**Brief Overview of the paper:** To test or not to test: Preliminary Assessment of normality when comparing two Independent Samples

According to the paper, the student’s t test is generally used for comparing the means of two independent samples. To control Type I Error Rate, we require the samples to meet the assumption under the null hypothesis, that is the two samples are drawn from a normally distributed population. In most cases, preliminary Shapiro Wilk is performed to check this condition and if the SW test is nonsignificant, we proceed to perform the t test, otherwise, the nonparametric equivalent, Wilcoxon test is performed.

**Main Objective:** The main objective of this paper is to evaluate the impact of normality assumption on the results of two-sample t-test.

**Method:**  A simulation study is conducted to assess the impact of preliminary test for normality on the results of two sample t tests. This is done using two strategies. In one strategy, they performed the t test on the samples if both samples passed Shapiro wilk test and in the other strategy, the t test was performed if the collapsed set of residuals from both samples passed Shapiro wilk test for normality. Finally, they estimated the conditional and overall Type I Error Rate for the parametric and non-parametric procedures for both strategies.

In the first run, equal sample sizes 10,20,30,40, and 50 for two groups were sampled from exponential, uniform, and normal distribution. The procedure is repeated until 10,000 pairs of the samples had passed the preliminary screening for normality (either strategy I or II), with .100, .050, .010, .005 or no pretest. Type I error rate was estimated as the number of significant t tests divided by 10,000. In the second run, the samples were generated until 10,000 pairs had failed preliminary screening for normality (either strategy I or II), and the conditional type I error rate was estimated as the number of significant Mann-Whitney’s U test divided by 10,000. Finally, 100,000 pairs of samples were generated from each distribution and depending on whether the preliminary Shapiro-Wilk test was significant or not a t test of Mann Whitney’ U test was conducted, and the unconditional type I error rate was estimated as the number of significant tests (t or U) divided by 100,000.

**Results:**  For strategy I, it is observed that conditional type I error rate for exponential distribution was inflated, and this was especially more pronounced for larger sample sizes and large pretest values. However, conditional type I error rate decreases below the nominal level especially for larger sample sizes and pretest values for uniform distribution. Conditional and unconditional Type I error rates were similar to the nominal significance level for normal distribution. For the nonparametric method, conditional Type I error rates for exponential were just a little above the nominal significance level. Type I error rates were more inflated for uniform and normal distribution.

For strategy I, it was shown that conditional Type I and unconditional Type I error rates for normal distribution were very close to the nominal significance level for all sample sizes and pretest values. There was significant inflation of Type I error rates for exponential and uniform distributions.

**Findings:**

1. Preliminary screening for normality is not always a problem if the population distribution is known. The problem is when the population distribution is unknown, which unfortunately is mostly the situation.
2. It is shown that conditional Type I Error Rates for both parametric and non-parametric tests were altered. This deviation was more evident with small sample sizes and low significance level of the pretest. The paper attributed this to selection mechanisms due to the preliminary test.
3. Preliminary testing for normality affects Type I error probabilities of the main test especially for nonnormal distributions.

**Discussion:**

1. Preliminary testing for normality is unnecessary and should therefore be avoided.
2. Just because the results of the Shapiro Wilk test are not significant does not suggest that the sample is normally distributed. A sample is a fixed set of “realizations”, It is not a random variable which could follow some distribution.
3. Preliminary testing for normality filters sample that are non-representative of the population, thus altering Type I error rates.