**TO TEST OR NOT TEST: PRELIMINARY ASSESSMENT OF NORMALITY WHEN COMPARING TOW INDEPENDENT SAMPLES**

**Brief Overview of the Paper**

**Main Objective:** The main objective of this paper is to evaluate the impact pf normality assumption on the results of two-sample t-test. It is aimed at investigating whether the t-test is robust to moderate deviations from normality and whether alternative nonparametric test is needed when normality is violated.

The authors began by outlining the increasing importance of statistical test in several research fields of study particularly in the medical field where most clinical trials yield data where the objective is to compare whether two independent samples differ from each other or not. They however decry the numerous statistical errors usually reported in these journals arising from using test procedures that require normally distributed data on data that is skewed (Olse,2003). The student’s t test procedure which is the most common procedure used in most situations assumes that the samples are drawn from a normally distributed population with equal variance and so if the assumptions are violated, the test statistic is compared with the wrong reference distribution, which may result in deviation of the actual Type I Error from the nominal significance level, and loss of power. They emphasize the need to check distributional assumptions before performing the parametric t test procedure on data resulting from these types of research. They noted that data resulting from medical research are usually not normally distributed making the parametric test procedures less reliant. The nonparametric, Mann-Whitney U test is usually recommended when distributional assumptions are not satisfied. Vickers, in 2005 also suggested that such data are best analyzed with analysis of covariance. Conventional statistical practice for comparing continuous outcomes from independent samples is to use a pretest for normality before testing the main hypothesis. This procedure has, however, been criticized for its theoretical drawbacks. First, the mere fact that there is not enough evidence to reject the null hypothesis does not suggest that the data are normally distributed. Schucany and Ng, 2006 emphasized that preliminary testing are usually about assumptions of characteristics of the population and not characteristics of the samples. This problem is more pronounced when the sample size is small, according to. Schucany and Ng, 2006. The paper also criticized that some preliminary tests require their own assumptions raising the question as to whether these assumptions also need to be examined. It is again noted that if the pretest fails, the actual test may still be robust for some data and so preliminary tests are not that necessary. According to the paper, some authors also argued that the fact that preliminary tests are applied to the same data as the subsequent test, can result in uncontrolled type I error rates. Rasch et al ,2011 argued that assumptions underlying the two-sample t test should not be pretested as pretesting leads to unknown final Type I and Type II Error Rates if the respective statistical tests are performed using the same set of observations.

In the current paper, the authors investigated the statistical properties of the student’s t test and Mann-Whitney’s U test by comparing two independent groups with different selection procedures. A simulation study is conducted by drawing equal samples of sizes 20, 30, 40, and 50 for two independent groups from three different distributions, exponential, Uniform, and Normal distributions. Two selection strategies were examined. In one strategy, the two samples t test was conducted if both samples had passed the preliminary Shapiro-Wilk test for normality, otherwise the Mann-Whitney’s U test was used whereas in the other strategy, the t test was used if the residuals of the collapsed samples and had passed Shapiro-Wilk test for normality otherwise the Mann-Whitney’s U test was used.

In the first simulation the procedure was repeated until 10000 pairs of samples had passed the preliminary screening for normality by either strategy I or II with or no pretest. The conditional Type I Error Rates were then estimated by the number of significant t tests divided by 10000. In the second simulation the procedure was repeated until 10000 pairs of samples had failed the preliminary screening for normality by either strategy I or II with or no pretest. The conditional Type I Error Rates were then estimated by the number of significant t tests divided by 10000. Lastly, 100000 pairs of samples were generated to assess the unconditional Type I Error Rate of the entire procedure. Depending on whether the preliminary Shapiro Wilk test was significant or not, Student’s t test of Mann-Whitney U test was performed for the main analysis. The Type I error rate of the entire two-stage procedure was estimated by dividing the number of significant tests (t or U) and divided by 100000.

**Results and Analysis**

**Estimated Type I Errors of two sample t test after a passed Shapiro-Wilk test for normality.**

The tables below contain the estimated type I error rate for the two-sample t test for samples drawn from uniform and normally distributed population at after both samples had passed Shapiro-Wilk test for normality. To the left are our simulations, and to the right are the estimates from the paper “To test or not to test.” Figures 1.1 and 1.2 below are the plot of Type I error rate against sample sizes for normal and uniform distribution.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estimates from Our simulation | | | | | |  | Estimates from the paper | | | | | |
| Uniform Distribution | | | | | |  | Uniform Distribution | | | | | |
| **α\_pre/n** | **10** | **20** | **30** | **40** | **50** |  | **α\_pre/n** | **10** | **20** | **30** | **40** | **50** |
| **0.1** | 0.042 | 0.044 | 0.039 | 0.035 | 0.036 |  | **0.1** | 0.043 | 0.044 | 0.039 | 0.039 | 0.036 |
| **0.05** | 0.044 | 0.044 | 0.042 | 0.038 | 0.037 |  | **0.05** | 0.043 | 0.037 | 0.04 | 0.04 | 0.037 |
| **0.01** | 0.049 | 0.046 | 0.046 | 0.044 | 0.044 |  | **0.01** | 0.049 | 0.05 | 0.046 | 0.045 | 0.041 |
| **0.005** | 0.052 | 0.049 | 0.047 | 0.045 | 0.045 |  | **0.005** | 0.052 | 0.05 | 0.048 | 0.044 | 0.043 |
| **w/o pretest** | 0.053 | 0.048 | 0.051 | 0.052 | 0.049 |  | **w/o pretest** | 0.058 | 0.047 | 0.052 | 0.047 | 0.05 |
| **Normal Distribution** | | | | | |  | **Normal Distribution** | | | | | |
| **α\_pre/n** | 10 | 20 | 30 | 40 | 50 |  | **α\_pre/n** | 10 | 20 | 30 | 40 | 50 |
| **0.1** | 0.05 | 0.053 | 0.05 | 0.048 | 0.05 |  | **0.1** | 0.049 | 0.053 | 0.05 | 0.049 | 0.05 |
| **0.05** | 0.044 | 0.049 | 0.047 | 0.05 | 0.047 |  | **0.05** | 0.049 | 0.05 | 0.05 | 0.053 | 0.046 |
| **0.01** | 0.05 | 0.049 | 0.054 | 0.049 | 0.048 |  | **0.01** | 0.05 | 0.05 | 0.047 | 0.048 | 0.051 |
| **0.005** | 0.046 | 0.051 | 0.046 | 0.053 | 0.05 |  | **0.005** | 0.047 | 0.047 | 0.05 | 0.054 | 0.05 |
| **w/o pretest** | 0.048 | 0.051 | 0.047 | 0.049 | 0.05 |  | **w/o pretest** | 0.051 | 0.053 | 0.049 | 0.053 | 0.05 |

