Available online at http://www.ajol.info

Int. J. Biol. Chem. Sci. 2(3): 373-388, 2008

International Journal of Biological and Chemical Sciences

ISSN 1991-8631

Review

http://indexmedicus.afro.who.int

Basic assumptions in statistical analyses of data in biomedical sciences

U.M. OKEH [1] and A.C. UGWU [2*]

[1] *Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Nigeria.*
[2] *Department of Medical Radiography and Radiological Sciences, Nnamdi Azikiwe University, Nnewi Campus, Anambra State, Nigeria.*
[*] *Corresponding author, E-mail: tonybullng@yahoo.ca, Tel: +2348076241297*

ABSTRACT

Every statistical procedure carries with it certain assumptions that must be at least approximately true before the procedure can produce reliable and accurate results. Researchers often apply a statistical procedure to their data without checking on the validity of the assumptions of the procedure. If one or more of the assumptions of a given statistical procedure are violated, then misleading results will be produced by the procedure. It is important that those who analyze data be fully aware of the details of the statistical procedure they are using, including its companion assumptions. If one or more assumptions are violated, an alternative procedure must be used to obtain valid results. This article aims at highlighting some basic assumptions in statistical analyses of data in biomedical sciences.

Keywords: samples, independence, non-parametric, parametric, statistical analyses.

INTRODUCTION

Assumptions refer to basic principles that are accepted on faith, or assumed to be true, without proof or verification. It is frequent and common experience that a researcher will apply a statistical method to a set of data without thoroughly checking that the assumptions of the method are valid. This may be especially true in diagnostic biomedical experiment because of the nature of the data involved. Measurements such as fetal abdominal circumference, fetal blood flow velocity, femur length, age of the fetus, biparietal diameter of the fetal head, ascending aortic cross-sectional area, etc, often follow a normal distribution as can be seen by the bell-shape of the histograms. For example, if a researcher wants to compare two means, he/she might apply the two sample t –test to the corresponding samples without checking the assumptions that go along with the test. Although the two sample t-test is designed to compare the means of two continuous populations, if the assumptions of the test are not true, then the results of the statistical analysis may be misleading. It is evident that when independent samples from two populations deviate drastically from normality (normality as in assuming that equal standard deviations exists and the sampled sized n_1 and n_2 are both less than 30) as a result of the fact that the samples (data) are ranked or ordinal in nature, then the use of t-test in testing for hypothesis about the difference between two population means may lead to false result.

Example 1: Presented in table 1 below are the lengths of time in seconds it took random samples of male and female students to perform a certain task. Should we conclude on the basis of these data that male students perform the task in shorter time than female students? Use 5% as level of significance to test the null hypothesis (Oyeka, 1996).

Table 1: Length of Time in Seconds Taken by Random samples of male and female students to perform a certain Task.

Male students	Female students
17	45
13	30
12	48
35	10
15	35
44	13
25	35
56	89
10	30
	21
	43
	85
	40

Source: Oyeka, 1996.

Solution

Test of hypothesis

H_0 = The median length of time taken by male students to perform the task is less than or equal to the median length of time taken by the female students.

H_1 = The median length of time taken by the male students is greater than that of female students.

When the sample sizes n_1 and n_2 are both 8 or more, the statistic U is approximately normally distribution with mean $U - \frac{n_1 n_2}{2}$

and standard deviation

$$\delta_\mu = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Hence, the corresponding Z-score for the Mann-Whitney U-statistic is calculated as

$$Z = \frac{U - U_u}{\delta_u} = U - \frac{n_1 n_2}{2} \div \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Z-test is preferred for use here because the t-test is only used for paired comparison when the sample size is small. Meanwhile the data presented in table 1 are not in pairs or not matched. Therefore, the assumption for the use of t-test is violated. In other words, applying the t-test on the data, will lead to result that is contrary as can be seen below.

t-test of hypothesis

$$H_0 : \mu_1 - \mu_2 \geq d_0$$
$$H_1 : \mu_1 - \mu_2 < d_0$$

Here $\mu_1 - \mu_2$ indicates difference between two population means. d_0 =some specified value that may be zero. From the example 1, represents male population mean while is female population mean. In particular when do is zero, this is equivalent to testing the null hypothesis $\mu_1 = \mu_2$ against the alternative hypothesis $\mu_1 \neq \mu_2$

To test this hypothesis, we select two independent random samples (male and female students), one of size n_1 from one population, and the other of size n_2 from the second population. We then compute their respective sample means; and $\bar{X}_1$ and $\bar{X}_2$ find the difference $d_0 = \bar{X}_1 - \bar{X}_2$. In using t-test, the population standard deviations δ_1 and δ_2 are unknown and so are computed from samples while the sample sized n_1 and n_2 are small (both are less than 30). If it is assumed that the two samples came from normally distributed populations with equal standard deviations, s_p. The t-test statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{s_p}^2}{n_1} + \frac{{s_p}^2}{n_2}}} = \frac{\bar{X}_1 - \bar{X}_2 - d_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where s_p is the pooled estimate of the common population standard deviation. If H_0 is true, then t-distribution has $n_1 + n_2 - 2$ degrees of freedom, where ${s_p}^2 = \frac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}$. For the purpose of analysis, recall from table 1 above the lengths of time in seconds taken by male students represented as X_1 and female students as X_2.

The calculation of the standard deviations for male and female students is shown in table 2.

Table 2: Calculation of the standard deviations for male and female students.

X_1	X_2	$\left(X_1-\bar{X}_1\right)$	$\left(X_2-\bar{X}_2\right)$	$\left(X_1-\bar{X}_1\right)^2$	$\left(X_2-\bar{X}_2\right)^2$
17	45	-8.22	4.69	67.57	21.99
13	30	-12.22	-10.31	149.33	106.29
12	48	-13.22	7.69	174.77	59.13
35	10	9.78	-30.31	95.65	918.69
15	35	-10.22	-5.31	104.45	28.19
44	13	18.78	-27.31	352.69	745.84
25	35	-0.22	-5.31	0.0484	28.19
56	89	30.78	48.69	947.41	2370.71
10	30	-15.22	-10.31	231.65	106.29
	21		-19-31		372.87
	43		2.9		8.41
	85		44.9		2016.01
	40		-031		0.0961
Totals=227	**524**			**2123.58**	**6782.706**

Source: Oyeka, 1996.

where $\bar{X}_1 = \frac{227}{9} = 25.22$, $n_1 = 9$,

$$s_1^2 = \frac{\sum\left(X_1-\bar{X}_1\right)^2}{n_1} = \frac{2123.58}{9} = 235.95.$$

Also, $\bar{X}_2 = \frac{524}{13} = 40.31$,

$$s_2^2 = \frac{\sum\left(X_2-\bar{X}_2\right)^2}{n_2} = \frac{6782.706}{13} = 521.75.$$

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(9-1)235.95+(13-1)521.75}{9+13-2} = \frac{8148.6}{20}$$

$$s_p = 20.1849.$$

But $$t = \frac{\bar{X}_1-\bar{X}_2-d_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{25.22-40.31}{20.1849\sqrt{\frac{1}{9}+\frac{1}{13}}} = -1.7241$$

$|t| = 1.7241$. Since the hypothesis indicates a one-sided test, the null hypothesis is rejected at the given 5% level of significance when $|t| \geq t_{1-\alpha;\ n_1+n_2-2}$ otherwise, H_0 is accepted but $t_{1-\alpha;\ n_1+n_2-2}$ is $t_{1-0.05;9+13-2\ =\ 1.7247}$. Hence H_0 is, therefore accepted at the 5% level of significance since $|t| = 1.7241$ is less than 1.7247. We conclude that the median length of time taken by male students is at most equal to that of the female students. This result almost would have deviated (since the calculated absolute t that is 1.7241 is a little less than the tabulated t which is 1.7247) from the result obtained when Mann-Whitney U-statistic was applied on the data in which case the conclusion would have been on the contrary. Medical researchers are advised to take caution in using t-test as it may lead to invalid results when wrongly applied. The test above is called small sample test.

In order to ensure that the right test is applied, the data of table 1 requires analysis by ranking using a non-parametric test called (this will be explained in the next subheading) Mann-Whitney U-test because of the failure of some normality assumptions particularly that standard deviation/variances are not equal. To test the null hypothesis using the Mann–Whitney U-test we first combine the two samples and rank the observations from the smallest to the largest, assigning the rank 1 to smallest value, the rank 2 to the next smallest value, and so on. The ranks assigned to the observations in the two samples after their combined ranking are shown below (Table 3).

Table 3: Ranks of the observations in the two samples.

Rank of lengths of time for male students	Ranks of length of time for female students
7	18
4.5	10.5
3	19
13	1.5
6	13
17	1.5
9	13
20	22
1.5	10.5
	8
	16
	21
	15
Sum $(R_1) = 81$	$R_2 = 172$
$n_1 = 9$	$n_2 = 13$

Source: Oyeka, 1996.

Since the sum of the ranks assigned to the observations in the male sample is calculated as $R_1 = 81$, we evaluate the Mann-Whitney U-statistic as

$$U = n_1 n_2 + n_1 \frac{(n_1+1)}{2} - R_1$$

$$= (9)(13) + \frac{(9)(10)}{2} - 81 = 81$$

The mean of the U-statistic is

$U_\mu = \frac{n_1 n_2}{2} \frac{(n_1+1)}{2} = 58.5$ and its standard deviation is

$$\delta_\mu = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} = \sqrt{\frac{(9)(13)(9+13+1)}{12}} = 14.9$$

Hence, the normal z-score corresponding to U = 81, is calculated from

$$z = \frac{U - U_u}{\delta_u} = \frac{81 - 58.5}{14.97} = 1.50$$

Because the alternative hypothesis implies a one-sided test, the null hypothesis cannot be rejected since 1.50 is less than 1.64, the critical value for the standard normal distribution corresponding to a 0.05 significance level. We therefore, conclude that the median lengths of time taken by male students to perform the task is at most equal to the median time have for female students. This result may have been on the contrary if t-test was used without checking to know if the data does not violate some of the normality assumptions. Some researchers always apply 3 sigma rule to check for normality of data. In this situation, for a normally distributed data which would benefit from a parametric test, the range of data would be approximately equal to mean $\pm$ 3SD (Ugwu , 2007).

Every statistical test or estimation procedure has some basic assumptions that go along with it. It is important, when conducting statistical procedures, to check the validity of these assumptions. If one or more of the assumptions of a given statistical procedure are not true, then an alternative procedure or approach must be taken to obtain valid results. Basic assumptions must be attached to each statistical procedure because the assumptions enable researchers to formulate a coherent and effective method of analyzing the data; otherwise, a practical method of analyzing the data would not be achievable.

PARAMETRIC AND NON-PARAMETRIC TEST

There is a set of statistical methods called non-parametric which do not require assumptions of any form of probability distribution from which measurements come. They are otherwise called distribution - free

method because inferences are not about population parameters, their knowledge or estimation is not required and no assumptions are made about them. However as noted above, parametric tests often require certain assumptions about the populations from which the samples are drawn. For example, the use of the t distribution to test the statistical significance of the difference between two population means, when the samples are small, requires that the two samples be independently drawn from populations with normal distribution and equal variances. Similarly, the use of the F distribution to compare the means of several populations requires that the samples be independently drawn from normally distributed populations with equal variances. Although these tests are sufficiently robust (insensitive to departures from the assumptions that underlie them) and can still be used even when some of the assumptions are not satisfied, the assumptions are nevertheless restrictive. On the other hand, non-parametric tests do not require these restrictive assumptions and may therefore, be used in situations where we would be justified in employing parametric test as well as in instances where the parametric tests are not applicable because the necessary assumptions may not be satisfied by the data. However, if both the parametric and the non-parametric test are equally applicable, then the parametric test will be more powerful as in having higher statistical power (Schefler,1984) than the non-parametric test in that the later has a greater probability of accepting a false hypothesis (committing a Type II error). Non-parametric tests are also found most suitable for analyzing ranked or ordinal data which often deviate, drastically from normality and therefore, may not be suitable for parametric tests (Oyeka, 1996). Consider in a one-way analysis of variance where the parametric F-test may be used to test the null hypothesis that several group means are equal. In this case, F-test requires that the sampled populations be normally distributed with equal variances which are rather restrictive assumptions in the sense that when they are not satisfied may invalidate the results of the F-test. In a situation of this nature, an alternative non-parametric procedure for one-factor analysis that may be used when the assumptions of the F- test are not satisfied and also when the assumptions are satisfied is the Kruskal-Wallis one-way analysis of variance by ranks. The primary difference in procedure between this test and the usual F-test for one-way analysis is that the Kruskal-Wallis test is based on a test statistic computed from ranks determined for pooled sample observations rather than from the observations themselves. It is an extended version of the Mann-Whitney U-test used when there are only k=2 groups involved in the one-way analysis of variance.

Example 2: The data below (Table 4) are weights of food (in grams) consumed per week by four species of fish (Oyeka, 1996). Test, at the 5 percent significance level, the null hypothesis that food consumption is the same for the four fish species.

The null hypothesis of no difference between quantities of food consumed by the four fish species will be tested using the Kruskal-Wallis one-way analysis of variance test procedure (Table 5). The above observations are first combined into one sample and ranked from the largest to the smallest, assigning the largest values a rank of 1 and the smallest value a rank of 22, since there are altogether total of n=5+6+5+6=22 observations. Tied observations are assigned their mean ranks. Table 3 shows the ranks assigned to the observations in each of the four fish species (Oyeka, 1996).

Table 4: Food consumptions (in gram) by species.

Species 1	Species 2	Species 3	Species 4
340	350	290	285
345	360	300	296
336	335	275	258
358	360	240	230
	385	305	295
	287		348

Source: Oyeka, 1996.

Table 5: Ranks of fish food consumption data. An illustration of the Kruskal-Wallis One-Way Analysis of Variance test.

Species 1	Species 2	Species 3	Species 4
9	5.5	16	18
8	2.5	13	14
10	11	19	20
4	2.5	21	22
5.5	1	12	15
	17		7
Total R_i , R_1 =36.5	R_2 =39.5	R_3 = 81	R_4 =96

Source: Oyeka, 1996.

The Kruskal-Wallis test statistic, H, is calculated from Tables 3 and it is given by

$$H = \frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1)$$

$$= \frac{12}{22(23)}\frac{(36.5)^2}{5} + \frac{(39.5)^2}{6} + \frac{(81)^2}{5} + \frac{(96)^2}{6} - 3(23)$$

$$= (0.024)(3374.69) - 69$$

$$= 80.99 - 69 = 11.99$$

Since there are at least 5 observations in each of the K =4 groups, H may be assumed to be approximately chi-square distributed with 4-1=3 degrees of freedom under the null hypothesis. Therefore, given 5 percent significance level, the critical chi-square value is $x^2_{0.95,3} = 7.82$. Since H= 11.99 >7.82, we reject H_o at the 5% significance level and conclude that food consumption by the four fish species are different. From the above example, it is seen that Kruskal-Wallis test takes account of both the direction and magnitude of observation thus, utilizing more information about the data of interest than another non-parametric test called median test which can be extended to situations in which there are more than two groups and can be used in place of Kruskal-Wallis test. Median test is less powerful than Kruskal-Wallis test because it considers only the direction of observations (Oyeka, 1996).

It is in general recommended that non-parametric tests be employed in such situations when the assumptions for parametric tests are seriously violated, or when the nature of the data makes it unadvisable to apply parametric tests. In the following sections, the basic assumptions of the standard parametric statistical procedures are discussed, a description of how to check those assumptions is given, and recommendations of what to do if the assumptions are not met are provided. A great deal of the Mathematical theory of statistics assumes the existence of populations and further assumes that for any given variable 'true' values of certain measures used to characterize a population exists. These 'true' values are called parameters. It is note worthy that one seldom has data for entire populations, and nearly always has to rely on samples to draw conclusions about populations. Hence, parameters are rarely ever calculated or exactly known. However, by drawing representative samples of populations, parameters can be estimated very well. An estimate of a population parameter is called a sample statistic, or sample. A statistic is calculated on the basis of only those data that are actually observed. It is conventional in statistics to represent population parameters by Greek letters (e.g. population mean represented as μ) and sample statistics by Roman letters (e.g. sample mean as $\bar{X}$). Parameters are constant for a given population, but statistics vary from sample to sample for repeated samples drawn from the same population. Thus, there could be several sample statistics corresponding to a given population parameter. But since we use sample statistics as estimates of population parameters and to draw conclusions about populations, it is important that we select 'best' estimates in the sense that they possess certain desirable properties (Oyeka, 1996).

Example 3: Find the arithmetic mean of the marks obtained by 10 students of a class in mathematics in a certain examination. The marks obtained are: 25, 30, 21, 55, 47, 10, 15, 17, 45 and 35 (Oyeka, 1996).

Solution:

Let $\bar{x}$ be the average mark obtained. This is otherwise called the sample statistic or the estimate of the population parameter. Hence sum of all the observations $\sum x = 25+30+30+21+55+47+1015+17+45+35 = 300$

Number of students n=10

$$\bar{x} = \frac{\sum x}{n} = \frac{300}{10} = 30$$

REPRESENTATIVE SAMPLE

One of the most important assumptions for any statistical procedure is that the sample be representative of the population being studied. In this way, the characteristics of the sample will be typical of the characteristics possessed by the study population. When the sample is not representative of the population under study, conclusions derived or drawn from the sample will not be valid for the study population. An example of this kind of error can be found in an article by Pearl (1929) where a negative correlation between presence of cancer and presence of tuberculosis in a sample of autopsy cases was assumed to be valid for live patients. So Pearl conceived a study to treat patients with terminal cancer with tuberculin (the protein of the tubercle bacillus) thinking the cancer would be arrested. The experiment failed; Pearl did not realize that an association found in autopsy cases should not be extrapolated to live patients unless all deaths are equally likely to be autopsied.

In principle, one can ensure a representative sample of the study population by taking a simple random sample of the population, namely, one for which every individual (or, more generally, experimental unit) has the same chance of being selected in the sample. The best way to ensure a simple random sample is to use a random number generator or a random number table (Nwabuokei, 1989; McClave et al., 1997). In many instances, a true random sample is not feasible, so researchers obtain a "convenience" sample.For instance, a researcher in Abakaliki, Ebonyi State may study the average fetal blood flow for Nigerian mothers by obtaining a sample of mothers in Abakaliki area hospitals. To the extent that hospitalized Abakaliki area mothers are not representative of all Nigerian mothers**,** because not only that the sample size is very small and inadequate, it will give biased information about the entire Nigerian mothers, and the conclusions drawn from the sample are invalid for Nigerian mothers, since only one state can never represent the entire Nigeria in terms of opinion.

However, inadequate randomization is often times experienced in the Lottery method of ensuring randomness of sample selection from the population of interest (Fienberg, 1970). These biased results and personal prejudice are often noticed if the slips are not of identical size, shape and colour. Lottery method is a very popular method of taking a random sample. Under this method, all items of the universe are numbered or named on separate slips of paper of identical size and shape. These slips are then made of the number of slips required to constitute the desired sample size. The selection of items thus depends entirely on chance. The method would be quite clear with the help of an example. If we want to take a sample of 10 persons out of a population of 100, the procedure is to write the names of the persons on separate slips of paper, fold these slips, mix them thoroughly and then make a blindfold selection of 10 slips (Gupta, 2001).

Generally, whether a collection of observations is a population or a sample depends on the study objectives. For example, a farmers' association may wish to compare the costs of pest control to the costs of other farm management and services. The purpose of the study determines whether the populations would comprise costs for an entire country, for a state or simply, for a small administrative area or community. If the population of costs covers all farms treated during a particular month in a calendar year, then the costs for all farms treated on a single day would constitute a sample. The costs for a single farm treated during the entire month would also constitute a sample, of the same population. If the farmers association is interested only in the costs for farms treated on a particular date, then those data would constitute populations rather than samples. Similarly, if the costs at different farms are to be compared, then the costs at any one farm would be a separate population.

