Comparison of Some Test Statistics for testing the Process Capability Index: An Empirical Comparison

B. M. GOLAM KIBRIA
Department of Mathematics and Statistics
Florida International University
Miami, Florida, USA

SHIPRA BANIK
Department of Physical Sciences
Independent University, Bashundhara, Dhaka 1229
BANGLADESH

Abstract: This paper considers fifteen different test statistics for testing the population process capability index. To assess the performance of the test statistics, empirical sizes and powers are calculated at the 5% nominal level and compared with the classical statistic under both symmetric and skewed distributions. It is evident from the simulation study that some of our proposed tests have better size and power properties as compared to the existing approaches.

Key-Words: Hypothesis testing; Monte Carlo simulation; Power of the test; Process capability index; Symmetric and skewed distributions.

Received: September 21, 2021. Revised: June 17, 2022. Accepted: July 25, 2022. Published: September 2, 2022.

1 Introduction

Statistical process control (SPC) is frequently used in production process or manufacturing to measure how consistently a product performs according to its given specifications. It helps us to monitor, control, and improve the process by eliminating special cause variation in a process (Porter and Oakland[1]). The level of deviation of a process relative to its specification limits is measured by the process capability index (PCI) that can be measured using various statistics in literature. The most commonly used PCI is Cp (Juran[2], Kane[3], and Zhang[4]), which is the fraction of the range between the process specifications to the spread of the process values, as measured by six standard deviation units. In fact, Cp gives the size of the range over which the process actually differs. In this paper we will empathize Cp which is suggested by Kane (1986) and defined as follows:

$$C_p = \frac{USL-LSL}{6\sigma}, \quad (1.1)$$

where USL is the upper specification limit, LSL is the lower specification limit and σ is the process standard deviation. The numerator of Cp provides the size of the range over which the process capacities can differ. The denominator offers the size of the range over which the process essentially differs (Kotz and Lovelace (1998)). Since both USL and LSL are predetermined by the practitioners, Cp mostly depends on the value of the process standard deviation σ. Let X_1, X_2, …,X_n be a random sample of size n from a distribution with finite mean μ and variance σ^2. The estimated Cp can be obtained follows,

$$\hat{C}_p = \frac{\text{USL}-\text{LSL}}{6S}, \quad (1.2)$$

where $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ is the sample standard deviation. More on Cp can be found in Montgomery (2020) among others. An overview of the process capability indices is given by Porter and Oakland (1991) and a compact survey and brief interpretations and comments on 170 publications on capability indices between 1992 to 2000 are given by Kotz and Johnson (2002).

Although the point estimator of Cp can be a suitable measure, however, it is obvious that the interval estimates and test of the hypothesis of Cp are also crucial. Numerous researchers have considered several techniques for estimating Cp by the confidence interval method. For examples, Kocherlakota and Kocherlakota, (1994), Abu-Shawiesh et al. (2020 a,b), Hummel and Hettmansperger (2004), Panichkitkosolkul (2014, 2016), Zhang, (2010) and very recently Kibria and Chen (2021) among others. However, the literature on the test statistics for testing the Cp is very inadequate.

We can see from (1.2) that the estimated Cp heavily depends on the value of the sample standard deviation S. For a skewed distribution, the median describes the center of the distribution better than the mean. Thus for skewed data, it make sense to define the sample standard deviation in terms of the median rather than the mean (Shi and Kibria, 2007). Wright (1995) and Chang et al. (2002) concluded that the process capability index is sensitive to the skewness. Therefore, the objective of this paper is to consider and also to propose some new test statistics based on the modified standard deviation for testing the population Cp and compare them with some existing tests under both normal and non-normal distributional conditions. It is important to note that tests with correct sizes and good powers are essential, particularly in finite samples. The organization of the paper is as follows. Several test statistics are provided in section 2. To compare the performance of the test statistics, a simulation study has been conducted in Section 3. Some concluding remarks are drawn in Section 4.

2 Statistical Methodology

We will review and propose some new test statistics for testing the null hypothesis H_0: $C_p \le C_{p0}$ (the process is not capable) against the alternative hypothesis H1: $C_p > C_{p0}$ (the process is capable) in this section.

2.1 Classical test

Suppose $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$, then it can be shown that under the null hypothesis, the ancillary statistic $\frac{(n-1)C_{po}^2}{\hat{C}_p^2}$ has a Chi-square distribution with (n-1) degrees of freedom. Thus, to test H0: $C_p \leq C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\hat{C_p^2}^2}, \qquad (2.1)$$

where $\hat{C}_p$ is the sample estimate of population Cp. At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,n-1}$ and $\chi^2_{1-\alpha,n-1}$ is the upper 1− α quintile of the central chi-squared distribution with n-1 degrees of freedom.

2.2 Test based on adjusted degrees of freedom

Hummel and Hettmansperger (2004) proposed an estimate for the degrees of freedom using the method of matching. It depends on the fact that the sample variance is a sum of squares and, for sufficiently large samples, is approximated as a chi-square estimate with the appropriate degrees of freedom. They matched the first two moments of the distribution of sample variance with that of a random variable X, which is distributed as $c\chi^2_r$. The solution for r and c is solved using the following systems of equations:

1) σ^2 = cr and

2) $\frac{\sigma^4}{n}\left(\kappa - \frac{n-3}{n-1}\right)$=2rc^2

where κ is the kurtosis of the distribution.

Following Panichkitkosolkul (2016), to test H0: $C_p \leq C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{\hat{r}\,(C_{p0})^2}{\widehat{Cp}^2} \qquad (2.2)$$

where $\hat{r} = \frac{2n}{\hat{\gamma}+2n/(n-1)}$ and $\hat{\gamma} = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{S^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$.

At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,\hat{r}}$, where $\chi^2_{1-\alpha,\hat{r}}$ is the upper 1− α quintile of the central chi-squared distribution with $\hat{r}$ degrees of freedom.

2.3 Test based on the large sample theory

If the normality assumption is invalid, then one can use the large sample theory, where $S^2 \sim N(\sigma^2, \frac{\sigma^4}{n}(\kappa_e + \frac{2n}{n-1}))$, κ_e is the excess kurtosis. Following, Panichkitkosolkul (2016), to test

H0: $C_p \leq C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$Z = \frac{2log\hat{c}_p - 2logC_{p0}}{\sqrt{A}} \qquad (2.3)$$

where $A = \frac{G_2++2n/(n-1)}{n}$ and $G_2 = \frac{n-1}{(n-2)(n-3)}[(n-1)g_2+6]$, $g_2 = \frac{m_4}{m_2^2} - 3$, $m_4 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^4$ and $m_2 = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. At α level of significance, the null hypothesis will be rejected when $Z > Z_\alpha$, where Z_α is the upper 1− α quintile of the standard normal distribution.

2.4 Test based on the augmented large sample theory

Burch (2014) considered a modification to the approximate distribution of log(S) by using a three-term Taylor's series expansion. Employing the large sample properties of S^2, and following Burch (2014) and Panichkitkosolkul (2016), to test H_0: $C_p \le C_{p0}$ vs. H_1: $C_p > C_{p0}$, the test statistic is defined as

$$Z = \frac{2log\ \hat{c}_p - 2logC_{p0} - C}{\sqrt{B}} \quad (2.4)$$

where B= $\widehat{var}\log(S^2) \approx \frac{1}{n}\left(\kappa_e + \frac{2n}{n-1}\right)\left(1 + \frac{1}{2n}\left(\kappa_e + \frac{2n}{n-1}\right)\right)$, $C = \frac{\hat{\kappa}_{e,5}+2n/(n-1)}{2n}$, $\hat{\kappa}_{e,5} = \left(\frac{n+1}{n-1}\right)G_2(1+\frac{5G_2}{n})$. At α level of significance, the null hypothesis will be rejected when $Z > Z_\alpha$, where Z_α is the upper 1− α quintile of the standard normal distribution.

