MATHEMATICAL SCIENCES

ROBUST PARAMETRIC MODIFICATIONS OF THE Z-TEST STATISTIC

Cvetkov Vasil
Chief Assist. Prof. PhD Eng.
University of Architecture, Civil Engineering and Geodesy.
1164, 1 Hristo Smirnenski Blvd.
Sofia, Bulgaria
ORCİD: 0000-0001-9628-6768

Abstract

The conventional z-test statistic, which is one of the most popular statistics, is based on the mean of a sample and the standard error of the mean. Consequently, in case of a violation of the normality of the data, the traditional z-test may lead to incorrect test conclusions. The main aim of this article is to present two robust parametric modifications of the traditional z-test statistic.

In order to minimize the effect of non-normality due to the presence of the outliers and some potential contaminated observations in a sample we use either the center or the C_2 statistic and their standard errors, respectively, instead of the mean and the standard error of the mean. The statistical power of one-sample z-tests based on the mean, center and C_2 statistics were compared by generating of random number samples with different sizes and known expectations. These samples were derived from some popular distributions. The comparison shows that the z-tests based on the center and the C_2 statistics are more powerful and efficient than the traditional one.

In addition, real-data illustrations by implementing both one-sample z-test and two-sample z-test are also provided.

Keywords: one sample z-test; two sample z-test; type II error; statistical power; systematic error in precise levelling

Introduction

The traditional z-test statistic (1) is one of the most widely used statistical methods.

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad (1)$$

Statistic (1) is a quite accurate, effective and reliable when there is no lack of normality of the data. However, if the distribution of the analysed data is not normal, statistic (1) is likely to lead to incorrect conclusions. The main reason for this is the expanded value of the standard deviation σ in the denominator of (1). In addition, the mean $\bar{X}$ is strongly affected by the observations in the tails of the distribution, which probability of appearance is overweighted contrary to these observations which are close to the median. Due to this drawback of the traditional z-test statistic the statistical power of a hypothesis test decreases, i.e. the obtained type II error of the test is greater. That is to say, it is more likely to incorrectly accept the null hypothesis H_0 when an alternative hypothesis H_1 is true.

In order to strongly minimize or even fix this disadvantage of the conventional z-test statistic we can redefine (1) in (2) under $H_0 : \mu=\mu_0$.

$$Z_c = \frac{\check{X}-\mu_0}{\sigma_{\check{X}}} \quad (2)$$

In equation (2) by the $\check{X}$ are denoted either the center or C_2 statistic. The notation $\sigma_{\check{X}}$ means the standard error of the used statistic in the nominator. How both the center and C_2 statistics can be calculated, one can find in the study [1]. The proves, that the above statistics are unbased and more effective than the means are given in [2], where is also shown that (2) is a pivotal quantity and converges to a normal distribution. What is more, as the sample size $n \to \infty$ then the center statistic converges to C_2 statistic.

The main objective of the current article is to demonstrate the higher statistical power of (2) over (1).

Simulations and Results

In order to compare the statistical power of both (1) and (2) four different distributions are used, e.g. binomial, normal, gamma and uniform. Random samples with different sizes based on these distributions are generated 1000 times and the type II error for each iteration is calculated separately for the mean, the center and C_2 statistic. Based on these results the statistical power of (1) and (2) is obtained. All tests are performed with the same significance level α=0.05 under the null hypothesis $H_0 : \mu=\mu_0$. The alternative hypothesis is $H_1 : \mu\neq\mu_0$. The parameres of each distribution, the results concerning the distibutions and specific conditions as μ_0 are given in the titles of Table 1, Table 2, Table 3 and Table 4 below.

Table 1

The power of a One Sample Z-Test with use of a B (n, p = 0.9), $\mu_0 = np\text{-}npq/20$, $\alpha = 0.05$

n	Mean	Center	C_2
30	0.205	0.340	0.353
40	0.222	0.384	0.394
50	0.228	0.398	0.407
60	0.257	0.445	0.452
70	0.284	0.482	0.488
80	0.313	0.528	0.533
90	0.340	0.550	0.555
100	0.385	0.612	0.617
200	0.746	0.916	0.917
300	0.937	0.996	0.997
400	0.974	1.000	1.000
500	0.991	1.000	1.000

