APPROPRIATE STATISTICAL ANALYSIS FOR TWO INDEPENDENT GROUPS

OF LIKERT-TYPE DATA

By

Boonyasit Warachan

Submitted to the

Faculty of the College of Arts and Sciences

of American University

in Partial Fulfillment of

the Requirements for the Degree

of Doctor of Philosophy

In

Mathematics Education

Chair:

____
Mary Gray, Ph.D.

____
Monica Jackson, Ph.D.

____
Jun Lu, Ph.D.

____
Dean of the College of Arts and Sciences

____
Date

2011
American University
Washington, D.C. 20016

DEDICATION

This dissertation is dedicated to my beloved grandfather, Sighto Chankomkhai.

APPROPRIATE STATISTICAL ANALYSIS FOR TWO INDEPENDENT GROUPS OF LIKERT-TYPE DATA

BY

Boonyasit Warachan

ABSTRACT

The objective of this research was to determine the robustness and statistical power of three different methods for testing the hypothesis that ordinal samples of five and seven Likert categories come from equal populations. The three methods are the two sample t-test with equal variances, the Mann-Whitney test, and the Kolmogorov-Smirnov test. In additional, these methods were employed over a wide range of scenarios with respect to sample size, significance level, effect size, population distribution, and the number of categories of response scale. The data simulations and statistical analyses were performed by using R programming language version 2.13.2. To assess the robustness and power, samples were generated from known distributions and compared. According to returned p-values at different nominal significance levels, empirical error rates and power were computed from the rejection of null hypotheses.

Results indicate that the two sample t-test and the Mann-Whitney test were robust for Likert-type data. Also the t-test performed the best to control of Type I error for both 5-point and 7-point Likert scale. For the Kolmogorov-Smirnov test, the Type I error rate

was not controlled for all circumstances. This means the testing procedure computed from R was not robust for the ordinal Likert-type data because the Type I error rate of this test was lower than the minimum of the robustness criteria. Therefore, the Kolmogorov-Smirnov test was quite conservative. For the statistical power of the test, the Mann-Whitney test was the most powerful for most of the distributions, especially under highly skewed and bimodal distributions. The t-test obtained high statistical power or close to the power from the Man-Whitney test under the uniform, moderate skewed or symmetric distribution with large location shift.

ACKNOWLEDGMENTS

Its accomplishment is owed to my loving family: my parents Jomnian and Somchai, my wife Nisagorn, and my children Chayanit and Chayanon.

I would like to acknowledge and extend my heartfelt gratitude to Dr. Mary Gray for her support, patience, understanding, and encouragement which has made the completion of this dissertation possible.

I am thankful for the contributions of dissertation committee members Dr. Monica Jackson, and Dr. Jun Lu.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the dissertation.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

Chapter

1. INTRODUCTION 1

Statement of the problem 1

The purpose of the study 2

Likert-type data 2

Level of measurement 3

Testing for two independent groups 5

Research questions 6

Limitation of the study 7

2. LITERATURE REVIEW 8

Treating ordinal as interval data 8

Parametric and nonparametric procedures 9

Problem in statistical procedure's assumtions 10

Studies of the robustness and power 11

Likert-type data in data in education 12

3. RESEARCH METHODOLOGY.. 14

Simulation of population distributions... 14

Sample sizes... 18

Robustness and significance levels.. 19

Power and levels of effect size... 19

The number of Likert categories and the tests... 19

4. RESULTS 21

Robustness of the statistical tests... 21

Relative power of the statistical tests.. 21

5. CONCLUSION AND RECOMMENDATIONS 87

Conclusion 87

Recommendations... 88

Directions for the future research 88

APPENDIX A: R code for the Type I error rate.. 89

APPENDIX B: R code for the statistical power... 91

REFERENCES 93

LIST OF TABLES

Table

1. Five population distribution shapes of the 5-point response scale 17

2. Five population distribution shapes of the 7-point response scale 17

3. Type I error rate estimates of the uniform distribution 22

4. Statistical power estimates of the uniform distribution at α = 0.01 29

5. Statistical power estimates of the uniform distribution at α = 0.05 31

6. Statistical power estimates of the uniform distribution at α = 0.10 33

7. Type I error rate estimates of the moderate skewed distribution.......................... 35

8. Statistical power estimates of the moderate skewed distribution at α = 0.01. 42

9. Statistical power estimates of the moderate skewed distribution at α = 0.05 44

10. Statistical power estimates of the moderate skewed distribution at α = 0.10…....46

11. Type I error rate estimates of the highly skewed distribution 48

12. Statistical power estimates of the highly skewed distribution at α = 0.01............ 55

13. Statistical power estimates of the highly skewed distribution at α = 0.05............ 57

14. Statistical power estimates of the highly skewed distribution at α = 0.10............ 59

15. Type I error rate estimates of the unimodal symmetric distribution..................... 61

16. Statistical power estimates of the unimodal symmetric distribution at α = 0.01 .. 68

17. Statistical power estimates of the unimodal symmetric distribution at α = 0.05...70

18. Statistical power estimates of the unimodal symmetric distribution at α = 010....72

19. Type I error rate estimates of the bimodal symmetric distribution………………74

20. Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.01$ 81

21. Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.05$ 83

22. Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.10$ 85

LIST OF FIGURES

Figure

1. Five marginal distributions of 5-point response scale. 16

2. Five marginal distributions of 7-point response scale. 16

3. Type I error rate estimates of the uniform distribution at $\alpha = 0.01$ for the 5-point Likert scale…………………………………………………….23

4. Type I error rate estimates of the uniform distribution at $\alpha = 0.01$ for the 7-point Likert cale………………………………………………….... .. 24

5. Type I error rate estimates of the uniform distribution at $\alpha = 0.05$ for the 5-point Likert scale…………………………………………………….25

6. Type I error rate estimates of the uniform distribution at $\alpha = 0.05$ for the 7-point Likert scale…………………………………………………….26

7. Type I error rate estimates of the uniform distribution at $\alpha = 0.10$ for the 5-point Likert scale…………………………………………………….27

8. Type I error rate estimates of the uniform distribution at $\alpha = 0.10$ for the 7-point Likert scale…………………………………………………….28

9. Statistical power estimates of the uniform distribution at $\alpha = 0.01$. 30

10. Statistical power estimates of the uniform distribution at $\alpha = 0.05$. 32

11. Statistical power estimates of the uniform distribution at $\alpha = 0.10$. 34

12. Type I error rate estimates of the moderate skewed distribution at α = 0.01 for the 5-point Likert scale……………………………………………………….36

13. Type I error rate estimates of the moderate skewed distribution at α = 0.01 for the 7-point Likert cale…………………………………………………... .. 37

14. Type I error rate estimates of the moderate skewed distribution at α = 0.05 for the 5-point Likert scale……………………………………………………….38

15. Type I error rate estimates of the moderate skewed distribution at α = 0.05 for the 7-point Likert scale……………………………………………………….39

16. Type I error rate estimates of the moderate skewed distribution at α = 0.10 for the 5-point Likert scale……………………………………………………….40

17. Type I error rate estimates of the moderate skewed distribution at α = 0.10 for the 7-point Likert scale……………………………………………………….41

18. Statistical power estimates of the moderate skewed distribution at α = 0.01. 43

19. Statistical power estimates of the moderate skewed distribution at α = 0.05. 45

20. Statistical power estimates of the moderate skewed distribution at α = 0.10……47

21. Type I error rate estimates of the highly skewed distribution at α = 0.01 for the 5-point Likert scale……………………………………………………….49

22. Type I error rate estimates of the highly skewed distribution at α = 0.01 for the 7-point Likert cale…………………………………………………... .. 50

23. Type I error rate estimates of the highly skewed distribution at α = 0.05 for the 5-point Likert scale……………………………………………………….51

24. Type I error rate estimates of the highly skewed distribution at α = 0.05

for the 7-point Likert scale……………………………………………………….52
25. Type I error rate estimates of the highly skewed distribution at α = 0.10
for the 5-point Likert scale……………………………………………………….53
26. Type I error rate estimates of the highly skewed distribution at α = 0.10
for the 7-point Likert scale……………………………………………………….54
27. Statistical power estimates of the highly skewed distribution at α = 0.01........... 56
28. Statistical power estimates of the highly skewed distribution at α = 0.05........... 58
29. Statistical power estimates of the highly skewed distribution at α = 0.10…..…...60
30. Type I error rate estimates of the unimodal symmetric distribution at α = 0.01
for the 5-point Likert scale……………………………………………………….62
31. Type I error rate estimates of the unimodal symmetric distribution at α = 0.01
for the 7-point Likert cale……………………………………………………... .. 63
32. Type I error rate estimates of the unimodal symmetric distribution at α = 0.05
for the 5-point Likert scale……………………………………………………….64
33. Type I error rate estimates of the unimodal symmetric distribution at α = 0.05
for the 7-point Likert scale……………………………………………………….65
34. Type I error rate estimates of the unimodal symmetric distribution at α = 0.10
for the 5-point Likert scale……………………………………………………….66
35. Type I error rate estimates of the unimodal symmetric distribution at α = 0.10
for the 7-point Likert scale……………………………………………………….67
36. Statistical power estimates of the unimodal symmetric distribution at α = 0.01. . 69
37. Statistical power estimates of the unimodal symmetric distribution at α = 0.05. . 71
38. Statistical power estimates of the unimodal symmetric distribution at α = 0.10...73

39. Type I error rate estimates of the bimodal symmetric distribution at $\alpha = 0.01$ for the 5-point Likert scale……………………………………………………….75

40. Type I error rate estimates of the bimodal symmetric at $\alpha = 0.01$ for the 7-point Likert cale……………………………………………………... .. 76

41. Type I error rate estimates of the bimodal symmetric distribution at $\alpha = 0.05$ for the 5-point Likert scale……………………………………………………….77

42. Type I error rate estimates of the bimodal symmetric distribution at $\alpha = 0.05$ for the 7-point Likert scale……………………………………………………….78

43. Type I error rate estimates of the bimodal symmetric distribution at $\alpha = 0.10$ for the 5-point Likert scale……………………………………………………….79

44. Type I error rate estimates of the bimodal symmetric distribution at $\alpha = 0.10$ for the 7-point Likert scale……………………………………………………….80

45. Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.01$…... 82

46. Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.05$…... 84

47. Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.10$….86

CHAPTER 1

INTRODUCTION

Statement of the problem

In education, psychology, and social science study, survey research is often conducted to measure subjective feelings, attitudes, or opinions. The rating measurement that is widely integrated with survey questionnaires is the Likert rating scale. This rating scale typically contains ordinal multiple choices. Five-point Likert scales are probably most commonly used and can be coded as 1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, and 5=Strongly Agree. Thus, the labeling data should be described and analyzed with percentage and nonparametric statistics only. However, there are many researchers who agree to apply parametric statistical methods which treat Likert-type data as if it were interval scale data (Brown, 2000; Cliff, 1984; Hodgson, 2003).

There has been no clear conclusion until now as to whether the ordinal scale data can assumed to be interval scale data and thus subject to the use of higher level statistical methods to analyze it (Acock & Martin, 1974). An ordinal rating scale such as the Likert scale consists of discrete numbers that can be ranked from low to high rate with no continuous distances between any two adjacent numbers. The interval distances between any two values of ordinal data are necessary for calculating statistics including mean, standard deviation, correlation coefficient, etc. Therefore, any tests of hypotheses that require these statistics might not be appropriate with ordinal data (Miller, 1998).

Moreover, the mistreatment of data and the inappropriate analysis may lead some to question the accuracy of conclusions from ordinal data (Harwell & Gatti, 2001). Many researchers practically compare the means of Likert-type data from each question by using the familiar Student's t-test instead of the non-parametric such as the Mann-Whitney test, Kolmogorov-Smirnov test, and so on. Using the Student's t-test with Likert-type data, they should be more concerned about the data distribution, sample size, or the number of rating choices in order to avoid the pitfalls of the test.

The purpose of the study

The purpose of this research is to determine the robustness and the relative power of three statistical procedures for two independent ordinal samples of five-point and seven-point Likert-type data in simulated various scenarios and then consider which procedure is appropriate for each condition. The three methods to be implemented are the two sample t-test (Student, 1908), the Mann-Whitney test (Mann & Whitney, 1947; Wilcoxon, 1945), and the Kolmogorov-Smirnov test (Gibbons & Chakraborti, 1992).

Likert-type data

Measurement on a continuous scale is sometimes not available; in particular for those variables concerning feelings, attitudes, or opinions; therefore, researchers create rating instruments according to ordered categories. Thus, one can describe feelings, attitudes or opinions. Rensis Likert's dissertation created a new attitude-scaling technique from a survey of student attitudes. He published his study in 1932. This technique is to present a statement for response with one of five given choices such as strongly disagree, disagree, neutral, agree, and strongly agree. He recommended assigning numerical values one through five for these multiple choices for data analysis purpose. Also he reported very satisfactory reliability for the rating scale developed with this technique (Likert, 1932). The Likert-type attitude scale is considered to be a quite reliable and valid instrument for attitude measurement (Arnold, McCrosby, & Prichard, 1967). Currently the Likert attitude scale has been applied to various fields of study**,** and researchers still confirm its reliability and validity (Abdel-Khalek, 1998; Chow & Winzer, 1992; Maurer & Andrews, 2000).

Moreover, research has shown that the variance and the reliability of rating is normally highest when 5 or 7 point rating scales are used and rater bias is minimized when 5 rating points or above are used (Stennet, 2002). The Likert-scale**s** used by researchers employ at least 5 and preferably 7 categories. Most of them would not use a 3-point or 4-point scale because of the departure from the assumption of normal distribution required for parametric statistical methods (Garson, 2002). Also comparing each of Likert-type item between groups instead of summative form is involving with the lack of normality.

Level of measurement

According to Stevens' theory of measurement scale**s** (Stevens, 1946), research data can be classified into four different measuring scales**;** ranging from lowest to highest these are nominal, ordinal, interval, and ratio. Each of them contains different information that will determine the statistical analysis method. The first scale is the nominal scale. In this scale, the value of data can be assigned by name or number in order to identify it correctly. Nominal numbers imply nothing about the ordering. For example, the numbers on jerseys of basketball players can be designed to identify who are guards, the captain, and so on. Player's number "17" does not necessarily indicate that he performs better than a player whose number is "5". An additional example is Social Security numbers; they are assigned for identification purposes only. It does not mean that one is better or worse than another because the number is higher. However, researchers rarely make mistakes with nominal numbers.

The ordinal scale is the second. Unlike the nominal scale, it has an additional feature because one can tell which value is greater or less than the other. The ranking may be either increasing or decreasing depending on the application. For example, if a food spicy rating is from the less spicy,"1," to the most spicy, "5," rating number 4 is spicier than rating number 3. However, the difference of any two rating numbers in this scale is not necessarily equal. That is, the interval between rating number 3 and 4 is not necessarily the same as the interval between rating number 1 and 2. Thus, the ranking implies an ordering among items but nothing in between. Thus, any sequence coding can be used for examples: 0-to-4, or negative 2 to plus 2.

A third scale level is interval scale. It has more information than the nominal and ordinal scales since there are equal spaces between any two values. Temperature is one example. The interval from 30 and 40 degree F is the same amount as at 80 and 90 degree F. Interval scale data can be used in arithmetic operations such as subtraction, addition, and multiplication However, after adding interval scale data, one cannot infer about the ratio of them since it does not have the true zero.

The last scale is the ratio scale, which is similar to the interval scale except it has the property of ratios and a true zero. Corresponding ratios on different parts of a ratio scale have the same meaning. Height is an example; there is no value below the zero point. The value of zero is absolutely no height. So, a height of four inches is twice as high as a height of two inches. In conclusion the ratio scale is highest scale level and the most informative scale for statistical analysis.

There are many statistical methods to consider in order to conduct research, and so the question of how to select an appropriate one is always faced. A misconception often encountered in selecting statistical analysis arises from the failure to consider the scale of data measurement. That will lead to the question of accuracy of statistical inference for the research questions.

Testing for two independent groups

We have chosen to compare by testing for two independent samples. The statistical methods in this study; the t-test, Mann-Whitney test, and Kolmogorov-Smirnov test, will be examined for the robustness and statistical power in different circumstances by the simulated Likert-type data sets. In order to find the appropriate test for each condition, the empirical Type I error rate of all tests will be compared.

The mentioned statistical tests for testing two independent groups can help us to know whether both testing groups come from the same distribution or population or not. Student's t-test is the most popular method for interval data but in this study, it will be used with assumed equal distance scoring system for ordinal Likert-type data to test for the equality of the means. The Student's t-test is well known and popular to be used to analyze an attitude survey with ordinal data instead of an alternative nonparametric procedure which would be appropriate. The Mann-Whitney test, which is an alternative to the Student's t- test, is widely applied for comparing two independent ordinal samples. The Kolmogorov-Smirnov test is also the alternative test and can be used to test whether two samples come from the same distribution.

The simulation data of this study will be generated by using R which is a language and environment for statistical computing and graphics and be analyzed with all these tests in various testing combinations such as significance levels, effect sizes, sample sizes, and distributions, and to observe the robustness of the tests, including their testing power.

Research Questions

The questions specifically addressed in this study are as follows:

1) Which, if any, of the statistical procedures in this study can control Type I error rate?

2) If the Type I error rate is not controlled, under what circumstances are tests liberal or conservative?

3) Which tests give the highest statistical power in each scenario?

4) Is there an overall best statistical procedure to recommend? Or are there any specific situations that indicate which method should be preferred?

5) How do results from parametric and non-parametric tests compare on simulated Likert-type data?

Limitations of the study

In this study will be limited to only statistical factors that used practically more often for two independent samples. They are five specific data distributions, three different levels of significance, two kinds of Likert rating scale, three effect sizes, and twelve pairs of sample sizes. Therefore, the finding of the study will be necessarily defined in terms of the specific data situations analyzed. While suggestions for further studies can be made, one must be cautious in generalizing beyond the specific data situations investigated in this study.

CHAPTER 2

LITERATURE REVIEWS

Treating ordinal as interval

There are long endless debates about whether we can treat ordinal data as interval data. For example, the collected data from Likert scale surveys are originally ordinal scale. Since the Likert-type data have an inherent order, assuming that the interval difference between agreeing and strongly agreeing is the same as between disagreeing and being neutral is inappropriate. Mogey suggested using a median and mode (not a mean) to describe the ordinal data and to apply nonparametric methods to determine the difference between the comparison groups. (Mogey, 1999)

Traylor (1983) showed that in many situations, ordinal data can be treated as interval data without a great loss in accuracy and with a gain in interpretability by using equal-scaling techniques. However, this may not be an appropriate decision, justifying the use of parametric statistics for interpretation and conclusion if the true scoring system is a gross, nonlinear distortion of the equal-interval scale being used (Traylor, 1983). Goldstein and Hersen (1984) also agree that the Likert scale presumes that the alignment of the five responses will be equal (Goldstein & Hersen, 1984).

Clason and Dormody (1994) illustrated that a variety of statistical methods are being used to analyze individual Likert-type data in the *Journal of Agricultural Education* from Volume 27 through 32. From the total of 188 articles, they investigated 95 articles that used Likert scaling. The results showed that 54% of the articles reported the responses as descriptive statistics (e.g., means, standard deviations, frequencies/percentages by category), 13% of the articles tested hypotheses between two groups using nonparametric statistics (e.g. Chi-square homogeneity tests, Mann-Whitney tests, Kruskal-Wallis tests), and 34% of the articles compared by means of two groups using parametric statistical procedures (e.g. t-test, analysis of variance). They suggested that any statistical methods that meaningfully answer the research questions, maintain the information of the data, and are not subject to scaling debates should be the procedures of choice in analyzing Likert-type items (Clason & Dormody, 1994).

Parametric and nonparametric procedures

Most statistical procedures in research can be classified to two main types; parametric and nonparametric procedures. For this study, the Mann-Whitney test, and the Kolmogorov-Smirnov test are the nonparametric procedures considered. The t-test is a parametric method that usually requires interval or ratio data level.

Parametric statistical procedure will provide more power than the nonparametric method if all assumptions for parametric analysis are met. Assumptions about normality and interval scale level are necessary for the t-test. Whether these assumptions are met are the common questions in Likert-type data. Using the t-test with the ordinal data may decrease the power of the test. On the other hand, if the Chi-square test, which requires only nominal data, is used with ordinal data, this nonparametric test may also be inappropriate.

Since the Likert rating scale is an ordinal scale, in order to compare the difference of two independent groups of Likert data, the Mann-Whitney test, a nonparametric procedure, can be used (Mogey, 1999). The t-test for two independent groups should not be applied because the ordinal data are discrete with no continuous value in between; thus finding the mean and standard deviation for testing is inappropriate. However, in this research, we want to determine the power and robustness of the t-test as compared to the other nonparametric tests, so we assume Likert-type data are of interval scale; that is, they have equally quantified distances in between the values. Of course, frequently they are not and this calls into question the use that some researches make of these data.

Problem in statistical procedure's assumptions

Micceri (1989) studied the characteristics of distributions of 440 large-sample data sets pertaining to achievement and psychometric measures. He found that 6.8% of the all distributions exhibit both tail weight and symmetry approximating that expected in the Gaussian distribution, and 4.8% showed relative symmetry and tail weights lighter than that expected in the Gaussian distribution. Based on the symmetry or asymmetry, 30.7 % were classified as being extremely asymmetric. His findings indicate that real-world distributions do not always conform to normality as expected. (Micceri, 1989)

The effects on parametric statistical methods when assumptions of normality and homogeneity of variance are violated are studied by using computer simulation. The results indicate that nonparametric methods are not always acceptable substitutes for the parametric ones. Whether they are depends on the sample sizes and distribution shapes as well (Zimmerman, 1998).

Delancy and Vargha indicated the effects of non-normality on the t-test while maintaining the equal variance of populations for the two-sample t-test and Welch's robust t-test. The results revealed that the validity of both methods depends on whether the two distributions are skewed in the same direction. The Type I error rates are quite acceptable from both tests if their parents' distributions are skewed in the same direction even with relatively small samples. However, the Type I error rate can increase remarkably when the two parent distributions are skewed in opposite directions (Delaney & Vargha, 2000).

Studies of the robustness and power

Ramsey (1980) studied the robustness of the t-test in normal populations with unequal variances. His results show that even if the sample sizes of two comparison groups are the same, the t-test of equal variance groups is not always robust (Ramsey, 1980).

Also considering homogeneity of variance, Blair showed that the Type I error rate of the t-test can be deviated markedly by a small amount of group variation (Blair, 1983). The nonparametric statistical procedure, the Mann-Whitney test, is an alternative method for the parametric t-test, and can be applied to the ordinal scale data (Gibbons & Chakraborti, 1992). That the Mann-Whitney test is still valid for small sample sizes can be shown by using Monte Carlo simulations (Fahoome, 1999).

Zimmerman (1996) showed that rank transforming of scores can reduce variance heterogeneity when scores from different variances are merged and ranked as a single data set. However, the distortion of the Type I and Type II error could not be reduced (Zimmerman, 1996).

Sawilowsky and Blair (1992) used simulation of eight real world distributions in psychology and education research and found that the t-test can be robust to Type I error when the sample sizes of both groups are equal and large enough for either one-tailed or two-tailed tests under distributions with light skew. Type II error is robust under these nonnormal situations (Sawilowsky & Blair, 1992).

Hogarty and Kromrey (1998) investigated Type I error control and the statistical power of four testing methods; the Student's t-test, the Chi-square test, the delta statistic, and a cumulative log model in ordinal data. The effects of the number of categories of response variable, sample size, population shape, and the effect model were determined. They found that the Student's t-test obtained the best control of Type I error rate, but was not the most powerful. The Chi-square test is the most powerful for the 5-point response scale. For the 7-point response scale, the results of the testing power are varied among these procedures (Kromrey & Hogarty, 1998).

Likert-type data in education

Using Likert scale to measure attitude is very practical and easy to use for questionnaire survey. Some research surveys need to know general attitudes of people on a specific topic, but the data from other survey may be used for a sensitive consequent such as for promotion or tenure decisions. The student evaluation of teaching survey is one of the examples. In interpreting and testing such the data, we must be very careful because the Likert-type data vary in their distribution; not all of them are normal distributed. Therefore, the mean is not always a good statistic to describe all situations. If the data shapes are skewed or bimodal, the mode or median would be a better measure. However, the accuracy of interpretation not only depends on the appropriate statistic itself, but also on the quality of the rating data. There are many rating surveys on the online where people can post their attitudes freely (Michals, 2003). For example, students can rate any teachers' or professors' performance of teaching across the country with any additional comments on the Internet. If all students from the class of the professor who will be rated evaluate their instructor, the result will be reliable. On the other hand, if only a few students from the class post their rating on the website, it might lead to the selection bias and the result may be inaccurate. The issues of advantage and disadvantage of student evaluation of teaching and the accuracy for interpretation of the data have been discussed in many studies (Crumbley, Henry, & Kratchman, 2001; Gray & Bergmann, 2003; Kolitch & Dean, 1999; Liaw & Goh, 2003; Nerger, Viney, & Riedel, 1997; Panasuk & LeBaron, 1999; Simpson & Siguaw, 2000; Worthington, 2002).

However, the traditional interpretation of the evaluation of teaching data still depends on only the mean as if the data are interval scale. Also the numbers of students in class who rate and the distribution of the data were always less considered. Therefore, the appropriate descriptive and inferential statistics should be reviewed.

CHAPTER 3

RESEARCH METHODOLOGY

The objective of the research presented in this study was to determine the relative robustness and power of three different methods for testing the hypothesis that ordinal samples of five and seven Likert categories come from equal populations. The three methods are the two sample t-test with equal variances, the Mann-Whitney test, and the Kolmogorov-Smirnov test. Although theoretically inappropriate, the t-test applied here assumes the equal distance for ordered responses of the Likert scales.

Since statistical robustness and power are a function of the Type I and Type II errors, all three tests were examined with respect to both. In addition, to obtain a general and practical understanding of the robustness and power, methods are employed over a wide range of scenarios with respect to population distribution, sample size, effect size, and the number of categories of the Likert scale.

Simulation of population distributions

The population distributions will be simulated using R, the language programming software version 2.13.2, to provide an empirical comparison of the Type I error rate, and the power of the tests of the three statistical procedures. To simulate the observations for both testing groups, the function "sample" in R will be applied for generating all data according to the desired distributions or shapes. Five population distribution shapes as shown in Figure 1 and Figure 2 were investigated.

Figure 1. Five population distribution sharps of the 5-point Likert response scale

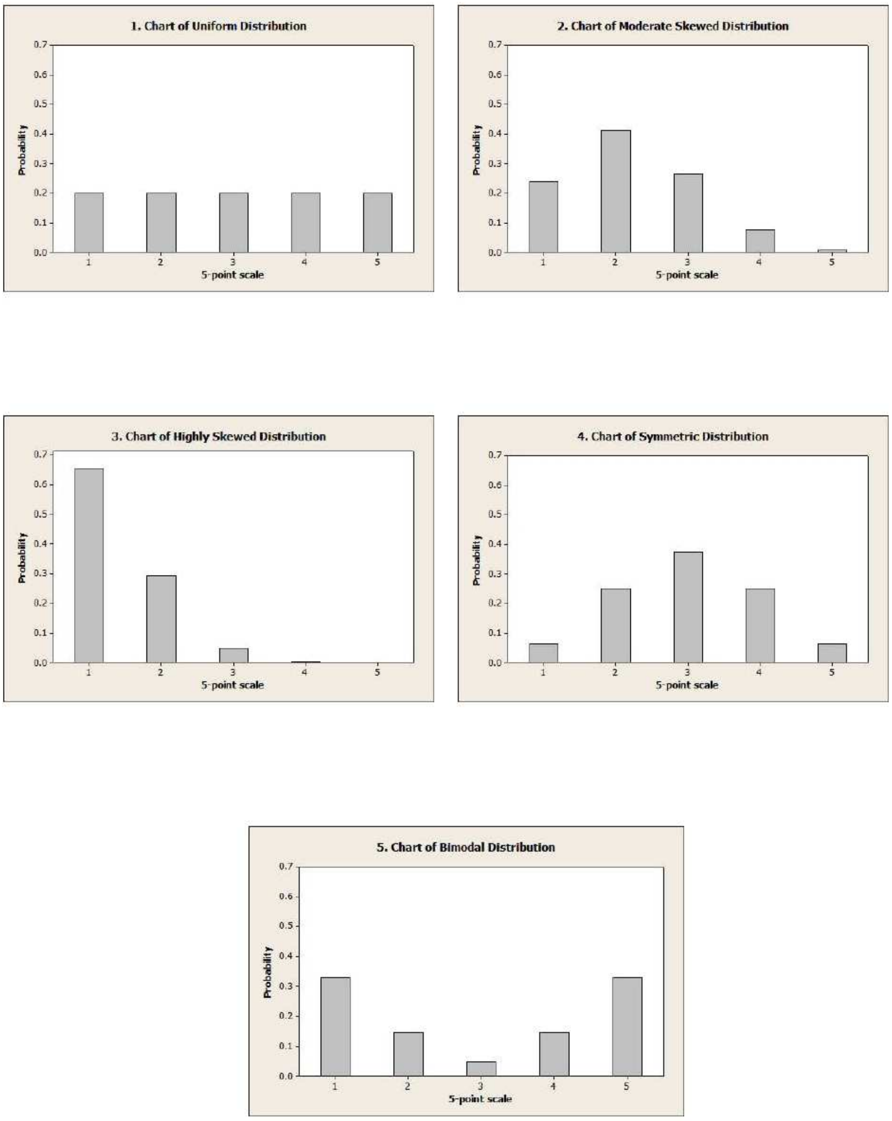
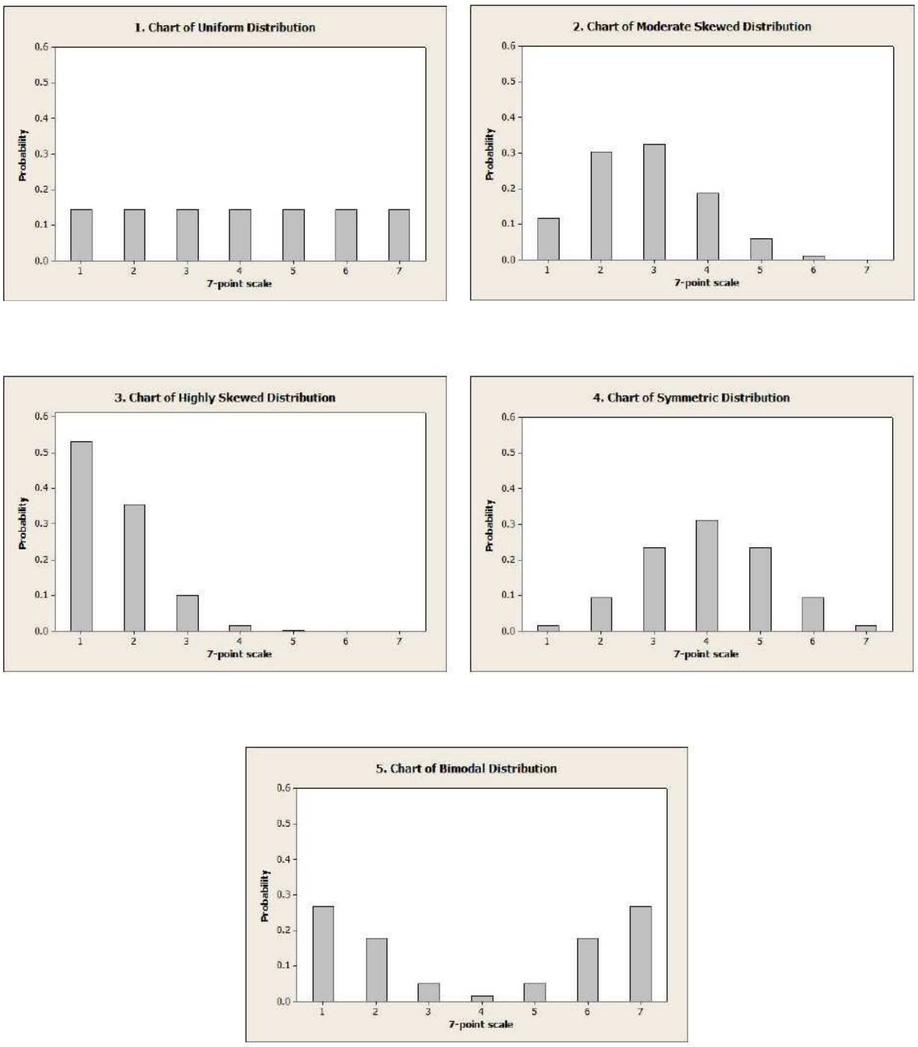


Figure 2. Five population distribution sharps of the 7-point Likert response scale



They are a uniform response distribution, a moderately skewed distribution, a highly skewed distribution, a unimodal symmetric distribution, and a bimodal symmetric distribution with their marginal probabilities in Table 1 and Table 2 for the 5-point Likert scale and the 7-point Likert scale respectively.

Table 1. Five marginal distributions for the 5-point response scale

5-point scale	Uniform	Moderate Skew	Highly Skew	Symmetric	Bimodal
1	0.2000	0.2401	0.6561	0.0625	0.3281
2	0.2000	0.4116	0.2916	0.2500	0.1476
3	0.2000	0.2646	0.0486	0.3750	0.0486
4	0.2000	0.0756	0.0036	0.2500	0.1476
5	0.2000	0.0081	0.0001	0.0625	0.3281

Table 2. Five marginal distributions for the 7-point response scale

7-point scale	Uniform	Moderate Skew	Highly Skew	Symmetric	Bimodal
1	0.142857	0.117649	0.531441	0.015625	0.265721
2	0.142857	0.302526	0.354294	0.093750	0.177174
3	0.142857	0.324135	0.098415	0.234375	0.049815
4	0.142857	0.185220	0.014580	0.312500	0.014580
5	0.142857	0.059535	0.001215	0.234375	0.049815
6	0.142857	0.010206	0.000054	0.093750	0.177174
7	0.142857	0.000729	0.000001	0.015625	0.265721

The uniform marginal distribution can be generated by using at equal proportions at each level of the response scale. For example of the 5-point scale, the uniform distribution was produced by generating data in which 20% of the observations were in each category of the response variable. Similar to the 7-point scale, each of response values will be simulated about 14.3% of all observations for the uniform. All five distribution data were generated using function “sample” in R programming language (For R coding detail, please see Appendix part).

Sample sizes

The total of twelve sample sizes was examined for the difference of testing groups. The sample sizes chosen for this study were as follows (10, 10), (10, 30), (10, 50), (30, 30), (30, 50), (30, 100), (50, 50), (50, 100), (50, 300), (100, 100), (100, 300), and (300, 300). These particular sample sizes were chosen because of the design conditions of a broad range of sample sizes and both balanced and unbalanced data. The conditions should be included because some statistical tests may behave differently under these circumstances.

Robustness and significance levels

For the robustness analysis, we consider the significance level**s** of 0.01, 0.05, and 0.10 which they are most commonly used for statistical testing. The Type I error occurs when a test incorrectly rejects a true null hypothesis. Type I error rate is the fraction of times that a type I error is made.

The empirical Type I error rate of the tests will be computed by comparing with the given significance level (α) for the robustness following the modified criteria from Bradley, 1978.

At nominal $\alpha = 0.01$, an observed Type I error rate within 40% of this rate, i.e., from 0.006 to 0.014, is considered robust.

At nominal $\alpha = 0.05$, an observed Type I error rate within 25% of this rate, i.e., from 0.0375 to 0.0625, is considered robust.

At nominal $\alpha = 0.10$, an observed Type I error rate within 20% of this rate, i.e., from 0.08 to 0.12, is considered robust.

Power and levels of effect size

Statistical power is defined as the probability of rejecting the null hypothesis given the alternative hypothesis is true. In order to evaluate the statistical power of the tests we need to specify the effect size. The effect size refers to the magnitude of the effect of the alternative hypothesis. If the effect size is large enough, the alternative hypothesis will be true and the null hypothesis of equality is false. Therefore, there is a real difference between both testing groups. In this study the effect sizes of 0.10, 0.30, and 0.50 will be examined.

The number of categories of the Likert scale and the tests

Likert scales vary in the number of response in the scale. The 5-point scale is the most common following with 7-point response scale but some Likert scales have 4-point response scales, eliminating the neutral/undecided category. In this study the 5-point and 7-point Likert response scale that often found in survey researches will be examined for the robustness and power of the tests.

The two independent groups with equal variances will be considered in the simulation and analyzed by the two sample t-test with equal variances, the Mann-Whitney test, and the Kolmogorov-Smirnov test.