2.5 Robust test

The sample mean and the sample variation can be influenced by the outliers or extreme values of the distribution. To overcome the extreme value problem, the trimmed technique is very useful (Burch (2014), Tukey (1948), and Dixon and Yuen (1974) among others). To modify the variance of the trimmed mean, Sindhumol et al. (2016) recommended an amendment, which is multiplying the variance of the trimmed mean with a fine-tuning constant. This technique can be described as follows: Consider $X_i \sim N(\mu,\sigma^2)$, i=1,2,…,n. Assume that the order statistics of above random samples is denoted by $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$. Then the r-times symmetrically trimmed sample is obtained by reducing both bottommost and uppermost r values. Then the trimmed sample mean and the trimmed sample standard deviation is defined respectively as follows: $\bar{X}_T = \frac{1}{n-2r}\sum_{i=r+1}^{n-r}X_{(i)}$ and $S_T = \sqrt{\frac{1}{n-2r-1}\sum_{i=r+1}^{n-r}(X_{(i)} - \bar{X}_T)^2}$, where $r = [\alpha n]$, trimming is done for 100α% (0 ≤ α ≤ 0.5) of n. The modified trimmed standard deviation, suggested by Sindhumol et al. (2016)) and is defined as follows: $S^*_T = 1.4826 S_T$ (For details, see Abu-Shawiesh et al. (2020a, b)). Now, in the view of equations (2.1) and (2.2), to test H_0: $C_p \le C_{p0}$ vs. H_1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\hat{C}_p^{*2}} \quad (2.5)$$

where $\hat{C}_p^* = \frac{\text{USL} - \text{LSL}}{6\,S_T^*}$ is the sample estimate of population Cp. At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,\hat{r}}$, where $\chi^2_{1-\alpha,\hat{r}}$ is the upper 1− α quintile of the central chi-squared distribution with $\hat{r}$ degrees of freedom.

2.6 Bootstrap test

Efron (1979) introduced the Bootstrap technique, which involves no assumptions about the primary population and can be applied to a range of situations. The accurateness of the bootstrap statistic relies on the number of bootstrap samples. If the number of bootstrap samples is large enough, the estimate may be precise. A bootstrap method is summarized as follows: Let $X^{(*)} = X_1^{(*)}, X_2^{(*)}, \ldots, X_n^{(*)}$, where the i^{th} sample is denoted $X^{(i)}$ for i=1,2 ,…, B, and B are the number of bootstrap samples. The number of bootstrap samples is naturally between 1000 and 2000. Following, Panichkitkosolkul (2014) to test H_0: $C_p \leq C_{p0}$ vs. H_1: $C_p > C_{p0}$, the test statistic is defined as

$$t_0^* = \frac{1}{2}\left(\frac{{C_{p0}}^2}{\widehat{Cp}^2} S^2 k_1 - k_1\right), \qquad (2.6)$$

where $k_1 = \sqrt{2n-2}$. We will reject the null hypothesis, when $t_0^* > \hat{t}_{1-\alpha}^*$, where $\hat{t}_{1-\alpha}^*$ is the quintiles of the following statistic, $T^* = \frac{{s^*}^2 - S^2}{\sqrt{\widehat{var}\left({s^*}^2\right)}}$, where ${s^*}^2$ is a bootstrap replication of the statistic s^2, $\widehat{var}\left({s^*}^2\right) = \frac{1}{n}\left(\hat{\mu}_4^* - \frac{n-3}{n-1}{s^*}^4\right)$ and $\hat{\mu}_4^* = \frac{1}{m}\sum_{i=1}^{m}(X_i^* - \bar{X}^*)^4$.

2.7. Proposed Modified (new) Test Statistics

Motived by the robust test statistic in section 2.5, we would like to use the sample median which is more resistant to outliers and skewed distribution, to define the sample standard deviation. The benefit of modified standard deviation for showed population was found in Shi and Kibria (2007). Now, we will propose following some new test statistics by modifying Equations (2.1), (2.2), (2.3) and (2.4) just by replacing each S by S_M, where

$$S_M = \sqrt{\frac{\sum_{i=1}^{n}(X_i - Md)^2}{n-1}} \qquad (2.7)$$

and Md is the median of the observations of X_1, $X_2, \ldots, X_n$.

2.7.1 Modified Classical test

To test H_0: $C_p \leq C_{p0}$ vs. H_1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\left(C_p^M\right)^2}$$

(2.8)

where $C_p^M = \frac{USL-LSL}{6S_M}$ is the modified sample estimate of population Cp. At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,n-1}$ and $\chi^2_{1-\alpha,n-1}$ is the upper 1− α quintile of the central chi-squared distribution with n-1 degrees of freedom. It is noted that we have assumed the distribution of the test statistic will be approximately chi-squared as classical test. Since, our objective is to compare the performance of the test statistics using empirical power, the critical values from chi-squared distribution does not effect that much as long as it attained the nominal level of the test. The same explanation is applicable for the following modified test statistics.

2.7.2 Modified Test based on adjusted degrees of freedom

To test H0: Cp≤C_{p0} vs. H1: Cp>C_{p0}, the test statistic is defined as

$$\chi^2 = \frac{\hat{r}\,(C_{p0})^2}{\left(C_p^M\right)^2} \qquad (2.9)$$

where $\hat{r} = \frac{2n}{\hat{\gamma}+2n/(n-1)}$ and $\hat{\gamma} = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\frac{\sum_{i=1}^{n}(X_i - Md)^4}{S_M{}^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$.

At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,\hat{r}}$ and $\chi^2_{1-\alpha,\hat{r}}$ is the upper 1− α quintile of the central chi-squared distribution with $\hat{r}$ degrees of freedom.

2.7.3 Modified Test based on the large sample theory

To test H0: $C_p \le C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as\

$$Z = \frac{2logC_p^M - 2logC_{p0}}{\sqrt{A}} \qquad (2.10)$$

where A = $\frac{G_2 + +2n/(n-1)}{n}$ and $G_2 = \frac{n-1}{(n-2)(n-3)}[(n-1)g_2 + 6]$, $g_2 = \frac{m_4}{m_2^2} - 3$, $m_4 = n^{-1}\sum_{i=1}^{n}(X_i - Md)^4$ and $m_2 = n^{-1}\sum_{i=1}^{n}(X_i - Md)^2$. At α level of significance, the null hypothesis will be rejected when $Z > Z_\alpha$, where Z_α is the upper 1− α quintile of the standard normal distribution.

2.7.4 Modified Test based on the augmented large sample theory

To test H0: $C_p \le C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$Z = \frac{2log\,C_p^M - 2logC_{p0} - C}{\sqrt{B}} \qquad (2.11)$$

where B= $\widehat{var}\log(S_M) \approx \frac{1}{n}\left(\kappa_e + \frac{2n}{n-1}\right)\left(1 + \frac{1}{2n}\left(\kappa_e + \frac{2n}{n-1}\right)\right)$, C = $\frac{\hat{\kappa}_{e,5} + 2n/(n-1)}{2n}$, $\hat{\kappa}_{e,5} = \left(\frac{n+1}{n-1}\right)G_2(1 + \frac{5G_2}{n})$. At α level of significance, the null hypothesis will be rejected when $Z > Z_\alpha$, where Z_α is the upper 1− α quintile of the standard normal distribution.