Table 2

The power of a One Sample Z-Test with use of a Gamma(r = 2, θ = 1), $\mu_0 = 1.91$, $\alpha = 0.05$

n	Mean	Center	C_2
30	0.185	0.312	0.327
40	0.179	0.311	0.324
50	0.179	0.299	0.310
60	0.181	0.316	0.326
70	0.188	0.319	0.328
80	0.186	0.324	0.333
90	0.202	0.333	0.338
100	0.188	0.341	0.349
200	0.249	0.387	0.392
300	0.280	0.457	0.461
400	0.308	0.502	0.506
500	0.323	0.585	0.588

Table 3

The power of a One Sample Z-Test with use of a N ($\mu = 0$, $\sigma = 1$), $\mu_0 = 0.2$, $\alpha = 0.05$

n	Mean	Center	C_2
30	0.300	0.437	0.449
40	0.341	0.478	0.486
50	0.366	0.517	0.524
60	0.394	0.543	0.548
70	0.440	0.596	0.601
80	0.440	0.616	0.620
90	0.496	0.661	0.664
100	0.525	0.691	0.694
200	0.718	0.855	0.856
300	0.856	0.941	0.942
400	0.927	0.977	0.978
500	0.964	0.993	0.993

Table 4

The power of a One Sample Z-Test with use of a U (a = 0, b = 3.464), $\mu_0 = 1.93$, $\alpha = 0.05$

n	Mean	Center	C_2
30	0.285	0.413	0.422
40	0.324	0.444	0.450
50	0.361	0.487	0.492
60	0.385	0.503	0.505
70	0.441	0.562	0.565
80	0.459	0.573	0.575
90	0.476	0.581	0.583
100	0.497	0.602	0.604
200	0.726	0.795	0.795
300	0.841	0.880	0.880
400	0.901	0.928	0.930
500	0.959	0.971	0.971

Graphical representations of the results given in Table 1, Table 2, Table 3 and Table 4 are shown in Figure 1 below.

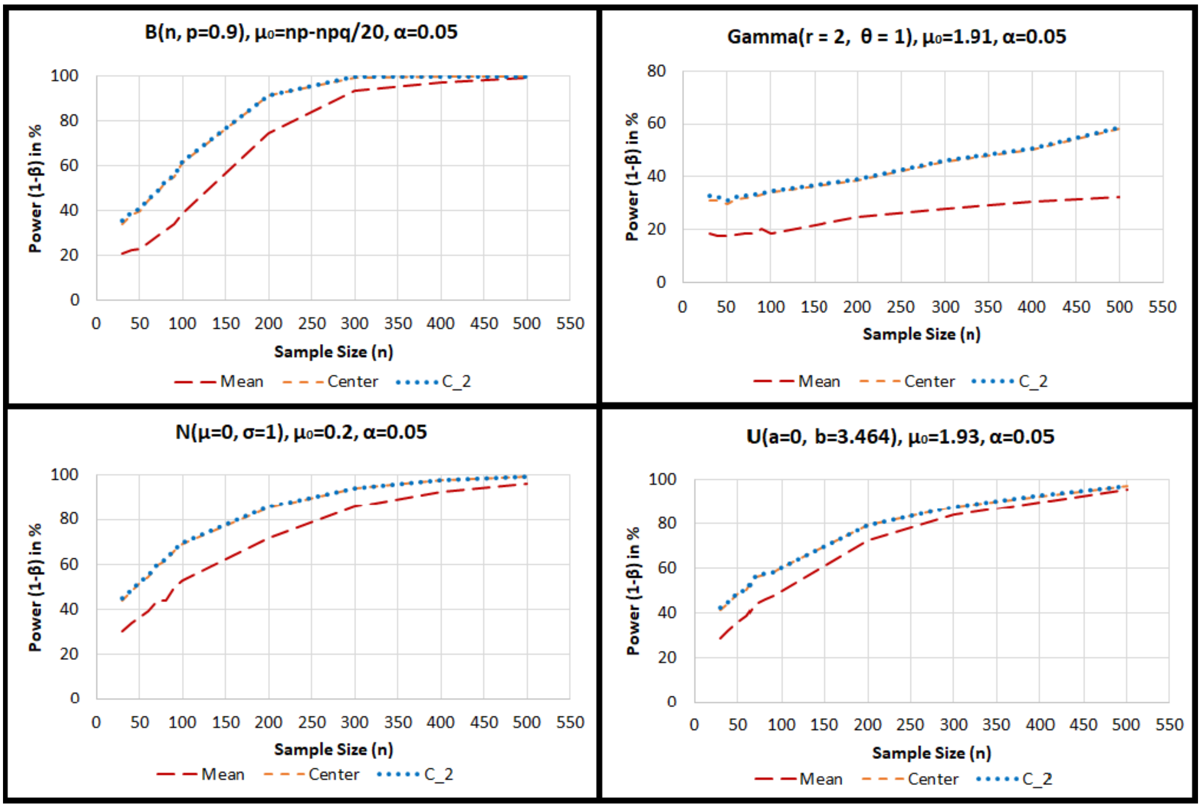

Figure 1: Graphs of the data given in Tables 1-4.