The t-test provided a test of the null hypothesis of equivalent of population means. The data being tested should be continuous; interval or ratio data, with a normal distribution and equal variances in the two groups. if these assumptions are violated, the nonparametric Mann-Whitney test or the Kolmogorov-Smirnov test may be used instead.

The Mann-Whitney test is used for a non-parametric test of location shift between the population distributions. The observations from both groups are combined and ranks, with the average rank assigned in the case of ties. The number of ties should be small relative to the total number of observations. If the populations are identical in location, the ranks should be randomly mixed between two samples.

The Kolmogorov-Smirnov test is a non-parametric test which based on the maximum absolute difference between the cumulative distribution functions for both testing groups. When this difference is significantly large the two distributions are considered difference.

CHAPTER 4

RESULTS

Three ordinal sample comparison methods were assessed for robustness and power at a wide range of scenarios (total of 1080 combinations). They are 5 population distributions, 12 sets of sample sizes, 3 effect sizes, 3 significance levels, and 2 types of Likert scale. The testing data were generated for 10000 samples in each of the combinations. The comparison methods were the two sample t-test, the Mann-Whitney test, and the Kolmokorov-Smirnov test. All obtained p-values from those tests will be examined for the type I error rate and the power of the tests.

The Type I error is defined as rejecting a null hypothesis when it is, in fact, true. A Type II error is defined as accepting a null hypothesis when it is, in fact, false. Thus, the statistical power can defined as rejecting a null hypothesis when it is, in fact, false. In this study, the error rate was obtained from the mean of all rejections results for each scenario.

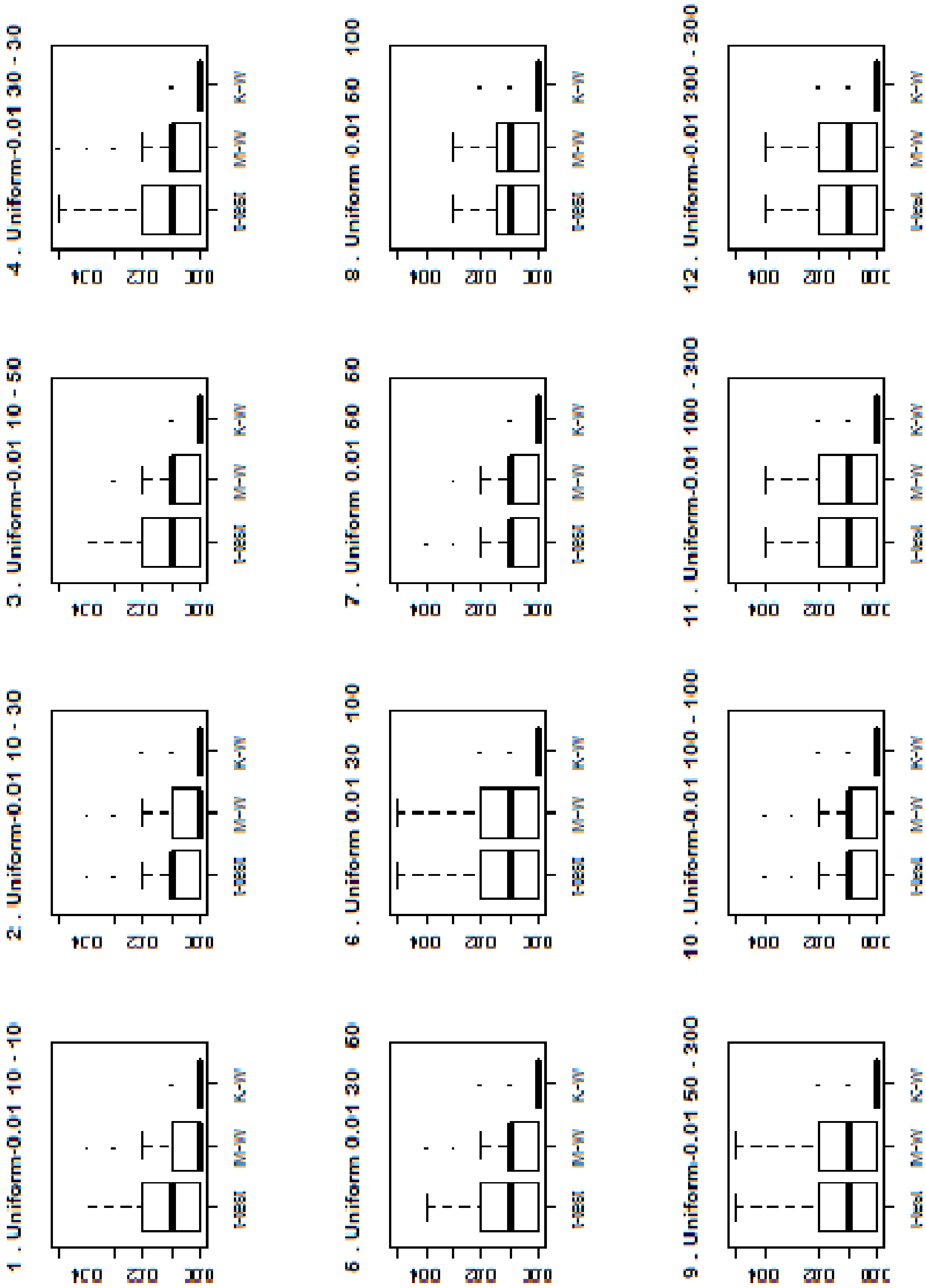
If the empirical Type I error rate is close to the given significance level (α) following the criteria of Bradley (1978) then the testing procedure is considered robust or the test is relevant. For the power, if the rejection rate of the null hypothesis is large (close to one), the more powerful the test. However, the desired test should meet the robustness and power properties.

Table 3. Type I error rate estimates of the uniform distribution

Significance Level	Sample Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
0.01	10, 10	0.012	0.007	0.001	0.011	0.007	0.001
	10, 30	0.008	0.006	0.002	0.008	0.007	0.002
	10, 50	0.010	0.008	0.000	0.010	0.008	0.002
	30, 30	0.011	0.009	0.001	0.011	0.010	0.002
	30, 50	0.011	0.010	0.002	0.013	0.012	0.003
	30, 100	0.012	0.011	0.002	0.011	0.010	0.002
	50, 50	0.009	0.008	0.001	0.010	0.009	0.001
	50, 100	0.010	0.009	0.001	0.009	0.008	0.002
	50, 300	0.011	0.011	0.002	0.009	0.009	0.002
	100, 100	0.008	0.008	0.001	0.010	0.010	0.002
	100, 300	0.010	0.010	0.002	0.012	0.012	0.002
	300, 300	0.011	0.010	0.002	0.012	0.012	0.004
0.05	10, 10	0.048	0.042	0.003	0.051	0.043	0.006
	10, 30	0.052	0.048	0.012	0.049	0.047	0.017
	10, 50	0.053	0.050	0.009	0.046	0.042	0.011
	30, 30	0.046	0.045	0.011	0.050	0.049	0.013
	30, 50	0.048	0.046	0.011	0.052	0.051	0.015
	30, 100	0.051	0.049	0.012	0.048	0.046	0.014
	50, 50	0.054	0.053	0.012	0.050	0.049	0.014
	50, 100	0.050	0.048	0.011	0.050	0.049	0.013
	50, 300	0.054	0.053	0.008	0.052	0.053	0.015
	100, 100	0.050	0.049	0.009	0.049	0.048	0.013
	100, 300	0.050	0.049	0.009	0.051	0.051	0.013
	300, 300	0.050	0.050	0.010	0.051	0.051	0.014
0.10	10, 10	0.103	0.096	0.020	0.104	0.096	0.024
	10, 30	0.095	0.095	0.022	0.107	0.104	0.029
	10, 50	0.100	0.099	0.021	0.104	0.103	0.027
	30, 30	0.100	0.100	0.022	0.102	0.097	0.028
	30, 50	0.105	0.103	0.026	0.100	0.098	0.030
	30, 100	0.097	0.096	0.024	0.102	0.102	0.036
	50, 50	0.102	0.102	0.023	0.103	0.102	0.028
	50, 100	0.096	0.096	0.022	0.102	0.102	0.027
	50, 300	0.100	0.100	0.025	0.095	0.096	0.028
	100, 100	0.098	0.099	0.023	0.098	0.099	0.027
	100, 300	0.103	0.104	0.024	0.095	0.095	0.027
	300, 300	0.100	0.101	0.032	0.097	0.097	0.034

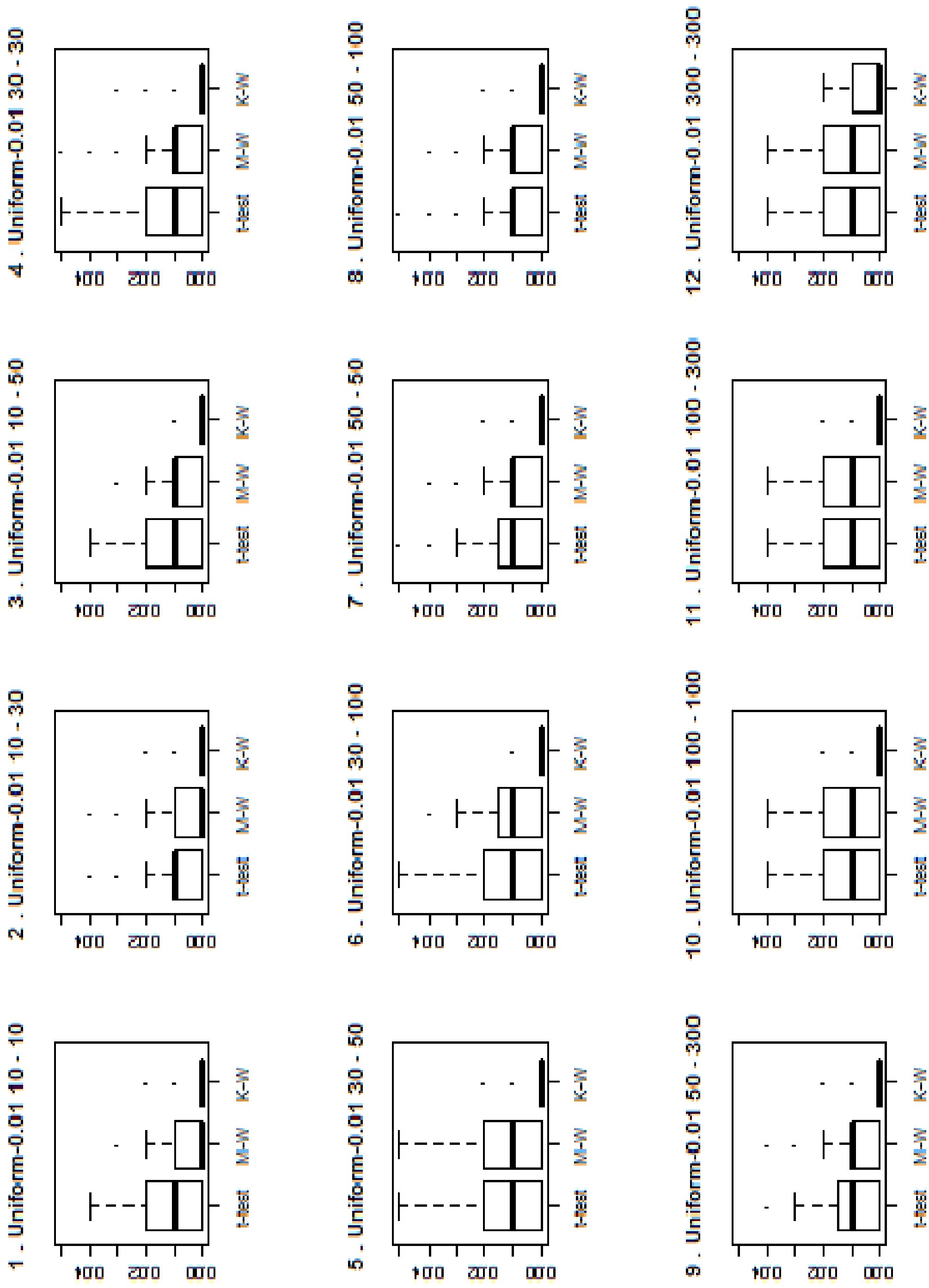
Table 3 shows that, given the uniform distribution, the empirical Type I error rates from the t-test and Mann-Whitney (MW) test are close to the nominal significance levels and followed the robustness criteria. However, the Kolmogorov-Smirnov (KS) test is not robust since the error rate is outside the criteria's range for all circumstances.

Figure 3. Distribution of Type I error rate estimates of the uniform distribution at α = 0.01 for the 5-point Likert scale



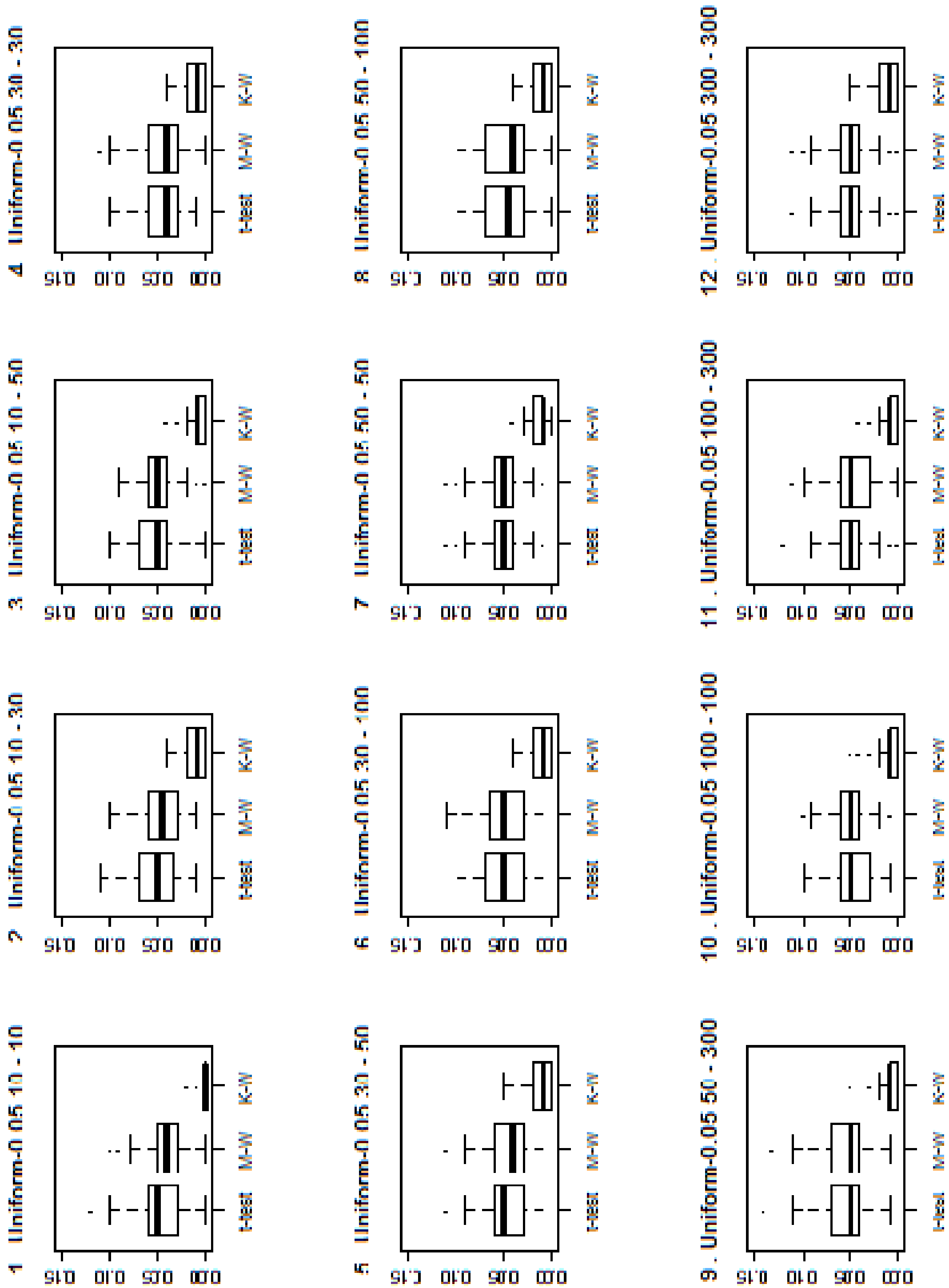
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 4 Distribution of Type I error rate estimates of the uniform distribution at α = 0.01 for the 7-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

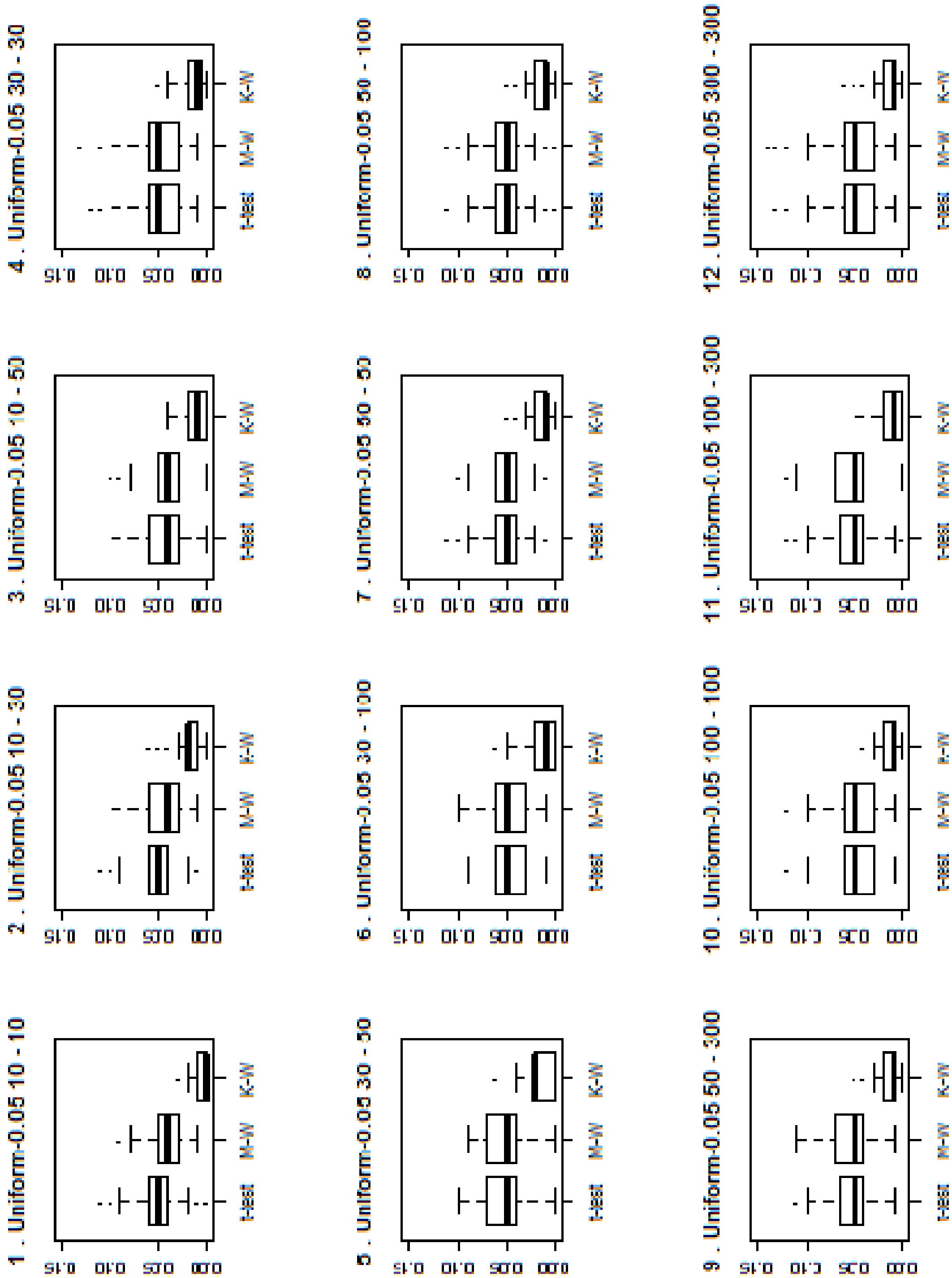
Figure 5. Distribution of Type I error rate estimates of the uniform distribution at α = 0.05 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

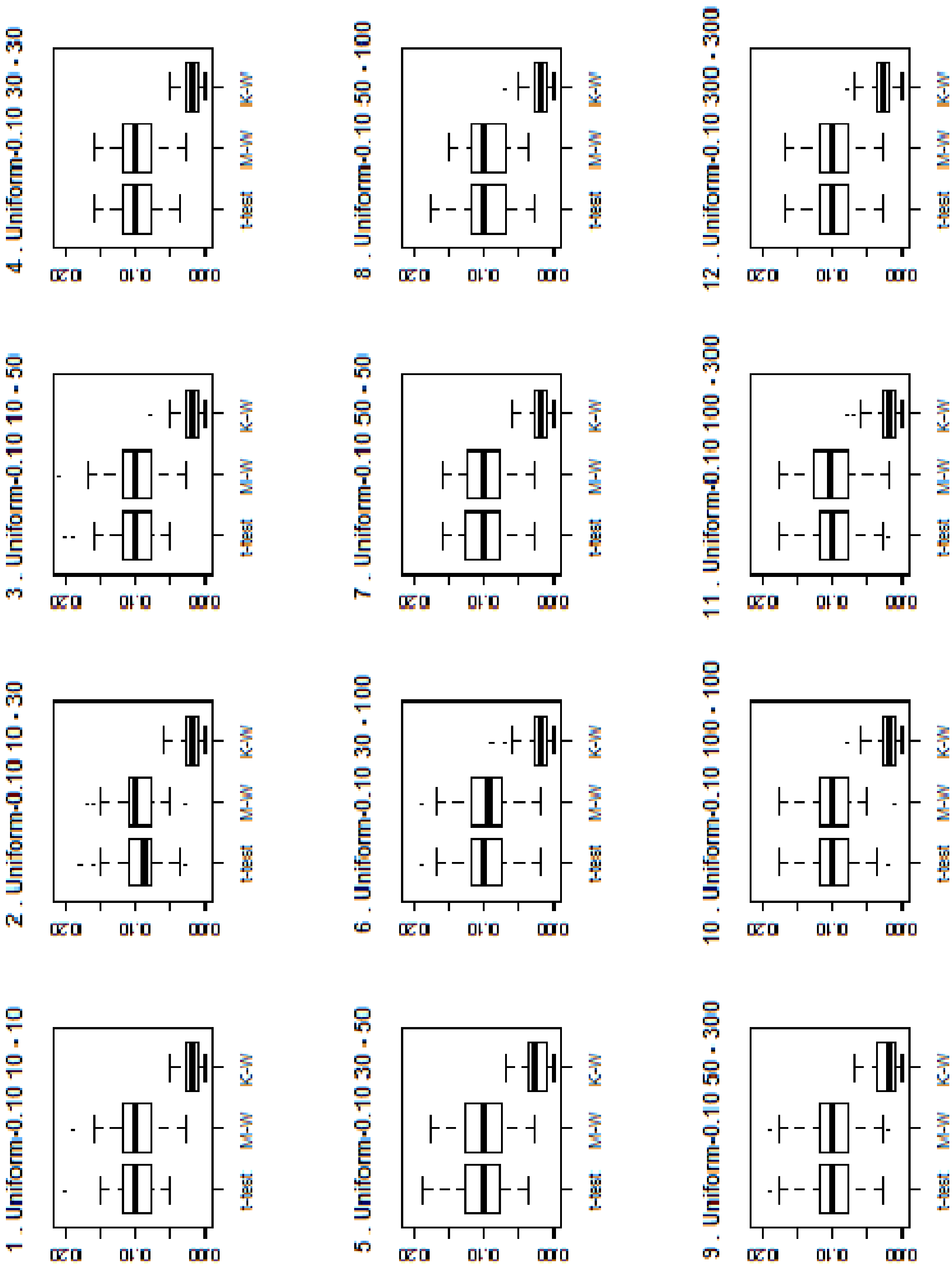
Figure 6. Distribution of Type I error rate estimates of the uniform distribution at $\alpha = 0.05$

for the 7-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 7. Distribution of Type I error rate estimates of the uniform distribution at α = 0.10 for the 5-point Likert scale

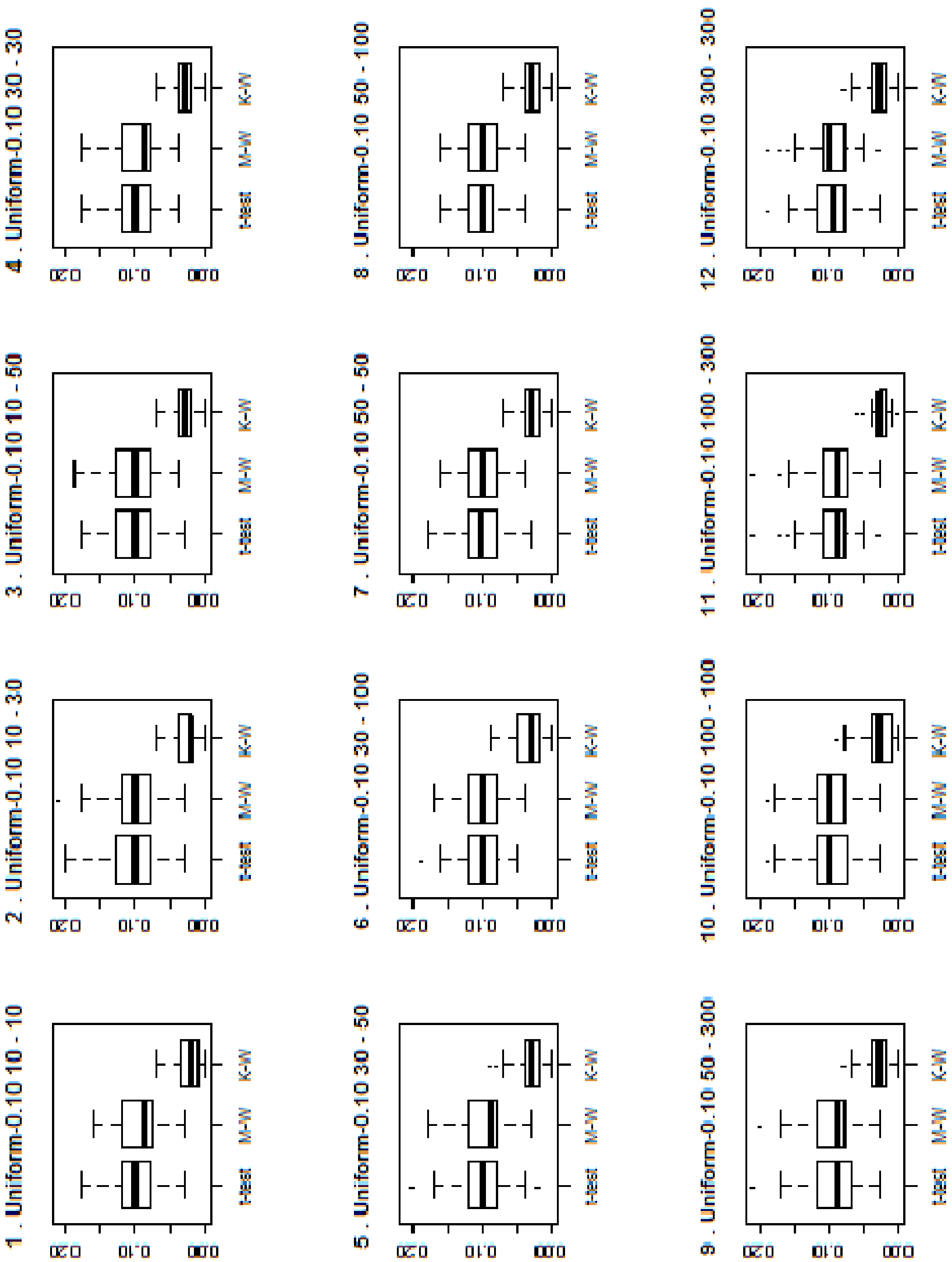


From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

.

Figure 8. Distribution of Type I error rate estimates of the uniform distribution at α = 0.10

for the 7-point Likert scale



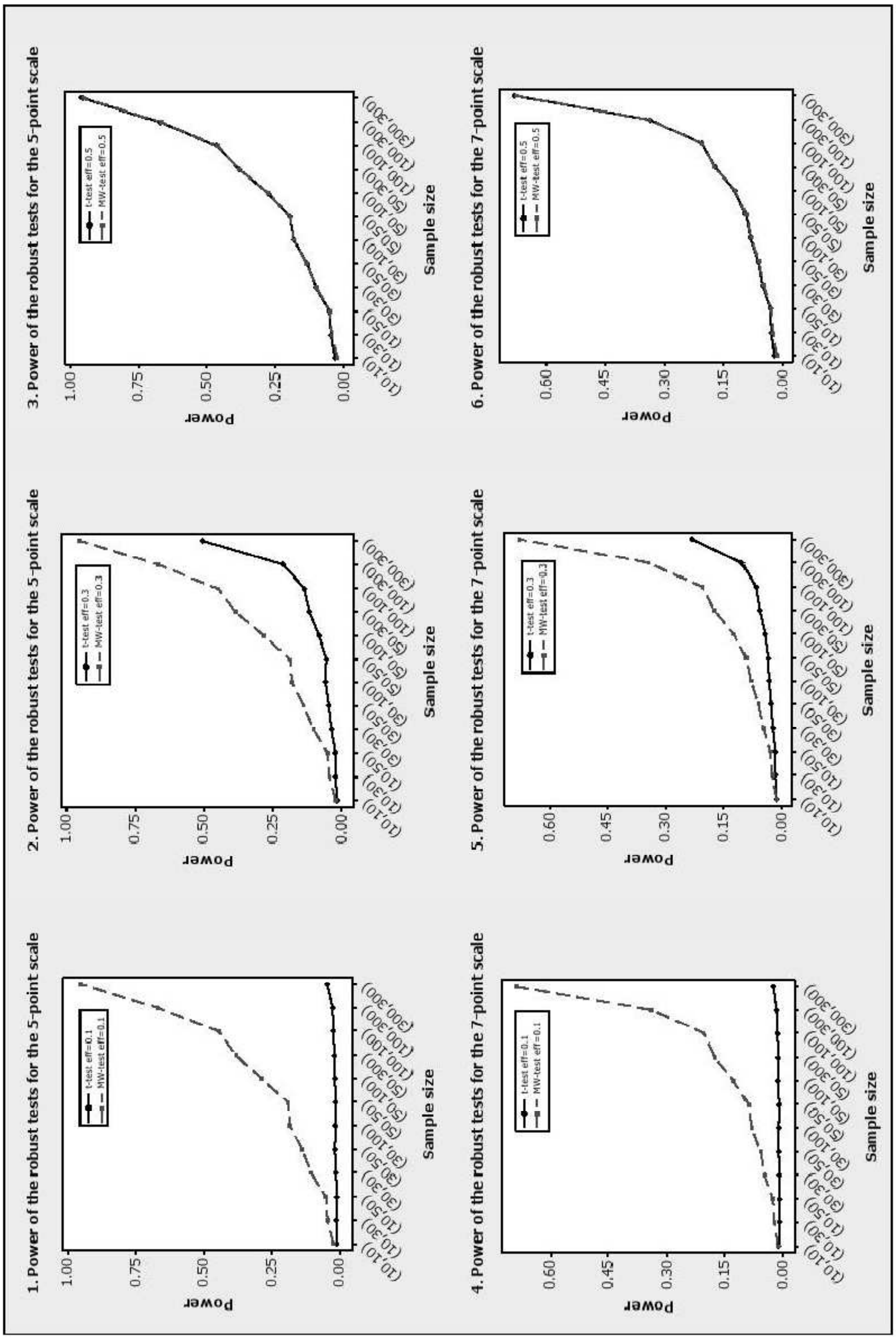
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Table 4. Statistical power estimates of the uniform distribution at $\alpha = 0.01$

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.012	0.023	0.006	0.012	0.013	0.006
10, 10	0.30	0.016	0.021	0.006	0.014	0.013	0.003
10, 10	0.50	0.033	0.023	0.006	0.022	0.014	0.005
10, 30	0.10	0.012	0.045	0.028	0.010	0.024	0.018
10, 30	0.30	0.022	0.045	0.030	0.016	0.024	0.017
10, 30	0.50	0.047	0.043	0.029	0.029	0.025	0.017
10, 50	0.10	0.011	0.051	0.027	0.010	0.027	0.014
10, 50	0.30	0.022	0.052	0.027	0.016	0.029	0.013
10, 50	0.50	0.053	0.051	0.026	0.031	0.030	0.013
30, 30	0.10	0.013	0.107	0.086	0.011	0.049	0.035
30, 30	0.30	0.037	0.103	0.080	0.023	0.047	0.036
30, 30	0.50	0.103	0.100	0.081	0.052	0.049	0.034
30, 50	0.10	0.017	0.141	0.115	0.012	0.059	0.044
30, 50	0.30	0.047	0.140	0.116	0.026	0.060	0.044
30, 50	0.50	0.135	0.137	0.112	0.063	0.061	0.044
30, 100	0.10	0.015	0.186	0.178	0.012	0.083	0.067
30, 100	0.30	0.059	0.184	0.176	0.032	0.080	0.067
30, 100	0.50	0.185	0.185	0.172	0.083	0.081	0.067
50, 50	0.10	0.015	0.193	0.180	0.011	0.090	0.066
50, 50	0.30	0.057	0.191	0.181	0.034	0.091	0.067
50, 50	0.50	0.200	0.199	0.190	0.094	0.092	0.070
50, 100	0.10	0.018	0.288	0.314	0.014	0.128	0.118
50, 100	0.30	0.083	0.285	0.318	0.042	0.123	0.110
50, 100	0.50	0.275	0.276	0.306	0.120	0.120	0.108
50, 300	0.10	0.019	0.383	0.493	0.014	0.176	0.192
50, 300	0.30	0.119	0.389	0.495	0.056	0.174	0.193
50, 300	0.50	0.385	0.385	0.494	0.172	0.173	0.195
100, 100	0.10	0.023	0.450	0.600	0.015	0.205	0.209
100, 100	0.30	0.138	0.452	0.612	0.066	0.204	0.215
100, 100	0.50	0.468	0.466	0.620	0.206	0.206	0.211
100, 300	0.10	0.026	0.673	0.953	0.018	0.339	0.479
100, 300	0.30	0.217	0.671	0.950	0.103	0.344	0.484
100, 300	0.50	0.673	0.668	0.948	0.337	0.338	0.474
300, 300	0.10	0.046	0.955	1.000	0.025	0.686	0.973
300, 300	0.30	0.508	0.955	1.000	0.232	0.677	0.972
300, 300	0.50	0.960	0.954	1.000	0.681	0.676	0.972

Table 4 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 9.

Figure 9. Statistical power estimates of the uniform distribution at $\alpha = 0.01$



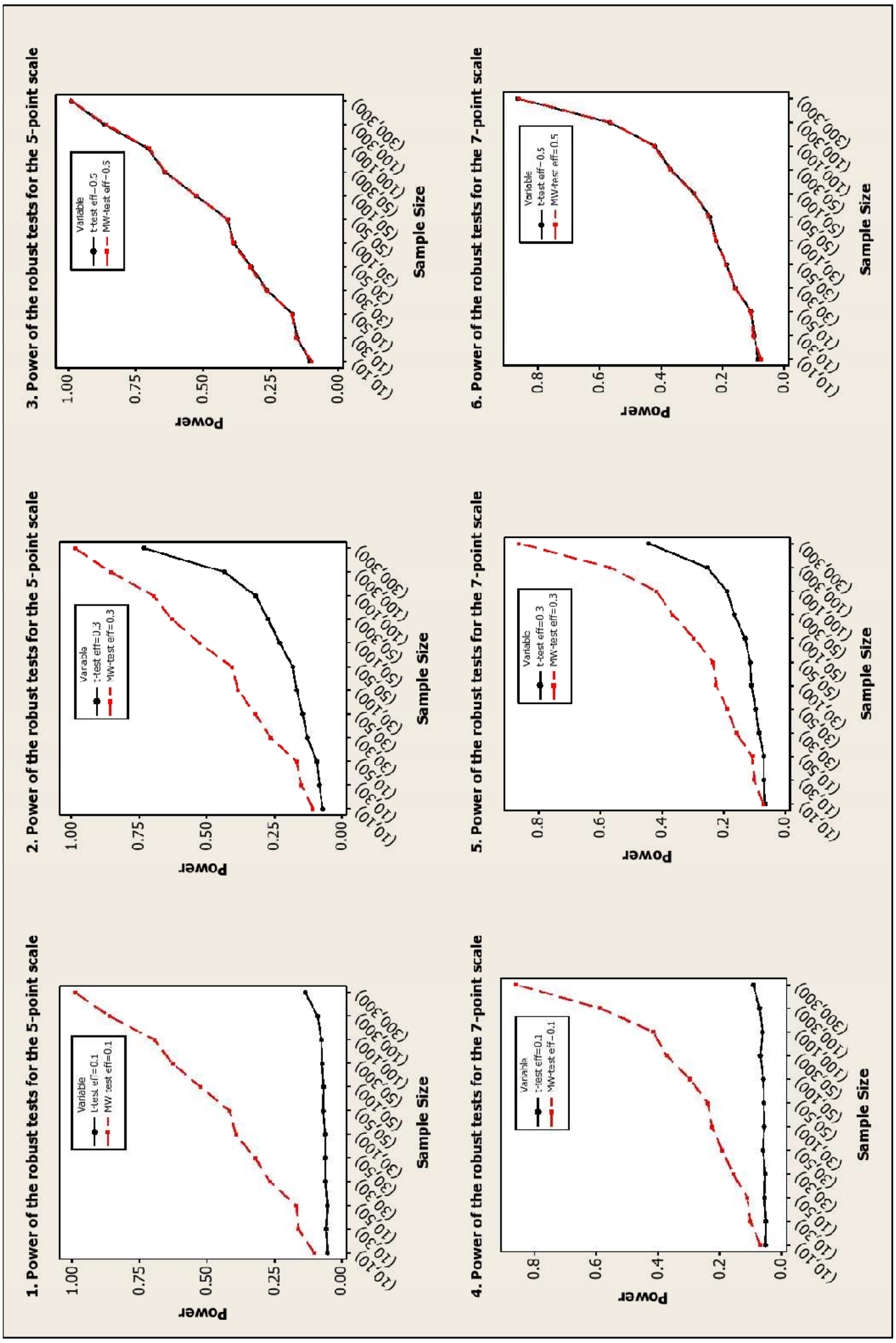
From Figure 9, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 5. Statistical power estimates of the uniform distribution at α = 0.05

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MS-test	KS-test
10, 10	0.10	0.053	0.102	0.034	0.051	0.068	0.022
10, 10	0.30	0.071	0.106	0.035	0.063	0.071	0.022
10, 10	0.50	0.106	0.099	0.034	0.084	0.075	0.025
10, 30	0.10	0.058	0.161	0.128	0.050	0.100	0.072
10, 30	0.30	0.082	0.149	0.118	0.068	0.099	0.080
10, 30	0.50	0.152	0.156	0.130	0.095	0.099	0.075
10, 50	0.10	0.052	0.169	0.122	0.054	0.110	0.074
10, 50	0.30	0.092	0.168	0.120	0.068	0.106	0.069
10, 50	0.50	0.170	0.171	0.121	0.106	0.107	0.069
30, 30	0.10	0.061	0.268	0.242	0.052	0.156	0.125
30, 30	0.30	0.128	0.263	0.239	0.084	0.158	0.126
30, 30	0.50	0.264	0.266	0.239	0.157	0.158	0.125
30, 50	0.10	0.061	0.322	0.312	0.059	0.191	0.157
30, 50	0.30	0.145	0.321	0.314	0.095	0.187	0.150
30, 50	0.50	0.322	0.328	0.312	0.186	0.187	0.155
30, 100	0.10	0.061	0.393	0.428	0.056	0.224	0.215
30, 100	0.30	0.168	0.384	0.415	0.108	0.222	0.210
30, 100	0.50	0.387	0.390	0.424	0.220	0.220	0.213
50, 50	0.10	0.068	0.417	0.484	0.057	0.238	0.235
50, 50	0.30	0.181	0.407	0.482	0.112	0.233	0.229
50, 50	0.50	0.410	0.412	0.478	0.240	0.244	0.242
50, 100	0.10	0.066	0.525	0.678	0.058	0.296	0.331
50, 100	0.30	0.230	0.528	0.679	0.130	0.295	0.332
50, 100	0.50	0.527	0.526	0.674	0.292	0.291	0.325
50, 300	0.10	0.073	0.628	0.821	0.068	0.371	0.430
50, 300	0.30	0.274	0.630	0.828	0.164	0.364	0.430
50, 300	0.50	0.643	0.641	0.833	0.368	0.370	0.427
100, 100	0.10	0.075	0.696	0.907	0.061	0.415	0.503
100, 100	0.30	0.319	0.697	0.906	0.188	0.417	0.502
100, 100	0.50	0.706	0.701	0.913	0.419	0.422	0.505
100, 300	0.10	0.089	0.864	0.999	0.071	0.588	0.805
100, 300	0.30	0.436	0.856	0.999	0.252	0.568	0.790
100, 300	0.50	0.867	0.860	0.999	0.566	0.563	0.794
300, 300	0.10	0.135	0.991	1.000	0.091	0.860	1.000
300, 300	0.30	0.734	0.989	1.000	0.442	0.863	0.999
300, 300	0.50	0.990	0.989	1.000	0.866	0.863	1.000

Table 5 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 10.

Figure 10. Statistical power estimates of the uniform distribution at α = 0.05



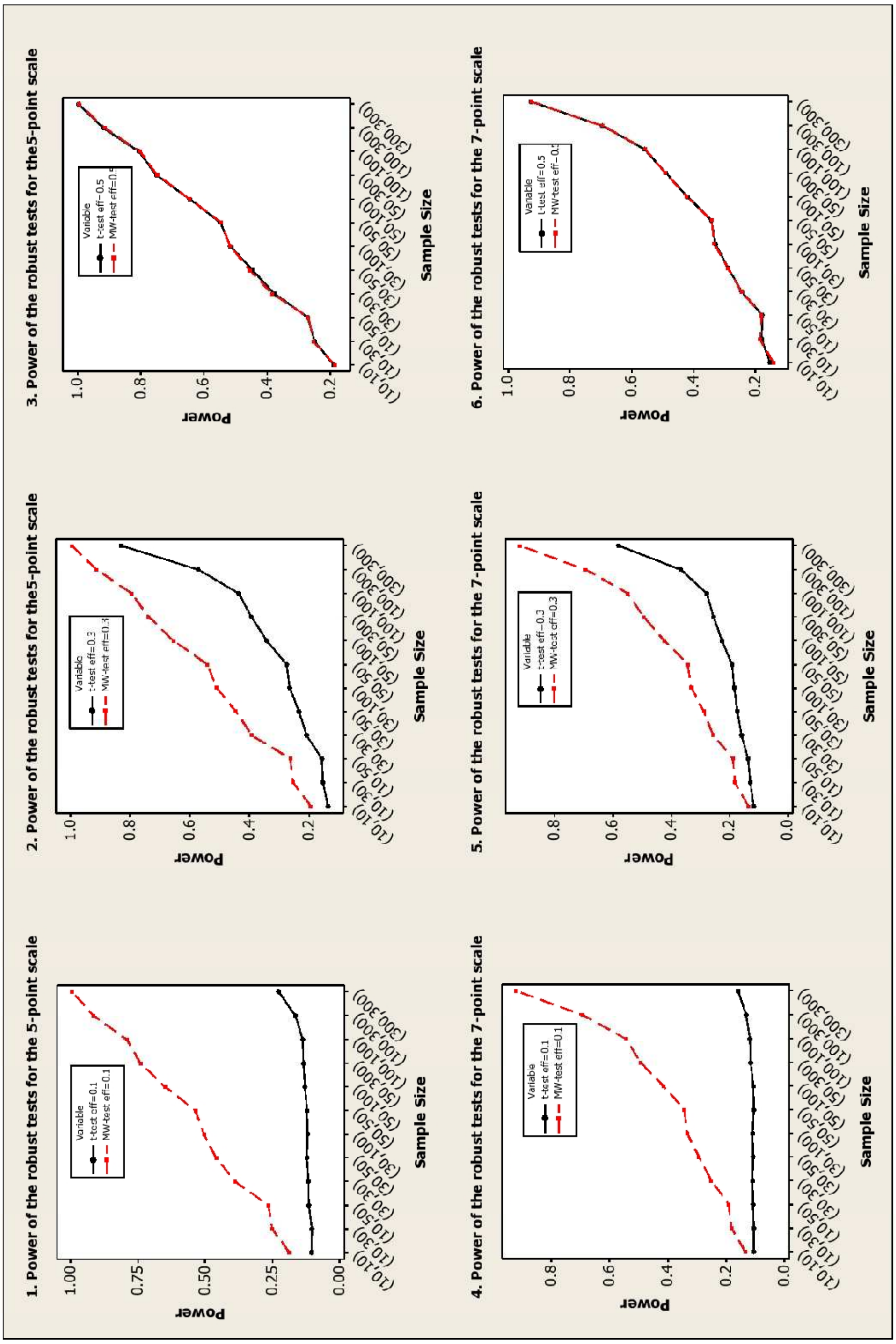
From Figure 10, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 6. .Statistical power estimates of the uniform distribution at α = 0.10

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.104	0.188	0.112	0.104	0.134	0.081
10, 10	0.30	0.137	0.194	0.118	0.117	0.135	0.082
10, 10	0.50	0.189	0.186	0.113	0.152	0.141	0.081
10, 30	0.10	0.103	0.251	0.169	0.105	0.181	0.116
10, 30	0.30	0.154	0.254	0.179	0.129	0.182	0.117
10, 30	0.50	0.250	0.254	0.177	0.178	0.184	0.121
10, 50	0.10	0.114	0.265	0.184	0.107	0.191	0.121
10, 50	0.30	0.157	0.263	0.179	0.137	0.188	0.128
10, 50	0.50	0.272	0.269	0.182	0.177	0.179	0.111
30, 30	0.10	0.116	0.390	0.376	0.108	0.250	0.209
30, 30	0.30	0.209	0.392	0.378	0.159	0.256	0.219
30, 30	0.50	0.378	0.385	0.370	0.244	0.245	0.206
30, 50	0.10	0.122	0.460	0.483	0.107	0.292	0.267
30, 50	0.30	0.235	0.446	0.469	0.173	0.288	0.268
30, 50	0.50	0.447	0.455	0.468	0.287	0.288	0.267
30, 100	0.10	0.119	0.505	0.585	0.108	0.333	0.338
30, 100	0.30	0.265	0.510	0.593	0.184	0.331	0.334
30, 100	0.50	0.517	0.517	0.597	0.328	0.332	0.328
50, 50	0.10	0.122	0.537	0.610	0.104	0.343	0.324
50, 50	0.30	0.276	0.542	0.611	0.192	0.343	0.327
50, 50	0.50	0.548	0.546	0.622	0.342	0.341	0.331
50, 100	0.10	0.129	0.650	0.815	0.106	0.414	0.452
50, 100	0.30	0.342	0.656	0.820	0.226	0.424	0.464
50, 100	0.50	0.646	0.645	0.811	0.419	0.421	0.463
50, 300	0.10	0.135	0.741	0.944	0.115	0.493	0.586
50, 300	0.30	0.394	0.738	0.948	0.255	0.495	0.591
50, 300	0.50	0.751	0.748	0.946	0.490	0.491	0.588
100, 100	0.10	0.137	0.792	0.974	0.118	0.544	0.689
100, 100	0.30	0.436	0.796	0.975	0.278	0.549	0.693
100, 100	0.50	0.807	0.803	0.974	0.559	0.557	0.701
100, 300	0.10	0.164	0.916	1.000	0.131	0.692	0.925
100, 300	0.30	0.572	0.913	1.000	0.368	0.694	0.924
100, 300	0.50	0.919	0.914	1.000	0.695	0.694	0.926
300, 300	0.10	0.227	0.996	1.000	0.158	0.919	1.000
300, 300	0.30	0.830	0.996	1.000	0.580	0.920	1.000
300, 300	0.50	0.997	0.996	1.000	0.926	0.924	1.000

Table 6 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 11.

Figure 11. Statistical power estimates of the uniform distribution at α = 0.10



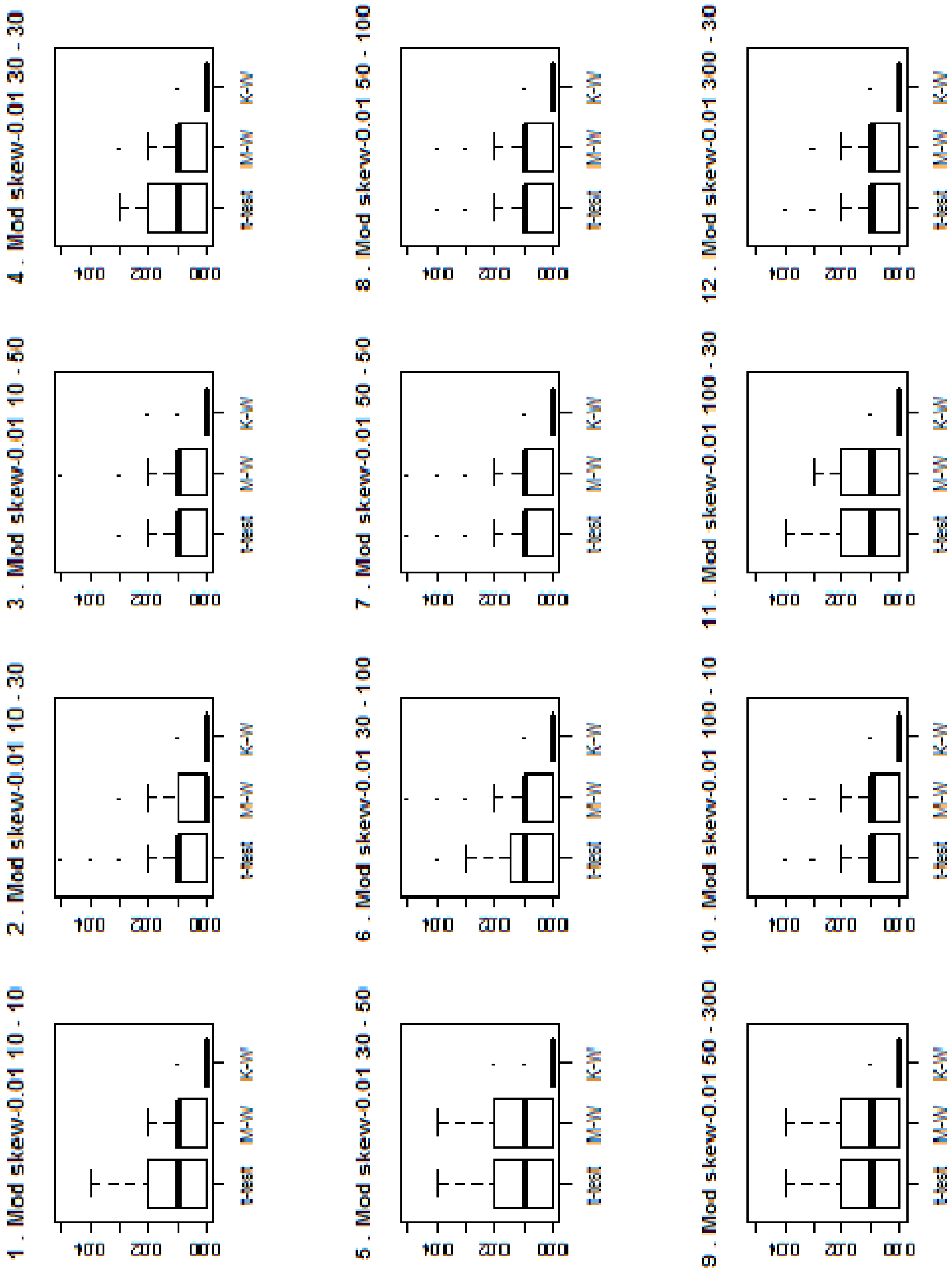
From Figure 11, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 7. Type I error rate estimates of the moderate skewed distribution

Significance Level	Sample Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
0.01	10, 10	0.011	0.006	0.000	0.010	0.007	0.001
	10, 30	0.009	0.006	0.001	0.011	0.008	0.001
	10, 50	0.011	0.009	0.001	0.010	0.008	0.001
	30, 30	0.011	0.010	0.001	0.012	0.010	0.001
	30, 50	0.010	0.010	0.001	0.010	0.009	0.001
	30, 100	0.011	0.010	0.001	0.010	0.010	0.001
	50, 50	0.010	0.009	0.001	0.010	0.008	0.001
	50, 100	0.009	0.008	0.001	0.009	0.010	0.001
	50, 300	0.010	0.010	0.000	0.010	0.009	0.001
	100, 100	0.011	0.010	0.001	0.010	0.010	0.001
	100, 300	0.009	0.009	0.000	0.010	0.009	0.001
	300, 300	0.010	0.008	0.001	0.010	0.009	0.001
0.05	10, 10	0.049	0.044	0.002	0.053	0.044	0.002
	10, 30	0.051	0.049	0.007	0.051	0.049	0.008
	10, 50	0.042	0.042	0.004	0.046	0.044	0.006
	30, 30	0.052	0.051	0.006	0.051	0.050	0.008
	30, 50	0.049	0.049	0.004	0.049	0.047	0.007
	30, 100	0.053	0.052	0.006	0.048	0.049	0.008
	50, 50	0.049	0.049	0.006	0.053	0.052	0.007
	50, 100	0.046	0.046	0.005	0.049	0.051	0.007
	50, 300	0.050	0.049	0.005	0.051	0.050	0.008
	100, 100	0.049	0.048	0.005	0.052	0.050	0.006
	100, 300	0.050	0.049	0.004	0.051	0.050	0.008
	300, 300	0.049	0.049	0.005	0.048	0.047	0.008
0.10	10, 10	0.096	0.088	0.011	0.102	0.093	0.014
	10, 30	0.101	0.098	0.011	0.098	0.094	0.016
	10, 50	0.106	0.106	0.013	0.099	0.098	0.015
	30, 30	0.101	0.100	0.013	0.094	0.094	0.016
	30, 50	0.096	0.096	0.012	0.099	0.098	0.019
	30, 100	0.098	0.098	0.013	0.098	0.097	0.018
	50, 50	0.095	0.096	0.010	0.100	0.098	0.014
	50, 100	0.103	0.102	0.013	0.103	0.100	0.015
	50, 300	0.099	0.100	0.014	0.096	0.098	0.016
	100, 100	0.100	0.101	0.011	0.101	0.098	0.018
	100, 300	0.099	0.101	0.013	0.105	0.104	0.018
	300, 300	0.098	0.101	0.017	0.099	0.100	0.018

Table 7 shows that, given the moderate skewed distribution, the empirical Type I error rates from the t-test and Mann-Whitney (MW) test are close to the nominal significance levels and followed the robustness criteria. However, the Kolmogorov-Smirnov (KS) test is not robust since the error rate is beyond the criteria's range for all circumstances.

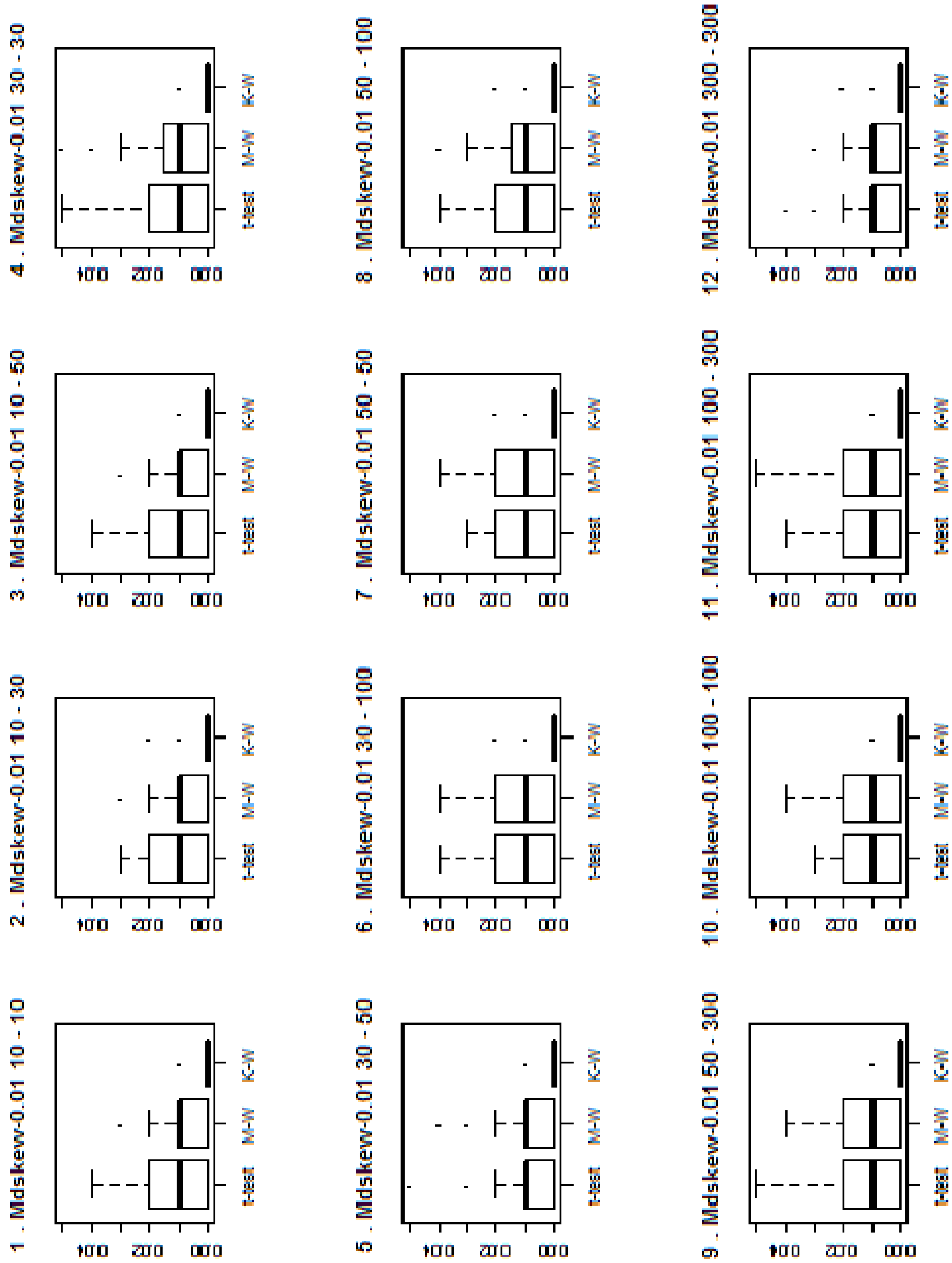
Figure 12. Distribution of Type I error rate estimates of the moderate skewed distribution at α = 0.01 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

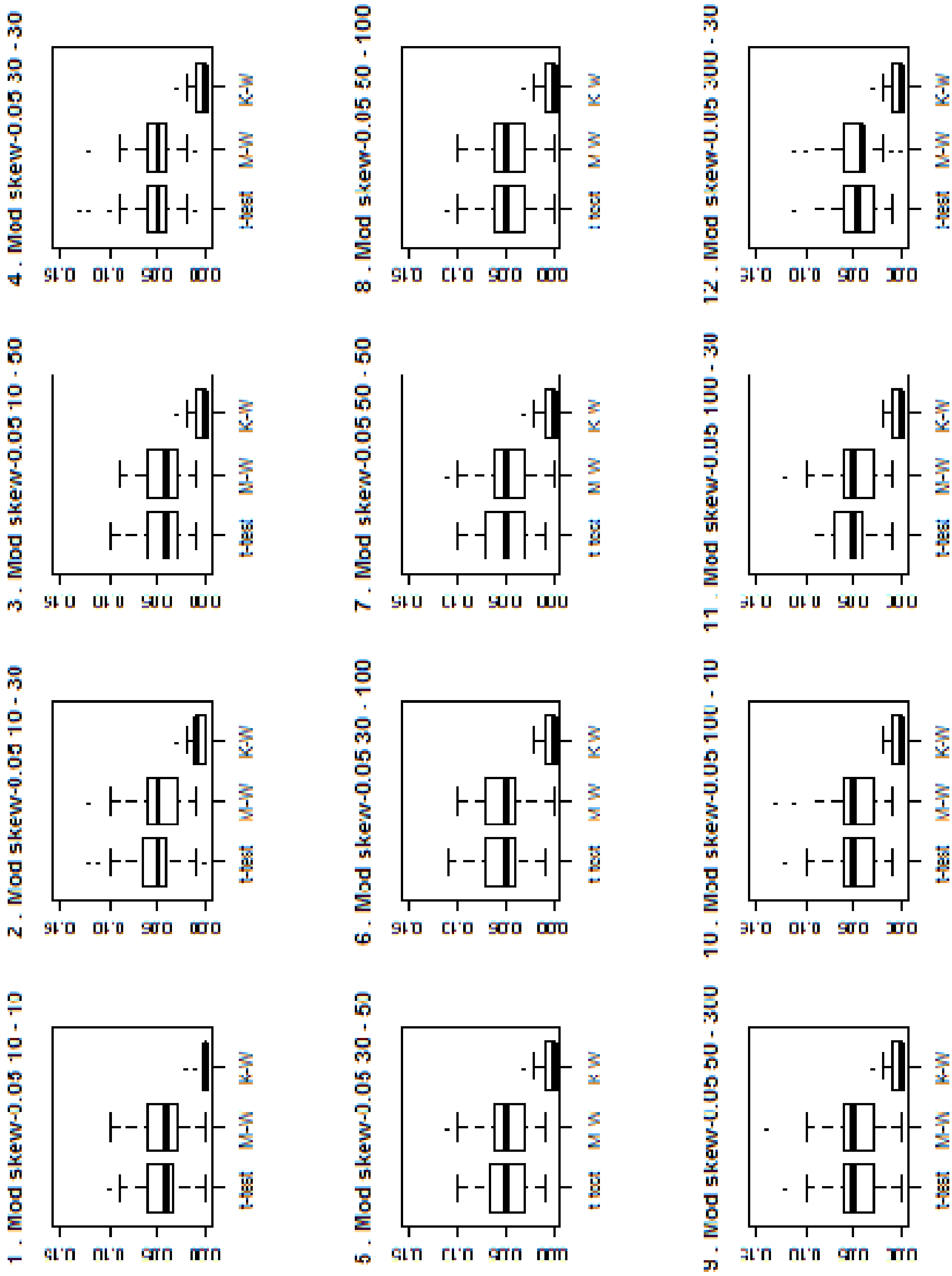
Figure 13. Distribution of Type I error rate estimates of the moderate skewed distribution

at α = 0.01 for the 7-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

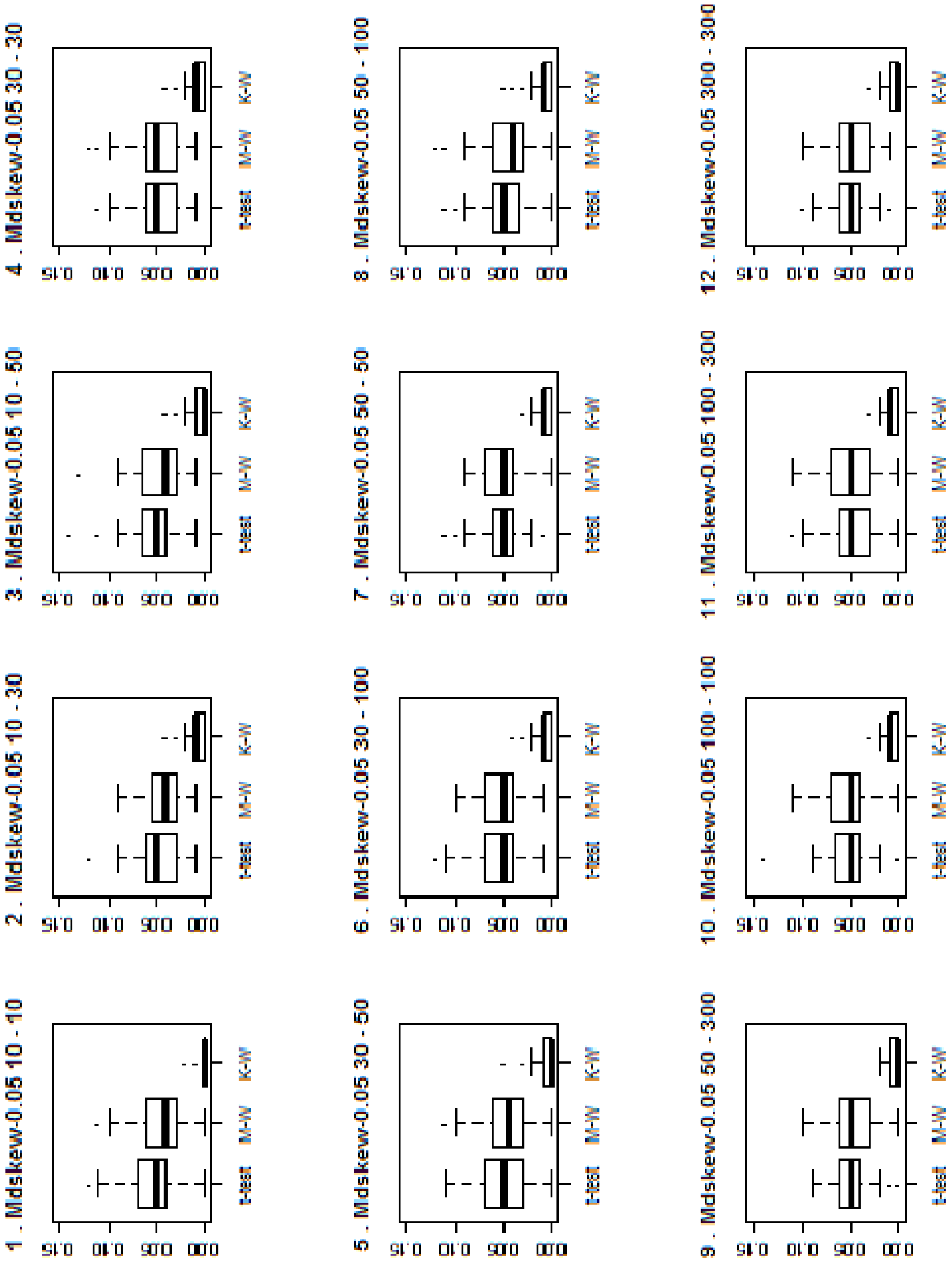
Figure 14. Distribution of Type I error rate estimates of the moderate skewed distribution at α = 0.05 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

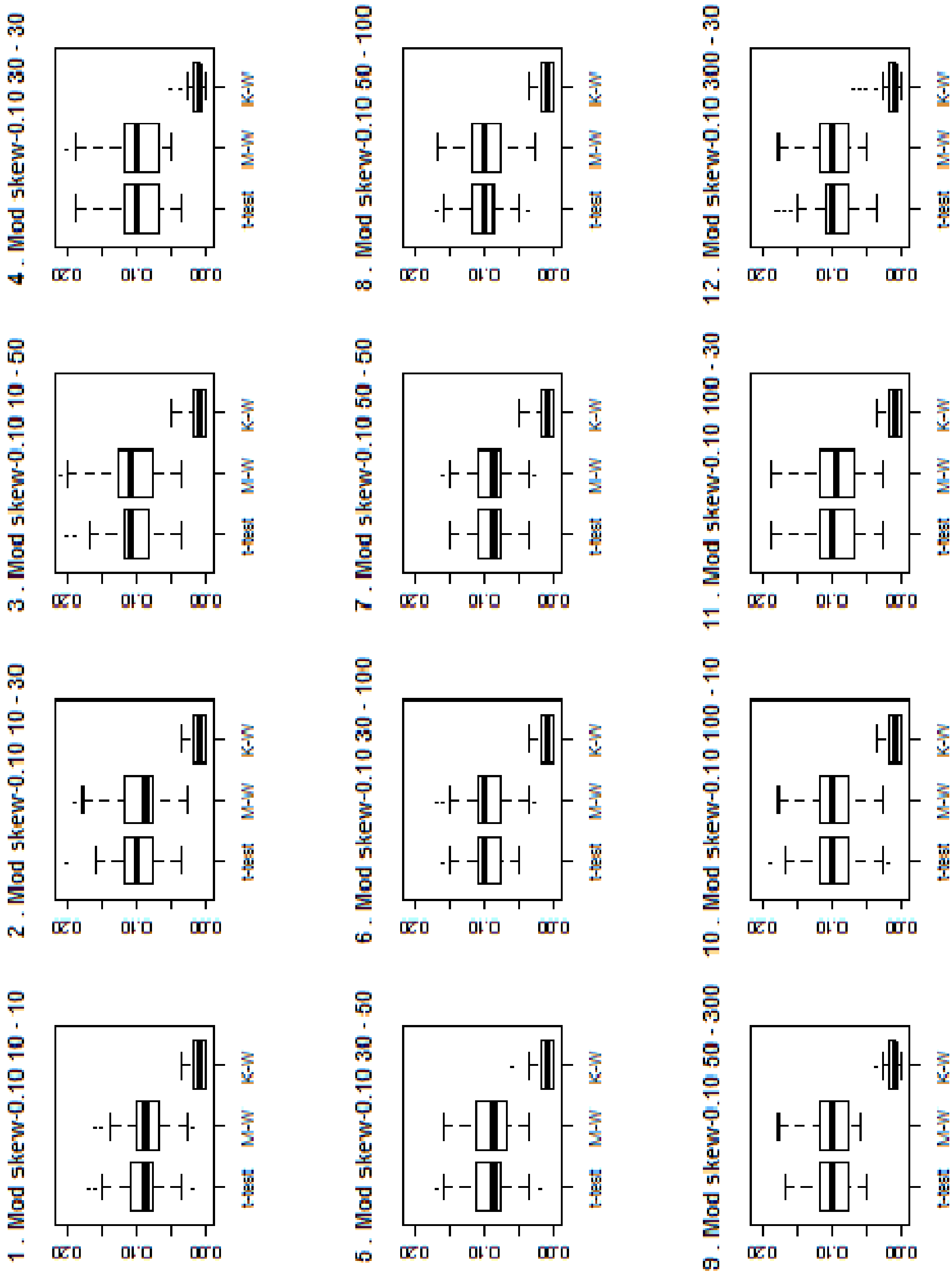
Figure 15. Distribution of Type I error rate estimates of the moderate skewed distribution

at α = 0.05 for the 7-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

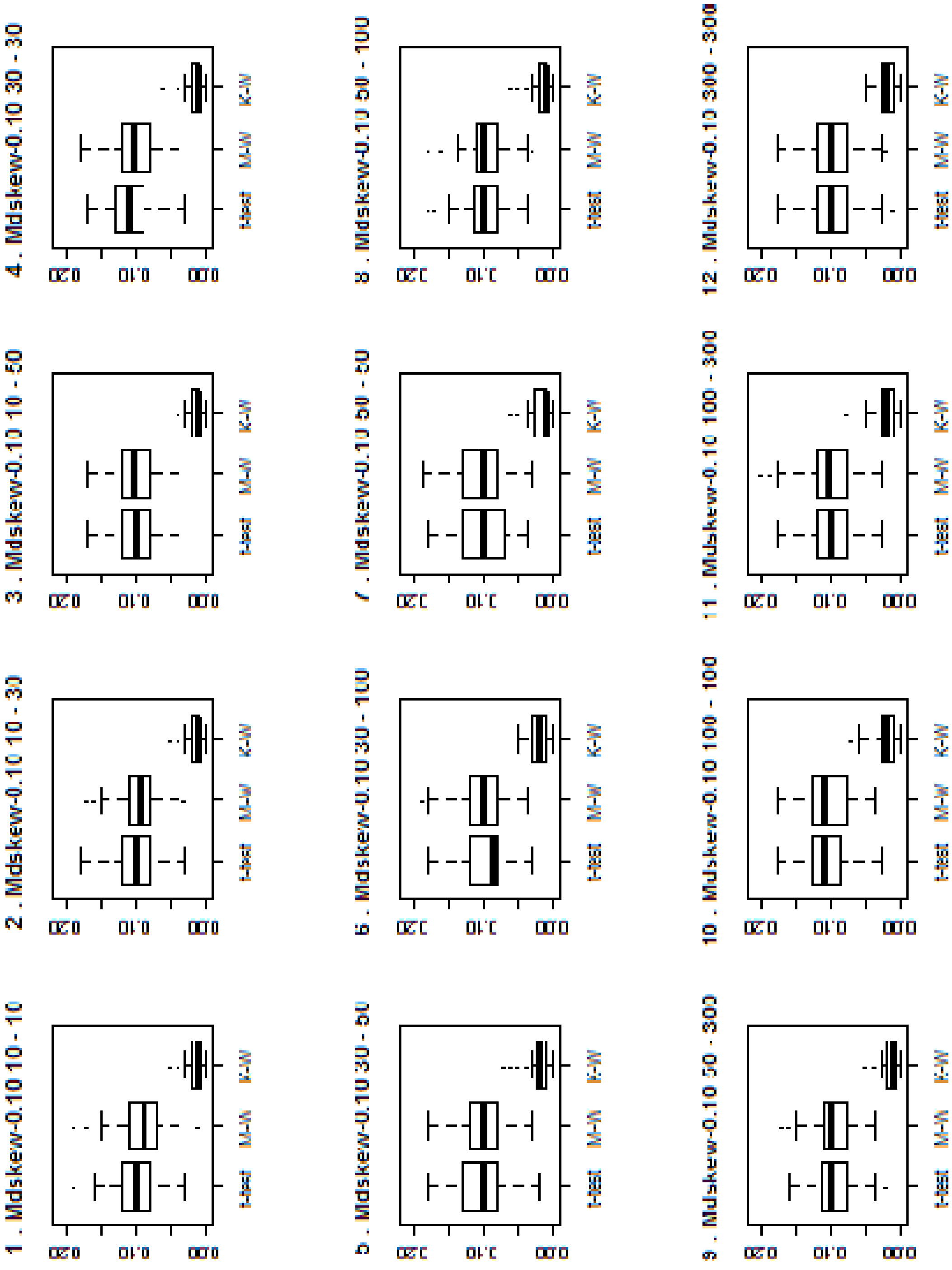
Figure 16. Distribution of Type I error rate estimates of the moderate skewed distribution at α = 0.10 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 17. Distribution of Type I error rate estimates of the moderate skewed distribution

at α = 0.10 for the 7-point Likert scale



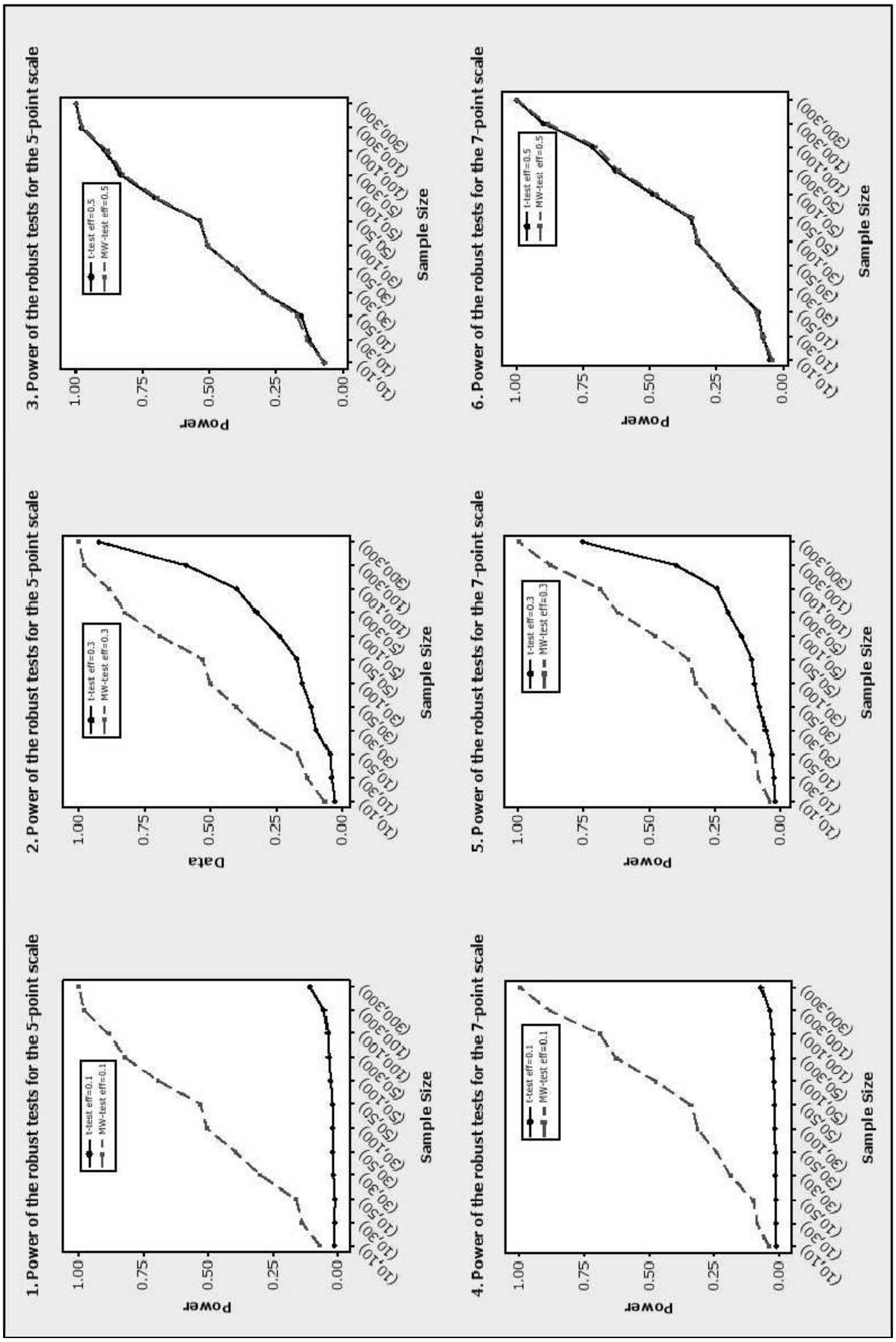
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Table 8. Statistical power estimates of the moderate skewed distribution at α = 0.01

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.012	0.065	0.041	0.011	0.039	0.020
10, 10	0.30	0.028	0.066	0.042	0.021	0.042	0.021
10, 10	0.50	0.069	0.066	0.041	0.050	0.041	0.019
10, 30	0.10	0.013	0.136	0.172	0.012	0.081	0.090
10, 30	0.30	0.040	0.131	0.164	0.026	0.082	0.089
10, 30	0.50	0.123	0.133	0.170	0.074	0.076	0.087
10, 50	0.10	0.011	0.158	0.170	0.012	0.096	0.090
10, 50	0.30	0.045	0.167	0.180	0.032	0.095	0.089
10, 50	0.50	0.153	0.169	0.179	0.089	0.096	0.091
30, 30	0.10	0.017	0.300	0.511	0.013	0.186	0.306
30, 30	0.30	0.096	0.308	0.506	0.056	0.175	0.296
30, 30	0.50	0.298	0.301	0.506	0.177	0.174	0.295
30, 50	0.10	0.020	0.392	0.649	0.013	0.239	0.416
30, 50	0.30	0.116	0.400	0.652	0.077	0.252	0.420
30, 50	0.50	0.398	0.398	0.655	0.243	0.242	0.422
30, 100	0.10	0.020	0.506	0.787	0.017	0.314	0.560
30, 100	0.30	0.150	0.501	0.781	0.096	0.324	0.574
30, 100	0.50	0.508	0.507	0.795	0.322	0.320	0.568
50, 50	0.10	0.020	0.531	0.831	0.018	0.339	0.598
50, 50	0.30	0.170	0.532	0.829	0.107	0.348	0.609
50, 50	0.50	0.537	0.533	0.836	0.345	0.340	0.601
50, 100	0.10	0.027	0.693	0.953	0.019	0.474	0.804
50, 100	0.30	0.237	0.690	0.950	0.146	0.475	0.809
50, 100	0.50	0.706	0.695	0.953	0.487	0.474	0.799
50, 300	0.10	0.033	0.822	0.990	0.023	0.626	0.932
50, 300	0.30	0.325	0.826	0.991	0.198	0.619	0.933
50, 300	0.50	0.836	0.825	0.990	0.627	0.615	0.933
100, 100	0.10	0.036	0.887	0.997	0.026	0.689	0.967
100, 100	0.30	0.398	0.886	0.998	0.241	0.686	0.965
100, 100	0.50	0.892	0.884	0.998	0.716	0.699	0.968
100, 300	0.10	0.051	0.980	1.000	0.036	0.884	0.999
100, 300	0.30	0.595	0.980	1.000	0.396	0.878	1.000
100, 300	0.50	0.982	0.976	1.000	0.896	0.882	0.999
300, 300	0.10	0.105	1.000	1.000	0.068	0.997	1.000
300, 300	0.30	0.921	1.000	1.000	0.754	0.996	1.000
300, 300	0.50	1.000	1.000	1.000	0.998	0.997	1.000

Table 8 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 18.

Figure 18. Statistical power estimates of the moderate skewed distribution at α = 0.01



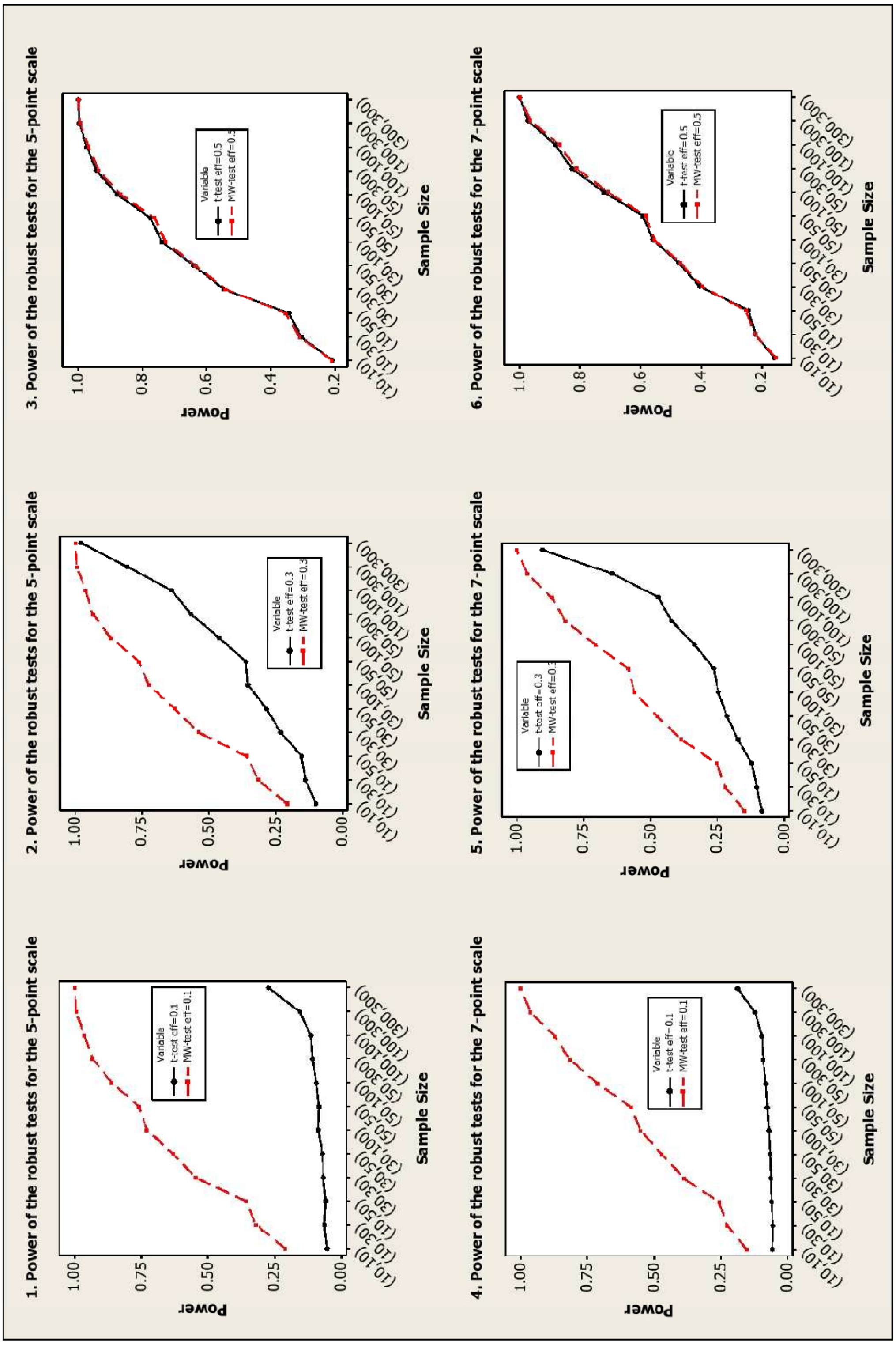
From Figure 18, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 9. Statistical power estimates of the moderate skewed distribution at α = 0.05

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.055	0.211	0.122	0.056	0.152	0.079
10, 10	0.30	0.101	0.208	0.124	0.082	0.148	0.073
10, 10	0.50	0.208	0.209	0.124	0.158	0.153	0.079
10, 30	0.10	0.064	0.321	0.382	0.054	0.225	0.260
10, 30	0.30	0.139	0.316	0.377	0.102	0.219	0.260
10, 30	0.50	0.305	0.311	0.382	0.220	0.222	0.257
10, 50	0.10	0.059	0.358	0.397	0.058	0.256	0.275
10, 50	0.30	0.155	0.360	0.399	0.121	0.251	0.269
10, 50	0.50	0.344	0.354	0.393	0.243	0.253	0.269
30, 30	0.10	0.068	0.547	0.745	0.062	0.388	0.556
30, 30	0.30	0.233	0.539	0.744	0.172	0.384	0.549
30, 30	0.50	0.547	0.541	0.736	0.405	0.396	0.557
30, 50	0.10	0.072	0.632	0.844	0.064	0.469	0.674
30, 50	0.30	0.286	0.628	0.842	0.214	0.474	0.676
30, 50	0.50	0.641	0.635	0.845	0.474	0.468	0.675
30, 100	0.10	0.087	0.729	0.935	0.068	0.550	0.794
30, 100	0.30	0.354	0.725	0.932	0.245	0.559	0.803
30, 100	0.50	0.740	0.728	0.934	0.560	0.555	0.798
50, 50	0.10	0.084	0.759	0.958	0.074	0.585	0.847
50, 50	0.30	0.362	0.762	0.960	0.262	0.583	0.854
50, 50	0.50	0.775	0.762	0.959	0.594	0.584	0.854
50, 100	0.10	0.095	0.861	0.993	0.080	0.711	0.947
50, 100	0.30	0.462	0.867	0.993	0.334	0.705	0.950
50, 100	0.50	0.879	0.868	0.993	0.721	0.708	0.944
50, 300	0.10	0.109	0.933	1.000	0.090	0.816	0.983
50, 300	0.30	0.568	0.934	1.000	0.420	0.817	0.985
50, 300	0.50	0.944	0.936	1.000	0.828	0.814	0.987
100, 100	0.10	0.115	0.965	1.000	0.094	0.871	0.997
100, 100	0.30	0.640	0.962	1.000	0.470	0.870	0.997
100, 100	0.50	0.973	0.968	1.000	0.882	0.870	0.996
100, 300	0.10	0.157	0.995	1.000	0.121	0.962	1.000
100, 300	0.30	0.805	0.994	1.000	0.643	0.960	1.000
100, 300	0.50	0.997	0.995	1.000	0.972	0.963	1.000
300, 300	0.10	0.274	1.000	1.000	0.187	1.000	1.000
300, 300	0.30	0.978	1.000	1.000	0.903	1.000	1.000
300, 300	0.50	1.000	1.000	1.000	1.000	1.000	1.000

Table 9 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 19.

Figure 19. Statistical power estimates of the moderate skewed distribution at α = 0.05



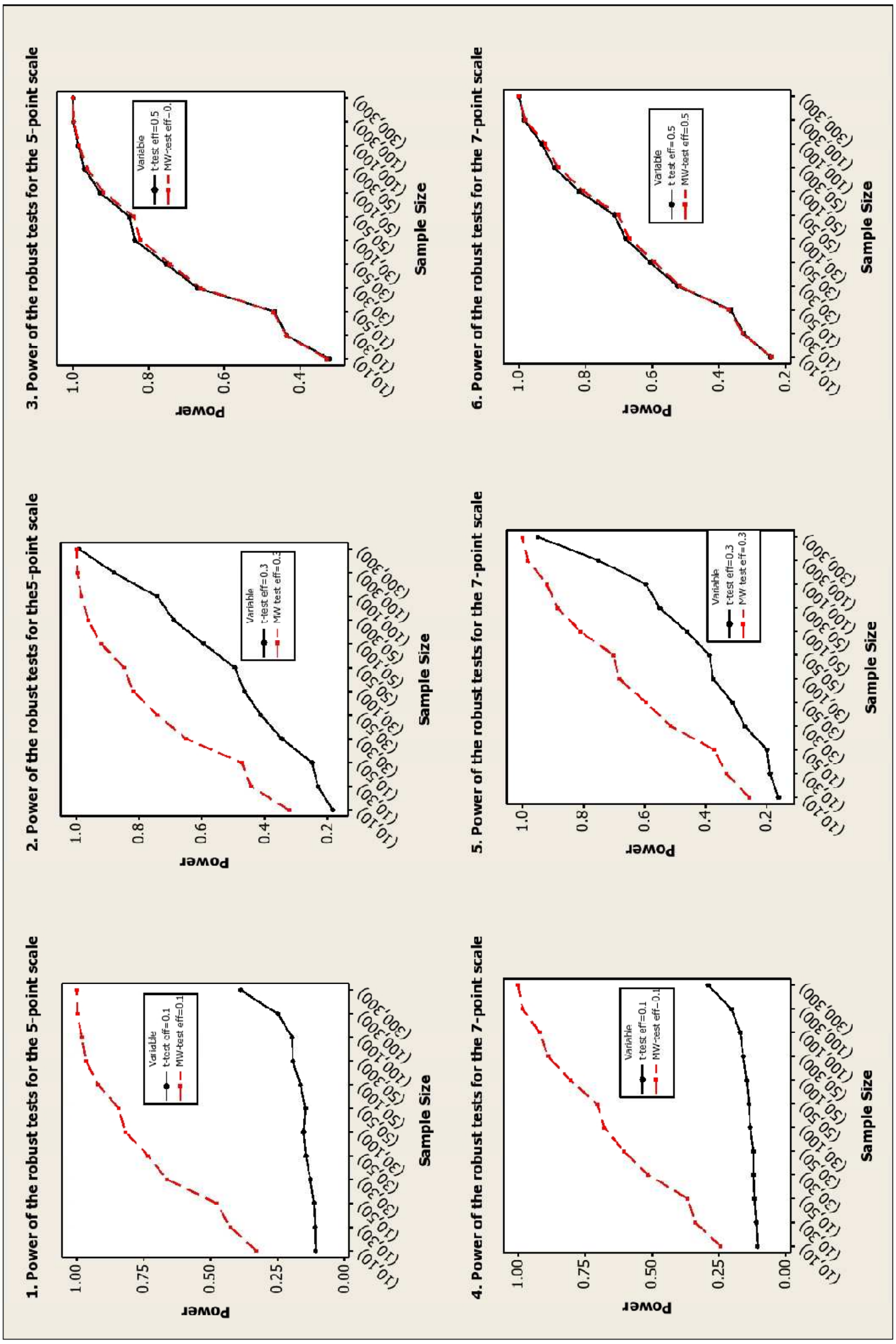
From Figure 19, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 10. Statistical power estimates of the moderate skewed distribution at α = 0.10

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.108	0.328	0.288	0.105	0.242	0.206
10, 10	0.30	0.183	0.322	0.283	0.160	0.256	0.210
10, 10	0.50	0.324	0.331	0.282	0.245	0.242	0.197
10, 30	0.10	0.109	0.427	0.456	0.109	0.338	0.343
10, 30	0.30	0.231	0.444	0.468	0.187	0.331	0.339
10, 30	0.50	0.437	0.438	0.460	0.325	0.328	0.339
10, 50	0.10	0.114	0.477	0.499	0.117	0.366	0.381
10, 50	0.30	0.249	0.473	0.504	0.198	0.370	0.382
10, 50	0.50	0.468	0.472	0.494	0.364	0.369	0.379
30, 30	0.10	0.128	0.663	0.839	0.120	0.514	0.685
30, 30	0.30	0.345	0.651	0.834	0.270	0.513	0.679
30, 30	0.50	0.673	0.665	0.837	0.524	0.517	0.684
30, 50	0.10	0.144	0.733	0.922	0.120	0.604	0.801
30, 50	0.30	0.412	0.742	0.919	0.310	0.594	0.801
30, 50	0.50	0.754	0.744	0.920	0.606	0.592	0.800
30, 100	0.10	0.152	0.817	0.969	0.133	0.677	0.886
30, 100	0.30	0.464	0.819	0.970	0.375	0.682	0.885
30, 100	0.50	0.837	0.823	0.969	0.680	0.668	0.882
50, 50	0.10	0.145	0.843	0.979	0.136	0.705	0.907
50, 50	0.30	0.495	0.849	0.979	0.386	0.702	0.908
50, 50	0.50	0.852	0.839	0.980	0.712	0.701	0.908
50, 100	0.10	0.164	0.919	0.998	0.145	0.805	0.975
50, 100	0.30	0.594	0.922	0.998	0.460	0.809	0.974
50, 100	0.50	0.928	0.919	0.998	0.819	0.805	0.975
50, 300	0.10	0.193	0.965	1.000	0.158	0.888	0.996
50, 300	0.30	0.690	0.963	1.000	0.549	0.884	0.995
50, 300	0.50	0.970	0.962	1.000	0.894	0.882	0.995
100, 100	0.10	0.195	0.980	1.000	0.170	0.919	1.000
100, 100	0.30	0.742	0.984	1.000	0.595	0.920	1.000
100, 100	0.50	0.987	0.984	1.000	0.931	0.921	0.999
100, 300	0.10	0.248	0.998	1.000	0.202	0.982	1.000
100, 300	0.30	0.882	0.998	1.000	0.750	0.982	1.000
100, 300	0.50	0.999	0.998	1.000	0.983	0.979	1.000
300, 300	0.10	0.387	1.000	1.000	0.290	1.000	1.000
300, 300	0.30	0.991	1.000	1.000	0.949	1.000	1.000
300, 300	0.50	1.000	1.000	1.000	1.000	1.000	1.000

Table 10 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 20.

Figure 20. Statistical power estimates of the moderate skewed distribution at α = 0.10



From Figure 20, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

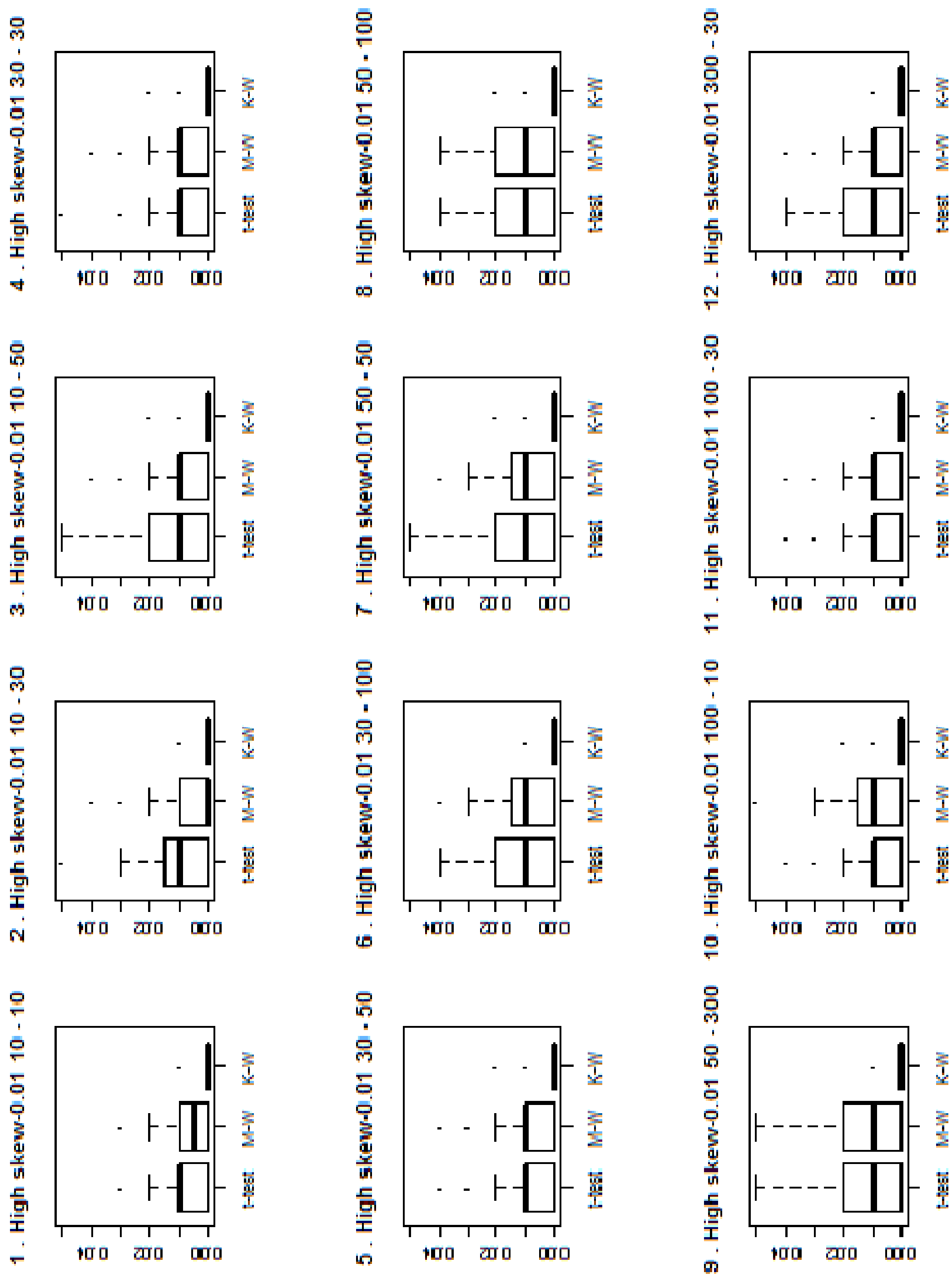
Table 11. Type I error rate estimates of the highly skewed distribution

Significance Level	Sample Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
0.01	10, 10	0.008	0.007	0.000	0.010	0.007	0.000
	10, 30	0.009	0.007	0.001	0.010	0.009	0.002
	10, 50	0.011	0.008	0.001	0.010	0.008	0.001
	30, 30	0.008	0.009	0.001	0.009	0.009	0.001
	30, 50	0.009	0.008	0.000	0.011	0.010	0.001
	30, 100	0.010	0.009	0.001	0.009	0.008	0.001
	50, 50	0.010	0.010	0.001	0.010	0.009	0.001
	50, 100	0.011	0.011	0.002	0.010	0.010	0.001
	50, 300	0.012	0.011	0.001	0.009	0.009	0.001
	100, 100	0.010	0.010	0.001	0.011	0.011	0.000
	100, 300	0.008	0.008	0.001	0.010	0.009	0.001
	300, 300	0.010	0.009	0.001	0.010	0.010	0.001
0.05	10, 10	0.050	0.040	0.002	0.050	0.048	0.003
	10, 30	0.050	0.049	0.009	0.049	0.047	0.007
	10, 50	0.050	0.049	0.005	0.049	0.046	0.005
	30, 30	0.047	0.045	0.006	0.049	0.050	0.005
	30, 50	0.054	0.054	0.008	0.050	0.050	0.005
	30, 100	0.055	0.055	0.008	0.047	0.047	0.005
	50, 50	0.049	0.049	0.007	0.052	0.050	0.006
	50, 100	0.050	0.049	0.007	0.049	0.050	0.005
	50, 300	0.051	0.048	0.006	0.047	0.045	0.005
	100, 100	0.052	0.050	0.005	0.050	0.048	0.006
	100, 300	0.051	0.050	0.006	0.050	0.049	0.005
	300, 300	0.051	0.050	0.005	0.050	0.051	0.007
0.10	10, 10	0.095	0.089	0.011	0.104	0.095	0.012
	10, 30	0.098	0.097	0.012	0.095	0.092	0.012
	10, 50	0.107	0.104	0.014	0.097	0.096	0.014
	30, 30	0.100	0.097	0.016	0.098	0.096	0.013
	30, 50	0.091	0.095	0.013	0.101	0.100	0.014
	30, 100	0.100	0.100	0.014	0.100	0.100	0.015
	50, 50	0.100	0.099	0.013	0.098	0.098	0.013
	50, 100	0.098	0.098	0.014	0.103	0.098	0.012
	50, 300	0.098	0.100	0.014	0.101	0.100	0.014
	100, 100	0.100	0.100	0.013	0.100	0.100	0.015
	100, 300	0.098	0.100	0.014	0.095	0.093	0.014
	300, 300	0.095	0.089	0.011	0.099	0.098	0.016

Table 11 shows that, given the highly skewed distribution, the empirical Type I error rates from the t-test and Mann-Whitney (MW) test are close to the nominal significance levels and followed the robustness criteria. However, the Kolmogorov-Smirnov (KS) test is not robust since the error rate is beyond the criteria's range for all circumstances.

Figure 21. Distribution of Type I error rate estimates of the highly skewed distribution

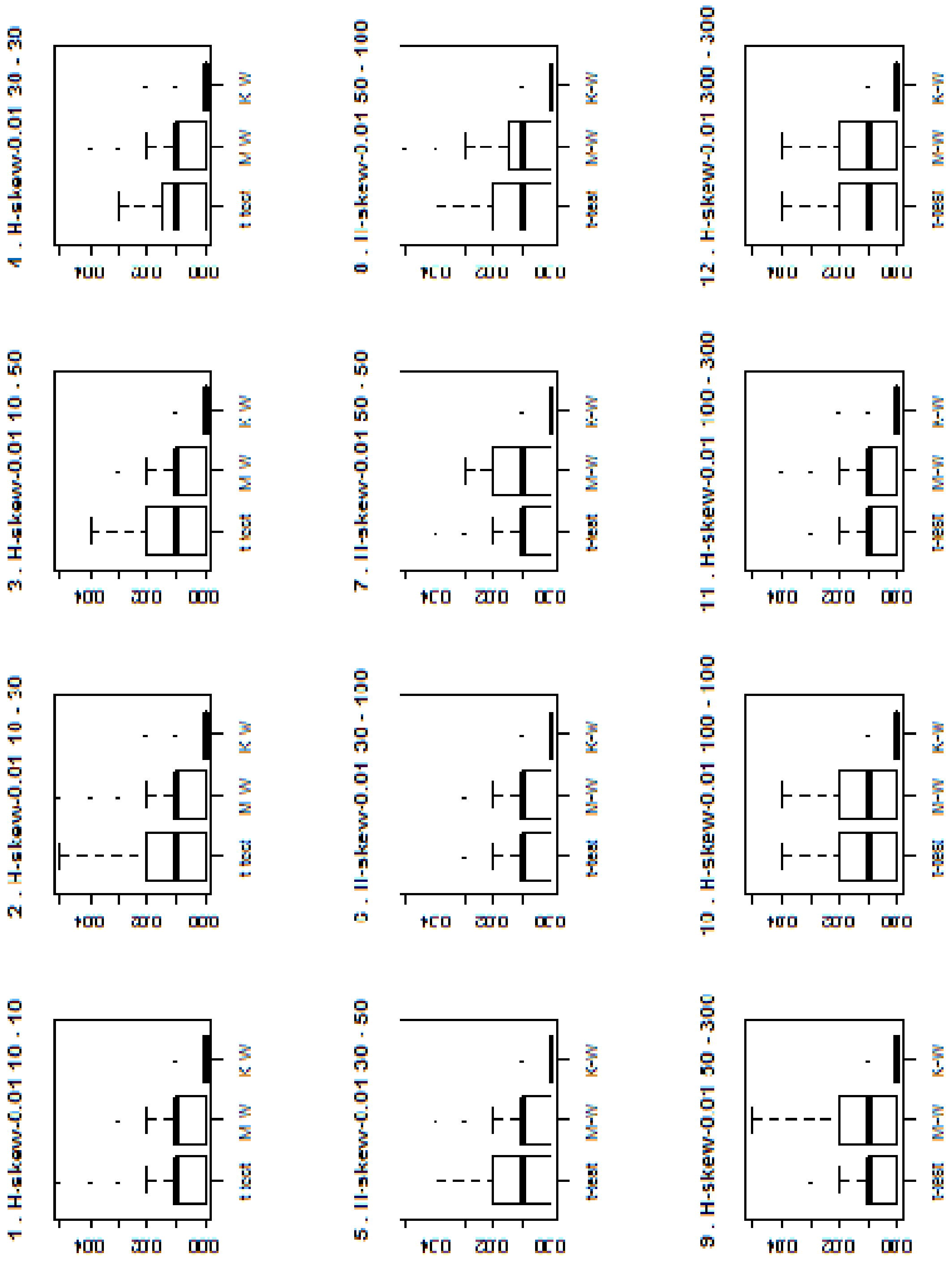
at α = 0.01 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

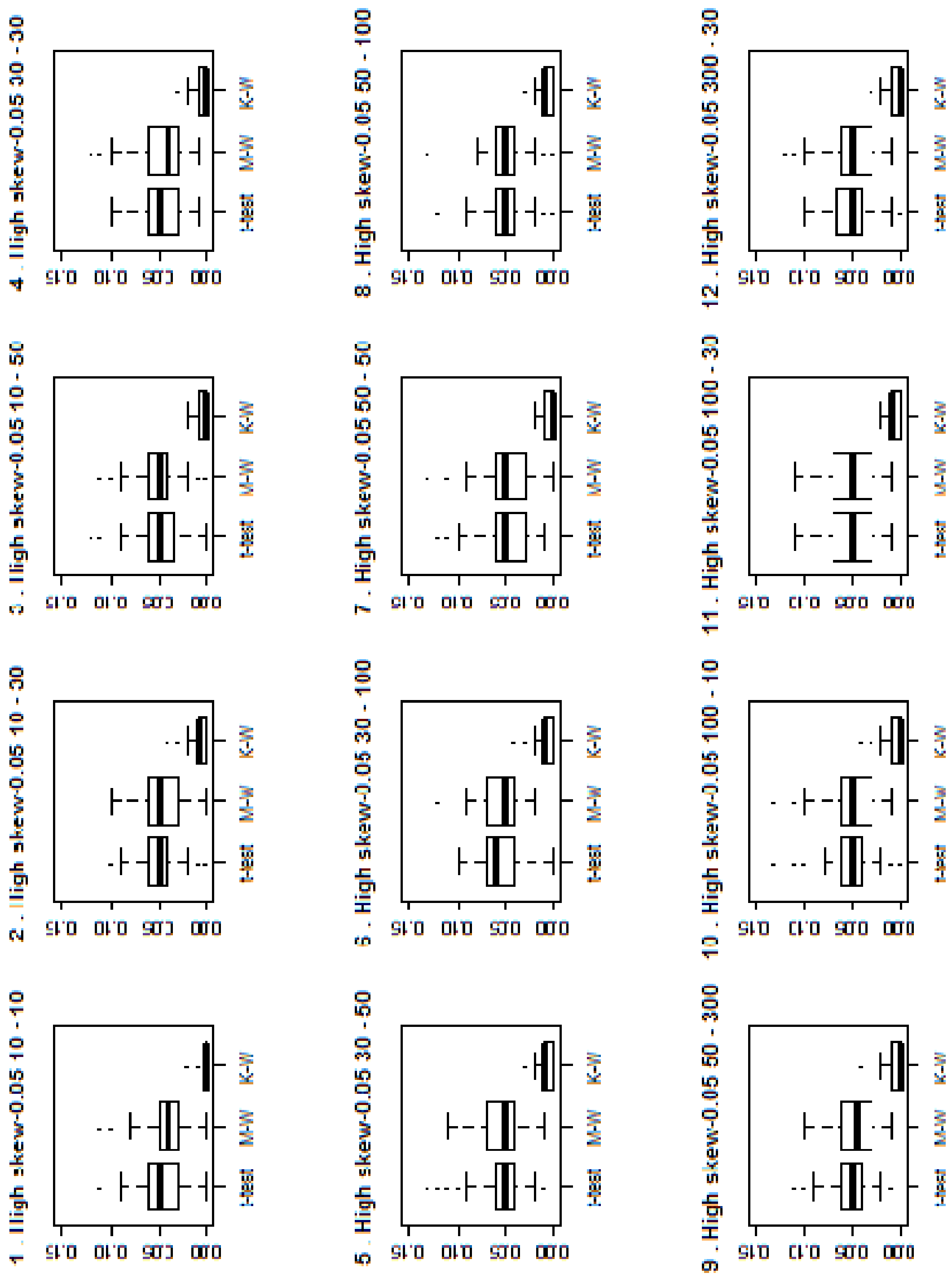
Figure 22. Distribution of Type I error rate estimates of the highly skewed distribution

at α = 0.01 for the 7-point Likert scale



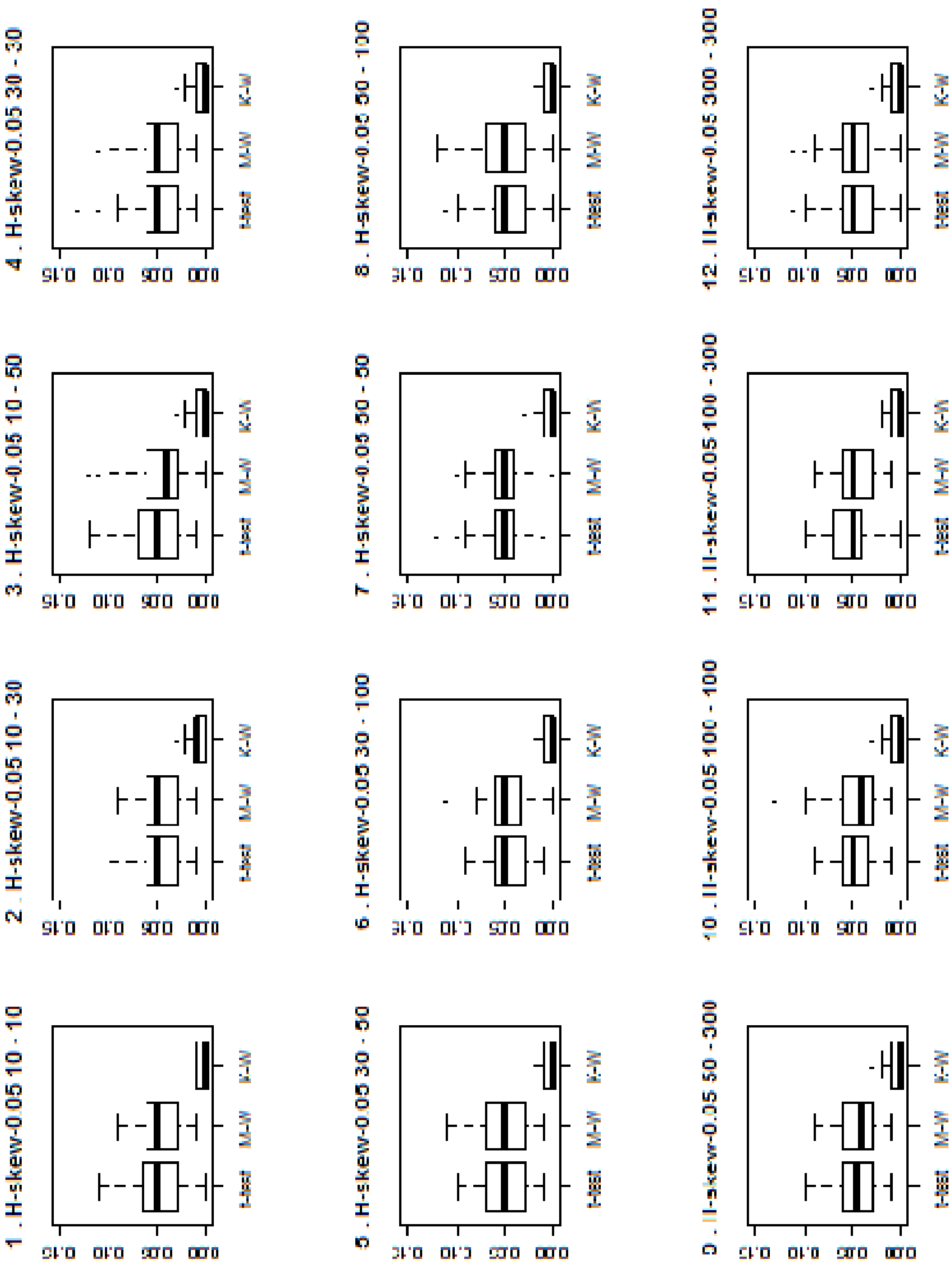
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 23. Distribution of Type I error rate estimates of the highly skewed distribution at α = 0.05 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

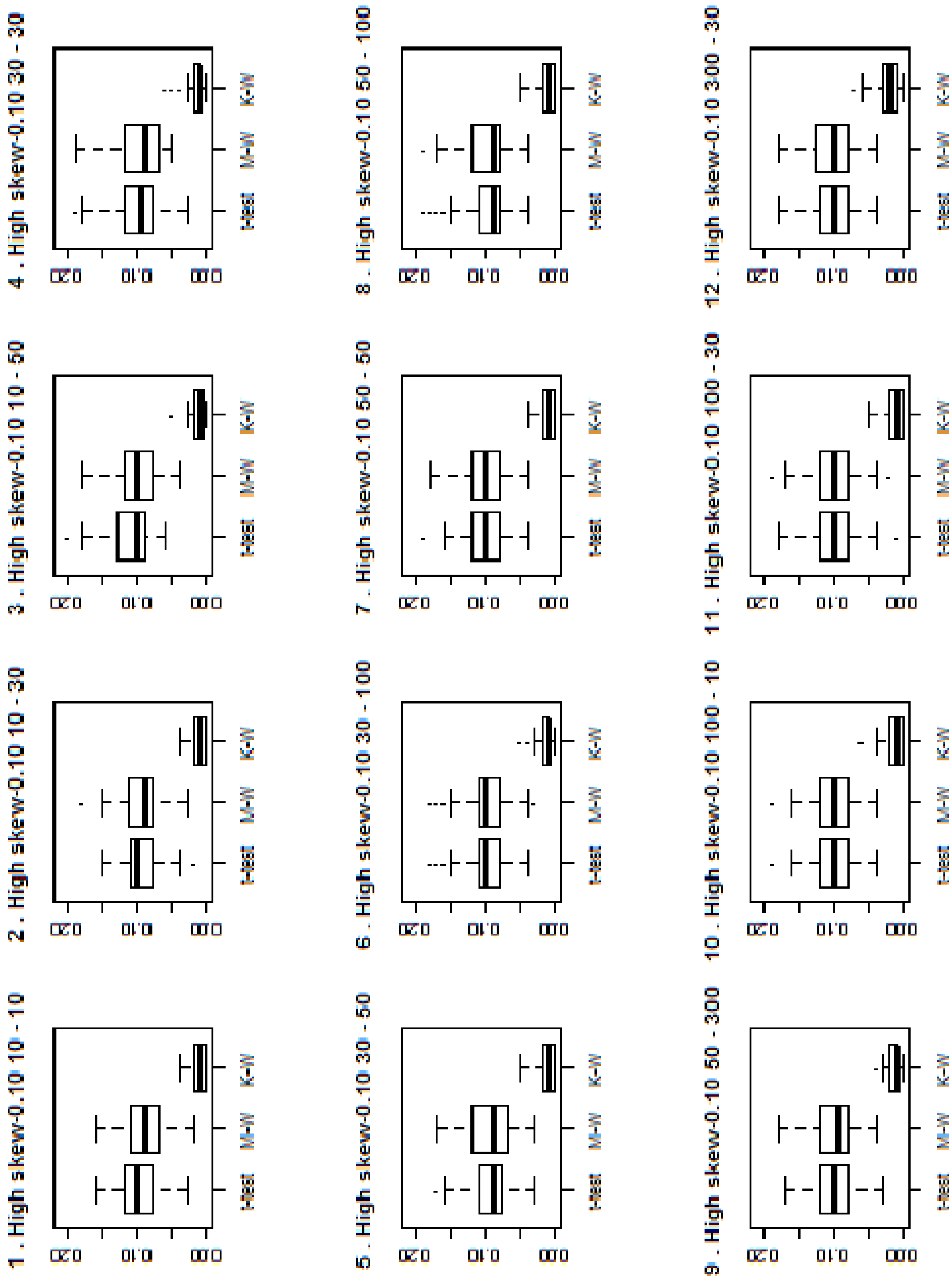
Figure 24. Distribution of Type I error rate estimates of the highly skewed distribution at α = 0.05 for the 7-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

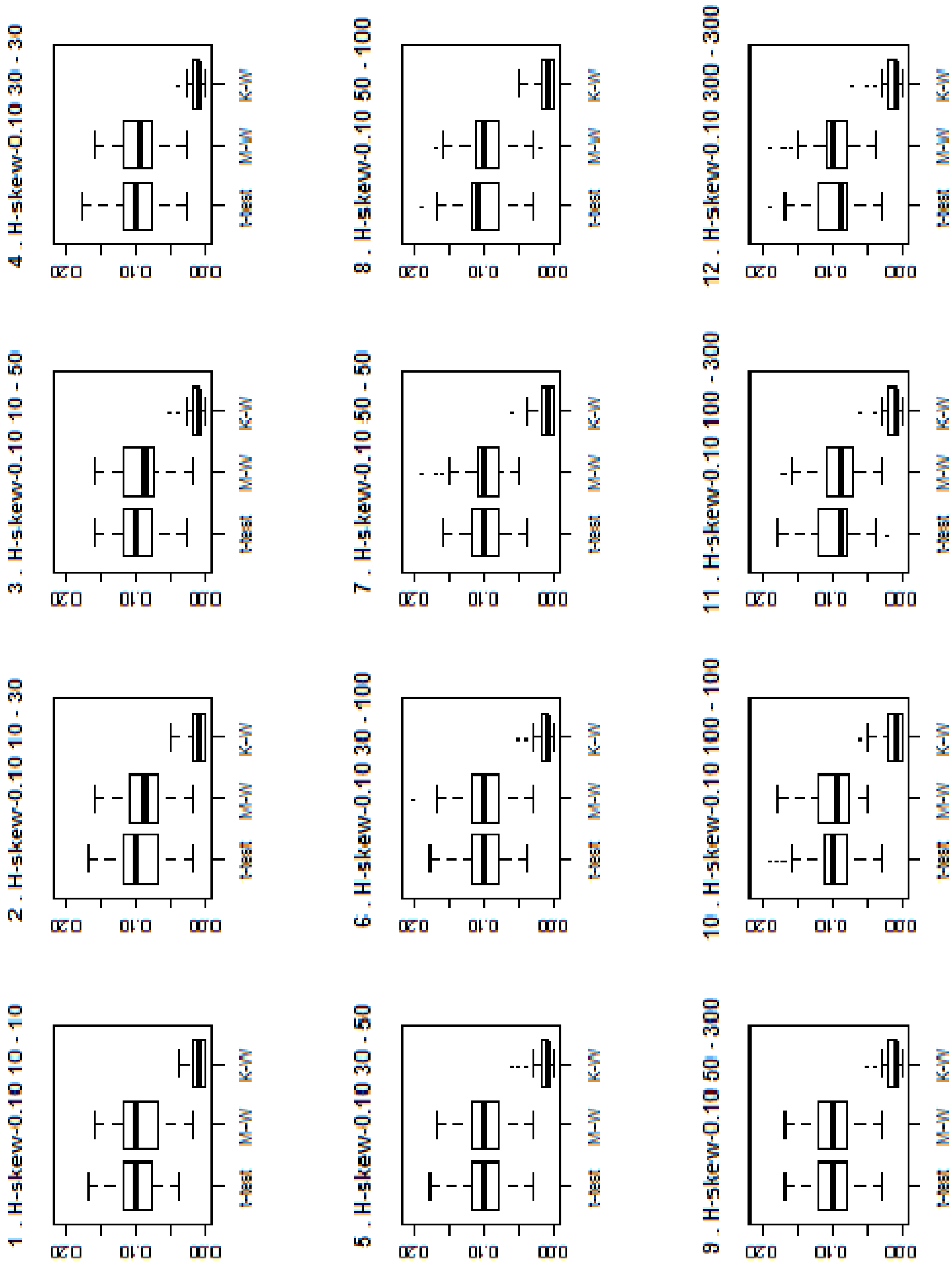
Figure 25. Distribution of Type I error rate estimates of the highly skewed distribution

at α = 0.10 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 26. Distribution of Type I error rate estimates of the highly skewed distribution at α = 0.10 for the 7-point Likert scale



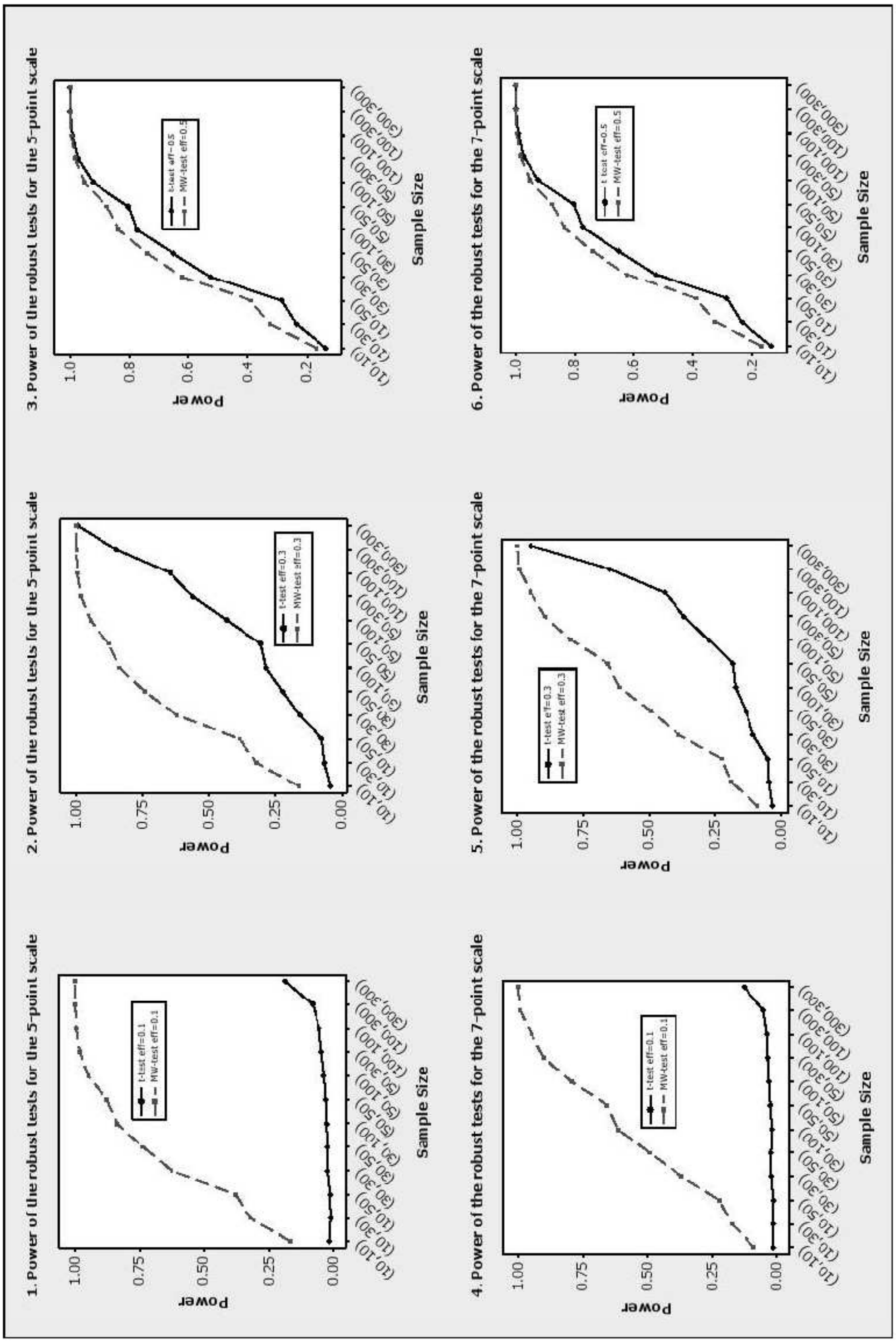
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Table 12..Statistical power estimates of the highly skewed distribution at α = 0.01

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.015	0.167	0.087	0.014	0.089	0.043
10, 10	0.30	0.042	0.163	0.088	0.031	0.087	0.041
10, 10	0.50	0.135	0.167	0.092	0.082	0.087	0.043
10, 30	0.10	0.011	0.321	0.446	0.012	0.173	0.215
10, 30	0.30	0.066	0.320	0.445	0.044	0.188	0.226
10, 30	0.50	0.236	0.325	0.448	0.136	0.176	0.215
10, 50	0.10	0.013	0.381	0.443	0.012	0.222	0.220
10, 50	0.30	0.076	0.388	0.447	0.048	0.223	0.224
10, 50	0.50	0.286	0.389	0.446	0.165	0.214	0.217
30, 30	0.10	0.022	0.627	0.903	0.020	0.368	0.577
30, 30	0.30	0.160	0.621	0.910	0.103	0.386	0.586
30, 30	0.50	0.526	0.622	0.910	0.332	0.381	0.583
30, 50	0.10	0.023	0.738	0.956	0.020	0.492	0.734
30, 50	0.30	0.224	0.743	0.956	0.132	0.491	0.729
30, 50	0.50	0.651	0.741	0.958	0.438	0.493	0.731
30, 100	0.10	0.025	0.836	0.984	0.018	0.611	0.860
30, 100	0.30	0.285	0.836	0.983	0.172	0.610	0.864
30, 100	0.50	0.774	0.836	0.983	0.567	0.615	0.863
50, 50	0.10	0.028	0.876	0.999	0.023	0.654	0.913
50, 50	0.30	0.306	0.877	0.998	0.181	0.653	0.918
50, 50	0.50	0.804	0.876	0.999	0.600	0.659	0.919
50, 100	0.10	0.036	0.950	1.000	0.029	0.792	0.980
50, 100	0.30	0.433	0.947	1.000	0.271	0.796	0.982
50, 100	0.50	0.923	0.952	1.000	0.747	0.784	0.982
50, 300	0.10	0.047	0.982	1.000	0.034	0.898	0.998
50, 300	0.30	0.562	0.983	1.000	0.366	0.893	0.998
50, 300	0.50	0.974	0.984	1.000	0.872	0.896	0.999
100, 100	0.10	0.056	0.997	1.000	0.037	0.947	1.000
100, 100	0.30	0.646	0.997	1.000	0.439	0.946	1.000
100, 100	0.50	0.992	0.997	1.000	0.925	0.947	1.000
100, 300	0.10	0.080	1.000	1.000	0.052	0.993	1.000
100, 300	0.30	0.849	0.999	1.000	0.646	0.992	1.000
100, 300	0.50	1.000	1.000	1.000	0.990	0.992	1.000
300, 300	0.10	0.188	1.000	1.000	0.125	1.000	1.000
300, 300	0.30	0.995	1.000	1.000	0.947	1.000	1.000
300, 300	0.50	1.000	1.000	1.000	1.000	1.000	1.000

Table 12 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 27.

Figure 27. Statistical power estimates of the highly skewed distribution at $\alpha = 0.01$



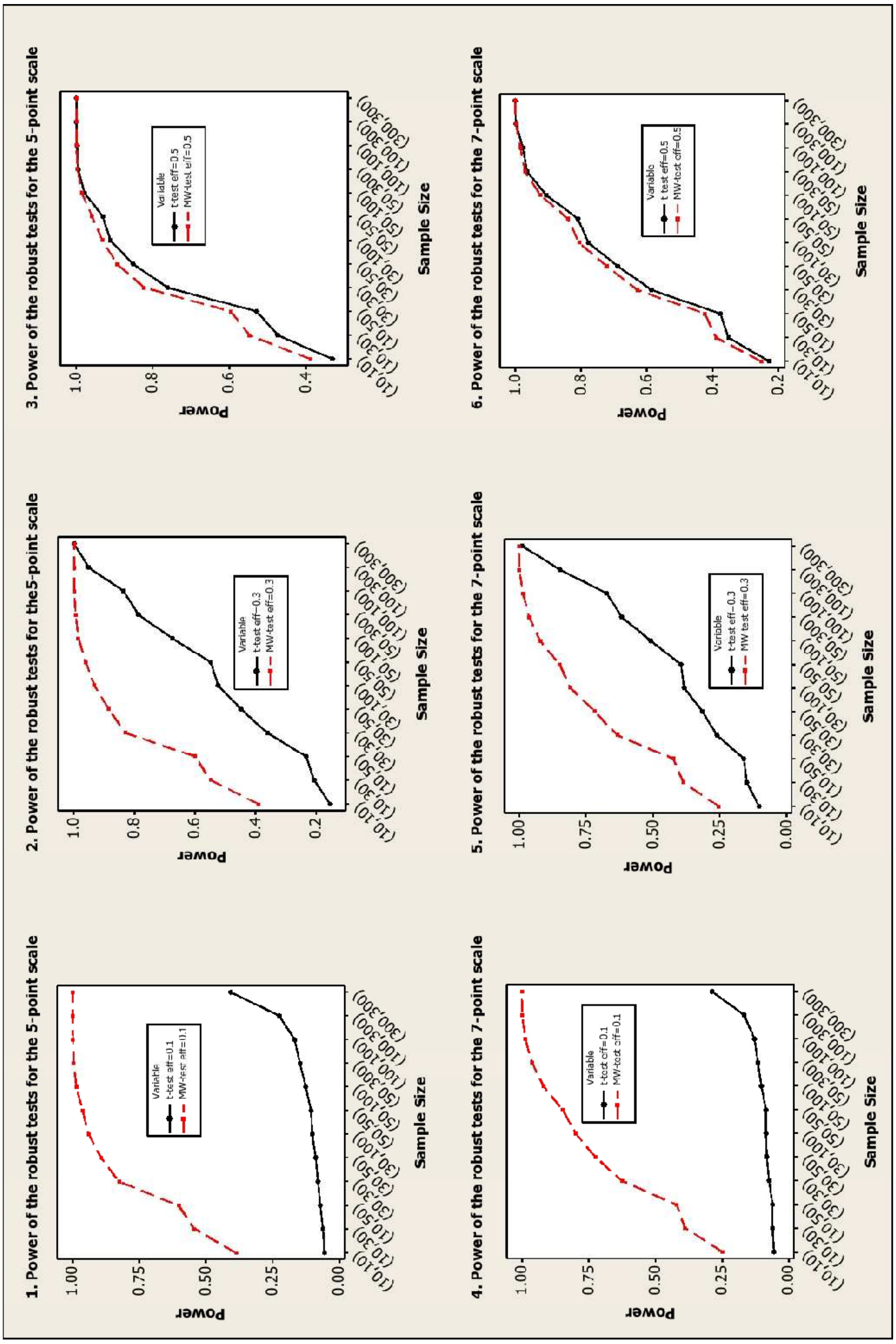
From Figure 27, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 13. Statistical power estimates of the highly skewed distribution at α = 0.05

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.056	0.388	0.254	0.054	0.247	0.130
10, 10	0.30	0.154	0.390	0.263	0.102	0.252	0.131
10, 10	0.50	0.331	0.389	0.260	0.228	0.253	0.132
10, 30	0.10	0.062	0.544	0.724	0.060	0.388	0.493
10, 30	0.30	0.207	0.549	0.723	0.148	0.385	0.486
10, 30	0.50	0.474	0.548	0.727	0.351	0.389	0.489
10, 50	0.10	0.071	0.603	0.724	0.060	0.420	0.485
10, 50	0.30	0.233	0.600	0.720	0.160	0.423	0.493
10, 50	0.50	0.529	0.597	0.719	0.376	0.423	0.491
30, 30	0.10	0.080	0.826	0.983	0.073	0.624	0.838
30, 30	0.30	0.361	0.830	0.986	0.259	0.631	0.838
30, 30	0.50	0.762	0.823	0.985	0.586	0.629	0.831
30, 50	0.10	0.089	0.894	0.996	0.082	0.725	0.919
30, 50	0.30	0.448	0.885	0.995	0.314	0.716	0.918
30, 50	0.50	0.852	0.894	0.996	0.688	0.721	0.924
30, 100	0.10	0.102	0.940	1.000	0.084	0.798	0.966
30, 100	0.30	0.524	0.933	1.000	0.383	0.808	0.972
30, 100	0.50	0.911	0.932	0.999	0.778	0.806	0.966
50, 50	0.10	0.106	0.963	1.000	0.085	0.847	0.992
50, 50	0.30	0.549	0.961	1.000	0.394	0.848	0.994
50, 50	0.50	0.931	0.959	1.000	0.809	0.840	0.992
50, 100	0.10	0.126	0.986	1.000	0.102	0.920	0.999
50, 100	0.30	0.675	0.986	1.000	0.507	0.921	0.999
50, 100	0.50	0.979	0.986	1.000	0.906	0.923	0.999
50, 300	0.10	0.148	0.996	1.000	0.114	0.962	1.000
50, 300	0.30	0.787	0.995	1.000	0.616	0.962	1.000
50, 300	0.50	0.995	0.997	1.000	0.964	0.970	1.000
100, 100	0.10	0.168	0.999	1.000	0.127	0.987	1.000
100, 100	0.30	0.837	1.000	1.000	0.672	0.985	1.000
100, 100	0.50	0.998	1.000	1.000	0.977	0.984	1.000
100, 300	0.10	0.225	1.000	1.000	0.166	0.999	1.000
100, 300	0.30	0.952	1.000	1.000	0.846	0.999	1.000
100, 300	0.50	1.000	1.000	1.000	0.998	0.998	1.000
300, 300	0.10	0.407	1.000	1.000	0.287	1.000	1.000
300, 300	0.30	1.000	1.000	1.000	0.988	1.000	1.000
300, 300	0.50	1.000	1.000	1.000	1.000	1.000	1.000

Table 13 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 28.

Figure 28. Statistical power estimates of the highly skewed distribution at α = 0.05



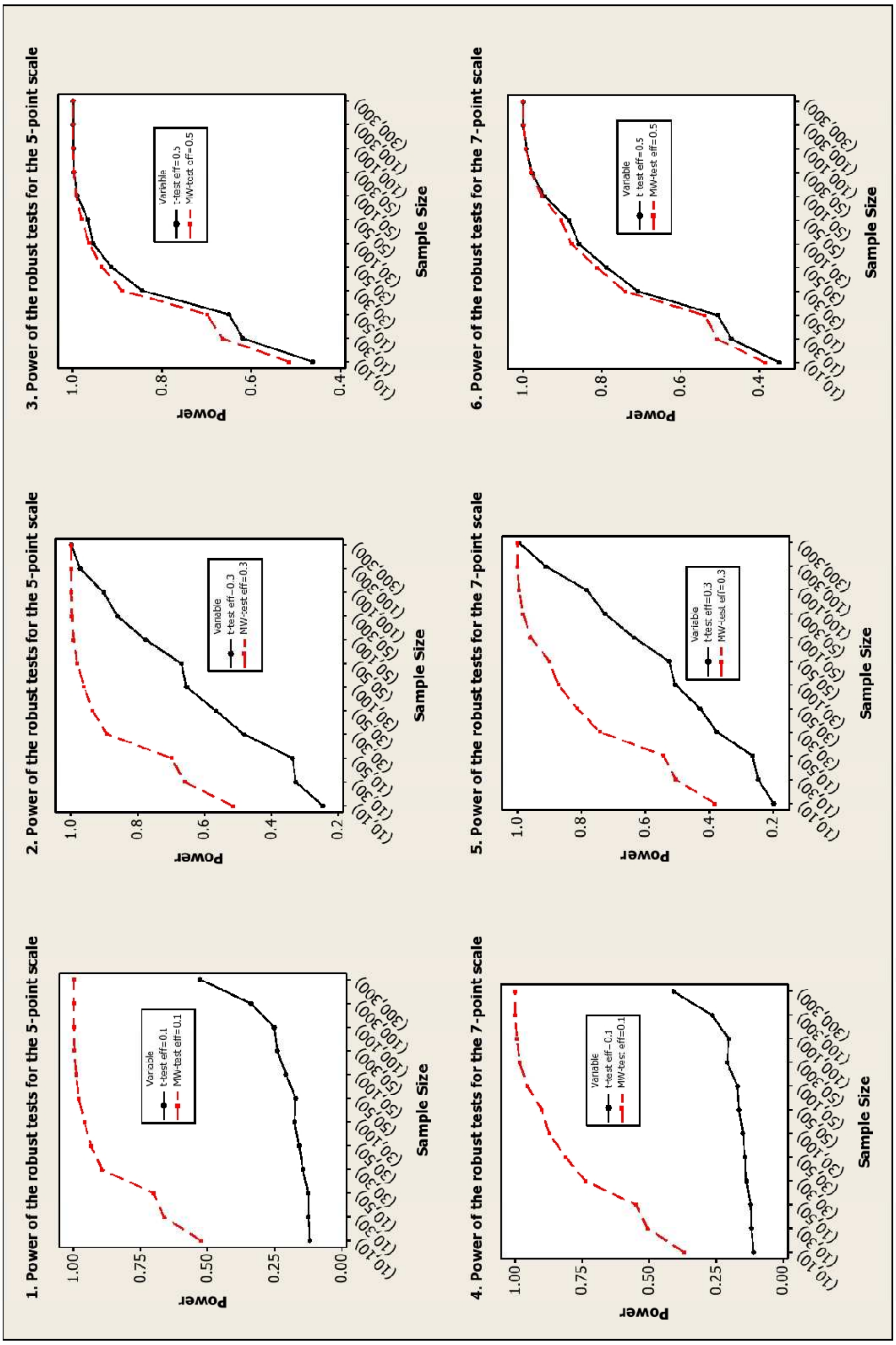
From Figure 28, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 14. .Statistical power estimates of the highly skewed distribution at α = 0.10

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KW-test	t-test	MW-test	KW-test
10, 10	0.10	0.120	0.526	0.526	0.110	0.369	0.315
10, 10	0.30	0.246	0.516	0.522	0.198	0.382	0.327
10, 10	0.50	0.462	0.516	0.514	0.348	0.384	0.323
10, 30	0.10	0.125	0.664	0.749	0.118	0.506	0.537
10, 30	0.30	0.327	0.661	0.749	0.246	0.504	0.540
10, 30	0.50	0.619	0.666	0.747	0.470	0.506	0.538
10, 50	0.10	0.125	0.703	0.753	0.121	0.550	0.569
10, 50	0.30	0.336	0.699	0.751	0.264	0.542	0.562
10, 50	0.50	0.650	0.700	0.762	0.505	0.538	0.560
30, 30	0.10	0.146	0.895	0.996	0.136	0.738	0.917
30, 30	0.30	0.482	0.892	0.996	0.375	0.742	0.918
30, 30	0.50	0.845	0.890	0.996	0.708	0.741	0.921
30, 50	0.10	0.160	0.936	1.000	0.142	0.811	0.967
30, 50	0.30	0.566	0.936	0.999	0.428	0.813	0.968
30, 50	0.50	0.915	0.935	1.000	0.788	0.813	0.967
30, 100	0.10	0.177	0.960	1.000	0.150	0.873	0.991
30, 100	0.30	0.655	0.961	1.000	0.506	0.872	0.989
30, 100	0.50	0.955	0.964	1.000	0.858	0.877	0.990
50, 50	0.10	0.173	0.982	1.000	0.164	0.900	0.997
50, 50	0.30	0.670	0.982	1.000	0.525	0.900	0.998
50, 50	0.50	0.967	0.980	1.000	0.883	0.904	0.998
50, 100	0.10	0.208	0.992	1.000	0.170	0.954	1.000
50, 100	0.30	0.777	0.993	1.000	0.634	0.958	1.000
50, 100	0.50	0.990	0.993	1.000	0.946	0.952	1.000
50, 300	0.10	0.241	0.998	1.000	0.208	0.983	1.000
50, 300	0.30	0.862	0.998	1.000	0.725	0.982	1.000
50, 300	0.50	0.998	0.998	1.000	0.978	0.979	1.000
100, 100	0.10	0.253	1.000	1.000	0.203	0.994	1.000
100, 100	0.30	0.904	1.000	1.000	0.783	0.994	1.000
100, 100	0.50	0.999	1.000	1.000	0.991	0.993	1.000
100, 300	0.10	0.339	1.000	1.000	0.264	1.000	1.000
100, 300	0.30	0.973	1.000	1.000	0.910	1.000	1.000
100, 300	0.50	1.000	1.000	1.000	1.000	0.999	1.000
300, 300	0.10	0.530	1.000	1.000	0.408	1.000	1.000
300, 300	0.30	1.000	1.000	1.000	0.995	1.000	1.000
300, 300	0.50	1.000	1.000	1.000	1.000	1.000	1.000

Table 14 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 29.

