Hypothesis Test

Classification of hypothesis test:

Parametric and non parametric test

Parametric Test:

If the information about the population is completely known by means of it parameters then statistical test is called parametric test

procedure for testing hypothesis about parameters in a population described by a specified distributional form, are called parametric test

eg: t-test, f-test, z-test, ANOVA

A t-test is used for testing the mean of one population against a standard or comparing the means of two populations if you do not know the populations’ standard deviation and when you have a limited sample ($n < 30$).

Assumption for t test

Sample must be random

observations indipendent

standard deviation is not known

Normal distribution of population

A z-test is used for testing the mean of a population versus a standard, or comparing the means of two populations, with large ($n \geq 30$) samples whether you know the population standard deviation or not.

It is also used for testing the proportion of some characteristic versus a standard proportion, or comparing the proportions of two populations.

F-test: The F distribution is often used to test the difference of variances of two normally distributed samples.

Non Parametric Test:

```
Chi-Square test

Wilcoxon Signed Rank Test

Mann Whitney U Test
```

Testing of Goodness of Fit:

A goodness of fit is a inferential procedure used to determine weather a frequency distribution follows a claimed distribution

Chi-Square test

Kolmogorov-Smirnov test

Questions The chi-square or KS -test test can be used to answer the following types of questions:

Are the data from a normal distribution?

Are the data from a log-normal distribution?

Are the data from a Weibull distribution?

Are the data from an exponential distribution?

Are the data from a logistic distribution?

Are the data from a binomial distribution?

Application of hypothesis testing on Goodness of fit:

Testing hypothesis:

Ho: the random variable follows the claimed distribution

H1: the random variable does not follow the claimed distribution

Testing the hypothesis is a procedure for deciding weather to accept or reject the hypothesis

The chi-square test

Steps to be followed:

Select the distribution whose adequacy is to be tested, and estimated its parameters from the sample data.

Divide the observed data into k klass intervals.

Count Ni, the number of observations falling in each class interval, i.

Determine the probability with which the RV lies in each of the class interval using the selected distribution

Claculate Ei, the expected number of observations in the class interval i, by multiplying the probability with the number of sample values (n).

Expected absolute frequency:

$$[P(\frac{U - mean}{std}) - P(\frac{L - mean}{std})] * n_{total}$$

formula for the chi -square test:

$$\chi_c^2 = \sum_{j=1}^{n} \frac{(O_j - E_j)^2}{E_j}$$

No. of class intervals, k = 8 No. of parameters, p = 2 Therefore ν = k – p – 1 = 8 – 2 – 1 = 5 Significance level α = 10% = 0.1

In [ ]:

```
The p, which is the number of parameters, also changes from distribution to dist
ribution,
exponential distribution, for example, has only one parameter lambda. So, p will
 be 1
log-normal distribution will have two parameters; gamma distribution will have t
hree
parameters and so on.
```

In [22]:

```
Image("chi.png")
```

Out[22]:

v	α 0.100	0.050	0.025	0.010	0.005	0.001
1	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
31	41.4217	44.9853	48.2319	52.1914	55.0027	61.0983
63	77.7454	82.5287	86.8296	92.0100	95.6493	103.4424
127	147.8048	154.3015	160.0858	166.9874	171.7961	181.9930
255	284.3359	293.2478	301.1250	310.4574	316.9194	330.5197
511	552.3739	564.6961	575.5298	588.2978	597.0978	615.5149
1023	1081.3794	1098.5208	1113.5334	1131.1587	1143.2653	1168.4972

In [24]:

```
Image("chi_test.png", width="700", height="200")
```

Out[24]:

From the Chi-square distribution table,

$$\chi^2_{0.1,5} = 9.24$$

ν \ α	0.9	0.1	0.05
3	0.584	6.25	7.81
4	1.06	7.78	9.49
5	1.61	**9.24**	11.07
6	2.20	10.64	12.59

Kolmogorov – Smirnov Goodness of fit test:

• Alternative to the Chi-square test.

• The test is conducted as follows

• The data is arranged in descending order of magnitude.

• The cumulative probability P(xi) for each of the observations is calculated using the Weibull's formula.

• The theoretical cumulative probability F(xi) for each of the observation is obtained using the assumed distribution.

• The absolute difference of P(x i) and F(x i) is calculated.

• The Kolmogorov-Smirnov test statistic Δ is the maximum of this absolute difference.

$$\Delta = maximum P(x_i) - F(x_i)$$

• The critical value of Kolmogorov-Smirnov statistic Δo is obtained from the table for a given significance level α.

• If Δ < Δo , accept the hypothesis that the assumed distribution is a good fit at significance level α.

In [27]:

Out[27]:

Table for Kolmogorov-Smirnov statistic Δ_o:

Size of sample	Significance Level α				
	0.2	0.15	0.1	0.05	0.01
5	0.45	0.47	0.51	0.56	0.67
10	0.32	0.34	0.37	0.41	0.49
20	0.23	0.25	0.26	0.29	0.36
30	0.19	0.20	0.22	0.24	0.29
40	0.17	0.18	0.19	0.21	0.25
50	0.15	0.16	0.17	0.19	0.23
Asymptotic formula (n > 50)	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

In [ ]: