STATISTICAL ANALYSIS ON EQUALITY OF IIP FOR FOOD, TOBACCO AND BEVERAGES PRODUCTION IN INDIA CONSIDERING BASE 2011-12 BY USING DISTRIBUTION FREE ADAPTIVE TEST

Chikhla Jun Gogoi [1]

[1] Assistant Professor, Department of Statistics Pandu College, Guwahati, Assam PIN-781012

Email: gogoi.chikhla8@gmail.com

Abstract

The index of industrial production (IIP) is as index that declares the performance of various industrial sectors of Indian economy. This is calculated and published by Central Statistical Organisation (CSO) every month. This is a composite indicator of the general level of industrial activity in the economy.

Now this study attempts to make a statistical analysis on equality of IIP for food, tobacco and beverages production in India. Secondary data is collected for the study. The conclusion and interpretation are made by the use of adaptive distribution free test.

Key words: Beta Generated family, Generalized Marshall-Olkin family, Power weighted moments, AIC, A and W.

1. INTRODUCTION

For the month of March 2023, the quick estimates of IIP with base 2011-12 stands at 150.5. According to the report food, tobacco and beverages in India was at 10.68% in 2019. This value is based on the World Bank Collection of development indicators and it was compiled from officially recognised sources. This chapter is based on secondary source and this India – food, tobacco and beverages actual values, historical data, forecasts and projections were sourced from the World Bank on June of 2023.

2. OBJECTIVE

The objective of the present study is to test whether the IIP for food, Tobacco and beverages production in India considering base 2011-12 is equal or not. Accordingly, this study is considering the null hypothesis that the IIP of all these three sectors in India is equal against the alternative that they are not equal.

3. METHODOLOGY

In this present study we are going to test our null hypothesis by using Adaptive test. This is a distribution free test. In this Adaptive test after collection of data we calculate the measure of skewness and tailweight of the distribution of the data set that we have collected. Then depending upon value of these two selection statistics we decide the test statistics which will help us to give the interpretation on the data set.

3.1 Selection Statistics

Here we will discuss the use of selection statistics S =(Q_1,Q_2), where Q_1 and Q_2 are Hogg's measure of skewness and tail weight defined by –

$Q_1 = \frac{\bar{U}_{5\%} - \bar{M}_{50\%}}{\bar{M}_{50\%} - \bar{L}_{5\%}}$ and $Q_2 = \frac{\bar{U}_{5\%} - \bar{L}_{5\%}}{\bar{U}_{50\%} - \bar{L}_{50\%}}$

Where, $\bar{U}_{5\%}$, $\bar{M}_{50\%}$ and $\bar{L}_{5\%}$ are the averages of the upper 5%, middle 50% and lower 5% of the order statistics in the combined sample. $\bar{U}_{50\%}$ and $\bar{L}_{50\%}$ are the averages of the upper 50% and lower 50% of the order statistics in the combined sample.

Table 3.1: Theoretical Values of Q_1 and Q_2 for Some Selected Distributions

Distributions	Q_1	Q_2
Uniform	**1**	**1.9**
Normal	**1**	**2.585**
Logistic	**1**	**3.204**
Double exponential	**1**	**3.302**
Exponential	**4.569**	**2.864**

Now let us define four categories of S-

$D_1 = \{S/0 \leq Q_1 \leq 2,\ 1 \leq Q_2 \leq 2\}$

$D_2 = \{S/0 \leq Q_1 \leq 2;\ 2 \leq Q_2 \leq 3\}$

$D_3 = \{S/\ Q_1 \geq 0;\ Q_2 > 3\}$

$D_4 = \{S/\ Q_1 > 2;\ 1 \leq Q_2 \leq 3\}$

This means that the distribution is short tailed or medium tailed if S falls in the category D_1 or D_2 respectively and is long tailed if S falls in the category D_3 and right skewed if it falls in the category D_4.

Buning (1996) proposed the following adaptive test A:

$$A = \begin{cases} G & if\ \ S \in D_1 \\ KW & if\ \ S \in D_2 \\ LT & if\ \ S \in D_3 \\ HFR & if\ \ S \in D_4 \end{cases}$$

Where G, KW, LT and HFR the Gastwrith, Kruskal-Wallis, Long-tailed and Hogg-Fisher-Randle statistics respectively describe in section 3.3.

3.3 Test Procedures

Let us consider, $X_{i1}, X_{i2}, \dots, X_{in_i}$, i=1, 2,……,c be independent random variables with absolutely continuous distribution function F(x-θ_i).

Here our null hypothesis is $H_0: \theta_1 = \theta_2 = \cdots = \theta_c$

and the alternative hypothesis is $H_1: \theta_r \neq \theta_s$ for at least one pair (r,s), r≠s.

3.3.1 F-Test

For normally distributed random samples with equal variance, for testing equality of means, the likelihood ratio F test is the best one. The test statistics defined as

$$F = \frac{(N-c)\sum_{i=1}^{C} n_i(\overline{X_i}-\bar{X})^2}{(C-1)\sum_{i=1}^{C}\sum_{j=1}^{n_i}(X_{ij}-\overline{X_i})^2}$$

where N=$\sum_{i=1}^{c} n_i$, $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $\bar{X} = \frac{1}{N}\sum_{i=1}^{c} n_i\bar{X}_i$

Under H_0 the test statistics follows F distribution with c-1 and N-c degrees of freedom.

3.3.2 Kruskal-Wallis (KW) Test

Let R_{ij} be the rank of the observation X_{ij} in the pooled sample. The Kruskal-Wallis test for two-sided alternative which based on the statistic

$$\textbf{KW}= \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{1}{n_i}\left[R_i - \frac{n_i(N+1)}{2}\right]^2$$

$$\text{KW} = \frac{12}{N(N+1)}\sum_{i=1}^{c}\frac{1}{n_i}[R_i - \frac{n_i(N+1)}{2}]^2$$

$$= \frac{12}{N(N+1)}\sum_{i=1}^{c}\frac{R_i^2}{n_i} - 3(N+1)$$

where $R_i = \sum_{j=1}^{n_i} R_{ij}$ and N $=\sum_{i=1}^{c} n_i$.

for sample size n_i and sample number k large, H_o is rejected if KW> $\chi^2_{\alpha,(k-1)}$. When k is small and n_i are small then exact distribution table for KW can be used.

3.3.3 Adaptive Test

Now let us define the suggested Adaptive test which involves, a linear rank statistics. Let us consider a the combined ordered sample $X_{(1),}X_{(2),\ldots.}X_{(N)}$ of $X_{11,}\ldots X_{1n_{1},\ldots.}X_{c1,}\ldots X_{cn_c}$ and indicator variables V_{ik} given by

$$V_{ik} = \begin{cases} 1 & if\ X_{(k)}\ \ belong\ to\ the\ ith\ sample\ , i = 1, \ldots, c, k = 1, \ldots, N \\ 0 & otherwise \end{cases}$$

where N=$\sum_{i=1}^{c} n_i$,

Let a(k) , k =1,2,…,N be real valued score with mean $\bar{a} = \frac{1}{N}\sum_{k=1}^{N} a(k)$. now we define for each sample a statistics A_i in the following way:

$$A_i = \frac{1}{n_i}\sum_{k=1}^{N} a(k)V_{ik}, \; 1\leq i\leq c$$

Then the linear rank statistics L_N is given by

$$L_N = \frac{(N-1)\sum_{i=1}^{c} n_i(A_i-\bar{a})^2}{\sum_{k=1}^{N}(a(k)-\bar{a})^2}$$

Under H_0, L_N is distribution free. It follows asymptotically chi- square distribution with c-1 degrees of freedom.

