Analyzing the Impact of Sample Size and Population Distribution on Statistical Test Power

Ashraf Hassan Idris Brama[1], Monastir Abbas Ahmed[2]

[1]Department of Management Information System & Product Management, Collage of Business & Economics, Qassim university, Buraydah, Kingdom of Saudi Arabia, a.brama[at]qu.edu.sa, stat.ashraf[at]gail.com.

[2]Department of Management Information System & Product Management, Collage of Business and Economics, Qassim university, Buraydah, Kingdom of Saudi Arabia, montasirstat[at]gmail.com

Abstract: ***This study investigates the influence of sample size and population distribution on the statistical power of tests, aiming to determine the optimal size that yields significant results. Utilizing a descriptive analytical approach, the researcher generated 10, 000 random units using the Pass12 program. Findings reveal that statistically normal distributions exhibit higher power, particularly for small sample sizes, compared to non-normal distributions. Regularly distributed data require larger samples for adequate power, while logistic distributions demonstrate increased statistical test power with larger sample sizes. The study underscores the significance of tailoring sample sizes to population distribution types to enhance test power, along with employing estimates to minimize standard deviation and mitigate skewness for improved statistical outcomes.***

Keyword: statistical power, sample size, population distribution, normal distribution, logistic distribution

1.Introduction

The statistical power is covered in detail in a number of texts (*Kraemer and Thiemann 1987, Cohen 1988, Lipsey 1990; see also a particularly clear paper by Muller and benignus 1992*). Briefly, the power of a test is the probability of rejecting the null hypothesis given that the alternative hypothesis is true. Power depends on the type of test, increases with increasing sample size, effect size, and higher α-level, and declines with increasing sampling variance. Effect size is the difference between the null and alternative hypotheses, and can be measured either using raw or standardized values. Raw measures, such as the slope in a regression analysis or difference between means in a t-test, are closer to the measurements that researchers take and so are easier to visualize and interpret. Standardized measures, such as the correlation coefficient or d-value (difference in means divided by the standard deviation), are dimensionless and incorporate the sampling variance implicitly, removing the need to specify variance when calculating power.

2.Statistical Power

Statistical power is defined as the probability of correctly rejecting the null hypothesis when it is false (Cohen, 1988), i. e., power (π) $=1-\beta$, where β represents the probability of a Type II error. Power can be affected by factors such as the significance criterion (α), sample size, number of groups or levels, effect size, and number of dependent variables. With respect to structural equation models, statistical power has been explored in two ways: the first investigates the power associated.

A concept closely aligned to type II error is statistical power. Statistical power is a crucial part of the research process that is most valuable in the design and planning phases of studies, though it requires assessment when interpreting results. Power is the ability to correctly reject a null hypothesis that is indeed false. Unfortunately, many studies lack sufficient power and should be presented as having inconclusive findings. Power is the probability of a study to make correct decisions or detect an effect when one exists.

The power of a statistical test is dependent on: the level of significance set by the researcher, the sample size, and the effect size or the extent to which the groups differ based on treatment. Statistical power is critical for healthcare providers to decide how many patients to enroll in clinical studies. Power is strongly associated with sample size; when the sample size is large, power will generally not be an issue. Thus, when conducting a study with a low sample size, and ultimately low power, researchers should be aware of the likelihood of a type II error. The greater the N within a study, the more likely it is that a researcher will reject the null hypothesis. The concern with this approach is that a very large sample could show a statistically significant finding due to the ability to detect small differences in the dataset; thus, utilization of p values alone based on a large sample can be troublesome .

It is essential to recognize that power can be deemed adequate with a smaller sample if the effect size is large. What is an acceptable level of power? Many researchers agree upon a power of 80% or higher as credible enough for determining the actual effects of research studies. Ultimately, studies with lower power will find fewer true effects than studies with higher power; thus, clinicians should be aware of the likelihood of a power issue resulting in a type II error. Unfortunately, many researchers, and providers who assess medical literature, do not scrutinize power analyses. Studies with low power may inhibit future work as they lack the ability to detect actual effects with variables; this could lead to potential impacts remaining undiscovered or noted as not effective when they may be.

Medical researchers should invest time in conducting power analyses to sufficiently distinguish a difference or association. Luckily, there are many tables of power values as well as statistical software packages that can help to determine study power and guide researchers in study design and analysis. If choosing to utilize statistical software to calculate power, the following are necessary for entry: the predetermined alpha level, proposed sample size, and effect size the investigator (s) is aiming to detect. By utilizing power calculations on the front end, researchers can determine adequate sample size to compute effect, and determine based on statistical findings; sufficient power was actually observed.