Table 1.1 Table 1.2

Chart

Description automatically generated with medium confidence Chart

Description automatically generated

Figure 1.1 Figure 1.2

**Estimated Type I Error probability of the two-stage procedure (Student’s t test or Mann-Whitney’s U test depending on preliminary Shapiro Wilk test for normality) for different sample sizes and**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Estimates from Our simulation | | | | | |  | Estimates from the paper | | | | | |
| Exponential Distribution | | | | | |  | Exponential Distribution | | | | | |
| **α\_pre/n** | **10** | **20** | **30** | **40** | **50** |  | **α\_pre/n** | **10** | **20** | **30** | **40** | **50** |
| **0.1** | 0.050 | 0.052 | 0.049 | 0.050 | 0.049 |  | **0.1** | 0.053 | 0.050 | 0.048 | 0.049 | 0.048 |
| **0.05** | 0.050 | 0.051 | 0.05 | 0.050 | 0.050 |  | **0.05** | 0.055 | 0.052 | 0.048 | 0.049 | 0.050 |
| **0.01** | 0.051 | 0.052 | 0.049 | 0.050 | 0.049 |  | **0.01** | 0.054 | 0.054 | 0.048 | 0.049 | 0.050 |
| **0.005** | 0.051 | 0.051 | 0.051 | 0.051 | 0.050 |  | **0.005** | 0.050 | 0.055 | 0.049 | 0.048 | 0.049 |
| **Normal Distribution** | | | | | |  | **Normal Distribution** | | | | | |
| **α\_pre/n** | 10 | 20 | 30 | 40 | 50 |  | **α\_pre/n** | 10 | 20 | 30 | 40 | 50 |
| **0.1** | 0.05 | 0.052 | 0.052 | 0.051 | 0.051 |  | **0.1** | 0.051 | 0.052 | 0.053 | 0.051 | 0.051 |
| **0.05** | 0.05 | 0.051 | 0.051 | 0.051 | 0.051 |  | **0.05** | 0.051 | 0.051 | 0.051 | 0.051 | 0.050 |
| **0.01** | 0.051 | 0.051 | 0.052 | 0.052 | 0.052 |  | **0.01** | 0.051 | 0.051 | 0.051 | 0.051 | 0.051 |
| **0.005** | 0.05 | 0.051 | 0.052 | 0.051 | 0.051 |  | **0.005** | 0.051 | 0.050 | 0.049 | 0.050 | 0.050 |
| **Uniform Distribution** | | | | | |  | **Uniform Distribution** | | | | | |
| **α\_pre/n** | 10 | 20 | 30 | 40 | 50 |  | **α\_pre/n** | 10 | 20 | 30 | 40 | 50 |
| **0.1** | 0.052 | 0.050 | 0.051 | 0.049 | 0.049 |  | **0.1** | 0.052 | 0.051 | 0.048 | 0.049 | 0.049 |
| **0.05** | 0.051 | 0.052 | 0.051 | 0.05 | 0.051 |  | **0.05** | 0.053 | 0.051 | 0.051 | 0.050 | 0.048 |
| **0.01** | 0.051 | 0.052 | 0.051 | 0.050 | 0.050 |  | **0.01** | 0.051 | 0.051 | 0.052 | 0.051 | 0.051 |
| **0.005** | 0.052 | 0.052 | 0.051 | 0.051 | 0.048 |  | **0.005** | 0.052 | 0.050 | 0.050 | 0.052 | 0.050 |

Table 1.3 Table 1.4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Exponential Distribution | | | | | |
| **α\_pre/n** | 5 | 10 | 15 | 20 | 25 |
| **0.1** | 0.033 | 0.075 | 0.119 | 0.142 | 0.155 |
| **0.05** | 0.033 | 0.067 | 0.096 | 0.118 | 0.135 |
| **0.01** | 0.03 | 0.05 | 0.073 | 0.092 | 0.105 |
| **0.005** | 0.029 | 0.049 | 0.065 | 0.079 | 0.1 |
| **w/o pretest** | 0.026 | 0.036 | 0.04 | 0.044 | 0.046 |

Table 1.5

**Observations:**

For data sampled from exponential distribution, have Type I error rate inflated above the nominal significance level of 0.05 by the preliminary test for normality. Type I error rate is rather lower than the nominal significance level of 0.05 without the pretest. This observation is the opposite for data drawn from uniform distribution. Type I error rates for normally distributed data are similar to the nominal significance level of 0.05. Type I error rates increase with an increase in sample size for exponential distribution but decreases with sample size for uniform distribution.

Conditional type I error rate decreases slightly with sample sizes but increases slightly with . For example, conditional Type Error rate decreases from 0.042 to 0.036 from sample sizes 10 to 50 but not in a consistent manner for uniform distribution. As decreases from 0.1 to 0.005, conditional type I error rate decreases from 0.053 to 0.042 for a sample size of 10 for the uniform distribution. Type I error rates for normally distributed data are similar to the nominal significance level of 0.05. Conditional Type I error rates for uniformly distributed data, however, are generally lower than the nominal significance level of 0.05 except for a few.

The two results are relatively the same with a few differences in decimal places. There are no results for exponential. It has a longer runtime after sample size 30. I verified that my code works with different sample sizes and the result is shown below. The plot of Type I error rate against various sample sizes are also shown in table 1.5 above.