Some of the scores to obtain more powerful test for types of distribution according to Buning(1991,1994) are as follows:

(i) Gastwrith Test G (Short Tailed):

$$a_G(k) = \begin{cases} k - \dfrac{(N+1)}{4} \, if k \leq \dfrac{(N+1)}{4} \\ 0 \quad if \, \dfrac{(N+1)}{4} \leq k \leq \dfrac{3(N+1)}{4} \\ k - \dfrac{3(N+1)}{4} \, if k \geq \dfrac{3(N+1)}{4} \end{cases}$$

(ii) Kruskal Wallis Test KW (Medium Tails):

If $a_{KW}(k) = k$, test transform to above KW test

(iii) LT Test (Long Tailed):

$$a_{LT} = \begin{cases} -(\left[\dfrac{N}{4}\right] + 1 & if k < \left[\dfrac{N}{4}\right] + 1 \\ k - \dfrac{(N+1)}{2} \, if \left[\dfrac{N}{4}\right] + 1 \leq k \leq [\dfrac{3(N+1)}{4}] \\ \left[\dfrac{N}{4}\right] + 1 & if k > [\dfrac{3(N+1)}{4}] \end{cases}$$

(iv) Hogg Fisher Randles Test HFR (Right Skewed):

$$a_{HFR} = \begin{cases} k - \dfrac{(N+1)}{2} & if k \leq \dfrac{(N+1)}{2} \\ 0 & if k > \dfrac{(N+1)}{2} \end{cases}$$

For left-skewed distributions we change the terms k- (N+1)/2 and 0 in the definition of the scores above.

Collected Data:

Index of Industrial Production

Base 2011-12

	Food	Tobacco	Beverages
Jan-21	111.4	72.4	54.8
Feb-21	111.3	74	82
Mar-21	130.1	96.5	85.2
Apr-21	110.3	95.6	85.3
May-21	117.3	86.9	86.3
Jun-21	118.8	95	86.8
Jul-21	115.3	81.7	89.9
Aug-21	116.2	79.4	90.2
Sep-21	127.8	70	91.5
Oct-21	145.1	108.2	91.5
Nov-21	118.1	75.8	91.8
Dec-21	151.8	101.8	92
Jan-22	147	93.3	92.9
Feb-22	139.9	79.9	94.4
Mar-22	136.7	79.2	94.7
Apr-22	115.6	94.6	95.3
May-22	149	103.7	95.4
Jun-22	155.5	90.6	98.1
Jul-22	113.2	72.2	102
Aug-22	157	92.5	104.3
Sep-22	145.8	77.9	104.7

Oct-22	138.3	114	110.7
Nov-22	150.5	107.3	116.1
Dec-22	132.7	85.4	118.2
Jan -23	117	113	119.5
FEB -23	141.7	96.4	119.9
Mar-23	122.7	87.9	126.2

4. CALCULATION AND CONCLUSION:

Q_1	0.838892
Q_2	1.738072

According to the values of Q1 and Q2 the collected data falls in the category of D1. That means the data follows the characteristics of a short tailed distribution and as per our theory Gastwirth test (or short tailed test) will be more preferrable for the data to proceed further.

The calculated value of test statistics is 208.41 and the tabulated value of the F test statistics with (2,24) df is 3.40 at 5% level of significance. Hence, we may reject our null hypothesis H_0 at 5% level of significance and conclude that the IIP of the Manufactured products namely Food, Tobacco and Beverages are not equal.

References

[1] Bean, R.(1974). Asymptotically efficient adaptive rank estimates in location models., Annals ofStatistics,2,63-74

[2] Buning, H.(1991). Robuste and Adaptive Tests . De Gruyter, Berlin .

[3] Buning, H.(1994). Robust and adaptive tests for the two sample location problem.– OR

Spektrum 16,33-39.