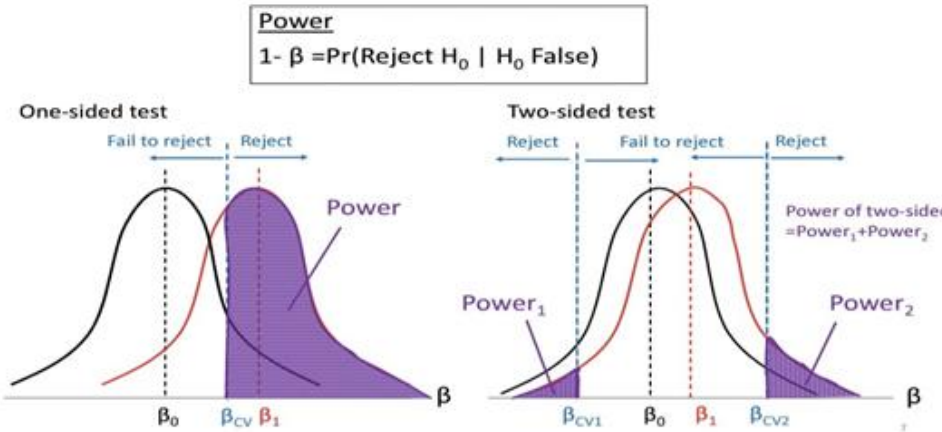

Figure 1: illustrate the power test

3.Type I and Type II Errors:

Type I and Type II errors can lead to confusion as providers assess medical literature. A vignette that illustrates the errors is the Boy Who Cried Wolf. First, the citizens commit a type I error by believing there is a wolf when there is not. Second, the citizens commit a type II error by believing there is no wolf when there is one.

A type I error occurs when in research when we reject the null hypothesis and erroneously state that the study found significant differences when there indeed was no difference. In other words, it is equivalent to saying that the groups or variables differ when, in fact, they do not or having false positives. An example of a research hypothesis is below:

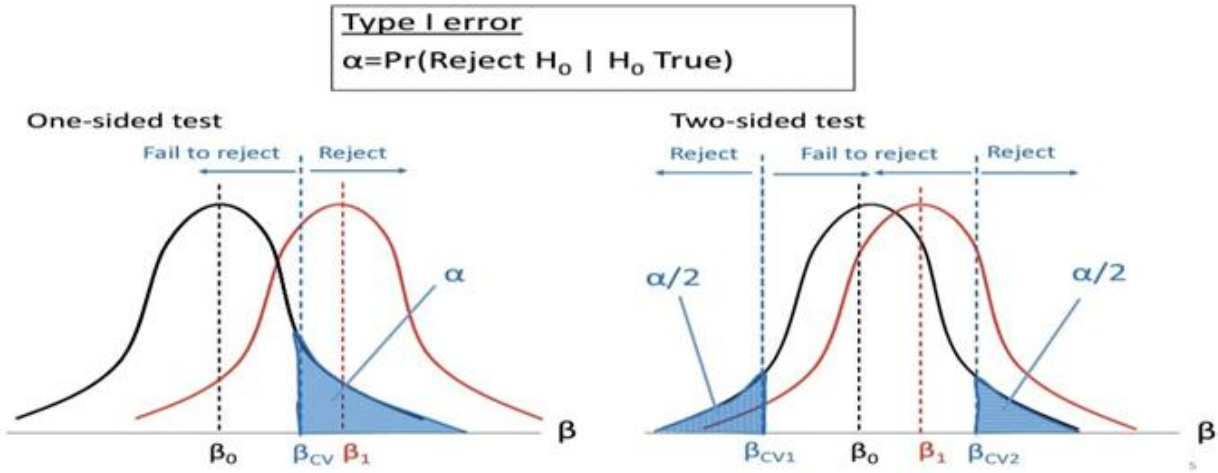

Figure 2: illustrate the type 1 error

For our example, if we were to state that Drug 23 significantly reduced symptoms of Disease A compared to Drug 22 when it did not, this would be a type I error. Committing a type, I error can be very grave in specific scenarios. For example, if we did, move ahead with Drug 23 based on our research findings even though there was actually was no difference between groups, and the drug costs significantly more money for patients or has more side effects, then we would raise healthcare costs, cause iatrogenic harm, and not improve clinical outcomes. If a p-value is used to examine type I error, the lower the p-value, the lower the likelihood of the type I error to occur.

4.Type II Error and Statistical Power of a Test:

What is the power of a test? The power of a statistical test is the probability that it will correctly lead to the rejection of a false null hypothesis (Greene 2000). The statistical power is the ability of a test to detect an effect, if the effect actually exists (High 2000). Cohen (1988) says, it is the probability that it will result in the conclusion that the phenomenon exists (p.4). A statistical power analysis is either retrospective (post hoc) or prospective (a priori). A prospective analysis is often used to determine a required sample size to achieve target statistical power, while a retrospective analysis computes the statistical power of a test given sample size and effect size.

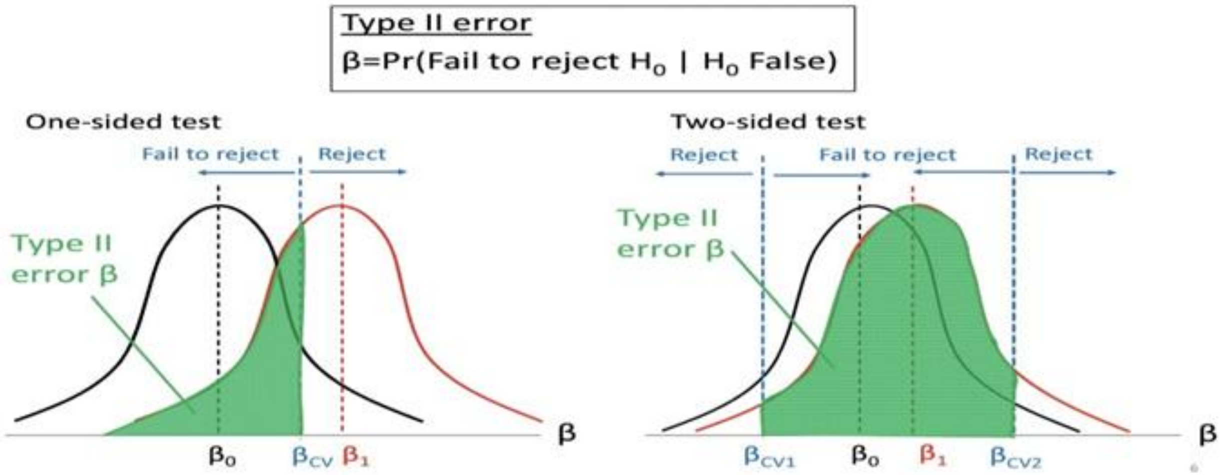

Figure 3: illustrate the type two error

Data analysis:

What is the effect of sample size on the independent sample t test?

Table 1: illustrate the independent sample t test with data that follow the uniform distribution

Power	N1	N2	Ratio	Alpha	Beta	Mean1	Mean2	S1 & s2
0.56201	10	10	1.000	0.05000	0.43799	0.0	1.0	1.0
0.86895	20	20	1.000	0.05000	0.13105	0.0	1.0	1.0
0.96771	30	30	1.000	0.05000	0.03229	0.0	1.0	1.0
0.99298	40	40	1.000	0.05000	0.00702	0.0	1.0	1.0
0.99861	50	50	1.000	0.05000	0.00139	0.0	1.0	1.0
0.99974	60	60	1.000	0.05000	0.00026	0.0	1.0	1.0
0.99995	70	70	1.000	0.05000	0.00005	0.0	1.0	1.0
0.99999	80	80	1.000	0.05000	0.00001	0.0	1.0	1.0
1.00000	90	90	1.000	0.05000	0.00000	0.0	1.0	1.0
1.00000	100	100	1.000	0.05000	0.00000	0.0	1.0	1.0

Source pass 12

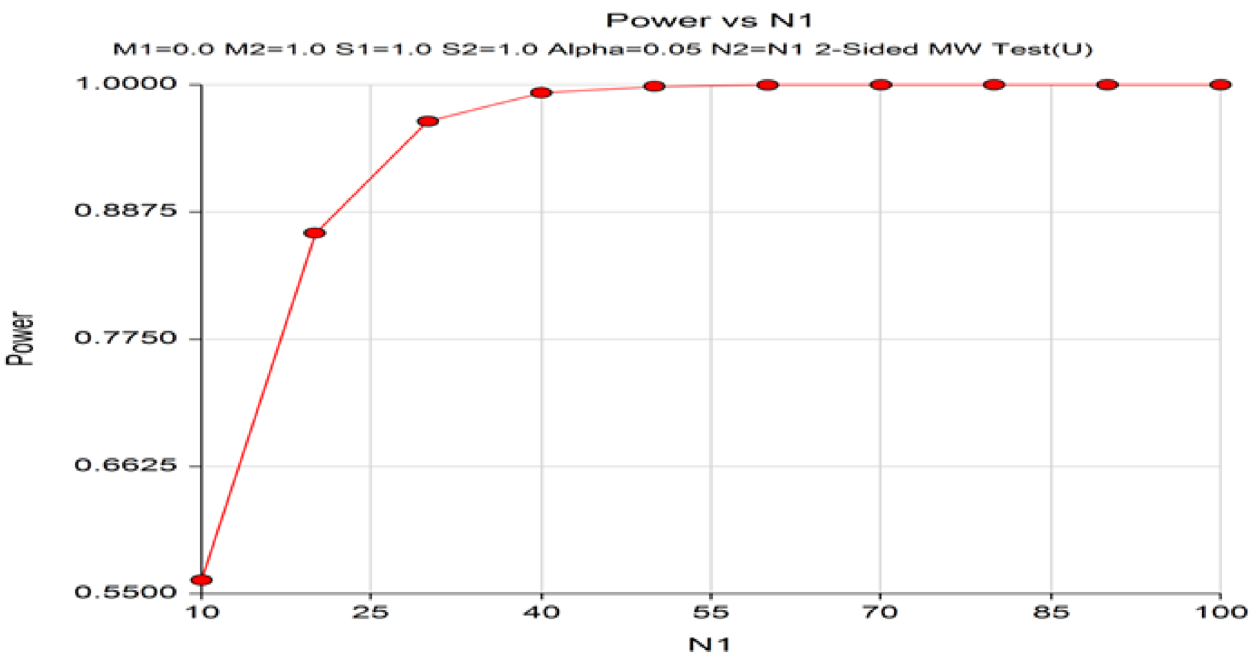

Figure 4: illustrate the relationship between the sample size and the power test

The previous results indicate that there is a significant effect of the sample size of the Test Force (t) for independent eyes of the data following the uniform distribution, where it was found that when the sample size was (10) the power test was

(0.56) and the beta value is (0.43799) and when the sample size was (20) the power test is (0.86) and the beta value is (0.13105) and when the sample size was (30) the power test is (0.96) and the beta value is (0.03229) and when the sample size 80) the test power is (0.99999) and the beta value is (0.00001) and when the sample size is (90) the test power is (1.000) and the beta value is (0.0000).

From the above, we can conclude that there is a significant effect of the sample size on the power of the Independent sample t test in the event that the data follow a uniform distribution and as the sample size increases, the power of the Independent sample t test samples increases, and the optimal size is (80) to give an optimal power.

Table 2: illustrate the independent sample t test with data that follow the logistic distribution

Power	N1	N2	Ratio	Alpha	Beta	Mean1	Mean2	S1 & s2
0.56201	10	10	1.000	0.05000	0.43799	0.0	1.0	1.0
0.88526	20	20	1.000	0.05000	0.11474	0.0	1.0	1.0
0.97600	30	30	1.000	0.05000	0.02400	0.0	1.0	1.0
0.99565	40	40	1.000	0.05000	0.00435	0.0	1.0	1.0
0.99929	50	50	1.000	0.05000	0.00071	0.0	1.0	1.0
0.99989	60	60	1.000	0.05000	0.00011	0.0	1.0	1.0
0.99998	70	70	1.000	0.05000	0.00002	0.0	1.0	1.0
1.00000	80	80	1.000	0.05000	0.00000	0.0	1.0	1.0
1.00000	90	90	1.000	0.05000	0.00000	0.0	1.0	1.0
1.00000	100	100	1.000	0.05000	0.00000	0.0	1.0	1.0

Source pass 12

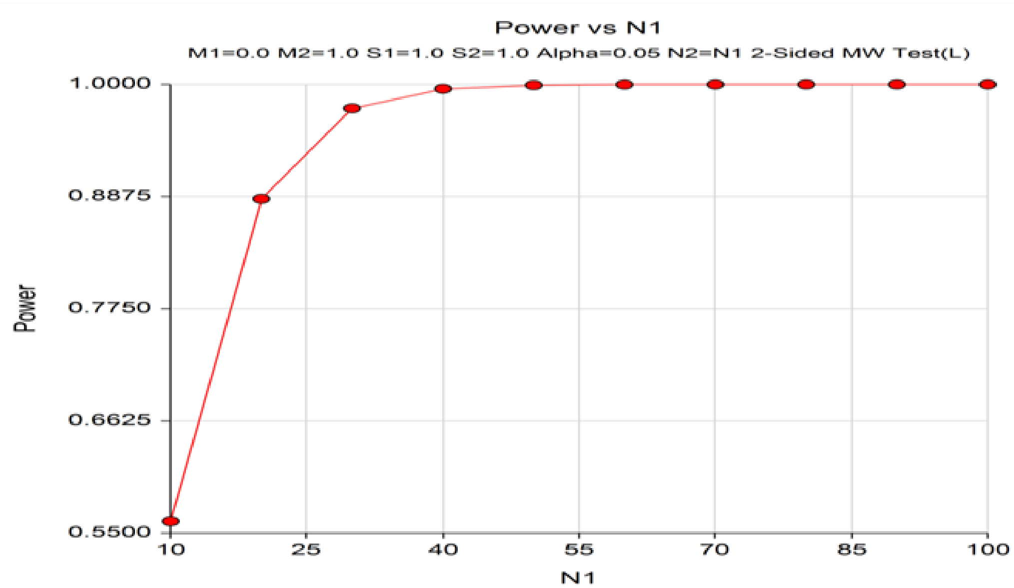

Figure 5: illustrate the relationship between the sample size and the power test

The previous results indicate that there is a significant effect of the sample size of the power of the independent sample t test eyes of the data that follow the regular distribution, where it was found that when the sample size was (10) the power test was (0.56) and the beta value is (0.43799) and when the sample size was (20) the power test is (0.88) and the beta value is (0.11474) and when the sample size was (30) the power test is (0.97) and the beta value is (0.02400) and when the sample size 70) the test power is (0.99) and the beta value is (0.00) and when the sample size is (80) the test power is (1.0) and the beta value is (0.00).

Table 3: illustrate the independent sample t test with data that follow the normal distribution

Power	N1	N2	Ratio	Alpha	Beta	Mean1	Mean2	S1 & s2
0.51336	10	10	1.000	0.05000	0.48664	0.0	1.0	1.0
0.85061	20	20	1.000	0.05000	0.14939	0.0	1.0	1.0
0.95677	30	30	1.000	0.05000	0.04323	0.0	1.0	1.0
0.99040	40	40	1.000	0.05000	0.00960	0.0	1.0	1.0
0.99772	50	50	1.000	0.05000	0.00228	0.0	1.0	1.0
0.99957	60	60	1.000	0.05000	0.00043	0.0	1.0	1.0
0.99991	70	70	1.000	0.05000	0.00009	0.0	1.0	1.0
0.99998	80	80	1.000	0.05000	0.00002	0.0	1.0	1.0
1.00000	90	90	1.000	0.05000	0.00000	0.0	1.0	1.0
1.00000	100	100	1.000	0.05000	0.00000	0.0	1.0	1.0