2.8 Test Based on IQR

To test H0: $C_p \le C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\left(C_p^{IQR}\right)^2} \qquad (2.12)$$

where $C_p^{IQR} = \frac{USL - LSL}{6S_{SIQR}}$ is the modified sample estimate of population *Cp* and SIQR = IQR/1.349. At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,n-1}$ and $\chi^2_{1-\alpha,n-1}$ is the upper 1− α quintile of the central chi-squared distribution with n-1 degrees of freedom.

2.9 Test Based on Sn

To test H0: $C_p \leq C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\left(C_p^{Sn}\right)^2} \qquad (2.13)$$

where $C_p^{Sn} = \frac{USL-LSL}{6S_n}$ is the modified sample estimate of population Cp and S_n estimator is proposed by Rousseeuw and Croux (1993) and is defined as the median of the n medians of the absolute differences between values,

At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,n-1}$ and $\chi^2_{1-\alpha,n-1}$ is the upper 1− α quintile of the central chi-squared distribution with n-1 degrees of freedom.

2.10 *Test Based on AAMD*

The AADM is a robust scale estimator that measures the deviation of the data from the sample median, MD, which is less influenced by outliers. It is defined as follows: S_{AADM}= $\frac{\sqrt{\pi/2}}{n}\sum_{i=1}^{n}|X_i - MD|$. The MD is best known for being insensitive to outliers and has a maximal 50% breakdown point (Rousseeuw and Croux, 1993). To test H0: $C_p \leq C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\left(C_p^{AAMD}\right)^2}$$

(2.14)

where $C_p^{AAMD} = \frac{USL-LSL}{6S_{AAMD}}$ is the modified sample estimate of population Cp. At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,n-1}$ and $\chi^2_{1-\alpha,n-1}$ is the upper 1− α quintile of the central chi-squared distribution with n-1 degrees of freedom*2*

2.11 *Test Based on MAD*

The MAD was first introduced by Hampel (1974) and is widely used in various applications as an alternative to S. MAD for a random sample is defined as follows: $S_{MAD} = 1.4826MD\{|X_i - MD|\}, i = 1,2,3,\dots,n.$ The 1.4826 factor given in S_{MAD} adjusts the scale for maximum efficiency when the data comes from a normal distribution.

To test H0: $C_p \leq C_{p0}$ vs. H1: $C_p > C_{p0}$, the test statistic is defined as

$$\chi^2 = \frac{(n-1)(C_{p0})^2}{\left(C_p^{MAD}\right)^2}$$

(2.15)

where $C_p^{MAD} = \frac{USL-LSL}{6S_{MAD}}$ is the modified sample estimate of population Cp. At α level of significance, the null hypothesis will be rejected when $\chi^2 > \chi^2_{1-\alpha,n-1}$ and $\chi^2_{1-\alpha,n-1}$ is the upper 1− α quintile of the central chi-squared distribution with n-1 degrees of freedom.

Since a theoretical comparison among the test statistics is not possible, a simulation study has been conducted in the section follow.

3 Simulation Study

3.1 Simulation Design

In order to determine the effectiveness of the test statistics under symmetric and skewed distributions, we have generated data from the following distributions:

I. Standard Normal distribution, N(50,1)
II. Chi-Square distribution, $\chi^2_{(1)}$
III. t(6) (t distribution with 6 DF)
IV. Beta (4,1) (will give skewness =-0.86, which is left skewed)

MATLAB R2018a programming language is used for all types of calculations. The number of simulation replications was 5000 for each case. Random samples were generated from each of the above-mentioned distributions with Cp =1.0 to calculate size of the test and C_p = 1.33 for power of the test respectively. We consider sample sizes, n=15, 30, 50, 80, 100 and 200 and B=2000 for bootstrap samples. The most common significance level (α=0.05) is used for estimating the size and power of the selected tests. Simulated results are tabulated in Tables 3.1 to 3.8 for selected random samples.

3.2. Results and Discussions

In this section, we will discuss the results of the simulation study, which test statistics have sizes close to the nominal level and also have good powers in finite samples. In Tables 3.1 to 3.4 (corresponding Figures 3.1 to 3.4 for better understanding), we have reported simulated sizes when data were generated from the N(50,1) $\chi^2_{(1)}$, $t_{(6)}$ and Beta(4,1) distributions respectively. In Table 3.1 (see Figure 3.1), we have reported estimated sizes for selected tests.

In Table 3.1, we have presented estimated sizes when data are generated from N(50,1) distribution. We have assumed USL = 53 and LSL = 47 to calculate the sample Cp. Our simulation results in Table 3.1 (Figure 3.1) show that when data are generating from N(50, 1) distribution, the classical test, adjusted classical test, augmented large sample, Bootstrapped test, our proposed modified classical, the modified adjusted degrees of freedom, the modified large samples tests, modified augmented large sample and classical AAMD tests have sizes close to the nominal level. When n increases, it is noticeable that estimated sizes are going to converge with the 5% nominal level for the classical test, adjusted classical test, augmented large sample, Bootstrapped test, our proposed modified classical, the modified adjusted degrees of freedom, the modified large samples tests. Overall, Large sample, Robust 5% and 10%, classical IQR, classical Sn and classical MAD tests have sizes much higher than the observed nominal 5% level [(.05+1.96*sqrt((.05*.95)/5000)=0.06] and therefore may not be suitable for testing when data are from Normal distribution.

Table 3.1: Empirical sizes for testing H_0: Cp≤1.0 vs. H_1: Cp>1.0 when data generated from the N(50,1) distribution with skewness 0 and Cp = 1.0

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical (C)	0.0532	0.0496	0.0484	0.0488	0.0448	0.0536	0.05
Adjusted Classical (AC)	0.0668	0.0576	0.0568	0.0516	0.0478	0.0534	0.06
Large sample (L)	0.0970	0.0800	0.0686	0.0684	0.0652	0.0588	0.07
Augmented large sample (ALS)	0.0138	0.0210	0.0270	0.0360	0.0474	0.0562	0.03
Robust_5% (R5)	0.2564	0.3006	0.4252	0.6830	0.7406	0.9222	**0.55**
Robust 10% (R10)	0.0530	0.1034	0.1002	0.0844	0.0768	0.0418	**0.08**
Bootstrap (B)	0.0068	0.0092	0.0210	0.0262	0.0394	0.0646	0.03
Modified classical (MC)	0.0642	0.0576	0.0556	0.0534	0.0492	0.0556	0.06
Modified Adjusted DF (MADF)	0.0674	0.0606	0.0602	0.0564	0.0500	0.0548	0.06
Modified large sample (MLS)	0.0774	0.0664	0.0600	0.0600	0.0600	0.0556	0.06
Modified augmented LS (MALS)	0.0138	0.0206	0.0280	0.0260	0.0374	0.0456	0.03
Classical_IQR	0.1348	0.1426	0.1508	0.1580	0.1372	0.1610	**0.15**
Classical_Sn	0.1446	0.1316	0.1246	0.1204	0.1098	0.1184	**0.12**
Classical_AAMD	0.0404	0.0440	0.0514	0.0528	0.0480	0.0582	0.05
Classical_MAD	0.1256	0.1202	0.1340	0.1382	0.1354	0.1518	**0.13**