Figure 29. .Statistical power estimates of the highly skewed distribution at α = 0.10



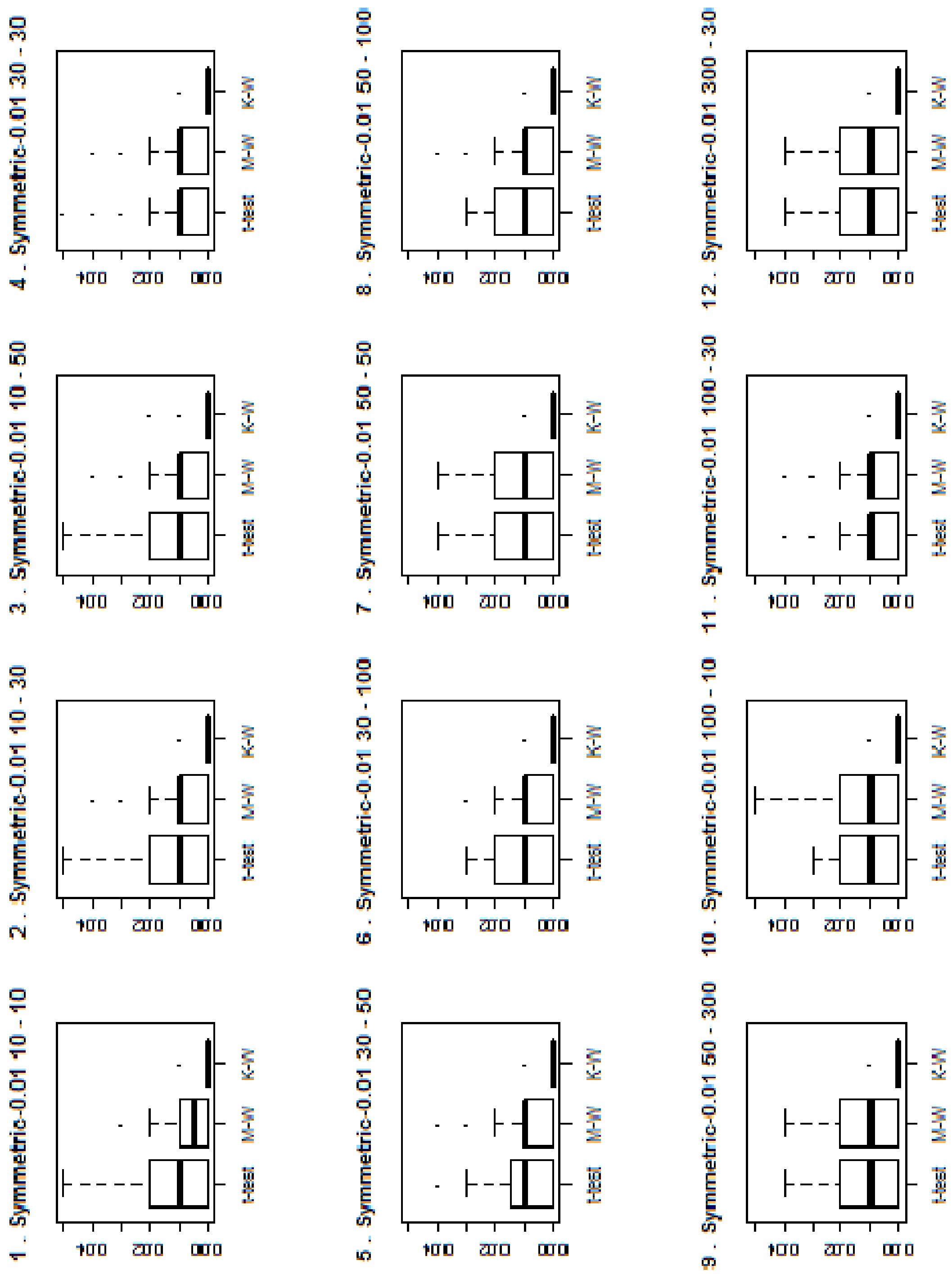
From Figure 29, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 15. Type I error rate estimates of the unimodal symmetric distribution

Significance Level	Sample Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
0.01	10, 10	0.012	0.007	0.000	0.010	0.007	0.000
	10, 30	0.011	0.008	0.001	0.010	0.008	0.001
	10, 50	0.011	0.010	0.002	0.010	0.008	0.001
	30, 30	0.009	0.008	0.000	0.011	0.009	0.001
	30, 50	0.010	0.009	0.001	0.011	0.009	0.001
	30, 100	0.010	0.008	0.001	0.011	0.009	0.001
	50, 50	0.011	0.010	0.001	0.009	0.009	0.001
	50, 100	0.011	0.010	0.000	0.010	0.008	0.000
	50, 300	0.011	0.010	0.001	0.011	0.010	0.001
	100, 100	0.010	0.010	0.000	0.009	0.009	0.001
	100, 300	0.008	0.008	0.001	0.010	0.010	0.001
	300, 300	0.012	0.011	0.001	0.009	0.010	0.002
0.05	10, 10	0.052	0.044	0.002	0.053	0.042	0.003
	10, 30	0.051	0.046	0.007	0.048	0.046	0.008
	10, 50	0.049	0.048	0.006	0.054	0.049	0.006
	30, 30	0.051	0.050	0.005	0.049	0.048	0.008
	30, 50	0.047	0.047	0.006	0.052	0.050	0.006
	30, 100	0.049	0.048	0.007	0.052	0.052	0.009
	50, 50	0.050	0.050	0.006	0.049	0.047	0.007
	50, 100	0.048	0.045	0.006	0.049	0.049	0.008
	50, 300	0.046	0.044	0.005	0.053	0.052	0.007
	100, 100	0.051	0.050	0.007	0.051	0.051	0.006
	100, 300	0.048	0.050	0.006	0.049	0.049	0.008
	300, 300	0.048	0.048	0.005	0.050	0.050	0.006
0.10	10, 10	0.102	0.094	0.012	0.100	0.095	0.016
	10, 30	0.101	0.095	0.013	0.097	0.093	0.017
	10, 50	0.106	0.104	0.013	0.105	0.102	0.017
	30, 30	0.097	0.096	0.014	0.100	0.097	0.018
	30, 50	0.099	0.099	0.015	0.104	0.101	0.018
	30, 100	2.000	0.101	0.016	0.097	0.097	0.019
	50, 50	0.103	0.103	0.013	0.103	0.102	0.016
	50, 100	0.105	0.102	0.014	0.097	0.098	0.018
	50, 300	0.100	0.099	0.013	0.098	0.098	0.017
	100, 100	0.105	0.106	0.016	0.096	0.094	0.015
	100, 300	0.102	0.104	0.018	0.095	0.094	0.017
	300, 300	0.099	0.100	0.019	0.100	0.097	0.020

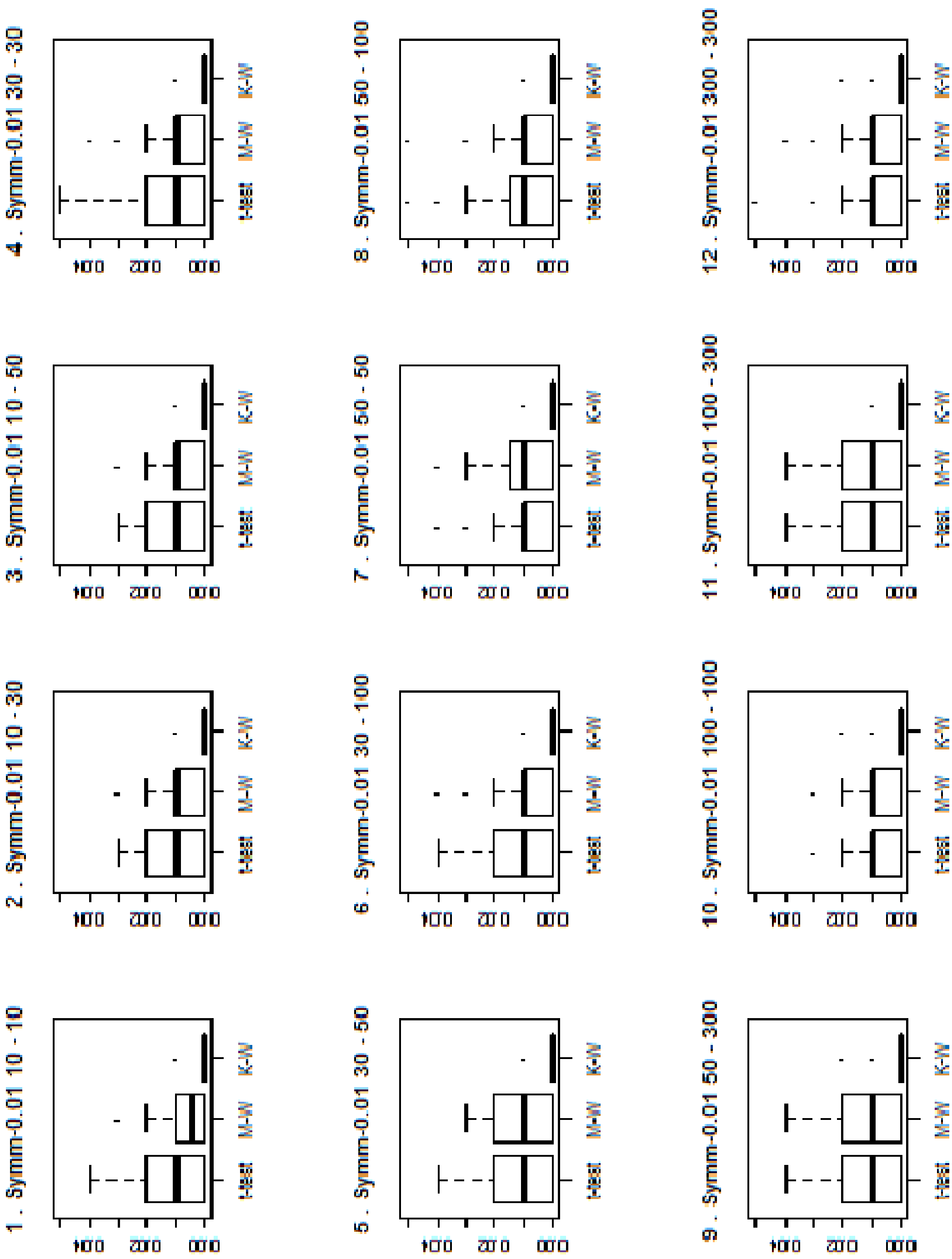
Table 15 shows that, given the unimodal symmetric distribution, the empirical Type I error rates from the t-test and Mann-Whitney (MW) test are close to the nominal significance levels and followed the robustness criteria. However, the Kolmogorov-Smirnov (KS) test is not robust since the error rate is beyond the criteria’s range for all circumstances.

Figure 30. Distribution of Type I error rate estimates of the unimodal symmetric distribution at α = 0.01 for the 5-point Likert scale



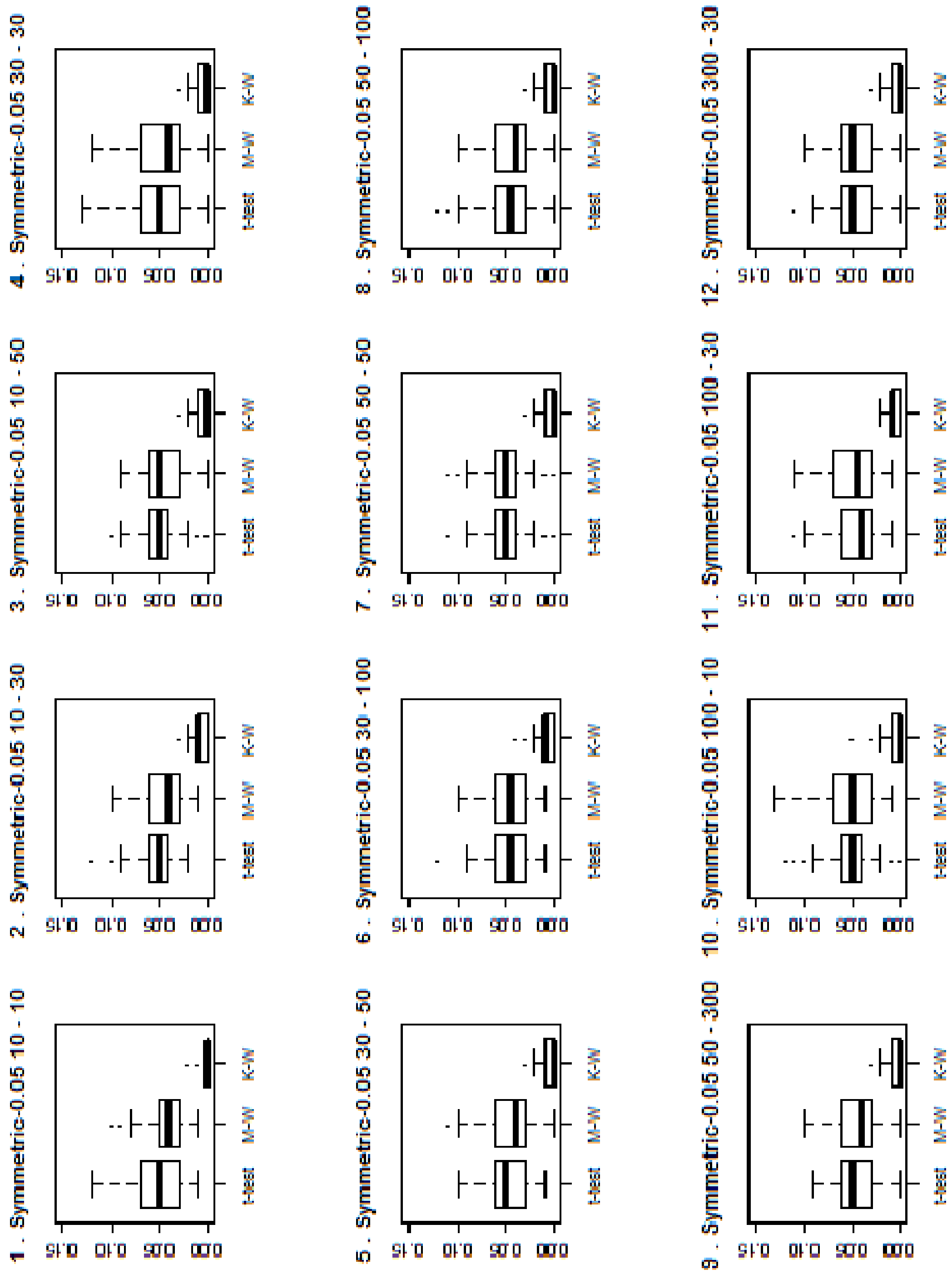
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 31. Distribution of Type I error rate estimates of the unimodal symmetric distribution at α = 0.01 for the 7-point Likert scale



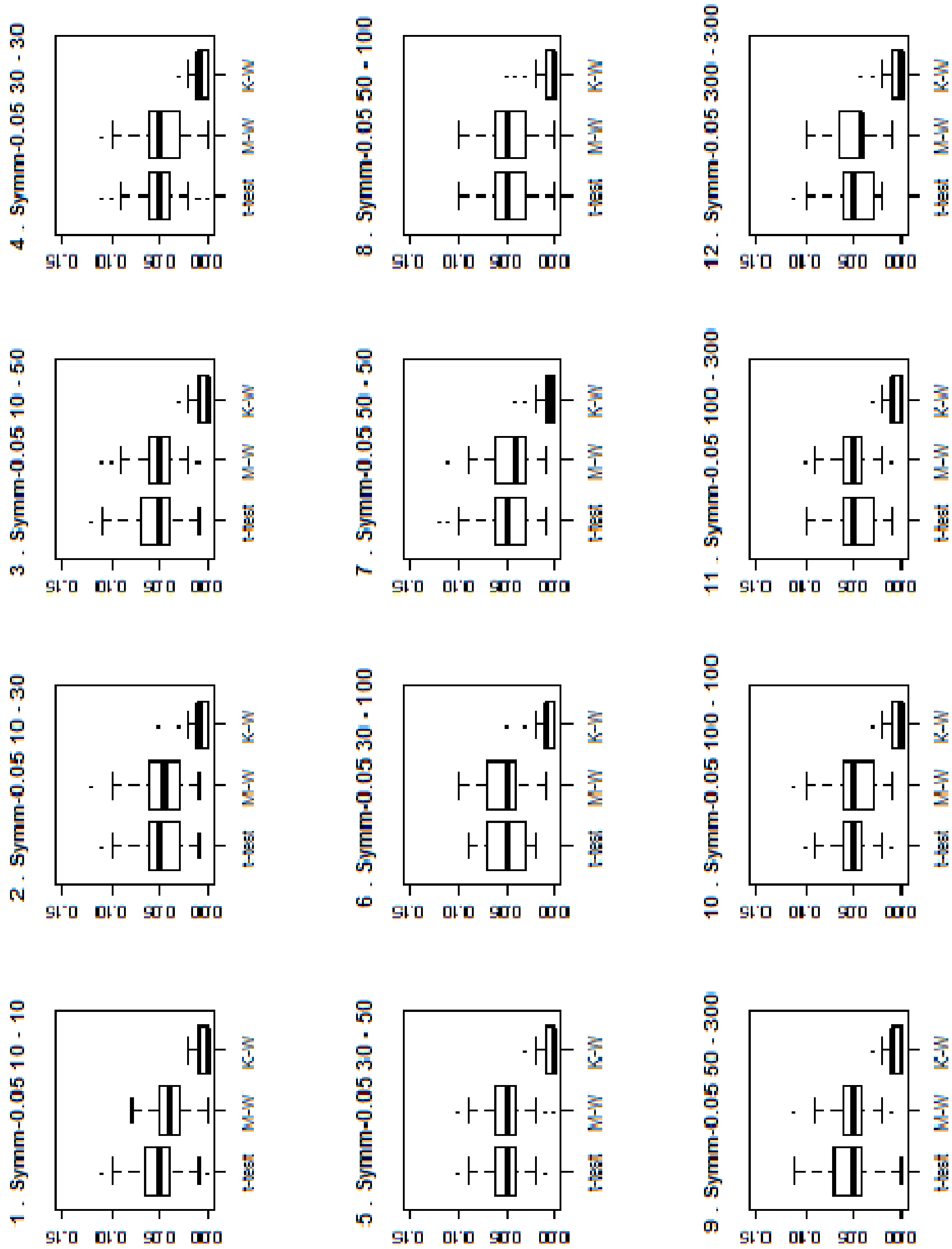
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 32. Distribution of Type I error rate estimates of the unimodal symmetric distribution at α = 0.05 for the 5-point Likert scale



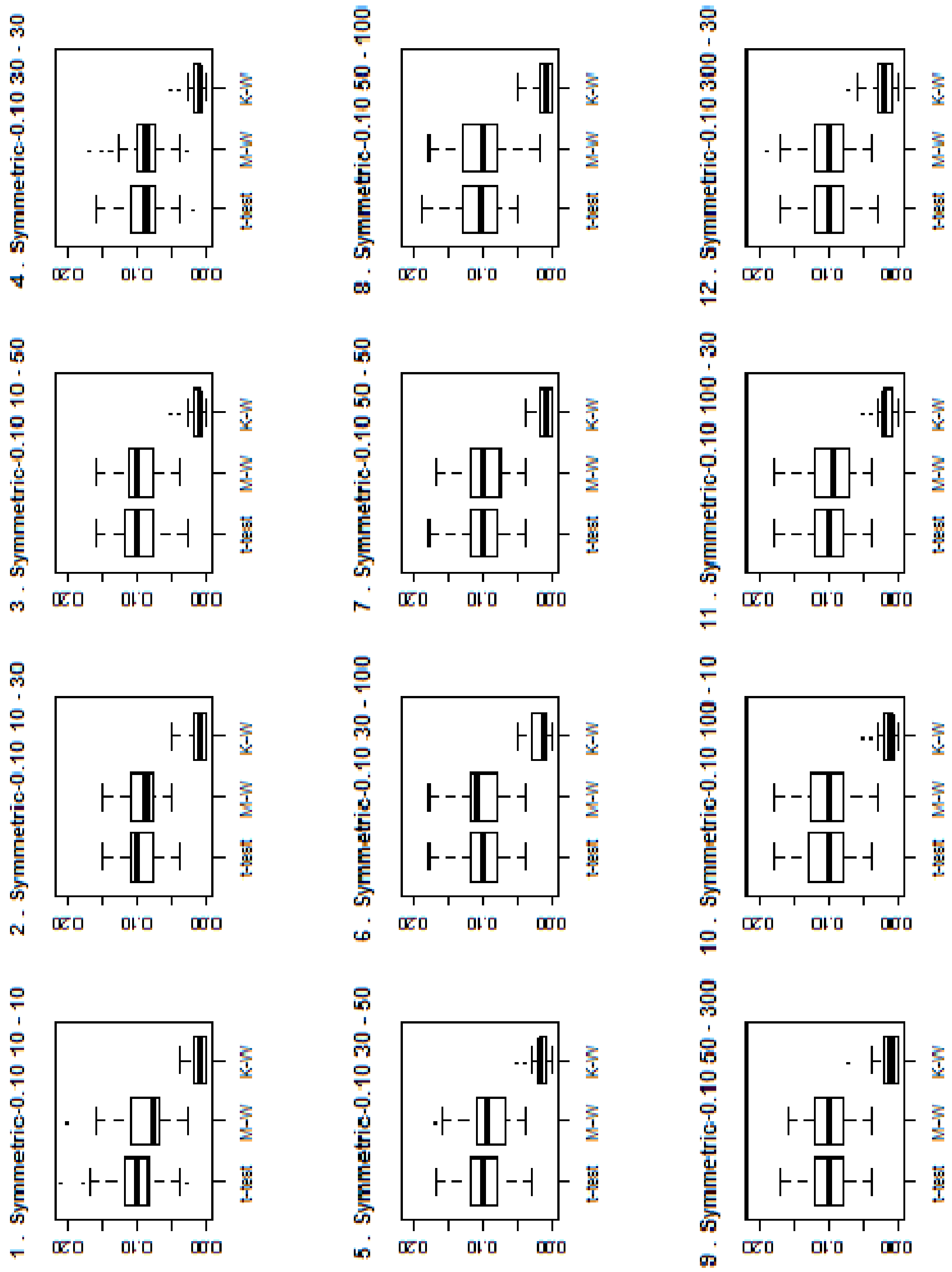
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 33. Distribution of Type I error rate estimates of the unimodal symmetric distribution at α = 0.05 for the 7-point Likert scale



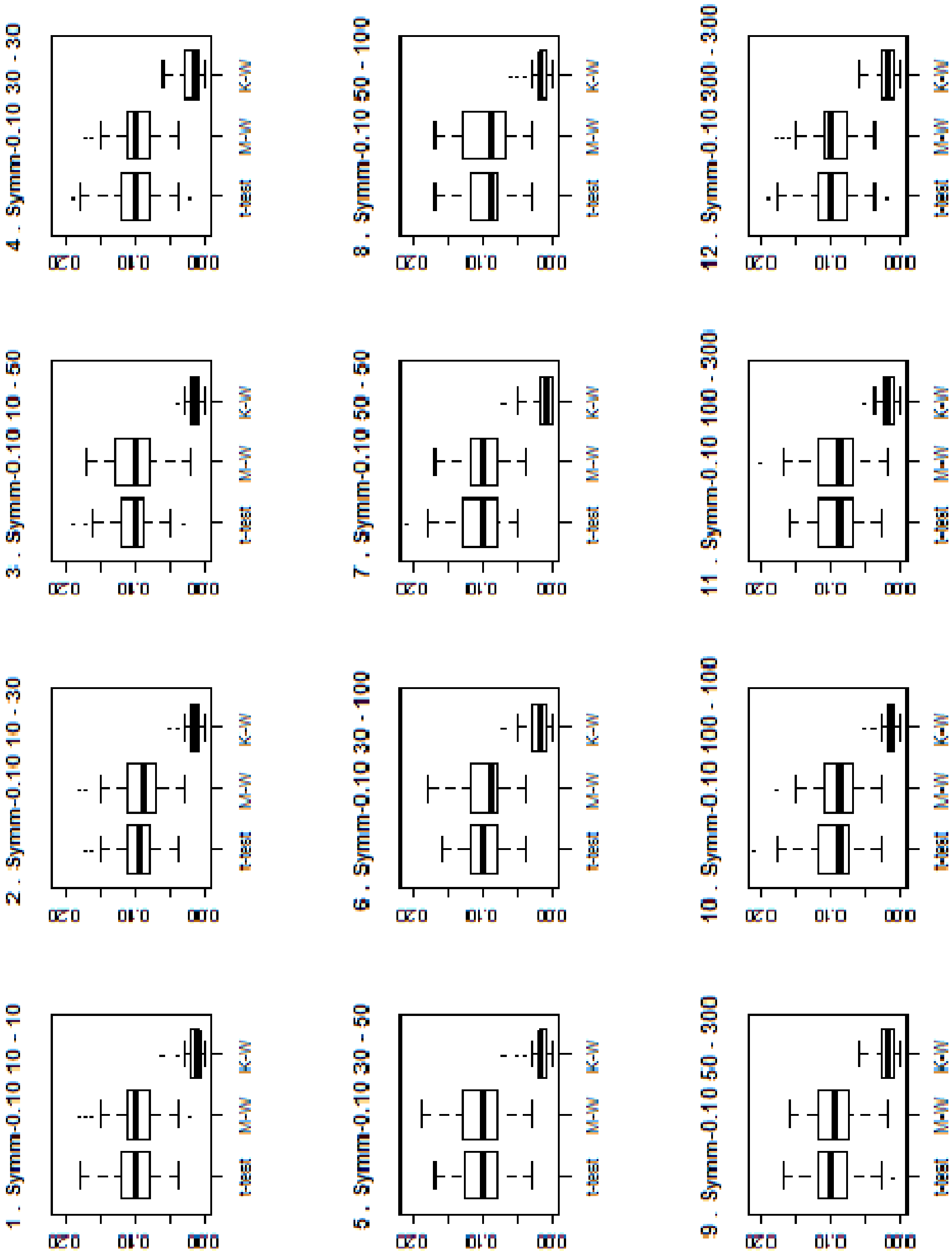
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 34. Distribution of Type I error rate estimates of the unimodal symmetric distribution at α = 0.10 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 35. Distribution of Type I error rate estimates of the unimodal symmetric distribution at α = 0.10 for the 7-point Likert scale



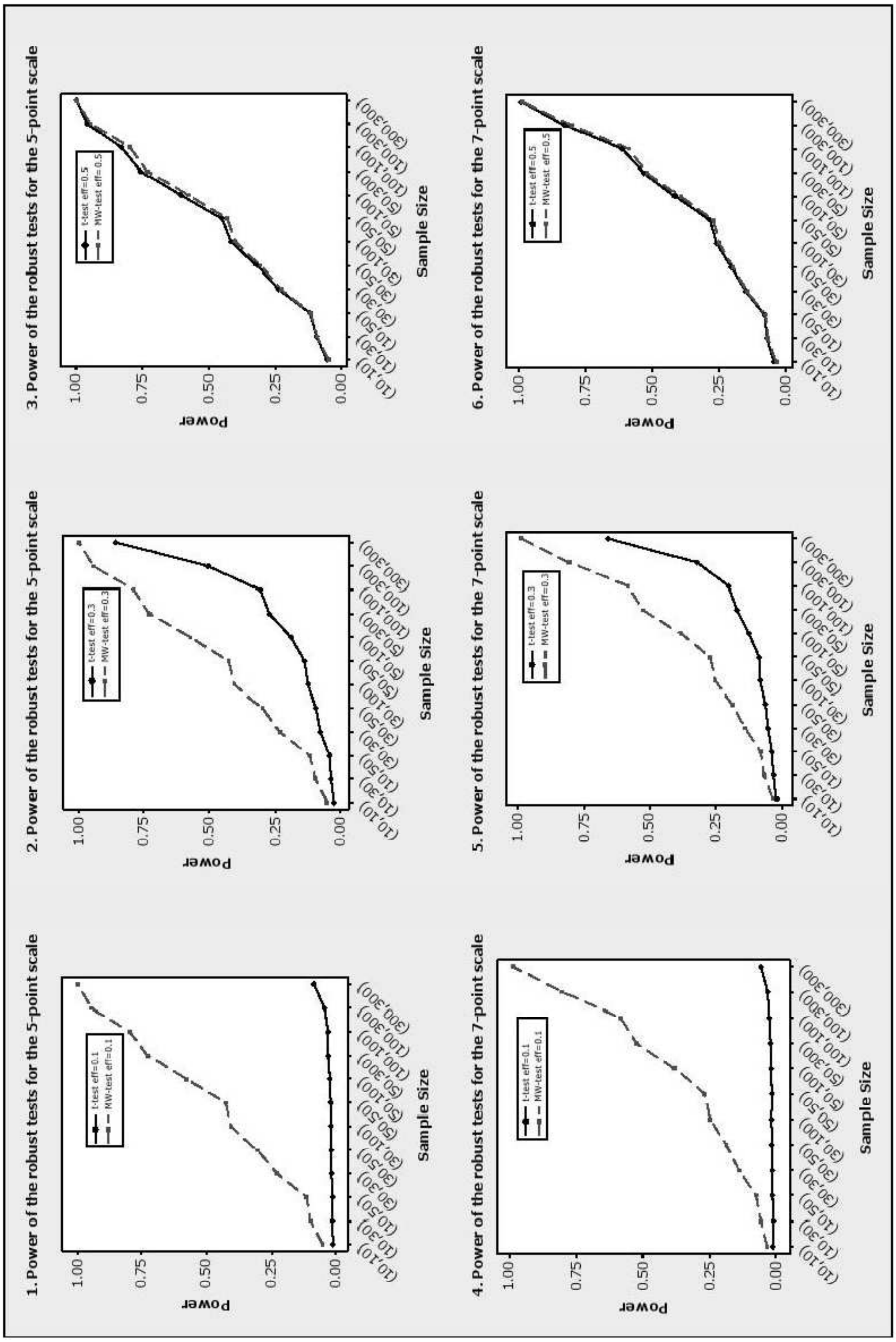
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Table 16. Statistical power estimates of the unimodal symmetric distribution at α = 0.01

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KW-test	t-test	MW-test	KW-test
10, 10	0.10	0.011	0.050	0.028	0.012	0.034	0.016
10, 10	0.30	0.024	0.051	0.029	0.018	0.031	0.016
10, 10	0.50	0.055	0.049	0.030	0.039	0.031	0.016
10, 30	0.10	0.013	0.098	0.122	0.010	0.057	0.063
10, 30	0.30	0.035	0.096	0.121	0.028	0.064	0.069
10, 30	0.50	0.095	0.093	0.112	0.065	0.062	0.067
10, 50	0.10	0.013	0.114	0.124	0.013	0.075	0.070
10, 50	0.30	0.042	0.118	0.120	0.034	0.077	0.071
10, 50	0.50	0.120	0.117	0.120	0.074	0.071	0.066
30, 30	0.10	0.016	0.229	0.395	0.014	0.140	0.225
30, 30	0.30	0.076	0.233	0.397	0.049	0.136	0.226
30, 30	0.50	0.240	0.230	0.395	0.146	0.141	0.230
30, 50	0.10	0.016	0.305	0.527	0.016	0.188	0.320
30, 50	0.30	0.094	0.300	0.522	0.059	0.186	0.321
30, 50	0.50	0.321	0.308	0.527	0.201	0.194	0.326
30, 100	0.10	0.018	0.405	0.682	0.017	0.249	0.452
30, 100	0.30	0.125	0.405	0.683	0.078	0.251	0.452
30, 100	0.50	0.416	0.400	0.685	0.256	0.247	0.454
50, 50	0.10	0.019	0.426	0.715	0.016	0.270	0.489
50, 50	0.30	0.138	0.427	0.715	0.082	0.270	0.486
50, 50	0.50	0.451	0.431	0.722	0.282	0.269	0.484
50, 100	0.10	0.022	0.580	0.892	0.018	0.382	0.702
50, 100	0.30	0.190	0.575	0.894	0.122	0.382	0.690
50, 100	0.50	0.607	0.579	0.892	0.411	0.392	0.696
50, 300	0.10	0.030	0.727	0.973	0.020	0.524	0.866
50, 300	0.30	0.274	0.730	0.974	0.167	0.524	0.861
50, 300	0.50	0.756	0.731	0.971	0.530	0.517	0.864
100, 100	0.10	0.030	0.795	0.989	0.026	0.584	0.911
100, 100	0.30	0.308	0.789	0.989	0.200	0.582	0.916
100, 100	0.50	0.828	0.796	0.990	0.612	0.586	0.916
100, 300	0.10	0.043	0.946	1.000	0.031	0.802	0.995
100, 300	0.30	0.505	0.945	1.000	0.320	0.806	0.996
100, 300	0.50	0.959	0.945	1.000	0.828	0.804	0.995
300, 300	0.10	0.085	0.999	1.000	0.057	0.987	1.000
300, 300	0.30	0.858	1.000	1.000	0.658	0.987	1.000
300, 300	0.50	1.000	1.000	1.000	0.992	0.987	1.000

Table 16 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 36.

Figure 36. Statistical power estimates of the unimodal symmetric distribution at α = 0.01



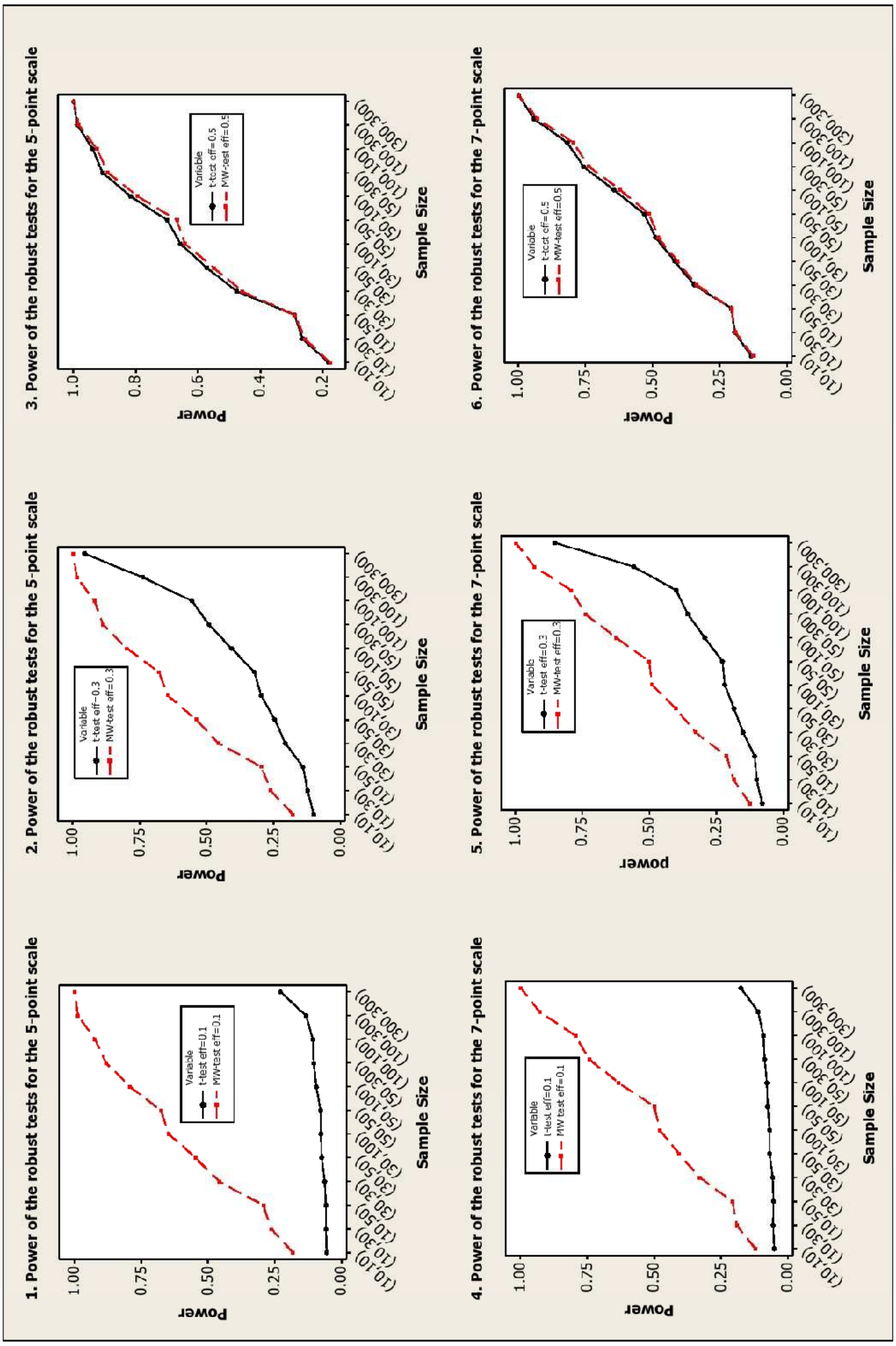
From Figure 36, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 17. Statistical power estimates of the unimodal symmetric distribution at α = 0.05

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.056	0.183	0.098	0.050	0.121	0.060
10, 10	0.30	0.099	0.177	0.101	0.078	0.124	0.061
10, 10	0.50	0.181	0.176	0.096	0.132	0.125	0.057
10, 30	0.10	0.058	0.263	0.310	0.054	0.190	0.207
10, 30	0.30	0.123	0.261	0.305	0.099	0.185	0.211
10, 30	0.50	0.264	0.257	0.298	0.193	0.191	0.210
10, 50	0.10	0.058	0.293	0.325	0.052	0.206	0.208
10, 50	0.30	0.140	0.295	0.326	0.106	0.210	0.214
10, 50	0.50	0.291	0.289	0.320	0.206	0.205	0.211
30, 30	0.10	0.063	0.458	0.642	0.056	0.328	0.465
30, 30	0.30	0.206	0.456	0.634	0.149	0.326	0.460
30, 30	0.50	0.477	0.459	0.638	0.345	0.336	0.460
30, 50	0.10	0.074	0.547	0.760	0.067	0.406	0.576
30, 50	0.30	0.246	0.540	0.759	0.183	0.400	0.570
30, 50	0.50	0.573	0.550	0.771	0.417	0.406	0.576
30, 100	0.10	0.078	0.648	0.877	0.067	0.478	0.715
30, 100	0.30	0.296	0.645	0.874	0.218	0.490	0.720
30, 100	0.50	0.658	0.643	0.873	0.488	0.479	0.714
50, 50	0.10	0.079	0.677	0.910	0.074	0.498	0.762
50, 50	0.30	0.322	0.679	0.913	0.226	0.500	0.756
50, 50	0.50	0.700	0.669	0.913	0.530	0.512	0.764
50, 100	0.10	0.095	0.792	0.978	0.077	0.632	0.896
50, 100	0.30	0.406	0.798	0.979	0.292	0.623	0.892
50, 100	0.50	0.818	0.795	0.978	0.643	0.620	0.891
50, 300	0.10	0.104	0.883	0.998	0.086	0.743	0.964
50, 300	0.30	0.491	0.888	0.996	0.357	0.739	0.962
50, 300	0.50	0.908	0.891	0.997	0.756	0.737	0.958
100, 100	0.10	0.107	0.924	0.999	0.090	0.794	0.984
100, 100	0.30	0.556	0.920	0.999	0.400	0.791	0.983
100, 100	0.50	0.940	0.926	0.999	0.816	0.794	0.986
100, 300	0.10	0.133	0.988	1.000	0.110	0.928	1.000
100, 300	0.30	0.739	0.985	1.000	0.558	0.929	1.000
100, 300	0.50	0.989	0.984	1.000	0.941	0.928	1.000
300, 300	0.10	0.230	1.000	1.000	0.175	0.998	1.000
300, 300	0.30	0.956	1.000	1.000	0.851	0.998	1.000
300, 300	0.50	1.000	1.000	1.000	0.998	0.997	1.000

Table 17 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 37.

