Statistics,
the Art of Mathematics
CENTRALITY TEST FOR
NON-NORMAL AND SKEWED POPULATION

DONNY SURYABUANAPUTRA

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
BOGOR AGRICULTURAL UNIVERSITY
BOGOR
2001
ABSTRACT

DONNY SURYABUANAPUTRA. Centrality test for non-normal and skewed population, with advisory committee ASEF SAEFUDIN and ANWAR FITRIANTO

The distribution of population is usually unknown. Central limit theorem state that if the sample size is large, $x \bar{x}$ will be normally distributed. When dealing with small sample size, with the assumption of normally distributed population and unknown $\sigma^2$, $t$ test is usually used to test the mean. This paper try to study effect of population skewness to t test, t modified 1, t modified 2 and two non-parametric test (Sign and Wilcoxon test). The result is t test can be use to test a sample from symmetric population, not only normal. The t modified 2 test is more robust than t test for skewed population. This test tolerate the population skewness from -1.2 to 1.2 if sample size is greater than 25. The sign test robust to the skewed population, but the value of the estimated $\alpha$ always close to the size of the test. Wilcoxon test is not robust to the population skewness. This test is more suitable for symmetric population

Keywords: t test, t modified test, centrality test, type I error, population skewness
CENTRALITY TEST FOR
NON-NORMAL AND SKEWED POPULATION

DONNY SURYABUANAPUTRA

A paper submitted to the Faculty of Mathematics and Natural Science of Bogor Agricultural University at Bogor
In partial fulfillment of the requirement for the degree of Sarjana Sains
Department of Statistics

DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
BOGOR AGRICULTURAL UNIVERSITY
BOGOR
2001
Title: Centrality Test for Non-Normal and Skewed Population
Name: Donny Suryabuanaputra
NRP: G03497060

Approved by:

[Signature]
Dr. Ir. Asep Saefuddin, MSc
Chair of Advisory Committee

[Signature]
Anwar Putrianto, SSI
Advisory Committee

Acknowledged by:

[Signature]
Dr. Ir. Asep Saefuddin, MSc
Head of Dept. of Statistics

Passed examination date: 30 AUG 2001
BIOGRAPHY

The author was born in Bogor, West Java on June 18th, 1979, as the first of three to Acep Andy Mulya and Enny Trisnawaty.

He was graduated from SD Bhayangkari Bogor in 1991 and SMPN 3 Bogor in 1994. In 1997, he was graduated from SMUN 1 Bogor and invited as a student in Department of Statistics, Bogor Agricultural University.

He was active in organization of statistics student, Gamma Sigma Beta and in 1999-2000 he was active as an instructor in Gamma Sigma Beta Computer Course.
ACKNOWLEDGMENT

All praise to Allah the Most Merciful and Compassionate for His blessing and mercy.
The author gratefully acknowledges Dr. Ir. Asep Saeffudin, MSc and Anwar Fitrianto, SSi as the
supervisors who always support and allocates their valuable time to have discussion in time of need.
For my family, Mama, Bapa, my sister Dewi, my brother Daniel thanks for loving me and encouraged
me all the time. For Dewi, thanks for your support and your love, without you... And for the rest of my
family who always beside me.
To all my friends in Statistics '34, thanks. Arief, you help me a lot. Baby, Ijul, Anik, Ade, Fitri,
Mulyana, Rizal, Wahyudi, Akmal, Fikri, if you read this I hope you always remember our friendship. For
all staff in Department of Statistics, thanks for the cooperation.

Bogor, September 2001

Donny Suryabuanaputra
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE</td>
<td>vii</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>Objective</td>
<td>1</td>
</tr>
<tr>
<td>THEORETICAL OVERVIEW</td>
<td>1</td>
</tr>
<tr>
<td>Beta Distribution</td>
<td>1</td>
</tr>
<tr>
<td>Gamma Distribution</td>
<td>1</td>
</tr>
<tr>
<td>Chi-square Distribution</td>
<td>2</td>
</tr>
<tr>
<td>Sign Test</td>
<td>2</td>
</tr>
<tr>
<td>Wilcoxon Test</td>
<td>2</td>
</tr>
<tr>
<td>T Test</td>
<td>2</td>
</tr>
<tr>
<td>T Modified 1</td>
<td>2</td>
</tr>
<tr>
<td>T Modified 2</td>
<td>2</td>
</tr>
<tr>
<td>Skewness</td>
<td>2</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>5</td>
</tr>
<tr>
<td>ILLUSTRATION</td>
<td>3</td>
</tr>
<tr>
<td>Data</td>
<td>3</td>
</tr>
<tr>
<td>Method</td>
<td>3</td>
</tr>
<tr>
<td>RESULT</td>
<td>4</td>
</tr>
<tr>
<td>CONCLUSION AND RECOMMENDATION</td>
<td>6</td>
</tr>
<tr>
<td>REFERENCE</td>
<td>6</td>
</tr>
</tbody>
</table>
TABLE