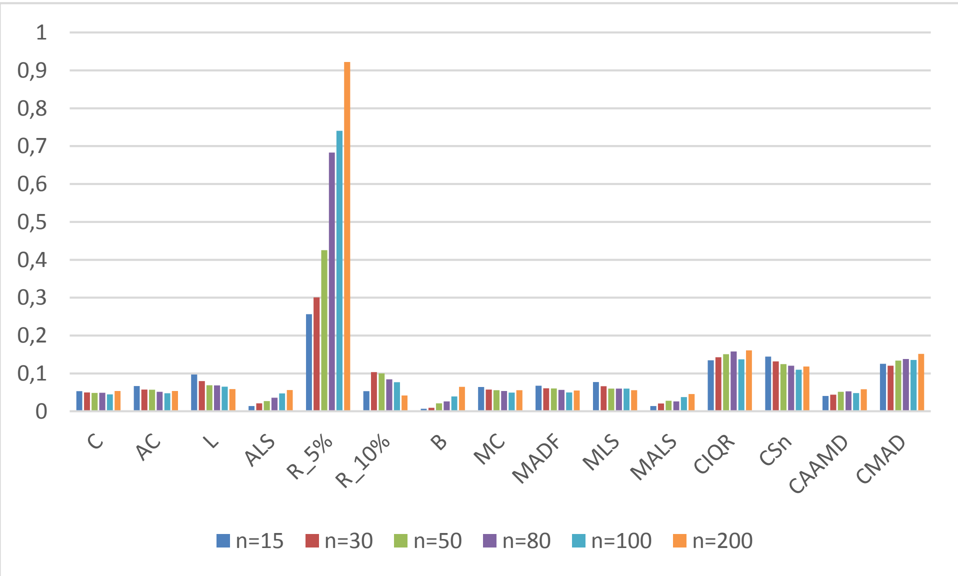

Figure 3.1: Estimated sizes for selected tests when data generated from the N(50,1) distribution

Table 3.2: Empirical sizes for testing H0: Cp≤1.0 vs. H1: Cp>1.0 when data generated from the $\chi^2_{(1)}$ distribution with skewness 2.828 and Cp = 1.0

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.1496	0.1844	0.1950	0.2072	0.2082	0.2404	**0.20**
Adjusted Classical	0.0828	0.0800	0.0706	0.0642	0.0572	0.0548	0.07
Large sample	0.2952	0.2156	0.1952	0.1726	0.1556	0.1208	**0.19**
Augmented large sample	0.1516	0.1150	0.1052	0.0986	0.0862	0.0770	**0.11**
Robust_5% (R5)	0.1438	0.1186	0.1116	0.1412	0.0652	0.0550	**0.11**
Robust 10% (R10)	0.0592	0.0598	0.0542	0.0518	0.0524	0.0513	0.05
Bootstrap	0.1370	0.1262	0.1180	0.1086	0.0982	0.0729	**0.11**
Modified classical	0.1200	0.1626	0.1724	0.1914	0.1937	0.1928	**0.17**
Modified Adjusted DF	0.0786	0.0783	0.0681	0.0623	0.0612	0.0532	0.07
Modified large sample	0.2042	0.1362	0.1052	0.0842	0.0658	0.0590	**0.11**
Modified augmented large sample theory	0.1314	0.1056	0.0988	0.0934	0.0816	0.0750	0.10
Classical_IQR	0.0410	0.0488	0.0474	0.0486	0.0493	0.0496	0.05
Classical_Sn	0.1246	0.1116	0.1046	0.1004	0.0981	0.0984	**0.11**
Classical_AAMD	0.0504	0.0449	0.0587	0.0513	0.0493	0.0482	0.05
Classical_MAD	0.1003	0.0901	0.0940	0.0932	0.0941	0.0910	0.09

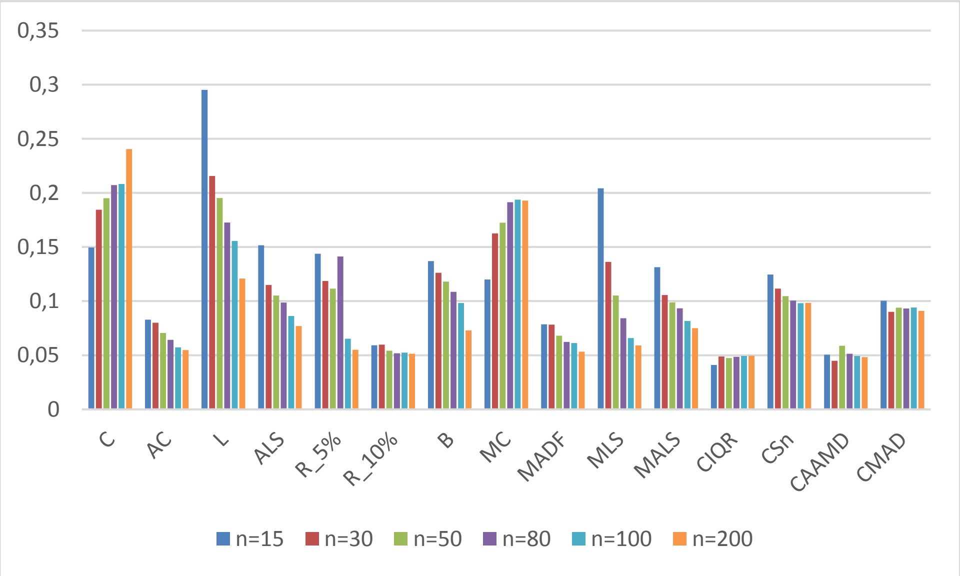

Figure 3.2: Estimated sizes for selected tests when data generated from the $\chi^2_{(1)}$ distribution

In the Table 3.2, we have reported estimated sizes when DGP is $\chi^2_{(1)}$. We have assumed USL = 5.243 and LSL = -3.243 to estimate the sample Cp. From Table 3.2 and Figure 3.2 we observed that the adjusted classical test, our proposed trimmed 10% robust test, the modified adjusted degrees of freedom, the classical IQR test and the AAMD tests have sizes close to the nominal level. When n increases, we have observed that the modified large sample test, the modified adjusted large sample test, the trimmed 5% robust test and the augmented large sample test sizes are converging to nominal level. Other considered tests sizes are observed higher than the nominal level. The estimated nominal sizes when random samples are drawn from the student's t distribution with 5 degrees of freedom are presented in Table 3.3 and graphical representation in Figure 3.3. We have assumed USL = 3.873 and LSL = -3.873 to find the estimated value of Cp. Both Table 3.3 and Figure 3.3 indicated that the adjusted classical test, augmented large sample, Robust 10%, all of our proposed tests but the modified large sample test have sizes close to the nominal level.

