- ComputeD^+ and D^- and get the maximum value
- Compute D that is D= max (D^+ , D^-)

Critical Region:

The critical value of $D_\alpha$is found from the K-S table values. where α is level of significance.

- Decision Rule:
- Where D calculated and D_α is tabulated

$$D < D_\alpha$$

Then we do not reject null hypothesis (H_0)

$$D > D_\alpha$$

Then we reject null hypothesis (H_0)

Example

Using the KS test check for the property of uniformity for the input set of random numbers 0.54,0.73,0.38,0.11 and 0.98

- Hypothesis:

  H_0: The distribution is uniform

  H_1:The distribution is not uniform

- Level of significance:

  α=0.05

Test Statistic:

- Arrange the data in ascending order then compute

  D= max (D^{+}, D^{-})

$$D^{+} = \max\left[\frac{i}{N} - R_i\right]$$

$$D^{-} = \max\left[R_i - \left(\frac{i-1}{N}\right)\right]$$

N= no .sample observations so N=5

Calculation:

i	$\frac{i}{N}$	$\boldsymbol{R_i}$	D^+	D^-
1	1/5=0.2	0.11	0.09	0.11
2	2/5=0.4	0.54	-0.14	0.34
3	3/5=0.6	0.68	-0.08	0.28
4	0.8	0.73	0.07	0.13
5	1	0.98	0.02	0.18

We need maximum of (D^+ , D^- ignore negative values) so see the column of D^+ and D^- and get the largest value that is

$$D^+ = \max\left[\frac{i}{N} - R_i\right] = 0.09$$

$$D^- = \max\left[R_i - \left(\frac{i-1}{N}\right)\right] = 0.34$$

D= max (D^+ , D^-) now the

D= max (D^+, D^-)

D=max(0.09,0.34)

so

D=0.34

- Critical Region:

D_{α}=0.565

Conclusion :

if D>D_{α} then reject the null hypothesis

- Conclusion:
- D_α=0.565
- D= 0.34

Observed value of D = 0.34 does not exceed the tabulated 0.565 so we do not reject null hypothesis (H_0).

Question no.2

Apply KS test for the data as follows 0.15,0.94,0.05,0.51 and 0.29 where

and check whether the distribution of data is uniform or not

Where D_α=0.565

Results

D= 0.31

D_{α}=0.565

Conclusion : do not reject the null hypothesis

Question no. 3

Apply KS test for the group of data as follows

1,3,5,7,9,10,12,14

Check whether the distribution of data is normal or not where

D_{α}=0.454