According to the results, which are illustrated above, the following conclusions can be made:

- Even for symmetric distributions as the normal, binomial and uniform the power efficiency of (2) over (1) is almost 0.50%. In case of strongly skewed distribution as the Gamma (2, 1) the power of (2) is approximately 10 times greater than that of (1).
- Apart from the kind of distribution, the statistical power of z-test statistic based on the center approximates these of C_2 statistic, especially when the size of a sample is greater than 100.

A real-data One-Sample Z-Test Example

In order to illustrate the use of (2) with real data let us conider the data given in Sample 1 below.

Sample 1

-0.376	0.046	-0.265	-0.956	0.959	-0.741	0.391	-0.604	-0.364	0.298
-1.068	-1.096	0.498	-0.478	-0.866	0.314	0.836	0.695	0.087	1.093
1.624	0.088	0.125	0.548	0.421	0.468	-1.003	-0.660	1.265	-0.831
0.464	-0.830	0.710	0.386	1.572	0.614	0.392	1.082	0.535	0.813
0.267	0.247	-0.180							

The real numbers in Sample 1 are discrepancies in the sections of the line 2 in the Third Levelling of Bulgaria divided by the length of the corresponding sections. The expectation of these discrepancies is zero, due to the fact that the levelling of each section starts and finishes in the same point and the orthometric height of the point is supposed to be unchangeable during the levelling. Consequently, our test $\mu_0 = 0$. We want to know whether there is any systematic effect in our measurements. We are also interested in if the systematic error is significant on a predefined confidence level $\alpha = 0.05$.

So, we form our null hypothesis $H_0 : \check{X} = \mu_0$ against the alternative hypothesis $H_1 : \check{X} \neq \mu_0$. Some impression about the distribution of the standardized discrepancies one can get from the bihistogram given in Figure 2. The results obtained by one-sample z-test by use of the (1) and (2) are given in Table 5.

Table 5

One-Sample Z-Test Results concerning Sample 1, μ_0=0

Description	Notation	Mean	Center	C_2
Expected Value	$\breve{X}$	0.1516	0.1968	0.1986
Standard Deviation	σ	0.7291	0.5522	0.5442
Standard Error	$\sigma_{\breve{X}}$	0.1112	0.0842	0.0830
Sample Size	n		43	
Z(two-tail, α=0.05)	$Z_{crit.}$		1.96	
Observed Z	$Z_{obs.}$	1.3633	2.3373	2.3928
p-Value($Z_{obs.}$, two-tail)	p-Value	0.1728	0.0194	0.0167
Type II Error	β	0.7241	0.3529	0.3327
Power	1-β	0.2759	0.6471	0.6673

According to the results shown in Table 5, the one-sample z-test based on (1) produced a mean $\bar{X}$ = 0.1516 and $Z_{obs.}$ = 1.3633. The critical z-test statistic Z (two-tail, α=0.05) = 1.96, which is greater than $Z_{obs.}$ = 1.3633. Consequently, the traditional one-sample z-test did not reject the null hypothesis at confidence level α=0.05. In other words, we can claim at 95 % confidence that the mean $\bar{X}$ = 0.1516 is not significantly different from 0 or there is not any significant systematic error in the analyzed discrepancies in line 2.

On the other hand, the results based on (2) by the use of either the center or C_2 statistic show a different picture. Based on their results, the null hypothesis must be rejected due to the fact that $Z_{obs.}$ = 2.3373 > Z (two-tail, α=0.05) = 1.96 and $Z_{obs.}$ = 2.3928 > Z (two-tail, α=0.05) = 1.96 for the center and C_2 statistic, respectively. According to these results, we can conclude that there is a significant systematic error in the discrepancies in the sections in line 2. The systematic error can be assessed to be equal to 0.20 mm/km and its standard error is 0.08 mm/km.

According to the produced type II errors, we should put trust in the results based on (2), the center and C_2 statistics. One can see that the power of (2) is approximately 2.3 times greater than that of (1). More information on how type II error is calculated, one can find in [5].