Table 1. The critical value (X) of Sign Test for α=0.05 and P (X≤x | n, 0.05) ......................... 5
Table 2. The model of population skewness (X) and estimated α (Y) for every sample size .......... 5
Table 3. The value of CV and R² for every model ................................................................. 5
Table 4. The interval of population skewness that can be tolerated by the t Test ................. 6

APPENDIX

Appendix 1. Simulation result with H₀:μ=μ₀, H₁:μ>μ₀ and Beta Distributed ..................... 7
Appendix 2. Simulation result with H₀:μ ≥μ₀ and H₁:μ<μ₀ and Beta Distributed ............... 8
Appendix 3. Simulation result with H₀:μ≤μ₀, H₁:μ>μ₀ and Gamma, Chi-square Distributed .... 9
Appendix 4. The design for sample size, distribution and the value of parameter .................... 10
Appendix 5. Graphic of simulation result .............................................................................. 11
Appendix 6. Graphic to compare all t Test ........................................................................... 12
Appendix 7. Residual analysis for every regression model ..................................................... 15
INTRODUCTION

Background

When dealing with statistical data, usually the distribution of the population is unknown. The central limit theorem states that if the sample size \( n \) is large, the sample distribution of 0 from any distribution will be approximately normal with mean \( \mu \) and standard deviation \( \sigma^2/n^{0.5} \) where \( \mu \) and \( \sigma^2 \) are mean and variance of population. But if the sample size is small then this theorem can't be applied. In order to test the mean from normal population with small sample size we usually use t test.

Clarke (1994) suggest to use t test for symmetrical population or nearly symmetrical, not only for normal population. Ling Chen (1995) said that population skewness effect the distribution of \( t \), more than kurtosis. Johnson (1978) made a modification for t test (t modified 1). This modification makes t test more robust to test a random sample from asymmetric population. Another modification for t test (t modified 2) was built by Ling Chen (1995) in order to make it more accurate.

There are other ways to test the centrality, like using non-parametric statistic procedure. The assumptions that have to be fulfilled in this procedure are more flexible. Some of the non-parametric statistic procedures, which used to test the centrality, are sign test and Wilcoxon rank test. Decision for sign test based on binomial distribution and the decision Wilcoxon rank test base on a unique distribution which can be approximate by normal distribution if the sample size is large (Daniel, 1989).

The purpose of the paper is to see the robustness of the five procedure (t test, t modified 1, t modified 2, Sign test and Wilcoxon test) with respect to non-normal assumption. This non-normal assumption may have a different shape. There is a possibility that the different distribution shape is unsuitable with the assumed distribution, which base on the test. If there is an unsuitability, then the possibility of rejecting a hypothesis will be unequal with the stated \( \alpha \) (P(reject \( H_0 \mid H_0 \) true)). This procedure will result in an error on decision making. The level of rejecting \( H_0 \), which is larger, than \( \alpha \) will cause the tendency of the test to reject \( H_0 \), while the level of rejecting \( H_0 \) which is smaller than \( \alpha \) will cause the tendency of the test to accept \( H_0 \). The best test is the test, which capable of rejecting \( H_0 \) appropriate with the stated \( \alpha \), because it shows that the statistic distribution base fit in with the distribution of it statistic.

Objective

The aims of this study are:
1. To know the robustness of t Test to test a sample from symmetric (non-normal) and skewed population.
2. To see the effectiveness of t modified 1, t modified 2, sign and Wilcoxon Test to test a sample from skewed population.

The robustness and the effectiveness of the test can be showed by the suitability of the rejected \( H_0 \) compared to the stated \( \alpha \).

THEORETICAL OVERVIEW

Beta Distribution

The beta density function with parameter \( \beta_1 \) and \( \beta_2 \) is

\[
f(x) = \frac{1}{B(\beta_1, \beta_2)} x^{\beta_1-1} (1-x)^{\beta_2-1}
\]

where

\[
0 \leq x \leq 1 \\
\beta_1, \beta_2 > 0
\]

\[
B(\beta_1, \beta_2) = \int_0^1 x^{\beta_1-1} (1-x)^{\beta_2-1} dx
\]

If the parameter \( \beta_1=\beta_2=1 \) then it will be uniformly distributed (0,1) (Nasoetion & Rambe, 1984) and if \( \beta_1=\beta_2 \) and more than one, it will be symmetric. But if the value of \( \beta_1=\beta_2 \) then it will have an asymmetric form.

Gamma Distribution

The Gamma density function is:

\[
f(x) = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{-x/\beta}
\]

With \( 0 < x < \infty \) and \( \alpha, \beta > 0 \).

The \( \Gamma(\alpha) \) is:

\[
\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx ; \quad \alpha > 0
\]

A random variable \( X \) from Gamma distribution will be wrote as \( X \sim \text{Gamma}(\alpha, \beta) \).