INDEPENDENCE

For many statistical procedures, the observations are assumed to be independent. That is, one cannot predict the outcome for one experimental unit given the outcome of any other experimental unit. Suppose, for example, that you wish to study the head circumference of a random sample of 1- year - old babies. If a set of twins is included in the sample, then the head circumference measurements for those two babies would be correlated, violating the independence of the observations. As another example, consider an article by Himes (1991) where the mean weights of three groups of men were compared using a one-way analysis of variance. The three groups were formed in such a way that many of the men were classified in two or all three of the groups, so their weights appeared in more than one group. Consequently, observations in one group were correlated with observations in another group. Because the one-way analysis of variance requires independent samples (i.e. observations in one group must be independent of observations in any other group) the results of the analysis are invalid. In biological experiments and field studies, correlated t-test or paired t-test are applicable. They can also be applied to paired data or two samples obtained from the same population at two different times and conditions (i.e. independent samples). Each individual gives a pair of observation. The paired or correlated t-test can otherwise be called t-test for two sample means. The t-test for correlated data provides the significance of the difference between the two correlated means (Rastogi, 2007).

Example 4: The percentage of water content in two varieties of watermelons was measured and the results obtained are presented in table 6. Find out whether there is significant difference in water content of two varieties.

Solution:

Identification of problem: To find out difference in the content of two varieties of watermelons.

Data Given:

$$n_1 = 12 \qquad n_2 = 15$$
$$s_2 = 15 \qquad s_2 = 19$$
$$\bar{x}_1 = 92 \qquad \bar{x}_2 = 84$$

Step 1: Calculation of standard deviation

$$s = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(12-1)(15)^2+(15-1)(19)^2}{12+15-2} = \frac{2475+5054}{25} = \frac{7525}{25} = 301.16$$

Step 2: calculation of t-value.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_1^2}{n_2}}} = \frac{92-84}{\sqrt{\frac{225}{12}+\frac{361}{15}}} = \frac{8}{\sqrt{15.75+2406}} = 1.22$$

Therefore, $t = 1.22$

step 3: critical value:

(1) Hypothetical value of t from distribution table =1.708

(2) Calculated value of t from observations=1.22

Inference: Observation of estimated t value 1.22 is less than hypothetical value of t which is 1.708. Therefore, it can be concluded that there is no significant difference in water content in two varieties of watermelons. Hence, null hypothesis is applicable or true for this case.

NORMALITY

Many statistical procedures require that the outcome variable be normally distributed. Theoretically this means that a histogram of the outcome variable for the entire study population must be bell shaped. A practical way of checking this assumption is to study the histogram of the sample. If it is not substantially non normal, then the parametric procedure will provide an adequate analysis. Of course, one may ask, "What constitutes a substantial lack of normality?" This is a subjective judgment to some extent: two statisticians may give two different

Table 6: The percentage of water content in two varieties of watermelons

Variety	Average water content	No of watermelons (n)	Standard Deviation (s)
A	92%	12	15%
B	84%	15	19%

Source: Rastogi, 2007

answers for the same data set. There are tests of normality called goodness of fit test (Oyeka, 1996), which can be used to help determine whether the underlying population is approximately normal, but experience at dealing with such questions or knowledge about the underlying population distribution is the best way to resolve such questions. One good example of test of normality using chi-square test is the goodness of fit test. Goodness of fit test indicates the closeness of observed frequency with that of the expected frequency. If the curves of these two distributions do not coincide or appear to diverge much, it is said that the fit is poor, if two curves do not diverge much the fit is less poor. Thus it helps to answer whether something (physical or chemical factors) did or did not have an effect. If observed and expected frequencies are in complete agreement with each other then the chi-square value will be zero. But it rarely happens in biological experiments. There is always some degree of deviation.

Example 5: In a chemical treatment, the patients were tested to see the effect of a potential hypertensive drug. The 50 patients were assigned to receive the active drug and other 50 the placebo at random. Their response to treatment was categorized as favorable or unfavorable (Rastogi, 2007). The data is given in table 7.

Test the hypothesis that drug has a significant effect. Use $\alpha = 0.05$

Solution:

Step 1: Problem identification: The hypertensive drug has a significant effect or not.

Step 2: Data: Given in the table. The attributes are arranged in two-way table or contingency table, i.e. two rows and two columns (Table 8).

Step 3: Hypotheses

1. Null hypotheses (H_0) stands for that drug does not have significant effect.
2. Alternative hypothesis (H_A) proposes that the effect of dug is significant

Step 4: level of significant =0.05 and Degree of freedom = (2-1) x (2-1) =1

Step 5: Calculation of expected frequency for each by using following formula: Expected frequency $(E) = \frac{\text{Row total x Column total}}{\text{Grand total}}$

Table 7: Result of the effect of hypertensive drug (observed frequency).

Treatment	Response		Total
	Unfavorable	Favorable	
Placebo	41	9	50
Drug	16	34	50
Total	57	43	100

Source: Rastogi, 2007.

Table 8: 2 X 2 contingency table showing expected frequencies.

Treatment	Effect of drug or response		Row total
	Unfavorable	Favorable	
Placebo	$57 \text{ X } 50/100$ = 28.5	$43 \text{ X } 50/100$ = 21.5	50
Drug	$57 \text{ X } 50/100$ = 28.5	$43 \text{ X } 50/100$ = 21.5	50
Column total	57.0	43.0	100

Source:Rastogi,2007.

