SM Assignment

Allena Venkata
Sai Abhishek
12202160109
MTech VSKP
DSS

Q1) Define Sign Test. Explain sign Test procedure?

Ans: * The sign test is the simplest form of the non-parametric test.

* It's name comes from the fact that, It is based on direction ('−' or '+' signs) based on the observations.

* In any problem of sign test, we count the
  (i) Number of positive signs
  (ii) Number of negative signs
  (iii) Number of zero's present.

* We take $H_0 : p = 0$ (Null hypotheses)

* The difference is 0 chances, effect the probability of a positive sign, for any particular pair is $\frac{1}{2}$ as the probability of a −ve sign

* If 's' is the number of times the number of time the frequency sign occours then 's' is the binomial distribution with $p = \frac{1}{2}$

The critical value for a two sided alternative when, $\alpha = 0.05$, can be

Scanned with CamScanner

02
12202116007 09

convenientey found by the expression.

$$k = \frac{(n-1)}{2} - (0.98)\sqrt{n}$$

⊛ Then the sign test is classified into two types.

1) One sample sign test
2) Paired Sample sign test.

⊛ One Sample Sign Test: In one sample sign test, we test tho null hypothesis (H_0) Against an approopriate alterenative based on a random sample of size n. We replace each sample value greater than M_0 with a positive sign and Each sample value less than M_0 with a negative Sample valu. And a discarded sample value exactly equal to other sample value (put 0)

Then the null Hypothesis (H_0) is testing for the '+' & '−' signs of a random variable following a binomial distribution with $P = \frac{1}{2}$ or not.

Scanned with CamScanner

03
12202160109

Paired Sample Sign Test:

* This test is again a very important application problems, involving paired data such as data related to the collections receivable before & after a new collation policy.
* In these problems each pair of sample value, can be replaced with a '+' sign if the first value is greater than the second & with a '–' sign if the second value is smaller than the first value or discarded with second value if both are equal.

Q2) Define Mann-Whitney U Test? Explain the procedure for Mann-Whitney U test.

Ans- Mann-Whitney U Test

It belongs to the family of Rank sum test family. With this test we can test the null hypothesis $\mu_1 = \mu_2$ without assuming whether the population sample has roughly the shape of a normal distribution.

The test of null hypothesis is that these two samples come from should be of identical populations maybe based on R_1 (sum of sample 1 ranks) and R_2 (sum of sample 2 ranks)

It maybe noted that, in practice it doesn't matter which sample we call sample 1 & sample 2 respectively

The sample sizes are taken as n_1, n_2 and sum of the samples are taken as R_1, R_2. Simply the sum of n_1 & n_2 positive integers which is known to be $\frac{(n_1+n_2)(n_1+n_2+1)}{2}$

This formula enables to find R_2 if we know & vice versa.

It is an alternative test for two sample test t-test. The decision was based on R_1 & R_2. Now the decision will be based upon either of the relative statistics ie.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Scanned with CamScanner

05
122021601009

Q3 The coin or loss of certain basket ball team for their least 50 consequitive games were as follows.

WWWWWW L WWWWWWL WL
WWWWLL WWWWL WWWLL
WWWWWWWWLL WW WWLL WLWWW

Ans- $n_1 = 36$
$n_2 = 14$
$r = 19$ (total number of runs)

$$\sigma = \sqrt{\frac{2 \times n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1+n_2)^2 * (n_1+n_2-1)}} = 2.8077$$

$$\mu = \frac{2n_1 n_2}{n_1+n_2} + 1 \qquad Z = \frac{r - \mu}{\sigma}$$

$$= -0.77$$

$$\mu = \frac{2(36)(14)}{36+14} \qquad |Z| = 0.77$$

$$\mu = \frac{529}{25} = 21.16$$

As $Z_{cal} < Z_{tab}$ at 5% level of significance (1.96)

$0.77 < 1.96$

Hence, we accept Ho. Thus, there is no real evidence to suggest that the evidence is random.

12202160100
06

Q4) The following data relate to the daily production of cement (in metric tonnes) a large point plant for 30 days.

11.5, 10.0, 11.2, 10.0, 12.3, 11.1, 10.2,
9.6, 8.7, 9.3, 9.3, 10.7, 11.3, 10.4, 11.4, 12.3,
11.4, 10.2, 11.6, 9.5, 10.8, 11.9, 12.4, 9.6
10.5, 11.6, 9.3, 10.4, 11.5

Use sign test to test the null hypothesis that plants average daily production of content is 11.2 metric tonnes against alternative hypothesis $\mu < 11.2$ metric tonns at 5% los Average = 11.2

>11.2 ('+'ve sign) <11.2 ('-'ve sign)

$= 11.2 \Rightarrow 0$

No. of Days	Production of cement (in tons)	sign.
1	11.5	+
2	10.0	−
3	11.2	0
4	10.0	−
5	12.3	+
6	11.1	−
7	10.2	−
8	9.6	−

122021601009
07

	8.7	−
9	9.3	−
10	9.3	−
11	10.7	−
12	10.3	+
13	10.4	−
14	11.4	+
15	12.3	+
16	11.4	+
17	10.2	−
18	11.6	+
19	9.5	−
20	10.8	−
21	12.4	+
22	9.6	−
23	10.5	−
24	11.6	+
25	8.3	−
26	9.3	−
27	10.4	−
28	11.5	+

number of the signs = 11
number of -ve signs = 18
no. of zeros = 1

$x = s = 11$
$n = 18 + 11 = 29$

Scanned with CamScanner

$p = \frac{1}{2} = 0.5$ $\quad q = 1-p$ 12202160100908

$= 1 - \frac{1}{2}$

$= 0.5$

$$Z = \frac{x - np}{\sqrt{npq}}$$

$$= \frac{11 - 39 * (\frac{1}{2})}{\sqrt{29 * (\frac{1}{2})(\frac{1}{2})}}$$

$$= \frac{-3.5}{\sqrt{29 * (\frac{1}{4})}}$$

$$Z = -1.299$$

$$= 1.30$$

$\mu_0 = 11.2$ $\mu < 11.2$ for Z table at 5% level of significance for L.T.T. Table value is -1.64

$Z_{cal} = -1.29$ $\quad Z_{tab} = -1.64$

$Z_{cal} > Z_{tab}$

Hence, we reject Ho Null Hypothesis ie, plant average production is less than 11.2 metric tonnes.