A real-data Two-Sample Z-Test Example

The previous section presented a hypothesis test for a single population parameter. This section extends the use of the modified z-test statistic, based on either the center or C_2 statistic, to the case of two independent populations. The real numbers in Sample 2 are discrepancies in the sections of line 8 in the Third Levelling of Bulgaria divided by the length of the corresponding sections. The expectation of these discrepancies is zero, due to the same reason given above for the discrepancies in line 2.

Sample 2

0.845	0.727	-0.746	-0.019	1.060	0.604	0.563	0.596	-0.042	0.766
-0.543	-0.678	0.378	-0.352	-1.394	-2.484	-0.623	0.328	-0.175	-0.296
0.246	-0.516	-0.771	-0.961	-0.079	-1.558	0.268	-0.285	0.018	0.073
-0.477	-0.531	-0.231	0.422	-0.612	2.460	-0.542	0.122	0.237	0.354
-0.556	0.772	-0.863	0.247	0.895	-0.021	-0.864	-0.342	-0.697	-0.432

Our task is to test whether the discrepancies in both lines 2 and 8 come from the same population with the location parameter μ=0. Therefore, our null hypothesis H_0 will be given as $H_0 : \mu_1 - \mu_2 = 0$, and the alternative hypothesis H_1 will be defined as $H_1 : \mu_1 - \mu_2 \neq 0$. Consequently, the two-sample z-test statistic can be written as (3).

$$Z_c = \frac{\breve{X}_1 - \breve{X}_2}{\sqrt{\sigma^2_{\breve{X}_1} + \sigma^2_{\breve{X}_2}}} \qquad (3)$$

In equation (3), the centers or C_2 statistics of both analyzed samples are denoted by $\breve{X}_1$, $\breve{X}_2$, and their standard errors are $\sigma_{\breve{X}_1}$ and $\sigma_{\breve{X}_2}$, respectively. A graphical comparison between Sample 1 and Sample 2 is done by the bihistogram pictured in Figure 2. The two-sample z-test results concerning the means, centers and C_2 statistics are given in Table 6.

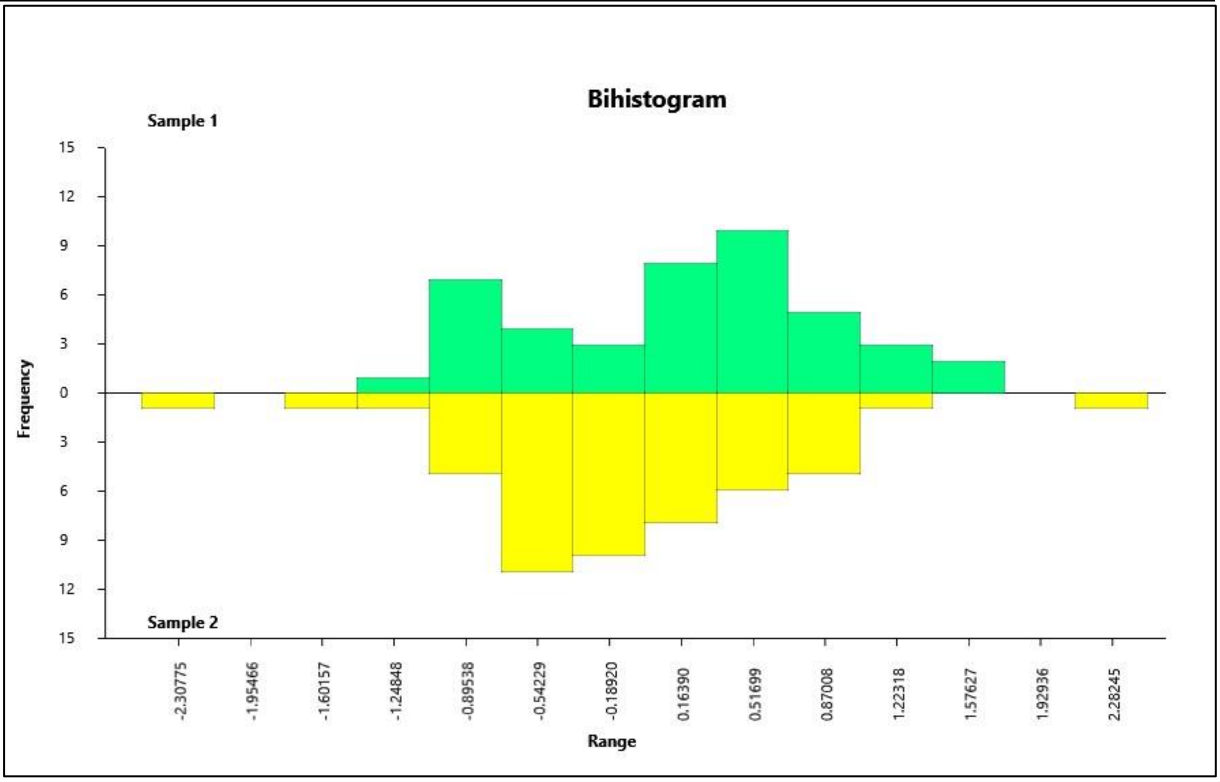

Figure 2: Bihistogram of Sample 1 and Sample 2.