Table 3.3: Empirical sizes for testing H_0: Cp≤1.0 vs. H_1: Cp>1.0 when data generated from the $t_{(5)}$ distribution with skewness 0 and Cp = 1.0

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.1048	0.1046	0.1208	0.1234	0.1400	0.1520	**0.12**
Adjusted Classical	0.0682	0.0558	0.0534	0.0464	0.0474	0.0440	0.05
Large sample	0.1802	0.1698	0.1488	0.1400	0.1328	0.1168	**0.15**
Augmented large sample	0.0410	0.0462	0.0436	0.0410	0.0382	0.0388	0.04
Robust_5%	0.1382	0.1116	0.1218	0.0956	0.1114	0.0956	**0.11**
Robust 10%	0.0188	0.0302	0.0070	0.0026	0.0012	0.0004	0.01
Bootstrap	0.0242	0.1242	0.1816	0.1768	0.3814	0.6348	**0.25**
Modified classical	0.1156	0.1134	0.1278	0.1300	0.1432	0.1554	**0.13**
Modified Adjusted Degrees of freedom	0.0696	0.0568	0.0550	0.0474	0.0488	0.0450	0.05
Modified large sample theory	0.1518	0.1516	0.1394	0.1316	0.1272	0.1134	**0.14**
Modified augmented large sample theory	0.0398	0.0454	0.0434	0.0410	0.0382	0.0388	0.04
Classical_IQR	0.0510	0.0334	0.0162	0.0090	0.0066	0.0002	0.02
Classical_Sn	0.0572	0.0326	0.0176	0.0102	0.0066	0.0002	0.02
Classical_AAMD	0.0354	0.0286	0.0248	0.0170	0.0152	0.0060	0.02
Classical_MAD	0.0476	0.0274	0.0112	0.0072	0.0064	0.0004	0.02

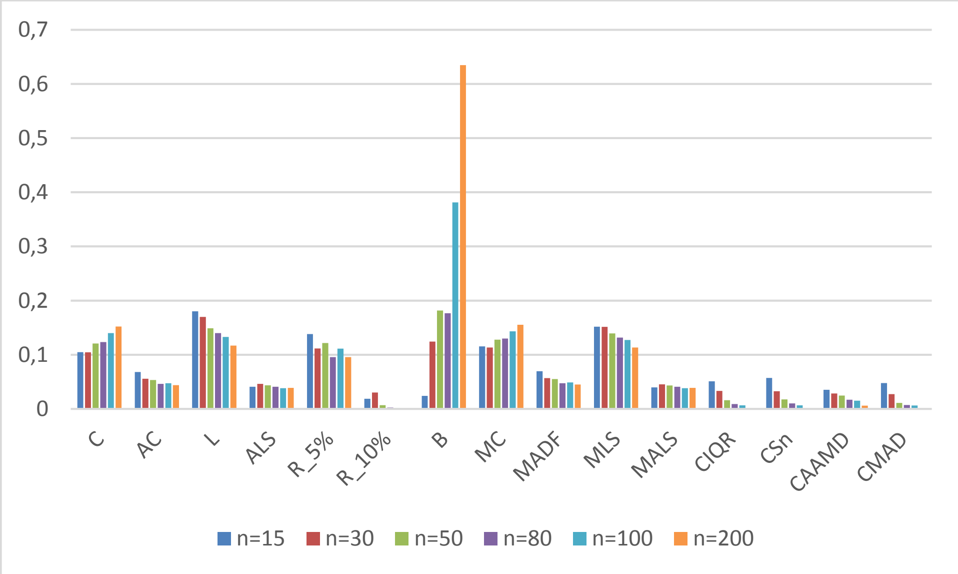

Figure 3.3: Estimated sizes for selected tests when data generated from the $t_{(6)}$ distribution

Table 3.4: Empirical sizes for testing H_0: Cp≤1.0 vs. H_1: Cp>1.0 when data generated from the Beta(4,1) distribution with skewness -0.86 and Cp = 1.0

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.0244	0.0212	0.0220	0.0210	0.0230	0.0244	0.02
Adjusted Classical	0.0704	0.0652	0.0646	0.0596	0.0552	0.0574	0.06
Large sample	0.0708	0.0550	0.0570	0.0558	0.0570	0.0574	0.06
Augmented large sample	0.0068	0.0042	0.0036	0.0016	0.0028	0.0008	0.00
Robust_5%	0.3314	0.4452	0.2124	0.6362	0.8512	0.3858	0.48
Robust 10%	0.0952	0.1053	0.2390	0.2690	0.2882	0.3942	0.23
Bootstrap	0.0090	0.0100	0.0150	0.0200	0.0220	0.0250	0.02
Modified classical	0.0424	0.0286	0.0300	0.0258	0.0280	0.0264	0.03
Modified Adjusted DF	0.0744	0.0690	0.0690	0.0632	0.0594	0.0612	0.07
Modified large sample	0.0544	0.0440	0.0430	0.0466	0.0504	0.0518	0.05
Modified augmented large sample theory	0.0066	0.0042	0.0024	0.0016	0.0028	0.0008	0.00
Classical_IQR	0.2058	0.2686	0.3422	0.3194	0.4442	0.5810	0.36
Classical_Sn	0.1850	0.2136	0.2366	0.2210	0.2850	0.3720	0.25
Classical_AAMD	0.0438	0.0570	0.0766	0.0712	0.0632	0.0612	0.06
Classical_MAD	0.1756	0.2380	0.3044	0.2920	0.4226	0.5612	0.33

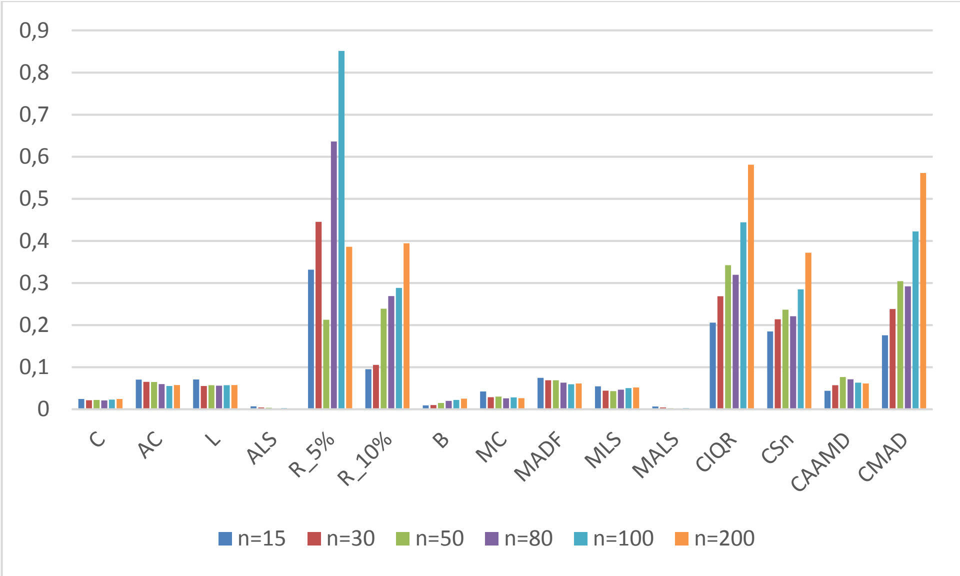

Figure 3.4: Estimated sizes for selected tests when data generated from the Beta(4,1) distribution

We have reported estimated nominal sizes when data are generated from the negative skewed distribution B(4,1) in Table 3.4 and graphical representation in Figure 3.4. We have assumed USL = 1.06 and LSL = -0.067 to estimate the sample Cp. Both Table 3.4 and Figure 3.4 evidenced that classical, adjusted classical, Large sample, Adjusted large sample, Boot strap, modified classical, modified adjusted degrees of freedom, modified large sample, modified augmented large sample, and classical AAMD tests have sizes close to the nominal level.

In the following Tables 3.5 to 3.8 we reported the empirical power of the tests when Cp=1.33.