Step 6 calculation of the difference between the observed and expected (0-E) values:

Unfavourable	Favourable
41- 28.5 =12.5	9- 21. 5 =12.5
16 –28.5 = 12. 5	34- 21.5 =12.5

Source:Rastogi, 2007.

Step 7 calculation of χ^2 value

Groups	**|0-E| -0.5**	$\left[\lvert O-E\rvert-0.5\right]^2$	$\frac{\left[\lvert O-E\rvert-0.5\right]^2}{E}$
1	12.5-0.5 =12	$(12)^2=144$	$\frac{144}{28.5}=5.05$
2	12.5-0.5 =12	$(12)^2=144$	$\frac{144}{28.5}=5.05$
3	12.5-0.5 =12	$(12)^2=144$	$\frac{144}{21.5}=6.70$
4	12.5-0.5 =12	$(12)^2=144$	$\frac{144}{21.5}=6.70$

Source:Rastogi,2007.

$$\chi^2=\sum\frac{\left[\lvert O-E\rvert-0.5\right]^2}{E}=5.05+5.05+6.70+6.70=23.50$$

$$\therefore\chi^2=23.50$$

Here the first two values represents unfavourable response while the last two represents favourable response .The general formula $\chi^2=\sum\frac{(O-E)^2}{E}$ was not used because the degree of freedom for the above problem is 1 since it is a 2x2 contingency table. We actually approximated the theoretical chi-square distribution (which is continuous) because we calculated the chi-square statistic, χ^2, from the observed frequencies which is discrete in nature (Oyeka,1996).This method is called Yates correction for continuity.

Step 8: Degree of freedom = (No of rows-1)(No of columns-1) i.e.

d.f = (2-1)(2-1)=1

Expected value of χ^2 from table at 0.05 level =3.84.

Value of χ^2 from calculations =23.50

Inference: The calculated or observed value of χ^2 is 23.50. It is much higher than the critical value of χ^2. Therefore, the null hypothesis is rejected. The conclusion is that the hyper drug has significant effect.

MEASURES OF DIVERGENCE FROM NORMALITY

Most statistical procedures involve the deviation of outcome variable from the normal distribution curve. These divergences can be studied by skewness and kurtosis. When the frequency distributions of the outcome variable, is asymmetrical, the distribution is known as skewed. Data may be skewed to the left or right. In normal distribution curve, skewness is zero. In skewed curve; mean, median and mode do not fall in the middle of the normal distribution curve and the skewness is called negative or positive respectively. Several instances are available in biomedical sciences where frequency distribution is skewed. Take the case of population density, which increases exponentially but not in time (Rastogi, 2007). In the same manner, kurtosis is the relative

flatness or peakedness of the frequency curve. According to kurtosis, the frequency curves can be platykurtic curve (when the frequency distribution curve is flatter than the normal bell shaped curve), Leptokurtic curve (when the frequency distribution curve is more peaked than the normal bell shaped curve) and Mesokurtic curve (i.e. the normal bell-shaped distribution curve).

POWER

In many statistical procedures, power is seen as the ability of the research design to defect relationships among variables. Precision contributes to the power of a design. Power is also increased when a large sample is used. One other aspect of a powerful design concerns the construction or definition of the independent variable. For both statistical and theoretical reasons, results are clearer and more conclusive when the differences between groups that are being compared are large. The aim here is to maximize group differences on the dependent variables by maximizing differences on the independent variable. In other words, the results are likely to be more clear-cut if the groups are as different as possible. This advice is more easily followed in experimental than in non-experimental processes. In experiments, the investigator can devise interventions that are distinct and as strong as time, money, ethics, and practicality permit. However, even in non experimental processes, there are frequently opportunities to operationalize the independent variables in such a way that power to detect differences is enhanced (Denise and Bernadette, 1995).

POWER ANALYSIS

Many published and unpublished studies in biomedical sciences result in non significant findings, that is, one or more of the researcher's hypotheses are not supported. Although standard statistical texts pay considerable attention to the problem of Type I errors (wrongly rejecting a true null hypothesis), little attention has been paid to Type II errors (wrongly accepting a false null hypothesis). Power analysis represents a method for reducing the risk of Type II errors and for estimating their occurrence. The statistician called the probability of committing a type I error as level of significance or alpha (α) while the probability of a type II error is beta (β), the complement of (I-β) is the probability of obtaining a significant result and is referred to as the power of a statistical test (Eze et al., 2005; Denise and Bernadette, 1995).

Example 6: Suppose we fix the level of significance (α) of a test at 5% and the mean of the random variable x follows normal distribution with mean, 80 and variance, 4, otherwise written as $\bar{x} \square N(80,4)$. We can obtain the critical value of the mean c at this level of α, the type Π error (β) and the power of the test $(1-\beta)$. Note that C which is called the critical region of the test is the subset of the sample space which in accordance with a prescribed test, leads to the rejection of the null hypothesis (H_0) under consideration (Eze et al., 2005).