Figure 37. Statistical power estimates of the unimodal symmetric distribution at $\alpha = 0.05$



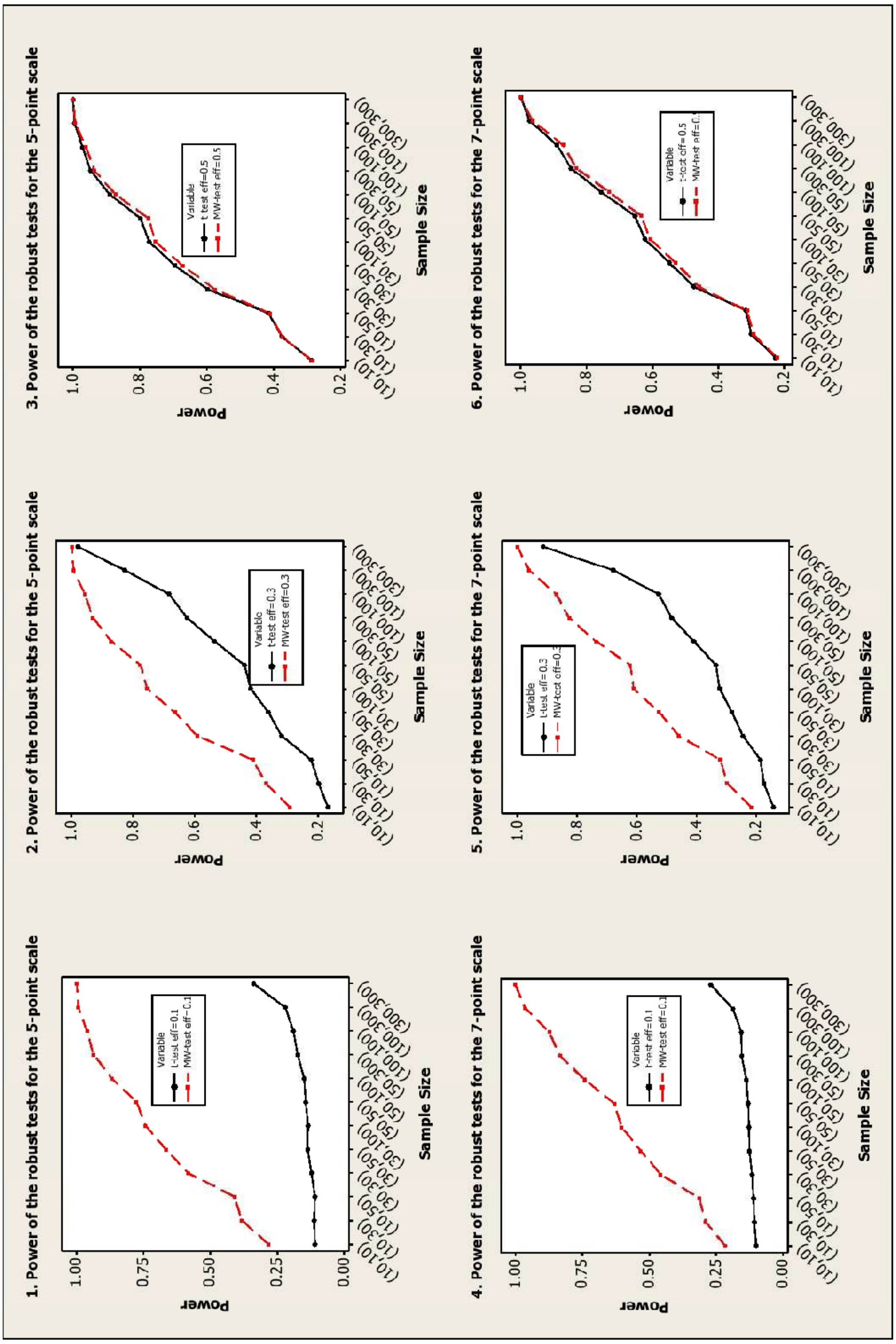
From Figure 37, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 18. Statistical power estimates of the unimodal symmetric distribution at α = 0.10

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.110	0.283	0.242	0.100	0.215	0.173
10, 10	0.30	0.168	0.291	0.246	0.141	0.217	0.175
10, 10	0.50	0.287	0.286	0.242	0.225	0.221	0.171
10, 30	0.10	0.112	0.382	0.398	0.107	0.288	0.286
10, 30	0.30	0.198	0.369	0.390	0.173	0.296	0.289
10, 30	0.50	0.376	0.375	0.392	0.299	0.292	0.285
10, 50	0.10	0.110	0.410	0.431	0.109	0.313	0.304
10, 50	0.30	0.223	0.412	0.430	0.184	0.319	0.305
10, 50	0.50	0.416	0.412	0.428	0.315	0.312	0.307
30, 30	0.10	0.123	0.584	0.767	0.115	0.457	0.600
30, 30	0.30	0.319	0.592	0.767	0.244	0.458	0.595
30, 30	0.50	0.598	0.576	0.751	0.473	0.457	0.596
30, 50	0.10	0.137	0.666	0.861	0.125	0.531	0.714
30, 50	0.30	0.362	0.666	0.864	0.281	0.525	0.707
30, 50	0.50	0.696	0.673	0.864	0.548	0.528	0.715
30, 100	0.10	0.135	0.742	0.935	0.127	0.603	0.815
30, 100	0.30	0.419	0.755	0.936	0.322	0.610	0.827
30, 100	0.50	0.773	0.754	0.941	0.622	0.606	0.818
50, 50	0.10	0.144	0.776	0.945	0.130	0.631	0.842
50, 50	0.30	0.439	0.777	0.952	0.333	0.622	0.833
50, 50	0.50	0.799	0.775	0.948	0.652	0.634	0.838
50, 100	0.10	0.150	0.868	0.991	0.136	0.740	0.942
50, 100	0.30	0.536	0.872	0.993	0.408	0.737	0.942
50, 100	0.50	0.891	0.874	0.993	0.755	0.732	0.942
50, 300	0.10	0.174	0.936	1.000	0.154	0.833	0.987
50, 300	0.30	0.626	0.932	0.999	0.483	0.827	0.986
50, 300	0.50	0.948	0.936	1.000	0.847	0.830	0.983
100, 100	0.10	0.190	0.961	1.000	0.155	0.874	0.996
100, 100	0.30	0.684	0.958	1.000	0.526	0.870	0.996
100, 100	0.50	0.972	0.962	1.000	0.890	0.871	0.997
100, 300	0.10	0.220	0.993	1.000	0.186	0.962	1.000
100, 300	0.30	0.829	0.995	1.000	0.679	0.961	1.000
100, 300	0.50	0.996	0.994	1.000	0.972	0.963	1.000
300, 300	0.10	0.338	1.000	1.000	0.269	1.000	1.000
300, 300	0.30	0.980	1.000	1.000	0.914	1.000	1.000
300, 300	0.50	1.000	1.000	1.000	0.999	0.999	1.000

Table 18 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 38.

Figure 38. Statistical power estimates of the unimodal symmetric distribution at α = 0.10



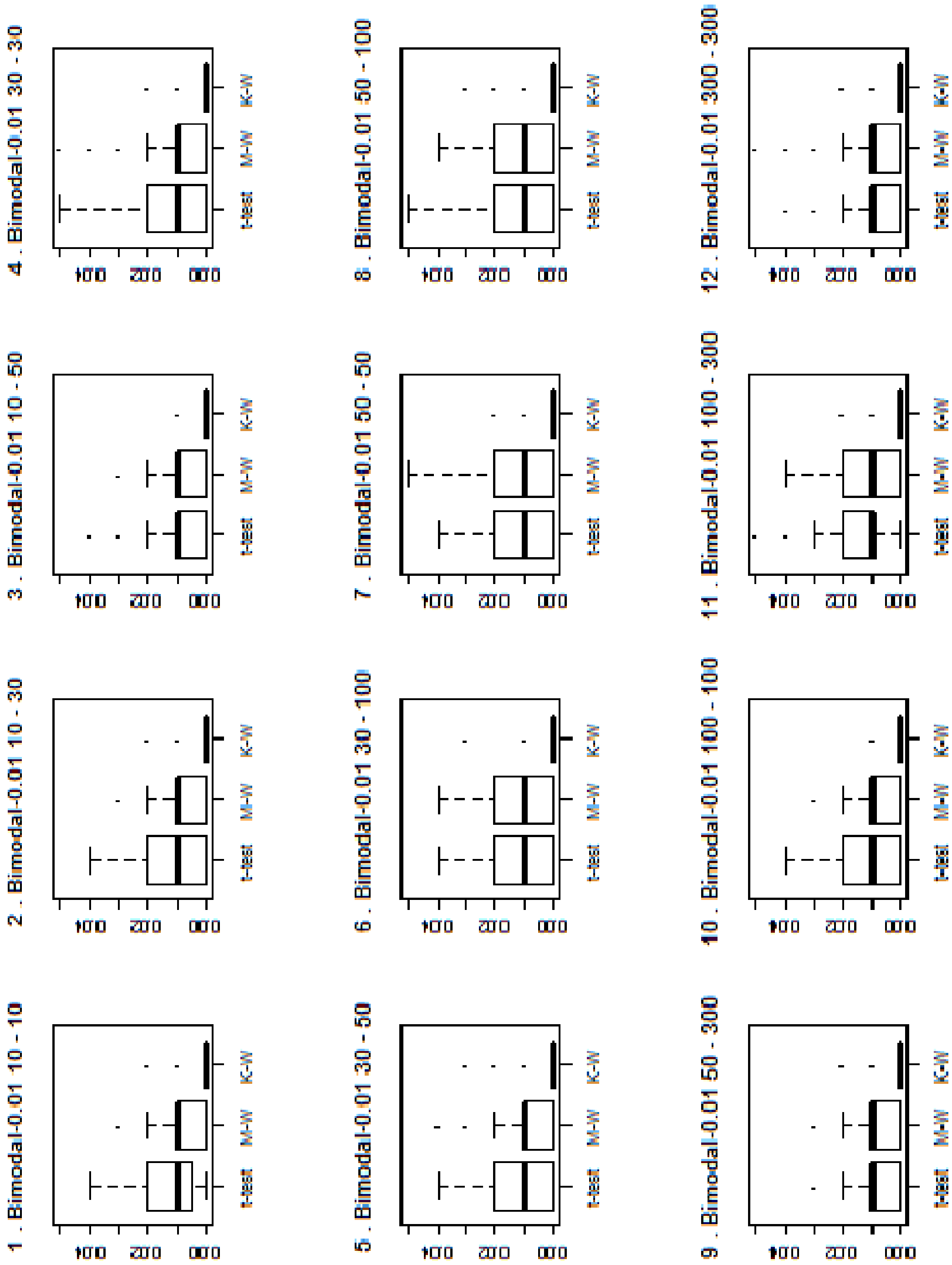
From Figure 38, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1 and 0.3. However, when effect size = 0.5 both statistical powers are almost the same for the 5-point and 7-point conditions.

Table 19. Type I error rate estimates of the bimodal symmetric distribution

Significance Level	Sample Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
0.01	10, 10	0.014	0.008	0.001	0.012	0.006	0.001
	10, 30	0.011	0.009	0.002	0.011	0.008	0.002
	10, 50	0.008	0.008	0.001	0.010	0.007	0.002
	30, 30	0.011	0.010	0.002	0.011	0.010	0.002
	30, 50	0.010	0.009	0.002	0.010	0.010	0.002
	30, 100	0.012	0.010	0.002	0.012	0.012	0.003
	50, 50	0.011	0.011	0.002	0.011	0.009	0.003
	50, 100	0.011	0.010	0.003	0.009	0.008	0.002
	50, 300	0.009	0.009	0.002	0.009	0.008	0.003
	100, 100	0.010	0.009	0.002	0.012	0.012	0.003
	100, 300	0.013	0.012	0.003	0.010	0.010	0.003
	300, 300	0.010	0.010	0.002	0.010	0.008	0.003
0.05	10, 10	0.053	0.047	0.006	0.052	0.042	0.006
	10, 30	0.050	0.048	0.015	0.049	0.046	0.017
	10, 50	0.048	0.047	0.012	0.049	0.048	0.014
	30, 30	0.053	0.051	0.014	0.049	0.047	0.015
	30, 50	0.053	0.051	0.014	0.050	0.049	0.013
	30, 100	0.051	0.049	0.013	0.050	0.050	0.015
	50, 50	0.054	0.053	0.015	0.046	0.045	0.015
	50, 100	0.046	0.048	0.012	0.052	0.052	0.017
	50, 300	0.048	0.047	0.014	0.048	0.051	0.016
	100, 100	0.048	0.048	0.012	0.050	0.049	0.016
	100, 300	0.052	0.051	0.015	0.051	0.050	0.015
	300, 300	0.051	0.053	0.014	0.053	0.050	0.014
0.10	10, 10	0.096	0.087	0.024	0.100	0.091	0.028
	10, 30	0.098	0.094	0.025	0.098	0.097	0.029
	10, 50	0.104	0.102	0.028	0.105	0.102	0.031
	30, 30	0.099	0.098	0.027	0.100	0.102	0.033
	30, 50	0.101	0.099	0.030	0.102	0.098	0.034
	30, 100	0.100	0.100	0.031	0.099	0.097	0.030
	50, 50	0.103	0.103	0.027	0.099	0.097	0.030
	50, 100	0.100	0.098	0.029	0.097	0.098	0.030
	50, 300	0.097	0.100	0.028	0.099	0.097	0.034
	100, 100	0.096	0.094	0.027	0.100	0.100	0.036
	100, 300	0.098	0.095	0.030	0.101	0.102	0.035
	300, 300	0.104	0.103	0.034	0.096	0.095	0.039

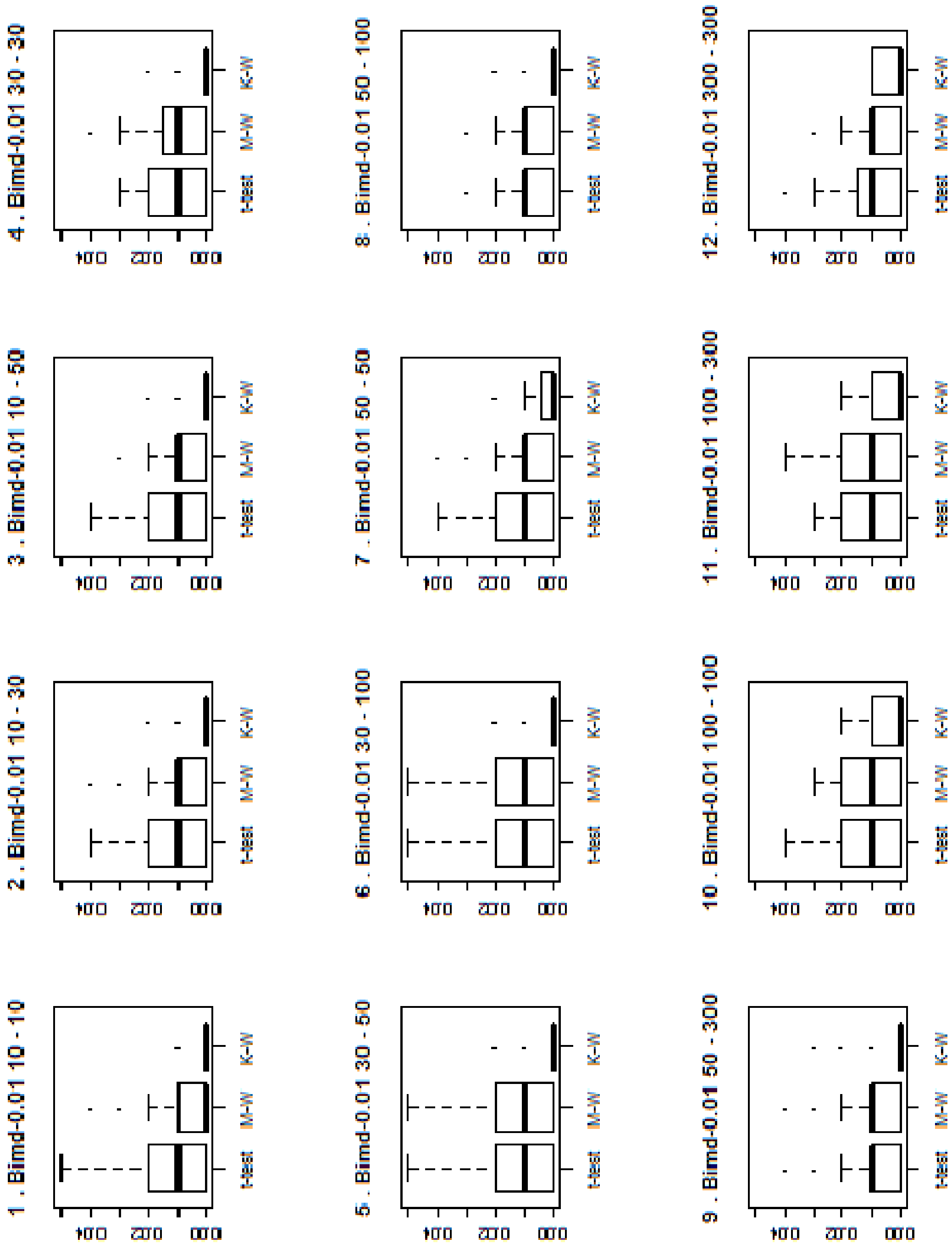
Table 19 shows that, given the bimodal symmetric distribution, the empirical Type I error rates from the t-test and Mann-Whitney (MW) test are close to the nominal significance levels and followed the robustness criteria. However, the Kolmogorov-Smirnov (KS) test is not robust since the error rate is beyond the criteria’s range for all circumstances.

Figure 39. Distribution of Type I error rate estimates of the bimodal symmetric distribution at α = 0.01 for the 5-point Likert scale



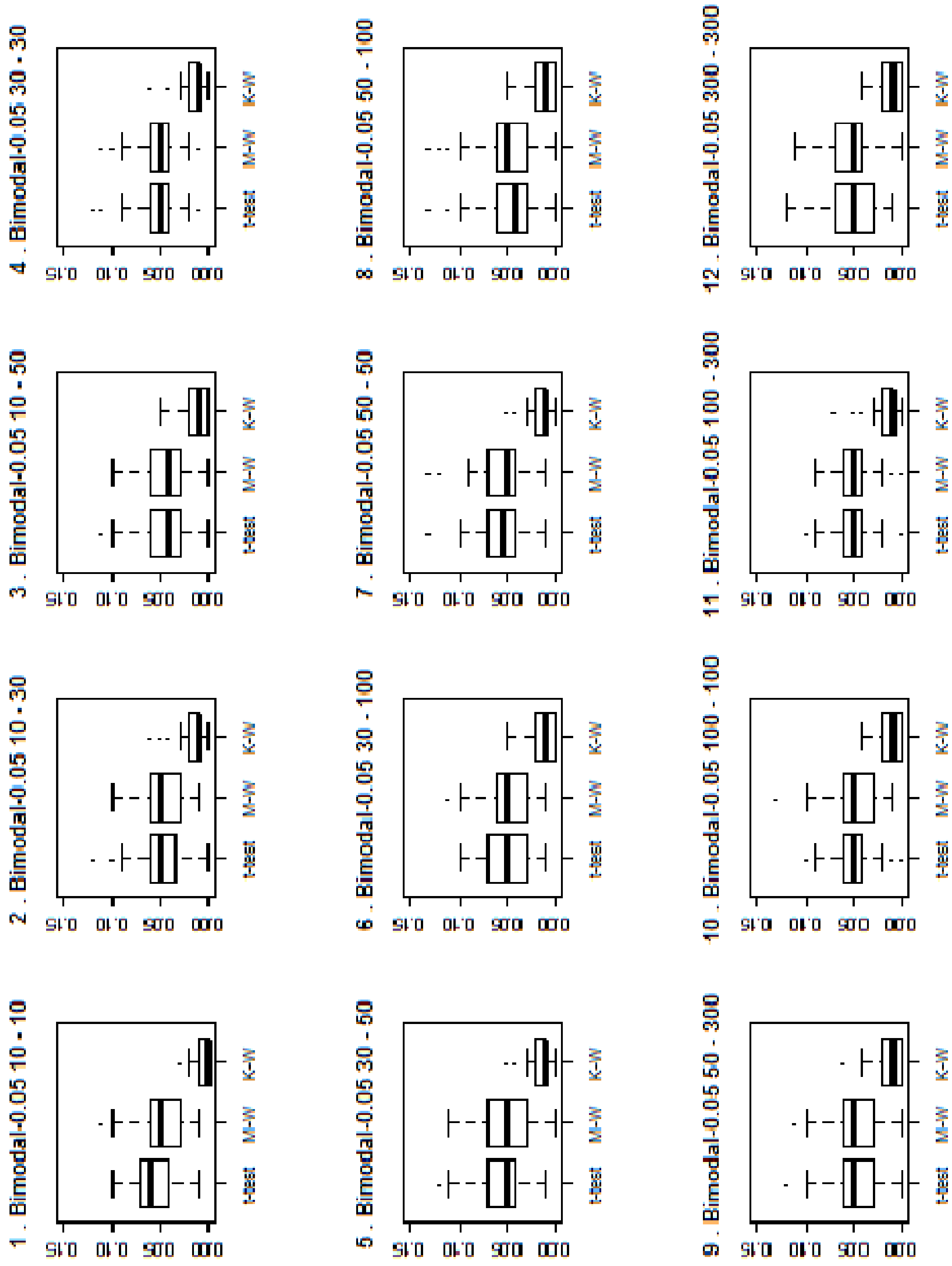
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 40. Distribution of Type I error rate estimates of the bimodal symmetric distribution at α = 0.01 for the 7-point Likert scale



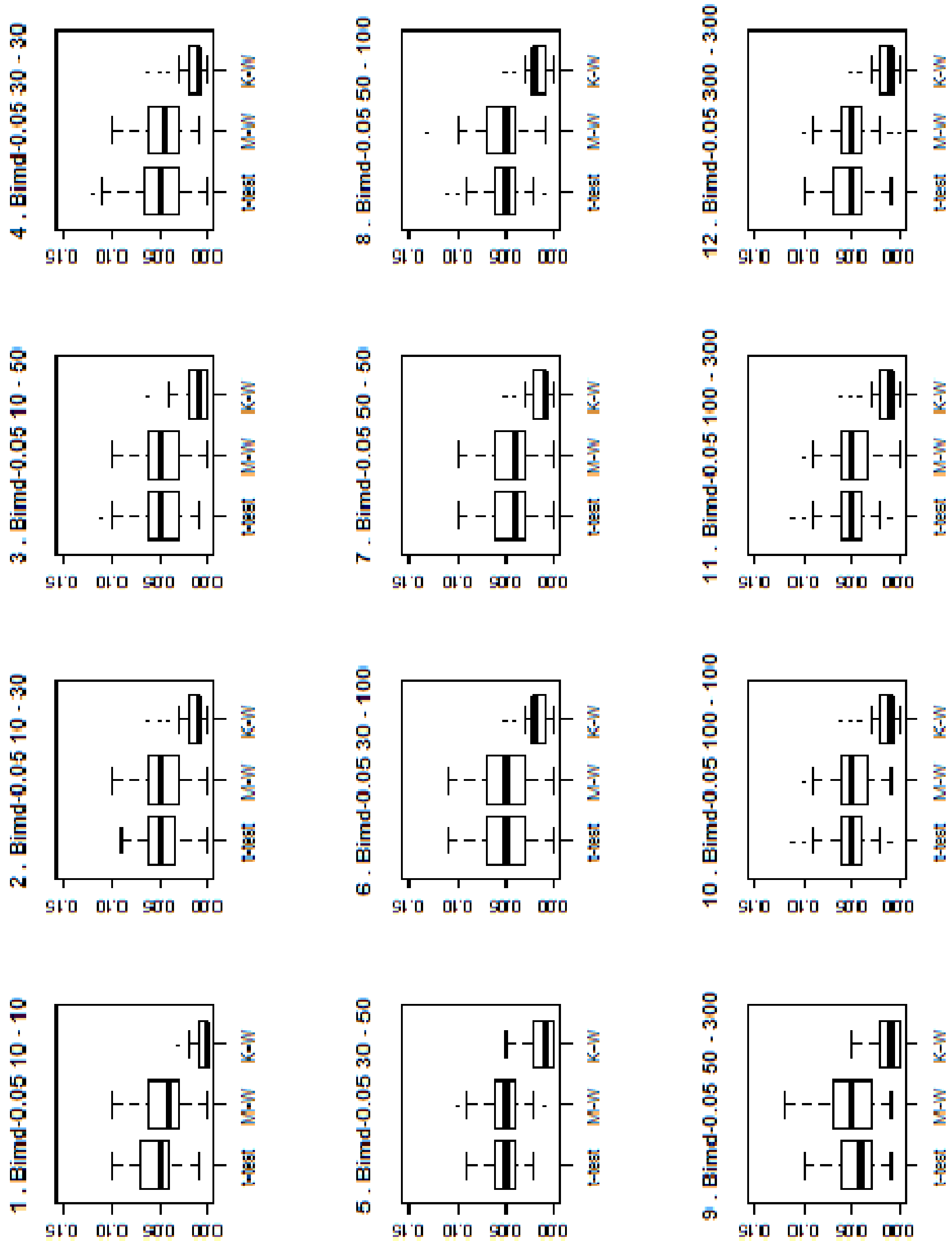
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 41. Distribution of Type I error rate estimates of the bimodal symmetric distribution at α = 0.05 for the 5-point Likert scale



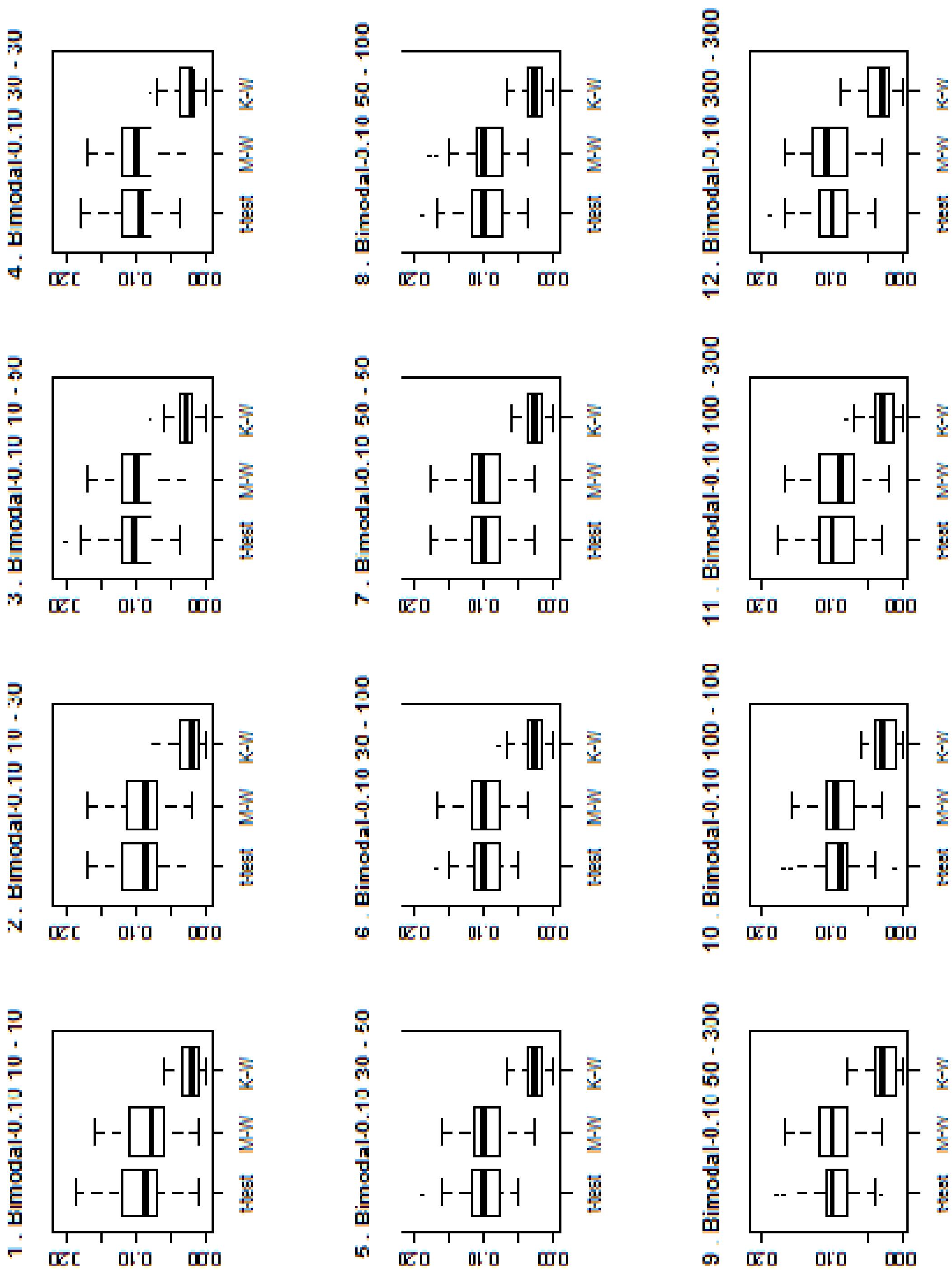
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 42. Distribution of Type I error rate estimates of the bimodal symmetric distribution at α = 0.05 for the 7-point Likert scale



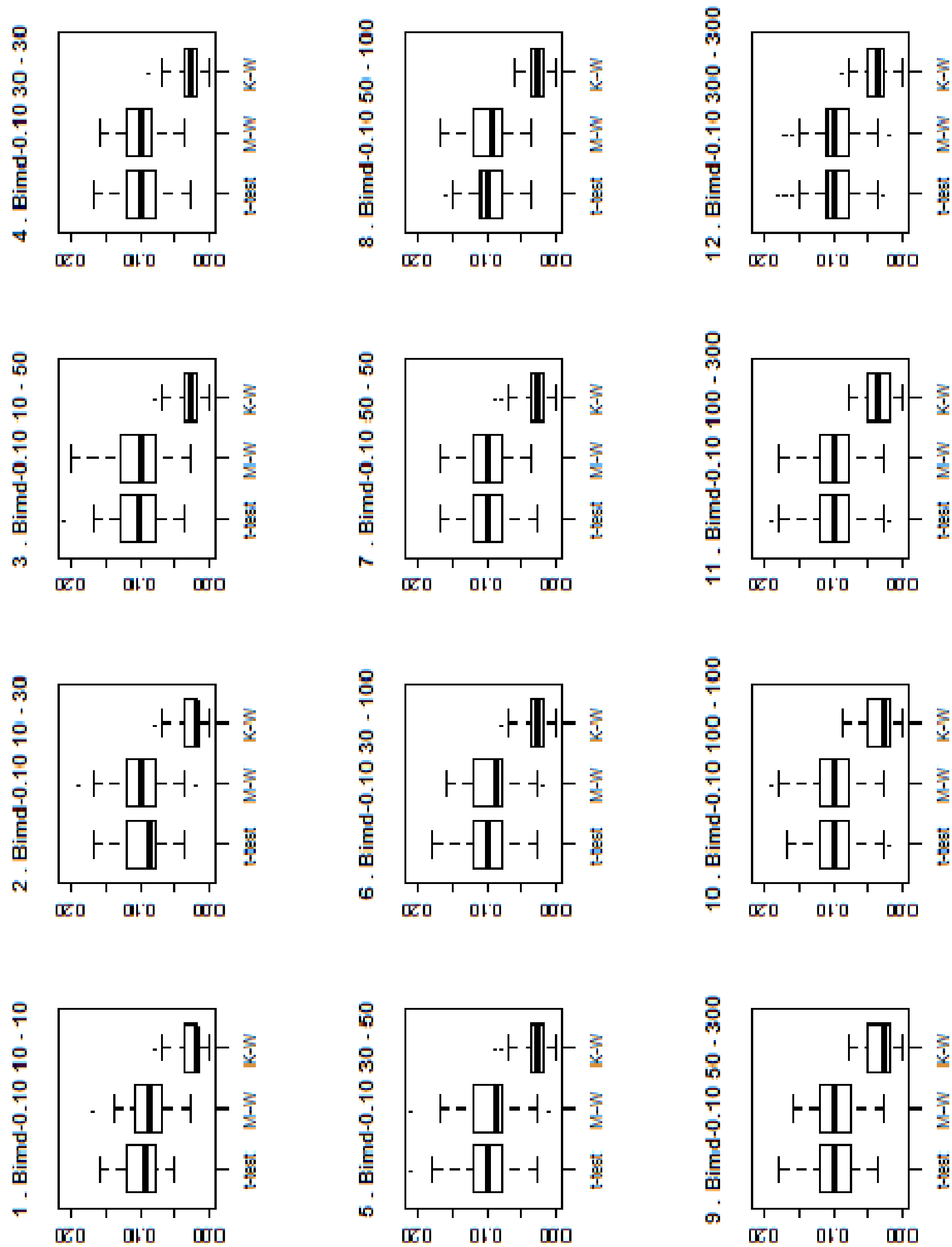
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 43. Distribution of Type I error rate estimates of the bimodal symmetric distribution at α = 0.10 for the 5-point Likert scale



From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Figure 44. Distribution of Type I error rate estimates of the bimodal symmetric distribution at α = 0.10 for the 7-point Likert scale



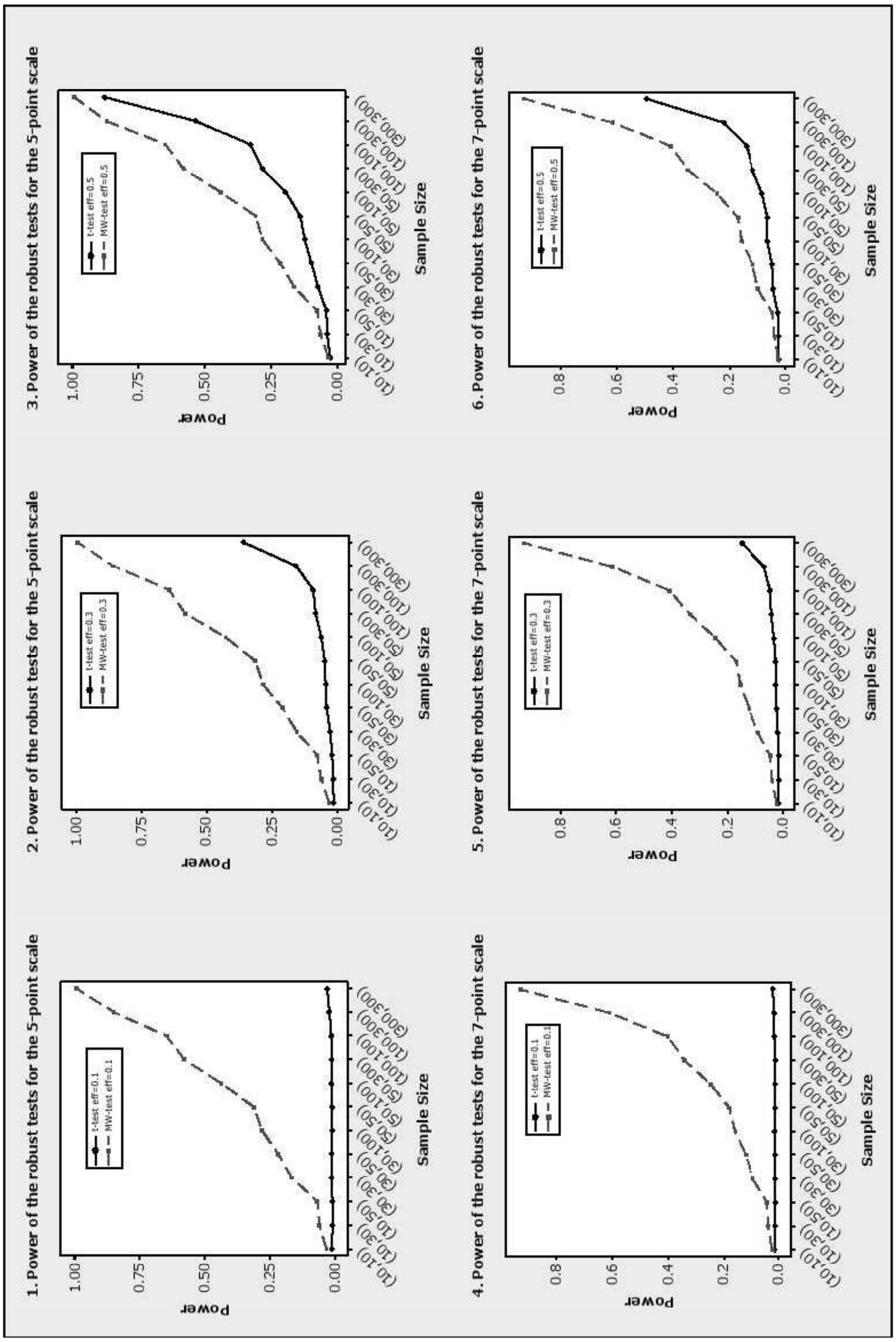
From boxplots, the medians of the error rates from t-test and Mann-Whitney test are closer to the given alpha than the median of Kolmogorov-Smirnov. Moreover, distribution shapes of the error rates from t-test are more symmetric than others in overall.