Source pass 12

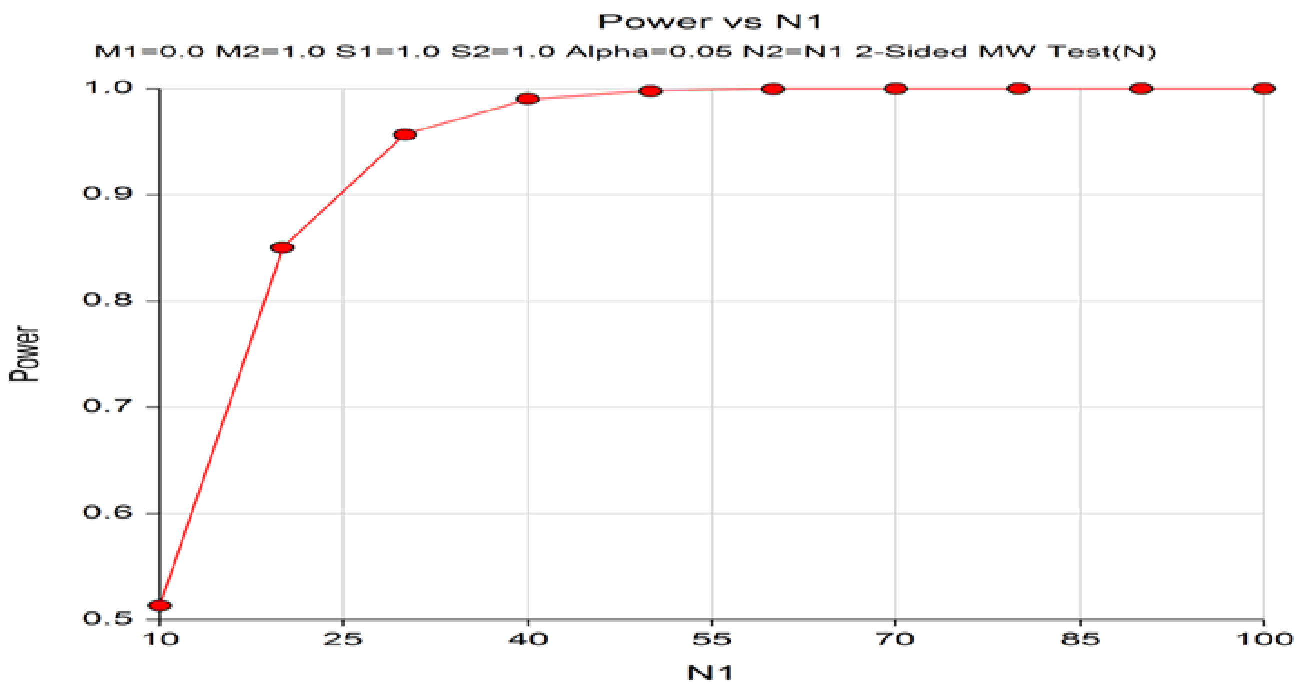

Figure 6: illustrate the relationship between the sample size and the power test

Previous results indicate that there is a significant effect of the sample size of the independent sample t test eyes of the data following the normal distribution, as it turned out that when the sample size was (10) the power test is (0.51) and the beta value is (0.48664) and when the sample size was (20) the power test is (0.85) and the beta value is (0.14939) and when the sample size was (30) the power test is (0.95) and the beta value is (0.04323) and when the sample size 80) the test power is (0.99999) and the beta value is (0.00001) and when the sample size is (90) the test power is (1.000) and the beta value is (0.000) From the above, we can conclude that there is a significant effect of the sample size on the power of the independent sample (t) test in the event that the data follows the normal distribution and with an increase in the sample size, the power of the independent sample (t) test increases and the optimal size is (80) to give an optimal power.

Table 4: illustrate the power of the ANOVA

Power	N	N	Actual Alpha	Beta	S. D. of mean	S. D. of data	S
0.092	2.0	6	0.045	0.909	0.5	0.1	0.1
(0.013)	0.079	0.104	(0.009)	0.035	0.054	0.1	0.1
0.149	3.0	9	0.049	0.852	0.5	0.1	0.1
(0.016)	0.133	0.164	(0.009)	0.040	0.058]	0.1	0.1
0.219	4.0	12	0.044	0.781	0.5	0.1	0.1
(0.018)	0.201	0.237	(0.009)	0.035	0.053]	0.1	0.1
0.292	5.0	15	0.048	0.708	0.5	0.1	0.1
(0.020)	0.272	0.312	(0.009)	0.039	0.057	0.1	0.1
0.359	6.0	18	0.054	0.642	0.5	0.1	0.1
(0.021)	0.337	0.380	(0.010)	0.044	0.063	0.1	0.1
0.401	7.0	21	0.052	0.600	0.5	0.1	0.1
(0.021)	0.379	0.422	(0.010)	0.042	0.062	0.1	0.1
0.463	8.0	24	0.043	0.537	0.5	0.1	0.1
(0.022)	0.441	0.485	(0.009)	0.034	0.052	0.1	0.1
0.504	9.0	27	0.053	0.497	0.5	0.1	0.1
(0.022)	0.482	0.525	(0.010)	0.043	0.062	0.1	0.1
0.592	10.0	30	0.046	0.409	0.5	0.1	0.1
(0.022)	0.570	0.613	(0.009)	0.036	0.055	0.1	0.1