In the Table 3.5 (see Figure 3.5), we have tabulated estimated powers when data are

generated from the normal distribution. It is observed from both Table 3.5 and Figure 3.5 that when the sample size increases, power of the tests are also increases. We also noticed that for sample sizes 100 and 200, all tests reach power 100% except the followings five tests and they are Large sample, Augmented large sample, Bootstrap, modified large sample and modified augmented large sample. For moderate sample sizes (50 and 80), classical test, adjusted classical, adjusted large, Robust 5% & 10%, moderate classical, moderate adjusted classical, CIQR, CSn and CAAMD have better powers than the rest. For small sample size (15, 30), Robust 5% performed the best. However, it fails to obtain the nominal level 0.05. However, among fifteen tests, the following six test statistics, classical, adjusted classical, modified classical, modified adjusted DF and classical AAMD performed very well and attained the nominal level 0.05. Our proposed modified classical test statistic performed the best.

Table 3.5: Empirical powers for testing H_0: Cp≤1.0 vs. H_1: Cp>1.0 when data generated from the N(50,1) distribution with skewness 0 and Cp = 1.33

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.5068	0.7372	0.8770	0.8896	0.9878	1.0000	0.83*
Adjusted Classical	0.5184	0.7396	0.8786	0.8838	0.9886	1.0000	0.83*
Large sample	0.0584	0.1036	0.3456	0.4563	0.5510	0.6634	0.36
Augmented large sample	0.0704	0.1324	0.3923	0.5012	0.6026	0.7654	0.41*
Robust_5%	0.7426	0.9042	0.9810	0.8340	1.0000	0.9974	0.91
Robust 10%	0.3794	0.7124	0.8340	0.9246	0.9574	1.0000	0.80
Bootstrap	0.0150	0.1170	0.1940	0.2616	0.5516	0.8836	0.34*
Modified classical	0.5432	0.7526	0.8880	0.8990	0.9894	1.0000	0.85*
Modified Adjusted DF	0.5242	0.7456	0.8828	0.8880	0.9894	1.0000	0.84*
Modified large sample	0.0656	0.1206	0.3867	0.4710	0.5745	0.6834	0.83*
Modified augmented large sample	0.0804	0.1652	0.4012	0.5018	0.6125	0.7800	0.42*
Classical_IQR	0.4926	0.6444	0.7602	0.7632	0.9154	0.9924	0.76
Classical_Sn	0.5476	0.7218	0.8446	0.8428	0.9640	0.9990	0.82
Classical_AAMD	0.4234	0.6796	0.8440	0.8512	0.9790	1.0000	0.80*
Classical_MAD	0.4482	0.6022	0.7442	0.7474	0.9070	0.9914	0.74

Note: Power with star (*) marks attained the nominal level 0.05

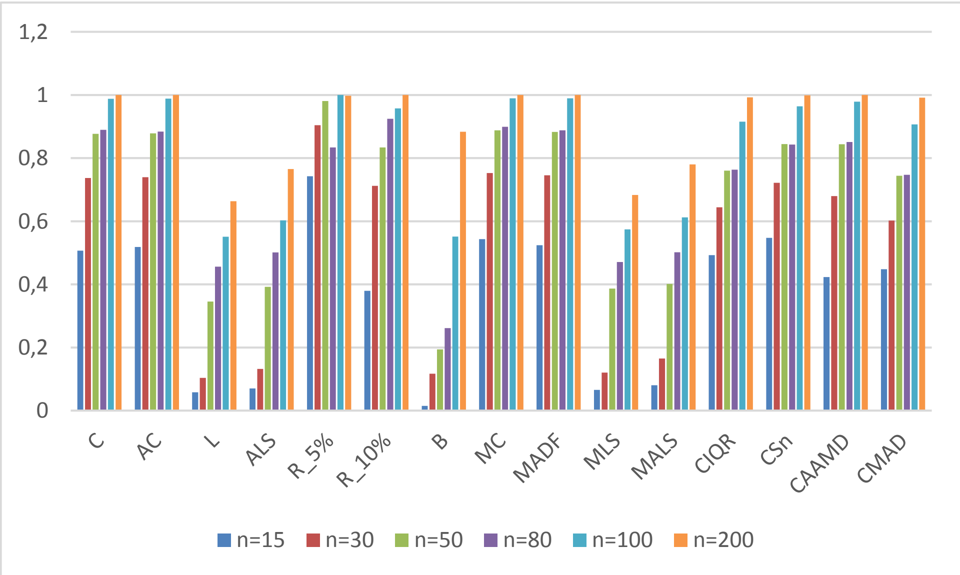

Figure 3.5: Estimated powers for all tests when data generated from the N(50,1) distribution

Next we have observed power properties of our proposed tests when data are generated from the positive skewed distribution, namely the chi-square distribution. Results are tabulated in the

Table 3.6 and the graphical representation in Figure 3.6. From both Table 3.6 and Figure 3.6, we observed that power depends on the size of the samples, ie. with increasing n, power increases for all tests. We have observed that our proposed tests have better power properties as compared to existing tests. Among all tests, for large n (50 and above) the proposed modified classical test performed the best followed by classical test, MADF, CAAMD, CMAD, CSn and so on. For small sample size (15, 30), the proposed CAAMD performed the best followed by CSn, MC and CAAMD. Overall, our proposed modified classical test performed the best followed by proposed classical AAMD, classical test and classical IQR and attained the nominal level 0.05.

Table 3.6: Empirical powers for testing H_0: Cp≤1.0 vs. H_1: Cp>1.0 when data generated from the $\chi^2_{(1)}$ distribution with skewness 2.828 and Cp = 1.33

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.3628	0.4998	0.6282	0.7424	0.7896	0.9288	0.66*
Adjusted Classical	0.2710	0.3368	0.4390	0.5402	0.5858	0.7968	0.49
Large sample	0.1388	0.2692	0.3310	0.4124	0.5092	0.6510	0.41
Augmented large sample	0.1572	0.2901	0.3826	0.4368	0.5130	0.6671	0.41
Robust_5%	0.3580	0.4104	0.5336	0.5674	0.6150	0.8780	0.56*
Robust 10%	0.1540	0.2170	0.2700	0.4814	0.5750	0.7380	0.41*
Bootstrap	0.1516	0.1910	0.4132	0.5508	0.5344	0.8896	0.46
Modified classical	0.4362	0.5958	0.7274	0.8420	0.8870	0.9798	0.75*
Modified Adjusted DF	0.3014	0.4000	0.5310	0.6604	0.7190	0.9088	0.59
Modified large sample	0.1860	0.2348	0.3536	0.4260	0.5128	0.6723	0.40
Modified augmented large sample	0.1492	0.2944	0.3936	0.4266	0.5328	0.6721	0.41
Classical_IQR	0.3438	0.4240	0.5358	0.6798	0.7642	0.8262	0.60*
Classical_Sn	0.5600	0.6021	0.6102	0.6587	0.7080	0.7778	0.65
Classical_AAMD	0.5534	0.5730	0.6510	0.7142	0.8220	0.8540	0.69*
Classical_MAD	0.4081	0.4641	0.5044	0.5623	0.7912	0.8610	0.60