Table 20. Statistical power estimates of the bimodal symmetric distribution at α = 0.01

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.012	0.033	0.008	0.014	0.022	0.005
10, 10	0.30	0.016	0.034	0.008	0.014	0.021	0.005
10, 10	0.50	0.028	0.035	0.008	0.021	0.023	0.005
10, 30	0.10	0.011	0.062	0.038	0.011	0.037	0.023
10, 30	0.30	0.017	0.063	0.040	0.014	0.037	0.022
10, 30	0.50	0.039	0.065	0.040	0.022	0.036	0.021
10, 50	0.10	0.011	0.069	0.033	0.011	0.041	0.019
10, 50	0.30	0.022	0.079	0.037	0.013	0.044	0.019
10, 50	0.50	0.042	0.077	0.036	0.025	0.045	0.020
30, 30	0.10	0.014	0.167	0.200	0.012	0.093	0.081
30, 30	0.30	0.029	0.157	0.189	0.017	0.089	0.077
30, 30	0.50	0.074	0.164	0.200	0.042	0.097	0.087
30, 50	0.10	0.014	0.219	0.330	0.012	0.120	0.117
30, 50	0.30	0.042	0.212	0.331	0.021	0.122	0.116
30, 50	0.50	0.098	0.217	0.334	0.045	0.115	0.113
30, 100	0.10	0.013	0.282	0.527	0.014	0.159	0.166
30, 100	0.30	0.046	0.285	0.534	0.024	0.154	0.165
30, 100	0.50	0.124	0.284	0.530	0.062	0.156	0.172
50, 50	0.10	0.012	0.313	0.568	0.012	0.180	0.232
50, 50	0.30	0.050	0.316	0.563	0.024	0.169	0.224
50, 50	0.50	0.140	0.310	0.568	0.062	0.170	0.234
50, 100	0.10	0.014	0.438	0.884	0.012	0.247	0.439
50, 100	0.30	0.063	0.429	0.879	0.031	0.245	0.435
50, 100	0.50	0.196	0.442	0.881	0.080	0.243	0.428
50, 300	0.10	0.014	0.581	1.000	0.011	0.341	0.831
50, 300	0.30	0.084	0.584	0.999	0.040	0.335	0.825
50, 300	0.50	0.283	0.583	0.999	0.116	0.346	0.831
100, 100	0.10	0.015	0.649	0.994	0.015	0.400	0.827
100, 100	0.30	0.093	0.645	0.994	0.045	0.408	0.832
100, 100	0.50	0.328	0.650	0.995	0.137	0.408	0.832
100, 300	0.10	0.022	0.852	1.000	0.015	0.610	1.000
100, 300	0.30	0.159	0.860	1.000	0.066	0.614	1.000
100, 300	0.50	0.536	0.872	1.000	0.220	0.613	1.000
300, 300	0.10	0.032	0.994	1.000	0.020	0.927	1.000
300, 300	0.30	0.362	0.995	1.000	0.148	0.930	1.000
300, 300	0.50	0.878	0.994	1.000	0.494	0.928	1.000

Table 20 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 45.

Figure 45 Statistical power estimates of the bimodal symmetric distribution at $\alpha = 0.01$



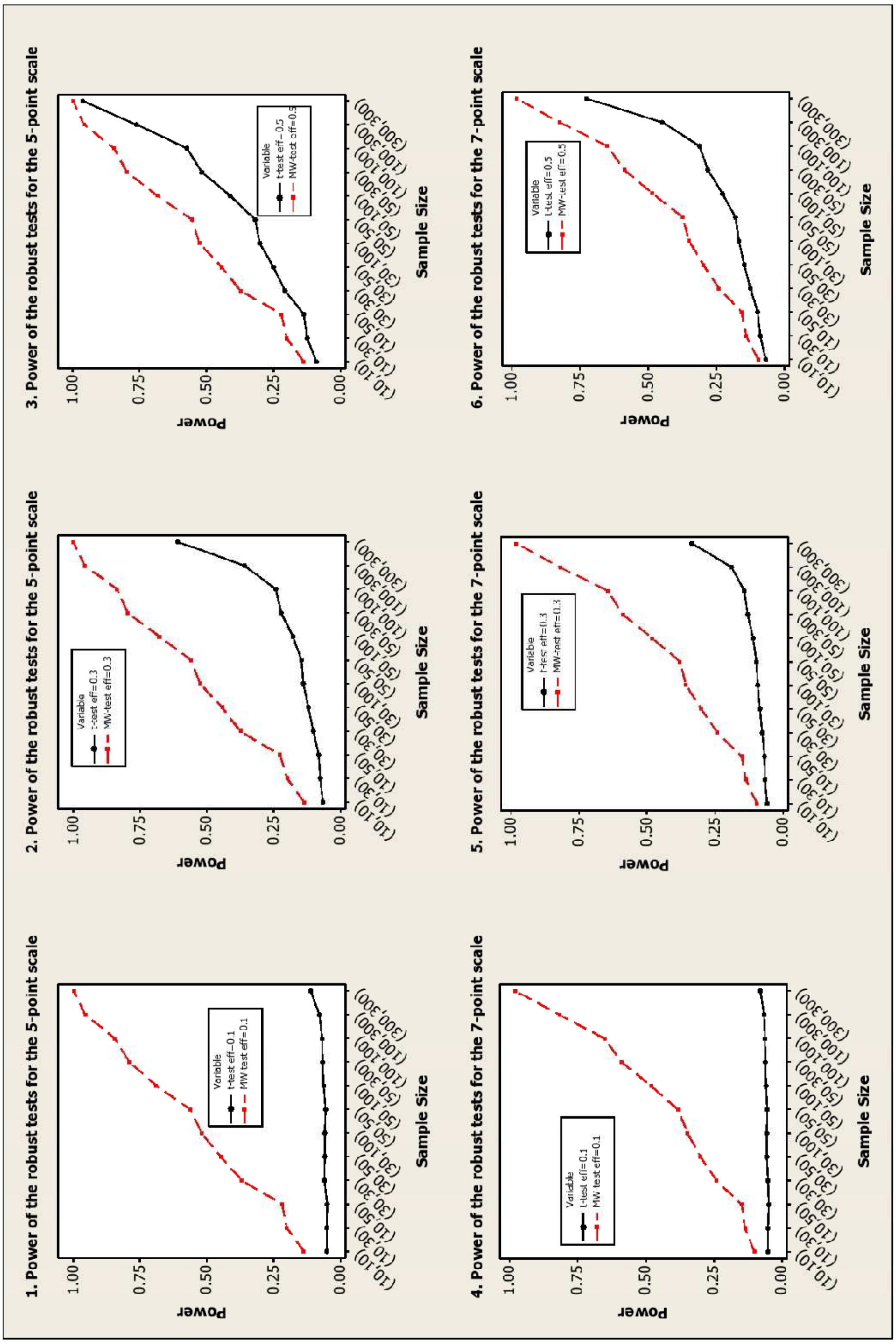
From Figure 45, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1, 0.3 and 0.5 for both the 5-point and 7-point conditions.

Table 21. Statistical power estimates of the bimodal symmetric distribution at α = 0.05

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.052	0.137	0.041	0.051	0.100	0.028
10, 10	0.30	0.066	0.136	0.038	0.058	0.095	0.028
10, 10	0.50	0.091	0.138	0.044	0.068	0.095	0.028
10, 30	0.10	0.052	0.201	0.166	0.050	0.135	0.100
10, 30	0.30	0.076	0.200	0.167	0.066	0.136	0.105
10, 30	0.50	0.125	0.201	0.172	0.087	0.139	0.103
10, 50	0.10	0.050	0.218	0.146	0.048	0.147	0.091
10, 50	0.30	0.082	0.226	0.151	0.068	0.151	0.096
10, 50	0.50	0.138	0.221	0.153	0.097	0.156	0.100
30, 30	0.10	0.060	0.370	0.510	0.051	0.239	0.266
30, 30	0.30	0.102	0.375	0.503	0.076	0.241	0.262
30, 30	0.50	0.208	0.374	0.508	0.124	0.239	0.260
30, 50	0.10	0.058	0.449	0.714	0.055	0.298	0.377
30, 50	0.30	0.121	0.444	0.720	0.084	0.300	0.384
30, 50	0.50	0.251	0.444	0.721	0.146	0.296	0.384
30, 100	0.10	0.059	0.520	0.894	0.054	0.345	0.509
30, 100	0.30	0.139	0.526	0.898	0.093	0.356	0.525
30, 100	0.50	0.303	0.528	0.898	0.166	0.349	0.514
50, 50	0.10	0.056	0.562	0.904	0.053	0.381	0.608
50, 50	0.30	0.147	0.559	0.900	0.097	0.378	0.605
50, 50	0.50	0.319	0.555	0.901	0.178	0.372	0.603
50, 100	0.10	0.063	0.690	0.994	0.057	0.480	0.862
50, 100	0.30	0.178	0.678	0.995	0.110	0.481	0.859
50, 100	0.50	0.412	0.685	0.993	0.227	0.483	0.862
50, 300	0.10	0.067	0.791	1.000	0.060	0.588	0.995
50, 300	0.30	0.223	0.798	1.000	0.129	0.589	0.995
50, 300	0.50	0.518	0.799	1.000	0.280	0.586	0.995
100, 100	0.10	0.068	0.845	1.000	0.061	0.651	0.982
100, 100	0.30	0.242	0.837	1.000	0.142	0.645	0.981
100, 100	0.50	0.574	0.846	1.000	0.311	0.649	0.985
100, 300	0.10	0.079	0.956	1.000	0.064	0.816	1.000
100, 300	0.30	0.360	0.959	1.000	0.189	0.820	1.000
100, 300	0.50	0.763	0.959	1.000	0.447	0.825	1.000
300, 300	0.10	0.112	0.999	1.000	0.078	0.979	1.000
300, 300	0.30	0.610	1.000	1.000	0.335	0.981	1.000
300, 300	0.50	0.963	0.999	1.000	0.724	0.980	1.000

Table 21 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 46.

Figure 46. . Statistical power estimates of the bimodal symmetric distribution at α = 0.05



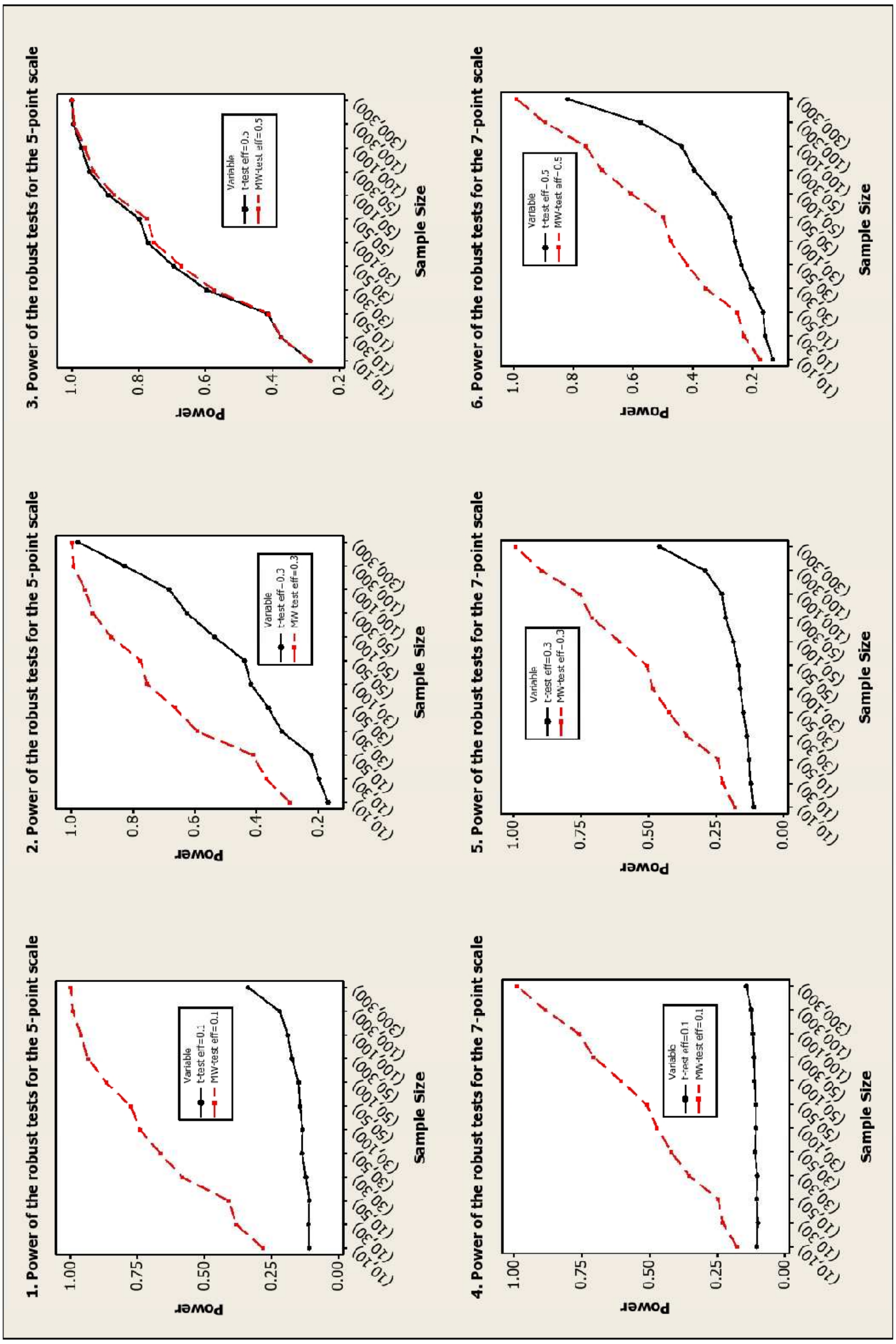
From Figure 46, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1, 0.3 and 0.5 for both the 5-point and 7-point conditions.

Table 22. Statistical power estimates of the bimodal symmetric distribution at α = 0.10

Sample Size	Effect Size	5 point Likert scale			7 point Likert scale		
		t-test	MW-test	KS-test	t-test	MW-test	KS-test
10, 10	0.10	0.110	0.283	0.242	0.103	0.176	0.106
10, 10	0.30	0.168	0.291	0.246	0.110	0.180	0.101
10, 10	0.50	0.287	0.286	0.242	0.131	0.174	0.104
10, 30	0.10	0.112	0.382	0.398	0.098	0.228	0.147
10, 30	0.30	0.198	0.369	0.390	0.120	0.224	0.151
10, 30	0.50	0.376	0.375	0.392	0.156	0.228	0.147
10, 50	0.10	0.110	0.410	0.431	0.103	0.244	0.142
10, 50	0.30	0.223	0.412	0.430	0.127	0.241	0.152
10, 50	0.50	0.416	0.412	0.428	0.163	0.250	0.156
30, 30	0.10	0.123	0.584	0.767	0.101	0.353	0.415
30, 30	0.30	0.319	0.592	0.767	0.134	0.356	0.417
30, 30	0.50	0.598	0.576	0.751	0.202	0.355	0.422
30, 50	0.10	0.137	0.666	0.861	0.108	0.418	0.539
30, 50	0.30	0.362	0.666	0.864	0.148	0.423	0.549
30, 50	0.50	0.696	0.673	0.864	0.236	0.417	0.550
30, 100	0.10	0.135	0.742	0.935	0.105	0.470	0.752
30, 100	0.30	0.419	0.755	0.936	0.160	0.483	0.762
30, 100	0.50	0.773	0.754	0.941	0.258	0.474	0.754
50, 50	0.10	0.144	0.776	0.945	0.106	0.509	0.744
50, 50	0.30	0.439	0.777	0.952	0.167	0.506	0.735
50, 50	0.50	0.799	0.775	0.948	0.275	0.499	0.740
50, 100	0.10	0.150	0.868	0.991	0.111	0.604	0.944
50, 100	0.30	0.536	0.872	0.993	0.185	0.606	0.945
50, 100	0.50	0.891	0.874	0.993	0.328	0.606	0.946
50, 300	0.10	0.174	0.936	1.000	0.113	0.705	0.999
50, 300	0.30	0.626	0.932	0.999	0.214	0.710	0.999
50, 300	0.50	0.948	0.936	1.000	0.394	0.704	1.000
100, 100	0.10	0.190	0.961	1.000	0.117	0.760	0.998
100, 100	0.30	0.684	0.958	1.000	0.228	0.753	0.997
100, 100	0.50	0.972	0.962	1.000	0.436	0.759	0.998
100, 300	0.10	0.220	0.993	1.000	0.122	0.885	1.000
100, 300	0.30	0.829	0.995	1.000	0.290	0.895	1.000
100, 300	0.50	0.996	0.994	1.000	0.574	0.895	1.000
300, 300	0.10	0.338	1.000	1.000	0.141	0.989	1.000
300, 300	0.30	0.980	1.000	1.000	0.459	0.991	1.000
300, 300	0.50	1.000	1.000	1.000	0.819	0.990	1.000

Table 22 indicates that the statistical power will be increased when effect size and sample size are increased in both 5-point and 7-point scale. For more details, Statistical power will be graphed only for the tests in which Type I error were controlled. See Figure 47.

Figure 47. Statistical power estimates of the bimodal symmetric distribution at α = 0.10



From Figure 46, the statistical power of the Mann-Whiney test seems to be superior to the t-test under the effect size = 0.1, 0.3 and 0.5 for the 5-point and 7-point scales. However, the power of the t-test is a bit higher at the effect size = 0.05 for the 5-point scale.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

Conclusion

The objective of this research was to determine the robustness and statistical power of three different methods for testing the hypothesis that ordinal samples of five and seven Likert categories come from equal populations. The three methods are the two sample t-test with equal variance, the Mann-Whitney test, and the Kolmogorov-Smirnov test. In additional, these methods were employed over a wide range of scenarios with respect to sample size, significance level, effect size, and population distribution.

The data simulations and statistical analyses were performed by using R programming language version 2.13.2. To assess the robustness and power, samples were generated from known distributions and compared. According to returned p-values at different nominal significance levels, robustness and power were computed from the error rate of rejecting the null hypotheses. For overall experiments, research questions could be answered as follows.

Results indicate that the two sample t-test and the Mann-Whitney test can control Type I error rate very well for all conditions according to the robustness criteria modified from Bradley (1978). Thus, these two procedures were robust for the ordinal Likert-type data. However, at nominal significance level = 0.01, for all distributions with small sample size the error rates from both tests were widely spread.

For the Kolmogorov-Smirnov test, the Type I error rate was not controlled for all circumstances. This means that this testing procedure computed from R was not robust for the ordinal Likert-type data because the Type I error rate of this test was lower than the minimum of the robustness criteria. Therefore, the Kolmogorov-Smirnov test from R was quite conservative.

For the statistical power of the test, the Mann-Whitney test was the most powerful for most of the distributions, especially under highly skewed and bimodal distribution. The t-test obtained high statistical power or close to the power from the Man-Whitney test under the uniform, moderate skewed or symmetric distribution with large location shift.

Recommendations

For the Likert-type data, if the sample size is small or midsize under any distribution shapes, the Mann-Whitney test will be the preferred procedure to be used because it will be robust and has high statistical power. The t-test is suitable to be used with large sample size ($n > 100$) under the uniform, moderate skewed or symmetric distribution. The Kolmogorov-Smirnov test is not recommended for the ordinal 5-point and 7-point Likert data with lot of ties because the lack of robustness property.

Directions of future research

Other statistical software and statistical procedures for two group samples could be used to confirm the robustness and power of the tests with various number of categories response scales. Especially interesting would be the use of a software package that has the parametric and nonparametric procedures with ties correction for p-value. Also, the use of Likert-type scale that applied to both groups of items and to single items could be compared for the robustness and the power. More research from simulated ordinal data with assumption of unequal difference (i.e. not interval) should be done, as Likert data cannot always be assumed to be like interval data.

APPENDIX A

R CODE FOR TYPE I ERROR RATE OF THE TESTS

```
# example1. R code for Type I error rate for the uniform dist for 5-point scale
# alphav is the significance level = 0.01, 0.05, or 0.10 for this function

type1.5p <- function(alphav)
  {
likert=1:5
ss1=c(10, 10, 10, 30, 30, 30, 50, 50, 50, 100, 100, 300)
ss2=c(10, 30, 50, 30, 50, 100, 50, 100, 300, 100, 300, 300)
#Five distribution shapes
#Uniform dist: use.. prob=c(.2, .2, .2, .2, ,.2)
#Moderately skewed dist: use..Binomial(4,0.3)
#Highly skewed dist: use..Binomial(4,0.1)
#Symmetric dist: use..Biomial(4,0.5)
#Bimodal dist: generated by dividing the density of
#    prob Binomial(4, 0.1) and Binomial(4, 0.9)  by 2
png(filename="D:/figures1.png", width=600,bg="white")
par(mfrow=c(3,4))
# prepare the prob vectors to generate the specific distributions
dist1=c(.2,.2,.2,.2,.2)
dist2=c(dbinom(0,4,.3), dbinom(1,4,.3),dbinom(2,4,.3), dbinom(3,4,.3),
    dbinom(4,4,.3))
dist3=c(dbinom(0,4,.1), dbinom(1,4,.1),dbinom(2,4,.1), dbinom(3,4,.1),
    dbinom(4,4,.1))
dist4=c(dbinom(0,4,.5), dbinom(1,4,.5),dbinom(2,4,.5), dbinom(3,4,.5),
    dbinom(4,4,.5))
  Lt=c(dbinom(0,4,.9), dbinom(1,4,.9),dbinom(2,4,.9), dbinom(3,4,.9),
    dbinom(4,4,.9))
  Rt=dist3
dist5=(Lt+Rt)/2
rej1=0
rej2=0
rej3=0
dat=0
#start sample size comparisons
for (k in 1:12)
    {
```

```
#repeat 100 times to find 100 values of Type I error rates
  for (i in 1:100)
          {
  ch1=0
  ch2=0
  ch3=0
#generating 100 times per testing method.
    for ( j in 1:100)
                {
    group1=sample(likert, size=ss1[k], replace=TRUE,
     prob=dist1)
    group2=sample(likert, size=ss2[k], replace=TRUE,
     prob=dist1)
    pv1<-t.test(group1,group2, var.equal=TRUE)$p.value
    if (pv1 < alphav) {ch1 = ch1+1}
    pv2<-wilcox.test(group1,group2)$p.value
    if (pv2 < alphav) {ch2 = ch2+1}
    pv3<-ks.test(group1,group2)$p.value
    if (pv3 < alphav) {ch3 = ch3+1}
                }
#compute Type one error rates
  rej1[i]=ch1/100
  rej2[i]=ch2/100
  rej3[i]=ch3/100
          }
#create output data
  dat <- data.frame(cbind(rej1,rej2,rej3))
  names(dat) <- c('t-test','M-W','K-S')
 mm=round(mean(dat),digits=3)
 print(mm)
boxplot(dat, ylim=c(0,0.21), main=paste(k, ".", "Uniform-0.10", ss1[k], "-" ,ss2[k]))
       }
}
# type   dev.off( )   after finish to get boxplots in file
```

```
# example2. R code for Type I error rate for the uniform dist for 7-point scale
# alphav is the significance level = 0.01, 0.05, or 0.10 for this function

type1.7p <- function(alphav)
   {
#Alpha values are 0.01, 0.05, and 0.10
#Seven point response scale

likert=1:7
#Twelve sample size combinations
ss1=c(10, 10, 10, 30, 30, 30, 50, 50, 50, 100, 100, 300)
ss2=c(10, 30, 50, 30, 50, 100, 50, 100, 300, 100, 300, 300)
#Five distribution shapes
#Uniform dist: use.. prob=c(.143, .143, .143, .143, ,.143,.143,.143)
#Moderately skewed dist: use..Binomial(6,0.3)
#Highly skewed dist: use..Binomial(6,0.1)
#Symmetric dist: use..Biomial(6,0.5)
#Bimodal dist: generated by dividing the density of
#    prob Binomial(6, 0.1) and Binomial(6, 0.9) by 2
png(filename="D:/figures2.png", width=600,bg="white")
par(mfrow=c(3,4))
# prepare the prob vectors to generate the specific distributions
dist1=c(.143, .143, .143, .143, .143,.143,.143)
dist2=c(dbinom(0,6,.3), dbinom(1,6,.3),dbinom(2,6,.3),
     dbinom(3,6,.3),dbinom(4,6,.3), dbinom(5,6,.3), dbinom(6,6,.3))
dist3=c(dbinom(0,6,.1), dbinom(1,6,.1),dbinom(2,6,.1), dbinom(3,6,.1),
     dbinom(4,6,.1), dbinom(5,6,.1), dbinom(6,6,.1))
dist4=c(dbinom(0,6,.5), dbinom(1,6,.5),dbinom(2,6,.5), dbinom(3,6,.5),
     dbinom(4,6,.5) , dbinom(5,6,.5) , dbinom(6,6,.5))
  Lt=c(dbinom(0,6,.9), dbinom(1,6,.9),dbinom(2,6,.9), dbinom(3,6,.9),
     dbinom(4,6,.9), dbinom(5,6,.9), dbinom(6,6,.9))
  Rt=dist3
dist5=(Lt+Rt)/2
rej1=0
rej2=0
rej3=0
dat=0
#start sample size comparisons
for (k in 1:12)
    {
#repeat 100 times to find 100 values of Type I error rates
 for (i in 1:100)
        {
```

```
        ch1=0
        ch2=0
        ch3=0
      #generating 100 times per testing method.
          for ( j in 1:100)
                {

          group1=sample(likert, size=ss1[k], replace=TRUE,
           prob=dist1)
          group2=sample(likert, size=ss2[k], replace=TRUE,
           prob=dist1)
          pv1<-t.test(group1,group2, var.equal=TRUE)$p.value
          if (pv1 < alphav) {ch1 = ch1+1}
          pv2<-wilcox.test(group1,group2)$p.value
          if (pv2 < alphav) {ch2 = ch2+1}
          pv3<-ks.test(group1,group2)$p.value
          if (pv3 < alphav) {ch3 = ch3+1}
                      }
      #compute Type I error rates
        rej1[i]=ch1/100
        rej2[i]=ch2/100
        rej3[i]=ch3/100
                }
      #create output data
        dat <- data.frame(cbind(rej1,rej2,rej3))
        names(dat) <- c('t-test','M-W','K-S')
       mm=round(mean(dat),digits=3)
       print(mm)
boxplot(dat, ylim=c(0,0.05), main=paste(k, ".", "Uniform-0.01", ss1[k], "-" ,ss2[k]))
            }
      }
 # type  dev.off( )  after finish to get boxplots in file
```

APPENDIX B

R CODE FOR POWER OF THE TESTS

```
# example3. R code of power for the uniform dist for 5-point scale
# alphav is the significance level 0.01, 0.05, or 0.10

power.5p <- function(alphav)
{
likert=1:5
effs=c(0.1,0.3,0.5)
mm=matrix(nrow=3,ncol=3)
ss1=c(10, 10, 10, 30, 30, 30, 50, 50, 50, 100, 100,300)
ss2=c(10, 30, 50, 30, 50, 100, 50, 100 ,300, 100, 300, 300)
dist1=c(.2,.2,.2,.2,.2)
dist2=c(dbinom(0,4,.3), dbinom(1,4,.3),dbinom(2,4,.3), dbinom(3,4,.3),
    dbinom(4,4,.3))
dist3=c(dbinom(0,4,.1), dbinom(1,4,.1),dbinom(2,4,.1), dbinom(3,4,.1),
    dbinom(4,4,.1))
dist4=c(dbinom(0,4,.5), dbinom(1,4,.5),dbinom(2,4,.5), dbinom(3,4,.5),
    dbinom(4,4,.5))
  Lt=c(dbinom(0,4,.9), dbinom(1,4,.9),dbinom(2,4,.9), dbinom(3,4,.9),
    dbinom(4,4,.9))
  Rt=dist3
dist5=(Lt+Rt)/2
rej1=0
rej2=0
rej3=0
dat=0
 for (k in 1:12)
 {
    for (h in 1:3)
    {
      for (i in 1:100)
      {
 ch1=0
 ch2=0
 ch3=0
        for ( j in 1:100)
        {
  group1=sample(likert, size=ss1[k], replace=TRUE,
```

```
     prob=dist1)
    group2=sample(likert, size=ss2[k], replace=TRUE,
    prob=dist1)
    group2p=group2+effs[h]
    pv1<-t.test(group1,group2p, var.equal=TRUE)$p.value
    if (pv1 < alphav) {ch1 = ch1+1}
    pv2<-wilcox.test(group1,group2p)$p.value
    if (pv2 < alphav) {ch2 = ch2+1}
    pv3<-ks.test(group1,group2p)$p.value
    if (pv3 < alphav) {ch3 = ch3+1}
                 }
  rej1[i]=ch1/100
  rej2[i]=ch2/100
  rej3[i]=ch3/100
          }
  dat <- data.frame(cbind(rej1,rej2,rej3))
  mm[h,]=round(mean(dat),digits=3)
         }
  print(mm)
       }
    }
```

```
# example4. R code of power for the uniform dist for 7-point scale
# alphav is the significance level 0.01, 0.05, or 0.10

power.7p <- function(alphav)
  {
likert=1:7
effs=c(0.1,0.3,0.5)
mm=matrix(nrow=3,ncol=3)
ss1=c(10, 10, 10, 30, 30, 30, 50, 50, 50, 100, 100, 300)
ss2=c(10, 30, 50, 30, 50, 100, 50, 100, 300, 100, 300, 300)
dist1=c(.143, .143, .143, .143, .143,.143,.143)
dist2=c(dbinom(0,6,.3), dbinom(1,6,.3),dbinom(2,6,.3),
      dbinom(3,6,.3),dbinom(4,6,.3), dbinom(5,6,.3), dbinom(6,6,.3))
dist3=c(dbinom(0,6,.1), dbinom(1,6,.1),dbinom(2,6,.1), dbinom(3,6,.1),
      dbinom(4,6,.1), dbinom(5,6,.1), dbinom(6,6,.1))
dist4=c(dbinom(0,6,.5), dbinom(1,6,.5),dbinom(2,6,.5), dbinom(3,6,.5),
      dbinom(4,6,.5) , dbinom(5,6,.5) , dbinom(6,6,.5))
   Lt=c(dbinom(0,6,.9), dbinom(1,6,.9),dbinom(2,6,.9), dbinom(3,6,.9),
      dbinom(4,6,.9), dbinom(5,6,.9), dbinom(6,6,.9))
   Rt=dist3
```

```
dist5=(Lt+Rt)/2
rej1=0
rej2=0
rej3=0
dat=0
for (k in 1:12)
  {
for (h in 1:3)
     {
   for (i in 1:100)
        {
  ch1=0
  ch2=0
  ch3=0
    for ( j in 1:100)
            {
    group1=sample(likert, size=ss1[k], replace=TRUE,
     prob=dist1)
    group2=sample(likert, size=ss2[k], replace=TRUE,
     prob=dist1)
     group2p=group2+effs[h]
    pv1<-t.test(group1,group2p, var.equal=TRUE)$p.value
    if (pv1 < alphav) {ch1 = ch1+1}
    pv2<-wilcox.test(group1,group2p)$p.value
    if (pv2 < alphav) {ch2 = ch2+1}
    pv3<-ks.test(group1,group2p)$p.value
    if (pv3 < alphav) {ch3 = ch3+1}
              }
  rej1[i]=ch1/100
  rej2[i]=ch2/100
  rej3[i]=ch3/100
            }
  dat <- data.frame(cbind(rej1,rej2,rej3))
  mm[h,]=round(mean(dat),digits=3)
           }
  print(mm)
      }
    }
```

REFERRENCES

Abdel-Khalek, A. M. (1998). Single-versus multi-item scales in measuring death anxiety. *Death Studies*, *22*(8), 763.

Acock, A. C., & Martin, J. D. (1974). The Undermeasurement controversy: Should ordinal data be treated as interval? *SO - Sociology & Social Research. 58, 4, 427-33, Jul 74.*

Arnold, W. E., McCrosby, J. C., & Prichard, S. V. O. (1967). The Likert Scale. *Today's Speech*, *15,* 31-33.

Blair, R. C. (1983). An Investigation of the robustness of the t test to unit of analysis violations. *Educational & Psychological Measurement, 43*(1), 69-80.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144-152.

Brown, J. D. (2000). What issues affect Likert-scale questionnaire formats? *Shiken:JALT Testing & Evaluatin SIG Newsletter*, *4*(1), 18-21.

Chow, P., & Winzer, M. M. (1992). Reliability and validity of a scale measuring attitudes toward mainstreaming. *Educational and Psychological Measurement, 52*(1), 223.

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education, 35*(4), 31-35.

Cliff, N. (1984). *Ordinal Measurement.* Paper presented at the Annual Meeting of the American Psychological Association (92nd, Toronto, Ontario, Canada, August 24-28, 1984).

Crumbley, L., Henry, B. K., & Kratchman, S. H. (2001). Students' perceptions of the evaluation of college teaching. *Quality Assurance in Education, 9*(4), 197.

Delaney, H. D., & Vargha, A. (2000). *The Effect of Nonnormality on Student's Two-Sample T Test.* Paper presented at the Annual Meeting of the American Educational Research Association (81st, New Orleans, LA, April 24-28, 2000). Sponsored by Hungarian OTKA, Hungarian FKFP, and the Open Society Support Foundation.

Fahoome, G. F. (1999). *A Monte Carlo Study of Twenty-one Nonparametric Statistics with Normal and non-normal Data.* Unpublished PhD. Dissertation, Wayne State University.

Garson, G. D. (2002, 09/21/02). *Data Levels and Measurement*, from http://www2.chass.ncsu.edu/garson/pa765/datalevl.htm

Gibbons, J. D., & Chakraborti, S. (1992). *Nonparametric statistical inference* (3rd ed.). New York: M. Dekker.

Goldstein, G., & Hersen, M. (1984). *Handbook of psychological assessment*. New York: Pergamon Press.

Gray, M., & Bergmann, B. R. (2003). Student teaching evaluations: Inaccurate, demeaning, misused. *Academe, 89*(5), 44.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research, 71*(1), 105-131.

Hodgson, P. (2003, 06/17/03). *Usability Test Data*, from http://www.userfocus.co.uk/articles/datathink.html

Kennedy, W. J., & Gentle, J. E. (1980). *Statistical computing*. New York: M. Dekker.

Kolitch, E., & Dean, A. V. (1999). Student ratings of instruction in the USA: Hidden assumptions and missing conceptions about 'good' teaching. *Studies in Higher Education, 24*(1), 27.

Kromrey, J. D., & Hogarty, K. Y. (1998). Analysis options for group difference on ordered categorical variables: An empirical investigation of Type I error control and statistical power. *Multiple Linear Regression Viewpoints, 25*, 70-82.

Liaw, S.-H., & Goh, K.-L. (2003). Evidence and control of biases in student evaluations of teaching. *The International Journal of Educational Management, 17*(1), 37.

Likert, R. (1932). A technique for the measurement of attitudes. *Achieves of Psychology, 140*, 5-53.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50-60.

Maurer, T. J., & Andrews, K. D. (2000). Traditional, Likert, and simplified measures of self-efficacy. *Educational and Psychological Measurement, 60*(6), 965.

Micceri, T. (1989). The Unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156.

Michals, L. (2003). Duluth, Minn., students use web site to rate teachers. *Knight Ridder Tribune Business News*, 1.

Miller, L. E. (1998). Appropriate analysis. *Journal of Agricultural Education, 39*(2), 1-10.

Mogey, N. (1999, 03/25/99). *So you want to use a Likert scale?*, from http://www.icbl.hw.ac.uk/ltdi/cookbook/info_likert_scale/

Moses, L. E. (1964). One sample limits of some two-sample rank tests. *Journal of the American Statistical Association, 59*(307), 645-651.

Nerger, J. L., Viney, W., & Riedel, R. G. I. (1997). Student ratings of teaching effectiveness: Use and misuse. *The Midwest Quarterly, 38*(2), 218.

Panasuk, R. M., & LeBaron, J. (1999). Student feedback: A tool for improving instruction in graduate education. *Education, 120*(2), 356.

Ramsey, P. H. (1980). Exact Type 1 error rates for robustness of Student's t test with unequal variances. *Journal of Educational Statistics, 5*(4), 337-349.

Sawilowsky, S. S., & Blair, R. C. (1992). A more realistic look at the robustness and Type II error properties of the t-test to departure from population normality. *Psychological Bulletin, 111*(2), 352-360.

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education, 22*(3), 199.

Stennet, B. (2002, 08/15/02). *Opinion Survey Rating Scale*, from http://www.assessmentplus.com/articles/opinion_survey_rating_scales.htm

Stevens, S. (1946). On the theory of scales of measurement. *Science, 103*, 677-680.

Student. (1908). Probable error of a mean. *Biometrika, 6*(1), 1-25.

Traylor, M. (1983). Ordinal and interval scaling. *Journal of Market Research Society, 25*(4), 297-303.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83.

Worthington, A. C. (2002). The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education. *Assessment and Evaluation in Higher Education, 27*(1), 49.

Zimmerman, D. W. (1996). A note on homogeneity of variance of scores and ranks. *Journal of Experimental Education, 64*(4), 351-362.

Zimmerman, D. W. (1998). Invalidation of parametric and nonparametric statistical tests by concurrent violation of two assumptions. *Journal of Experimental Education, 67*(1), 55-68.