Table 6

Two-Sample Z-Test Results concerning Sample 1 and Sample 2, Δ_0=0

Description	Notation	Mean	Center	C_2
Location 1	$\check{X}_1$	0.1516	0.1968	0.1985
Location 2	$\check{X}_2$	-0.1141	-0.1250	-0.1253
Difference in Locations	$\check{X}_1$-$\check{X}_2$	0.2657	0.3218	0.3239
Standard Error of $\check{X}_1$	$\sigma_{\check{X}_1}$	0.1112	0.0842	0.0830
Standard Error of $\check{X}_2$	$\sigma_{\check{X}_2}$	0.1103	0.0749	0.0736
Z(two-tail, α=0.05)	Zcrit.		1.96	
Observed Z	Zobs.	1.6964	2.8556	2.9196
p-Value($Z_{obs.}$, two-tail)	p-Value	0.0898	0.0043	0.0035
Type II Error	β	0.6038	0.1852	0.1686
Power	1-β	0.3962	0.8148	0.8314

According to the results given in Table 6, the traditional two-sample z-test based on the means did not reject the null hypothesis at confidence level α=0.05, owing to the result of $Z_{obs.}$ = 1.3633 < Z (two-tail, α=0.05) = 1.96. In other words, we can claim at 95 % confidence that the means of both samples 1 and 2 are equal. Moreover, there are no significant systematic errors in the discrepancies of the sections in both lines at used confidence level.

In contrast, the results based on either the center or C_2 statistic shows a different picture. According to these results, the null hypothesis must be rejected, due to the fact that $Z_{obs.}$ = 2.8556 > Z (two-tail, α=0.05) = 1.96 and $Z_{obs.}$ = 2.9196 > Z (two-tail, α=0.05) = 1.96 for the center and C_2 statistic, respectively. According to the results, there is a significant difference in the systematic errors of the section discrepancies in both lines 2 and 8, which even reflect on the different sign of the assessment of the expectations of both samples.

Analyzing Table 6, one can see that the power of the tests based on both the center and C_2 statistics is more than twice greater than that of the traditional variant concerning samples 1 and 2. More information how type II error is calculated, can be found in the book [5]. Additional comparison between traditional two-sample z-test and the proposed here robust variant based on a center statistic are given in [3]. In order to facilitate their calculations concerning z-tests, anyone is invited to test CBSTAT software [4].

Conclusions

In this paper, two more robust modifican of the conventional z-test were proposed. Both statistics the center and C_2 were used in these modifications, instead of the mean. The standard errors of the center and C_2 staristics, respectively were used, instead of the standard error of the sample mean. Through Monte Carlo simulations the statistical power of the traditional z-test and the proposed modifications were compared. It has been shown that the proposed methods overperform the classical one. The simulation results were fully supported by the real data, which were obtained in the Third Levelling of Bulgaria.

In order to use z-test, however, the sample size should be greater than 30 [2] or even greated than 40 [5] . When this requirement does not meet, one should use the robust modifications of t-test based on the center and C_2 statistics. The robust modifications of the one-sample t-test, the paired two-sample t-test and the independent two-sample t-test are going to be presented in future publications.

References

1. Cvetkov V. (2022). ALTERNATIVE MEASURES OF CENTRAL TENDENCY. Deutsche Internationale Zeitschrift Für Zeitgeössische Wissenschaft, 38, 4-8. https://doi.org/10.5281/zenodo.7002877 .

2. Cvetkov V. (2023). THE CENTRAL LIMIT THEOREM AND THE MEASURES OF CENTRAL TENDENCY. Deutsche Internationale Zeitschrift Für Zeitgenössische Wissenschaft, 49, 14-21. https://doi.org/10.5281/zenodo.7594733.

3. https://center-based-statistics.com/html/zTest.html , (visited on 14.03.2023)

4. https://center-based-statistics.com/html/download.php, (visited on 14.03.2023)

5. Montgomery, Douglas C.; Runger, George C. (2014). Applied Statistics and Probability for Engineers (6th ed.). Wiley. ISBN-13 9781118539712.