Solution: $p(\bar{x} > c / H_o) = \alpha$ =p(Rejecting H_o/H_o is true)= $p\left(Z > \frac{c-80}{2}\right) = 0.05$,

Thus $\varnothing$(c-80)/2=0.05 Where $\varnothing$ is the value of 5% from table and c-80/2 =1.65

$\Rightarrow$ C=80+2 (1.65)=83.3

Hence we reject H_o when

$$\bar{x} > 83.3 \ \ or \ \ \text{when}\left(\frac{\bar{x}-U}{\delta}\right)\sqrt{n} > 1.65$$

The performance of the test in terms of its ability to reach a correct decision is gauged by an evaluation of β - the probability of type $\prod$ error.

β = (accepting H_o/ H_o is false) = p($\bar{x}$ is in the acceptance region (A) when it is actually in (R))

$$= p(\bar{x} \in A / \bar{x} \in R)$$
$$= p(\bar{x} \le c / H_1 : U = 85)$$
$$= p(\bar{x} \le 83.3 / U = 85)$$
$$= p\left(Z \le \frac{83.3-85}{2}\right)$$
$$\beta = p(Z \le -0.85) = 0.1977$$

That is, the test will fail in about 19.77% to correctly discriminate between H_0 and H_1. In other words, the power of the test to yield a correct decision is given by 1-β =1-0.1977=0.8023.

In performing a power analysis, there are four components, at least three of which must be known to or estimated by the researcher or experimenter; power analysis solves for the fourth component. The four major factors are as follows:

1. The significance criterion and other things being equal, the more stringent this criterion, the lower is the power.
2. The sample size, n. As sample size increases, power increases.
3. The population effect size, gamma (γ). Gamma is a measure of how wrong the null hypothesis is, that is, how strong the effect of the independent variable is on the dependent variable in the population.
4. Power, or I –β. This is the probability of rejecting the null hypothesis. The two purposes for power analysis are to solve for the sample size needed in a study to increase the likelihood of demonstrating significant results and to determine the power of statistical test, after it has been applied.

SAMPLE SIZE

In most statistical procedures, the assumptions that must be made depend largely on the sample size available. For instances, if the sample size is quite large, then, as a result of the Central Limit Theorem (Parzen, 1960), the normality assumption need not be made in many procedures. But if the sample size is relatively small, the normality assumption must be true for the results to be valid. A number of formulae have been devised for determining the sample size depending upon the availability of information. A few formulae are given below: $n = \left(\frac{Z\delta}{d}\right)^2$

Where n= sample size, Z= value at a specified level of confidence or desired degree of precision.

S= Standard deviation of the population, d= Difference between population mean and Sample mean.

Example 7: Determine the sample size if $\delta = 6,$ population mean = 25, sample mean =23 and the desired degree of precision is 99 percent (Gupta, 2001).

Solution:

$$n = \left(\frac{Z\delta}{d}\right)^2$$

n=?, d=25-23=2,δ =6

Z =2.576(at 1% level the Z value is 2.576)

Substituting the values:

$$n = \left(\frac{2.576x6}{2}\right)^2 = 7.728^2 = 59.72$$

In the same manner, the sample size (n) can be determined from the formula for determining the standard error of mean i.e.

$$\delta_X = \frac{\delta}{\sqrt{n}} \Rightarrow \delta_X{}^2 = \frac{\delta^2}{n} \text{ or } n = \left(\frac{\delta}{\delta_X}\right)^2$$

Therefore, if δ is 10 and $\delta_X = 2.25,$ n shall be $n = \left(\frac{10}{2.25}\right)^2 = (4)^2 = 16$

Also from the formula for calculating standard error of proportion, the sample size can be determined from the fact that if

$$\delta_p = \sqrt{\frac{pq}{n}} \; or \; \delta_p{}^2 = \frac{pq}{n} \; or \; n = \frac{pq}{\delta_p{}^2}$$

If p=.5, q=.5 and

$$\delta_p = .005, \; n = \frac{0.5 \times .5}{(0.005)^2} = 10,000.$$

More details regarding the determination of sample size are seen in Gupta (2001) and Khamis (1988).

PARAMETER SPECIFICATIONS

Sometimes a statistical procedure makes assumptions about one or more parameters. For instance, in tests involving two or more populations, it may be required that the variance of the outcome variable be the same for all populations. There is a statistical test to help one make a judgment about the equality of variances and this is called hypothesis testing concerning two population variances – see hypothesis 10 in the next section.

In the one-way analysis of variance and linear regression analysis (see hypotheses 12

and 13 in the next section), the most common way to check assumptions is to study the residuals (the difference between the observed response and the response predicted from the model). These residuals can be used to check the normality assumption, the equal variance assumption, and even independent of the observations, depending on the available data. The details of residual analysis are beyond the scope of this article, but the interested reader is referred to McClave et al. (1997).

Specific assumptions in the elementary (parametric) statistical tests

For 13 of the elementary statistical test, the null hypothesis is specified along with other information about the sampling situation, followed by the assumptions of the parametric statistical test appropriate for the hypothesis. In the following list, μ represents a population mean, ρ a population proportion, δ^2 a population variance, and β a linear regression coefficient. A general discussion of the terminology and concepts of hypothesis testing can be found in articles by Khamis (1987) , Khamis and Warner (1997).