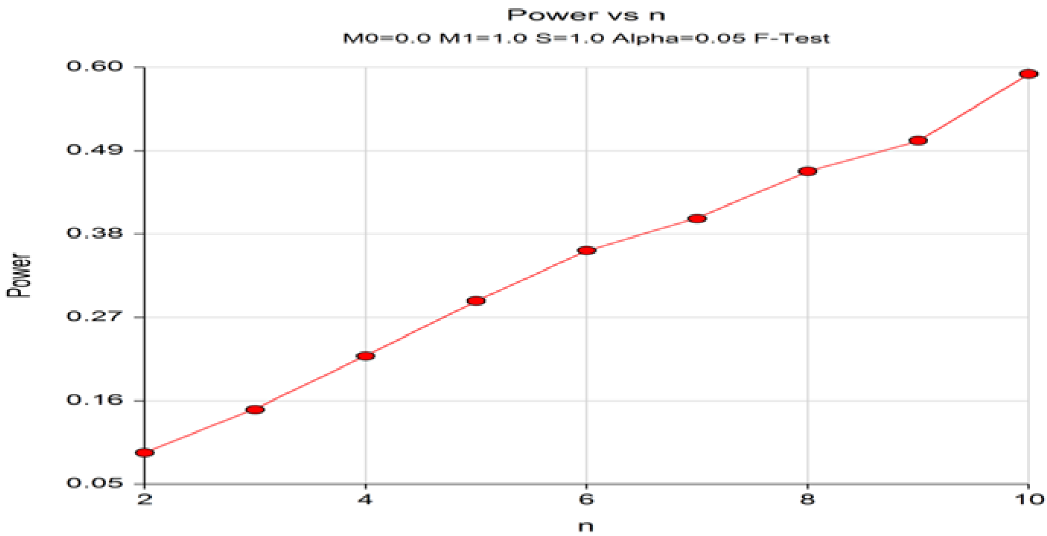

Figure 7: illustrate the relationship between the sample size and the power test

Previous results indicate that there is a significant effect of the sample size of the ANOVA contrast analysis power, as it turned out that when the sample size was (2) the test power was (0.09) and the beta value is (0.909) and when the sample size was (6) the test power is (0.35) and the beta value is (0.642) and when the sample size was (9) the test power is (0.50) and the beta value is (0.497) and when the sample size was (10) the test power is (0.59) and the beta value is (0.409). From the above, we can conclude that there is a significant effect of the sample size on the strength of the ANOVA contrast analysis test and with an increase in the sample size, the strength of the ANOVA contrast analysis test increases.

Table 5: illustrate the power of the chi-square test

Power	N	W	Chi-Square	DF	Alpha	Beta
0.27351	10	0.5000	2.5000	2	0.05000	0.72649
0.50367	20	0.5000	5.0000	2	0.05000	0.49633
0.68765	30	0.5000	7.5000	2	0.05000	0.31235
0.81542	40	0.5000	10.0000	2	0.05000	0.18458
0.89624	50	0.5000	12.5000	2	0.05000	0.10376
0.94401	60	0.5000	15.0000	2	0.05000	0.05599
0.97081	70	0.5000	17.5000	2	0.05000	0.02919
0.98521	80	0.5000	20.0000	2	0.05000	0.01479
0.99270	90	0.5000	22.5000	2	0.05000	0.00730
0.99647	100	0.5000	25.0000	2	0.05000	0.00353

Source: pass 12

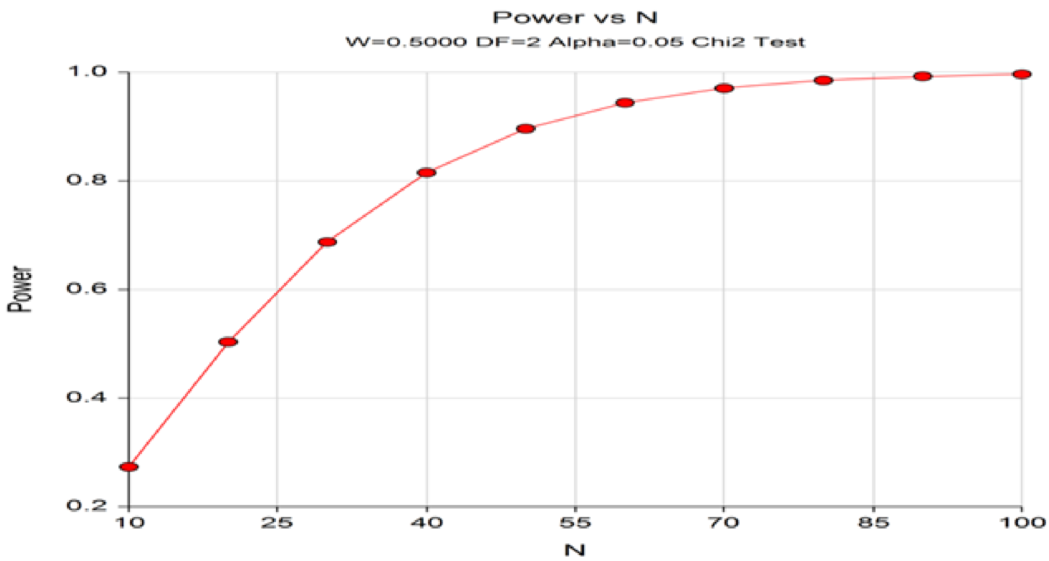

Figure 8: illustrate the relationship between the sample size and the power test

Previous results indicate that there is a significant effect of the sample size of the chi-square test, as it turned out that when the sample size was (10) the power teat was (0.27) and the beta value was (0.72649) and when the sample size was (20) the power teat is (0.50) and the beta value is (0.49633) and when the sample size was (50) the power teat is (0.89) and the beta value is (0.10376) and when the sample size was (70) the power teat is (0.97) and the beta value is (0.02919) when and when the sample size is (100) the test power is (0.99) and the beta value is (0.00353). From the above, it can be concluded that there is a significant effect of the sample size on the power of the chi-square test and with an increase in the sample size, the power of the chi-square test increases .

Table 6: shows the sample size, the optimum and the power of the test according to the distribution of the population.