Note: Power with star (*) marks attained the nominal level 0.05

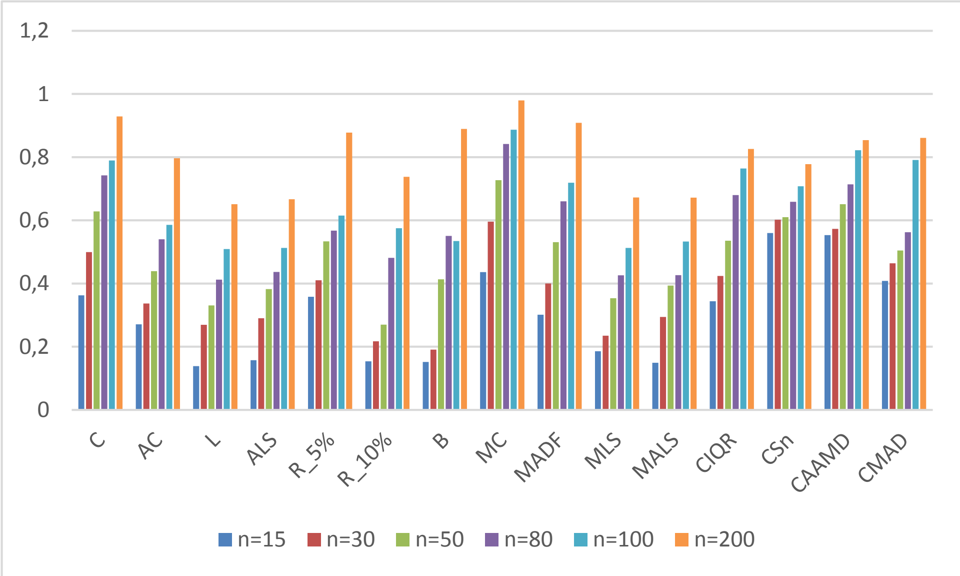

Figure 3.6: Estimated powers for selected tests when data generated from the $\chi^2_{(1)}$ distribution

Next, we reported the empirical power of the tests when data are generated from t-distribution in Table 3.7 and graphical representation in Figure 3.7. It appears from Table 3.7 and Figure 3.7 that all but the trimmed 10% robust test, the classical IQR test, CSn and the classical MAD tests are powerful for large sample size (say n=100 and 200). Overall, the augmented large sample and proposed modified augmented large sample test performed better than the rest of the tests and attained the nominal level 0.05. For small sample, proposed modified augmented large sample test performed better than the counterpart augmented large sample and for large sample, the augmented large sample test performed better than the counterpart proposed modified augmented large sample test.

To see the performance of the test statistics for negatively skewed distribution, we have generated data from the beta distribution with parameters values 4 and 1. Results are tabulated in the Table 3.8 and also for visual inspection in the Figure 3.8 respectively. It is found that classical test, adjusted classical test modified classical test and classical AAMD tests performed better than the rest of the estimators. Overall, it may be concluded that our proposed tests have good powers in finite samples and attained the nominal level 0.05.

Table 3.7: Empirical powers for testing H_0: Cp≤1.0 vs. H_1: Cp>1.0 when data generated from the $t_{(5)}$ distribution with skewness 0 and Cp = 1.33

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.4098	0.5834	0.5926	0.8782	0.9266	0.9924	0.73
Adjusted Classical	0.3728	0.5186	0.5206	0.8212	0.8886	0.9870	0.68*
Large sample	0.4256	0.6074	0.6472	0.8441	0.9304	0.9929	0.74
Augmented large sample	0.4426	0.6089	0.6606	0.8539	0.9421	0.9923	0.75*
Robust_5%	0.5356	0.6696	0.3826	0.9118	0.5494	0.9966	0.67
Robust 10%	0.3654	0.3856	0.4438	0.5012	0.5368	0.6349	0.48*
Bootstrap	0.1570	0.3358	0.2710	0.3288	0.7070	0.9666	0.46
Modified classical	0.4366	0.6018	0.6070	0.8838	0.9288	0.9926	0.74
Modified Adjusted DF	0.3784	0.5222	0.5252	0.8248	0.8918	0.9872	0.69*
Modified large sample	0.5194	0.6168	0.6654	0.8604	0.9002	0.9912	0.76
Modified augmented large sample	0.5226	0.6114	0.6606	0.8023	0.8972	0.9534	0.74*
Classical_IQR	0.2770	0.3252	0.3070	0.4424	0.4712	0.5992	0.40*
Classical_Sn	0.3322	0.3988	0.4048	0.5750	0.6258	0.7958	0.52*
Classical_AAMD	0.2914	0.4326	0.4328	0.7420	0.8158	0.9632	0.61*
Classical_MAD	0.2398	0.2956	0.2864	0.4216	0.4558	0.5944	0.38*

Note: Power with star (*) marks attained the nominal level 0.05

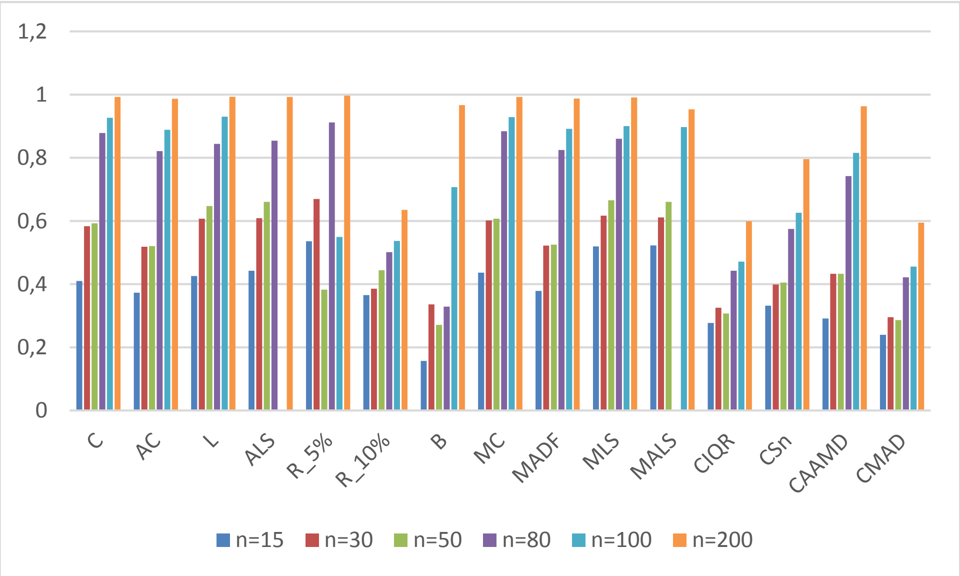

Figure 3.7: Estimated powers for selected tests when data generated from the $t_{(5)}$ distribution

Table 3.8: Empirical powers for testing H_0: $Cp \leq 1.0$ vs. H_1: $Cp > 1.0$ when data generated from the Beta(4,1) distribution with skewness -0.86 and Cp = 1.33

			Sample sizes				
Test Statistics	**n=15**	**n=30**	**n=50**	**n=80**	**n=100**	**n=200**	**Average**
Classical	0.5230	0.7806	0.9286	0.9878	0.9974	1.0000	0.87*
Adjusted Classical	0.5882	0.8132	0.9458	0.9906	0.9982	1.0000	0.89*
Large sample	0.3336	0.5602	0.5923	0.5989	0.6045	0.6342	0.55*
Augmented large sample	0.3606	0.5032	0.6032	0.6048	0.6123	0.6432	0.55*
Robust_5%	0.8228	0.9538	0.9958	0.9998	1.0000	1.0000	0.96
Robust 10%	0.5082	0.8452	0.9530	0.9886	0.9964	1.0000	0.88
Bootstrap	0.3526	0.3987	0.4015	0.5742	0.6543	0.8964	0.55*
Modified classical	0.5704	0.7970	0.9356	0.9884	0.9974	1.0000	0.88*
Modified Adjusted DF	0.5904	0.8170	0.9480	0.9908	0.9982	1.0000	0.89
Modified large sample	0.3524	0.4902	0.6312	0.6123	0.6328	0.6545	0.56*
Modified augmented large sample theory	0.3606	0.4876	0.6032	0.6231	0.6419	0.6662	0.56*
Classical_IQR	0.6000	0.7796	0.8942	0.9704	0.9880	0.9996	0.87
Classical_Sn	0.6408	0.8356	0.9440	0.9890	0.9962	1.0000	0.90
Classical_AAMD	0.4778	0.7660	0.9320	0.9878	0.9974	1.0000	0.86*
Classical_MAD	0.5288	0.7432	0.8864	0.9638	0.9840	0.9998	0.85