[4] Buning, H.(2000).Robustness and power of parametric ,nonparametric, robustified and adaptivetests the multi-sample location problem., Statistical Papers 41,381-407.–

[5] Buning, H.(2002). An adaptive distribution-free test for the general two-sample Problem.

Computational Statistics 17,297-313.

[6] Bunning, H.(1996).,Adaptive tests for the c- ple location problem-the case of two-Sidedsam

alternatives', Comm.Stat. Theo. Meth. 25,1569-1582.–

[7] Bunning, H. (1999).,AdaptiveJonckheere-Type Tests for Ordered Alternatives. Journal of

Applied Statistics ,26(5),541-51

[8] Buning, H. And Kossler,W.(1998).Adaptive tests for umbrella alternatives., Biometrical

journal,40,573-587.

[9] Buning, H. And Kossler,W.(1999). The asymptotic power of Jonckheere-Type tests for orderedalternatives. Asutralian and New Zealand Journal of Stagbtistics,41(1), 67-77.

[10] Buning, H. and Thadewald,T.(2000). An adaptive two-sample location-scale test of Lepage-type for symmetric distributions. Journal of Staistical Computation and Simuation., 65,287-310.

[11] Freidlin, B. And Gastwirth, J.(2000).Should the median test be retired from general use? TheAmerican Statistician,54(3),161-164.

[12] Freidlin, B.;Miao,W. And Gastwirth, J.L.(2003a). On the use of the Shapiro-Wilk test In two-stage adaptive inference for paired data from moderate to very heavy tailed distributions.

Biometrical Journal 45,887-900

[13] Gastwirth, J.(1965).Percentile Modification of Two-Sample Rank Tests,Journal of theAmericanStatistical Association, 60(312),1127-1141.,

[14] Hajek, J.(1962). Asymptotically Most Powerful Rank- Order Tests .Annuals of MathematicalStatistics ,33,1124-1147

[15] Hajek, J.(1970). Miscellaneous Problems of Rank Test Theory. In M.L.Puri , ed., Nonparametrictechniques in Statistical Inference, Cambridge, Mass.:Cambridge University Press.

[16] Hajek, J. ;Sidak, Z. And Sen, P.(1999). Theory of Ranks, 2nd edition. New York and London:Academic Press .

[17] Hall, P. And Padmanabhan, A.R.(1997). Adaptive inference for the two-sample scale Problem',Technometrics 39,412-422.

[18] Hao, Li and Houser, D.(2012). Adaptive procedures for Wilcoxon -Mann –Whitney Test.,SevenDecades of Advances.

[19] Hill, N.J., Padmanabhan,A.R. and Puri, M.L.(1988). Adaptive nonparametric procedures andapplications', Applied Statistics,37,205-218.

[20] Hogg, R.V.; Uthoff, V.A., Randles, R.H., and Devenport, A.S.(1972): 'On the selection of theunderlying distribution and adaptive estimation' ,Jour.Amer.Stat.Assoc.,67,597-600.

[21] Hogg, R.(1967). Some Observations on Robust Estimation.Journal of the Amarican StatisticalAssociation,62,1179-86.

[22] Hogg, R.(1974). Adaptive Robust Procedure: A Partial Review and Some Suggestions for

Future Applications and Theory. Journal of American Statistical Association,69(348),909-23.

[23] Hogg, R.V.(1976). A new dimension to nonparametric tests. Commun. Stat. Theo. Methods,A5,1313-1325.

[24] Hogg, R., Fisher, D. and Randles, R.(1975). A Two-Sample Adaptive Distribution Free Test.Journal of the American Statistical Association, 70(351),656-61.

[25] Hogg, R.V. and Lenth, R.V.(1984). A review of some adaptive statistical techniques.

Communications in Statistics- Theory and Methods.,13,1551-1579.

[26] Hogg, R.V. and Randales, R.H.(1975). Adaptive distribution free regression Methods.

Technometrics17,399-408