Scanned with CamScanner

12.2021 Wednesday

Q5) Explain ratio to trend method?

Ans. This method is an improvement over the simple average method and is based on the assumption that seasonal variation for any given month is a constant factor of trend. The measurement of seasonal variation by this method consists of various steps:

1. Compute the trend values by the principal of the least squares by fitting an appropriate mathematical curve.
2. Express the original data as the principal of trend values. Assuming the multiplicative model, these percentages will therefore will contain seasonal, cyclic specific values.
3. The cyclic & irregular components are than wiped out by averaging the percentages for different months if the data is monthly, thus leaving us with an indices of seasonal variations. Either Arthematic mean (or) Median can be used, but median

Scanned with CamScanner

is prefferred as mean or median can be used but median is prefferred as mean to AM since later gives a undue weightage to extreme values, which are primarily due to, seasonal swings if there are a few abnormal values, modified mean can be used with advantage

4. Finally, these Indices, obtained in step 3, are adjusted to a total of 1200, for monthly adata or 400 to a quaterly data. by multiplying them through by a constant 'k'

$$k = \frac{1200}{\text{Total of Indices}} \quad \& \quad k = \frac{400}{\text{Total of Indices}}$$

Q6) Explain Link Relative Method.

Ans- This method is also known as pearson's method, is based on the averaging of the link relatives link relative is the values of one seasons expressed as a percentage. Here season refers to time period it would mean monthly da

Link relative for any month = $\dfrac{\text{current month's value}}{\text{previous month's value}} \times 100$

1220216011 011

The steps involved may be summed up i.e.

1. Convert the original data, into link relatives using the above formula
2. Convert the average (Mean or Median) link relatives for each month of the data are monthly. Mean or median can be used for averaging.
3. As in the case of 'ratio to trend' method average the link relatives for each month of the monthly data

Chain Reaction (C.R) for any season is obtained on multiplied the chain relative of the processing season & dividing by 100.

$$\text{C.R for any season} = \frac{(\text{C R of that season})(\text{CR of preceeding season})}{100}$$

Thus for monthly data, the chain relative to first month.

4. Thus the adjustment, is done by subtracting correlation factor from each chain relative.

Scanned with CamScanner

12202160609

d= 1/12 [second (NEW) CR. for Jan - (01)]
then assuming linear method trend, the correlation for Feb, March .. Dec is d, 2d.... 11d respectively

5. Finally adjust the corrected chain, to total by expressing the corrected chain relatives as percentage of arithematic Mean. The resultant gives the adjusted monthly indices of seasonal variables.

Q7) In a certain Industry the production of certain commodity during 2004-2011 is given as:

Year	2004	2005	2006	2007	2008	2009
Production	66.6	84.9	88.6	78.0	96.8	105.2

2010	2011	2013	2014	2015
93.2	111.6	88.3	117.0	115.2

(1) The trend value from 2004- 2015 estimate production of the commodity during the years 2015, 2016 if present trend continues

12202160609
13

Year	Y_t	x	$x \cdot Y_t$	x^2
2004	66.5	-5	-333	25
2005	84.9	-4	-339.6	16
2006	88.6	-3	-265.8	9
2007	78.9	-2	-156	4
2008	96.8	-1	-96.8	1
2009	105.2	0	0	0
2010	93.2	1	93.2	1
2011	111.6	2	223.2	4
2012	88.3	3	264.9	9
2013	117.0	4	468	16
2014	115.2	5	576	25
	1045.4		43.1	135

The normal equation can be written as:-

$\Sigma y = na + b\Sigma t$

$\Sigma t.y = a\Sigma t + b\Sigma t^2$

We shift the origin to middle time period, 2009. let $x = t - 2009$

then least square of line Y_t.

$$Y_t = a + bx$$

Scanned with CamScanner

Normal equations 1220216009
14

$\sum y_t = na + b\sum x$ —— (2)

$\sum x \cdot y_t = a\sum x + b\sum x^2$ —— (3)

$1045.4 = 11a + b(0)$

$434.1 = a(0) + b(135)$

$1045.4 = 11a \Rightarrow a = \frac{1045.4}{11} = 95.0364$

$434.1 = 135b \Rightarrow b = \frac{434.1}{135} = 3.2155$

The straight line eq^n

$y_t = 95.0364 + 3.2155x$

$y_t = 95.3044 -$

$= 79.2508$

$= 83.1972$

$= 91.09$

$= 102.9292$

Production in the year 2015 & 2016

for the year 2015

$y_t = 95.0364 + 3.2155(t-2009)$

$= 95.0364 + 3.2155(6)$

$= \cancel{95}.\ 114.3297$

Scanned with CamScanner

simularly for 2016. 12202160109 15

$$y_t = 95.0364 + 3.2155\,(t-2009)$$

$$= 95.0364 + 3.2155\,(2016-2009)$$

$$= 95.0364 + 3.2155\,(7)$$

$$= 117.54525$$

Scanned with CamScanner