Note: Power with star (*) marks attained the nominal level 0.05

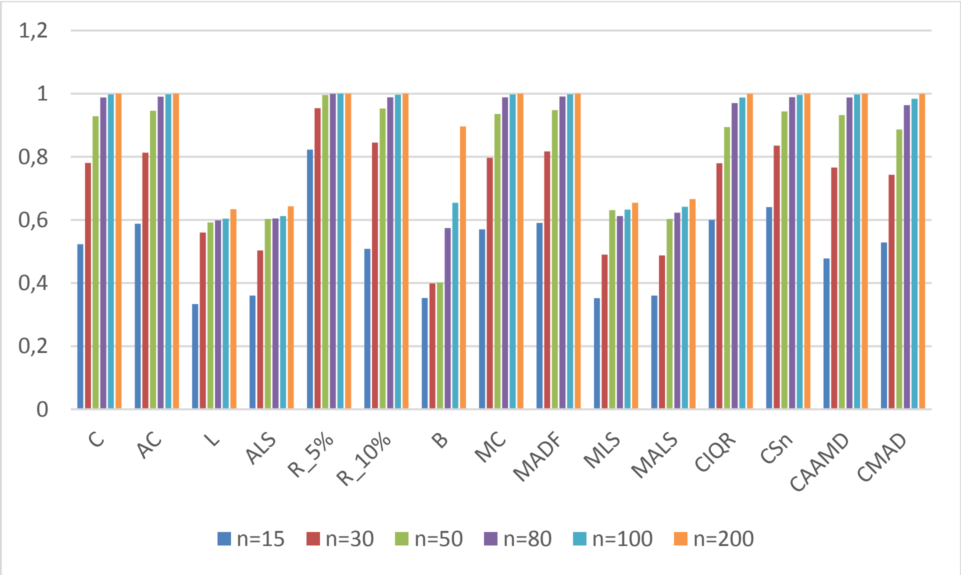

Figure 3.8: Estimated powers for selected tests when data generated from the Beta(4,1) distribution

4. Concluding remarks

This paper considers fifteen different test statistics (7 existing and 8 proposed) for testing the population process capability ratio. Since, a theoretical comparison among the tests is not possible, a simulation study has been conducted to compare the performance of the test statistics under various kinds of distribution such as symmetric and skewed distributions. Empirical size and power of the test were considered as performance criterion. Our simulation results show that some of test statistics have sizes close to the 5% nominal level and also have good powers in finite samples. We believe that the findings of this paper will contribute to process capability literature, and it will be helpful to choose a test statistic when some researchers are interested in testing the population process capability index. Since the conlcusions of the paper is based on the simulation study, for the definite statement about a specific test and for a specific distribution, we need more analysis.

References:

[1] Abu-Shawiesh, M.O.A, Srinivasan, M.R., Sindhumol, M.R., Kibria, B. M. G. (2020a), Performance of a Robust Confidence Interval for the Process Capability Index Cp based on the Modified Trimmed Standard Deviation Under Non-Normal Distributions, WSEAS Transactions on Mathematics, 19, 571-580. DOI: 10.37394/23206.2020.19.62.

[2] Abu-Shawiesh, M., Banik, S., Kibria, B M G., Akyüz, H. E., (2020b). A comparison of some modified confidence intervals based on robust scale estimators for process capability index, Production Engineering-Research and Development, 14 (2), 2020, 217 – 229.

[3] Banik, S. and Kibria, B. M. G. (2022). Size and Power Properties of the Tests for the Process Capability Index, Jahangirnagar University Journal of Statistical Studies.36, 1-14.

[4] Burch, B. (2014), Estimating kurtosis and confidence intervals for the variance under nonnormality, Journal of Statistical Computation and Simulation, 84, 2710–2720.

[5] . Chang, Y. S., Choi, I. S. and Bai, D. S. (2002) Process capability indices for skewed populations. Quality and reliability Engineering International, 18, 383-393.

[6] Dixon, W. J., and Yuen, K. K. (1974), Trimming and winsorization: A review, Statistische Hefte, 15, 157-170.

[7] Efron, B. (1979), Bootstrap methods: Another look at the jackknife, Annals of Statistics, 7, 1-26.

[8] Hummel, R.B.S. and Hettmansperger, T. (2004), Better confidence intervals for the variance in a random sample, Minitab Technical Report.

[9] Juran, J.M. (1974), Jurans Quality Control Handbook, New York, NY, USA: McGraw-Hill, USA.

[10] Kane, V.E. (1986), Process capability indices, Journal of Quality Technology, 18, 41-52.

[11] Kibria, B. M. G. and Chen, W. (2021). Comparison on some modified confidence intervals for estimating the Process Capability Index Cp: Simulation and Application. International Journal of Statistical Sciences, 21(2), 145-166.

[12] Samuel Kotz & Norman L. Johnson (2002) Process Capability Indices—A Review, 1992–2000, Journal of Quality Technology, 34:1, 2-19, DOI: 10.1080/00224065.2002.11980119

[13] Kotz, S. and Lovelace, C.R. (1998), Process Capability Indices in Theory and Practice. Arnold: London.

[14] Kocherlakota, S. and Kocherlakota, K. (1994). Confidence intervals for the process capability ratio based on robust estimators. Communications in Statistics-Theory and Methods. 23(1), 257-276.

[15] Montgomery, D. C. (2020). Introduction to statistical Quality Control-Eighth Edition. Wiley, Hoboken, NJ.

[16] Panichkitkosolkul (2014), Confidence interval for the process capability index Cp based on the bootstrap-t confidence interval for the standard deviation, Metodoloski Zvezki, 11, 79-92.

[17] Panichkitkosolkul (2016), Confidence intervals for the process capability index Cp based on confidence intervals for variance under non-normality, Malaysian Journal of Mathematical Sciences, 10, 101-115.

[18] Porter, L. J. and Oakland, J. S. (1991). Process capability indices-an overview of theory and practice. Quality and Reliability Engineering International. 7, 437-448.

[19] Shi, W. & Kibria, B. M. G. (2007) On some confidence intervals for estimating the mean of a skewed population, International Journal of Mathematical Education in Science and Technology, 38:3, 412-421,

[20] Sindhumol, M. R., Srinivasan, M. M., and Gallo, M. (2016), Robust control charts based on modified trimmed standard deviation and Gini's mean difference, Journal of Applied Quantitative Methods, 11, 18-30.

[21] Tukey, J. W. (1948), Some Elementary Problems of Importance to Small Sample Practice. Human Biology, pp.205-214.

[22] Wright, P. A. (1995). A Process Capability Index Sensitive to Skewness. Journal of Statistical Computation and Simulation 52, 195-203.

[23] Zhang, J. (2010), Conditional confidence intervals of process capability indices following rejection of preliminary tests. Ph.D. Thesis, The University of Texas at Arlington, USA.

Author Contributions:

B M Golam Kibria outlined the paper, developed statistical methodology and edit the paper.
Shipra Banik carried out the simulation study, write the discussion of the simulation results and review the paper.

Report potential sources of funding if there is any
Not applicable