Distribution	Independent sample	
	Sample size	Power test
Normal distribution	40	0.99
Logistic distribution	70	0.99
Uniform distribution	80	0.99

5.Results

1. By increasing the sample size, the power of the independent sample (t) test increases, in the case data with uniform distribution, the sample size from (20) to (80) gives optimal power and the beta value is low from (0.13) to (0.00001).
2. By increasing the sample size, the power of the independent sample (t) test increases, in the case data with logistic distribution, the sample size from (20) to (70) gives optimal power and the beta value is low from (0.11) to (0.00002).
3. By increasing the sample size, the power of the independent sample (t) test increases, in the case data with normal distribution, the sample size from (20) to (80) gives optimal power and the beta value is low from (0.48) to (0.00002).
4. By increasing the sample size, the power of the (f) test increases, the sample size from (50) to (70) gives optimal power and the beta value is low from (0.0001) to (0.00002).
5. By increasing the sample size, the power of the chi-square test increases, the sample size from (100)) gives optimal power and the beta value is low from (0.003).
6. In the case of data with normal distribution, the sample size that gives the optimal power is less than the data with the logistic and uniform distribution.

6.Recommendations

Based on the results, the paper recommends the following:

1. Using the appropriate sample size because it leads to tangible results and thus increases the power of statistical testing
2. Determining the distribution of the population or taking into account the naturalness of the study because this reduces the value of the skewness coefficient and thus increases the power of the statistical test.
3. The data that with normal distribution do not need to increase the sample size compared to the logistic and uniform distribution.
4. The power provides us with useful information about the test holistically.
5. If the results of the study are not statistically significant, it is assumed that the researcher should interpret that result and not just point out that it is not statistically significant.
6. Do not overdo it by increasing the sample size the limit that gives an appropriate test power.
7. The researcher must perform the result of the statistical power test in order to make sure that the statistical test used does not suffer from a decrease in power and that the failure to obtain the statistical significance was not a decrease in the power of the test used.
8. The power of statistical testing should be studied more broadly and comprehensively for its importance in statistical studies and conducted research on it more comprehensively.

7.Conclusions

Through the results of the analysis, it was found that the power of the statistical test is affected by an increase in the sample size and the distribution of the population, and that the data with normal distribution need a smaller sample size (40) to give the optimal power (0.99), the data with logistic distribution need a sample size (70) to give an optimal power (0.99) and the uniform distribution needs a sample size (80) to give an optimal power (0.99), as for the exceeds the sample size, as for the chi-Square test and the ANOVA-test, their power increases with an increase in the sample size as well.

References

[1] Wichal W. Trosset (2006) - *statistical inference*-collage of William & Mary

[2] Abramowitz, M and stegun. L. A (Eds) (1972). *Handbook of mathematical tables 9th printing* - new yourk. Dover, pp.836 - 937.

[3] Cochran, W. -*sampling techniques - 2nd* (1963). Ed. New yourk, john Wiley & sons, Inc.

[4] American psychological Association (2001). *Publication Manual of the American Psychological Association (5th ed.)* Washington, DC: Author.

[5] Aaron, A., & Aaron, E. N. (1997), Statistics for the behavioral and social sciences: A brief course. NJ: Prentice Hall.

[6] Maxwell, S and Delaney, H (2004). *Designing experiments and analyzing data: A model comparison perspective (2nd ed)*-Mahwah, NJ: Lawrence Erlbaum.

[7] Lieberman MD, Cunningham WA. Type I and Type II error concerns in fMRI research: re-balancing the scale. Soc Cogn Affect Neurosci.2009 Dec; 4 (4): 423-8. [PMC free article] [PubMed]

[8] Singh G. A shift from significance test to hypothesis test through power analysis in medical research. J Postgrad Med.2006 Apr-Jun; 52 (2): 148-50. [PubMed]
[9] Bezeau S, Graves R. Statistical power and effect sizes of clinical neuropsychology research. J Clin Exp Neuropsychol.2001 Jun; 23 (3): 399-406. [PubMed]
[10] Keen HI, Pile K, Hill CL. The prevalence of underpowered randomized clinical trials in rheumatology. J Rheumatol.2005 Nov; 32 (11): 2083-8. [PubMed]
[11] Gaskin CJ, Happell B. Power, effects, confidence, and significance: an investigation of statistical practices in nursing research. Int J Nurs Stud.2014 May; 51 (5): 795-806. [PubMed]
[12] Algina J, Olejnik S. Conducting power analyses for ANOVA and ANCOVA in between-subjects designs. Eval Health Prof.2003 Sep; 26 (3): 288-314. [PubMed]
[13] Ueki C, Sakaguchi G. Importance of Awareness of Type II Error. Ann Thorac Surg.2018 Jan; 105 (1): 333. [PubMed]
[14] Leon AC. Multiplicity-adjusted sample size requirements: a strategy to maintain statistical power with Bonferroni adjustments. J Clin Psychiatry.2004 Nov; 65 (11): 1511-4. [PubMed]