1. H_O: $\mu = \mu_o$, $n \geq 30$ (Z-test) Assumption: simple random sample from the study population
2. H_O: $\mu = \mu_O$, $n < 30$ (t-test) Assumptions: simple random sample from the study population, normality of study population.
3. Ho: P = Po (Z-test for a binomial proportion) Assumption: simple random sample from the study population of binomial ("Yes/No") responses: $Po \pm 3\delta_0$ does not include 0 or 1, where $\delta_{o=} (Po \mid I - Po \mid / +n)^{1/2}$
4. Ho: $\delta^2 = \delta^2_0$ (chi – squared test of variance) Assumptions: simple random sample from the study population: normality of the study population.
5. H_O: $\mu_1 = \mu_2 =$ Do, $n_1 \geq 30$ and $n_2 \geq 30$ (two-sample Z-test) Assumptions: random, independent samples from the two study +populations.
6. H_O: $\mu_1 = \mu_2 =$ Do, $n_1 < 30$ or $n_2 < 30$ (two–sample t-test or pooled–sample t-test) Assumptions: random independent samples from the two study populations; normality for each population: population variances are the same.
7. H_O: $\mu_1 = \mu_{2=}$ = Do, Observations from the two samples are paired according to one or more criteria. Number of paired differences is at least 30 (paired Z-test). Assumptions: paired difference form a random sample from the population of paired difference
8. H_O: $\mu_1 = \mu_2$ = Do, Observations from the two samples are paired according to one or more criteria: number of paired differences is less than 30 (paired t-test). Assumption: paired differences form a random sample from the population of paired differences: normality for the population paired differences.
9. Ho: $P_1 - P_2 = 0$ (Z-test for the equality of two binomial proportions). Assumptions: simples random samples from the study populations of binomial ("Yes/No") responses $p_1 \pm 3\delta_1$, does not include 0 or 1 where $\delta_1 = \{p_1 (I - p1) / n_1\}^{1/2}$ and $p_2 + 3\delta_2$ does not include 0 or 1 where $\delta_2 = [p_2 (1 - p_2)/n_2]^{1/2}$
10. Ho: $\delta_2 / \delta^2_2 = 1$ (F-test for equality of variances). Assumptions: random, independent samples from the study populations: normality for each population
11. Ho: $\mu_1 = \mu_2 = \mu_{3=} \ldots\ldots = \mu_k$, $n \geq 30$ for all K groups (large sample one-way analysis of variances for K groups) Assumptions: random, independent sample for all K groups
12. Ho: $\mu_1, = \mu_2 = \mu_{3=} \ldots\ldots = \mu_k$, $n < 30$ for at least one group (one-way analysis of variance for K groups). Assumptions: random, independent sample for all K groups. Normality of each population: the variances for all K groups are the same.
13. Ho: $\beta = \beta_o$ (t-test for linear regression: $Y = \alpha + \beta x + \varepsilon$). Assumptions: random sample of (x, y) pairs; normality and equal variances for Y values at each level of X.

What to do if one or more assumptions are violated

(Non-parametric Tests)

Each of the tests listed will be reviewed for the case in which one or more of the assumptions is violated. Because randomness and independence of observations is fundamental for any test procedure, they will continue to be assumed true. Most of the alternative tests listed below are non-parametric tests (McClave et al., 1997).

1. No alternative approach is needed because there are no assumptions beyond that of a simple random sample from the study population.
2. If the underlying populations are severely non-normal, use the Sign Test to determine whether the population median is equal to μ_o.
3. If $p_o + 3\delta_0$ contains 0 or 1 exact binomial probabilities must be computed (McClave et al., 1997).
4. No practical alternative.
5. No alternative approach is needed because there are no assumptions beyond that of simple random samples from the study populations.
6. If the normality assumptions or the equal variance assumption is violated, use the Wilcoxon Rank Sum Test to test for a difference in population distributions.
7. No alternative approach is needed because there are no assumptions beyond that of simple random samples from the study populations.
8. If normality of the population of paired difference is violated, use the Wilcoxon Signed Rank Test.
9. No practical alternative.
10. If normality of the populations is violated, use Levene's Test (Snedecor and Cochran, 1980).
11. No alternative approach is needed because there are no assumptions beyond that of simple random samples from the study populations.
12. If the normality or equal variance assumptions are violated, use the Kruskal-Wallis H-Test to compare the population distributions.
13. If the normality assumptions is violated, use a non-parametric regression technique (Birkes and Dodge, 1993) or the Spearman Rank correlation coefficient to determine the linear association between X and Y. By way of summary, the table 9 is a list of the parametric tests along with their non-parametric counterparts. Unless otherwise indicated, all tests are discussed in McClave et al. (1997). Note that the above 13 assumptions and information as to what will be done when these assumptions are violated were discussed in greater details in those references contained in them.

Conclusion

In the standard statistics classes, students are told which statistical procedure to use for each kind of data analysis problem. For instance, to compare two population means, you may use the two-sample t-test to compare several means, you may use a one-way analysis of variance, to compare two variances, you may use an F-test, etc. An easy way to identify the correct statistical procedure for each of the more common data

Table 9: Parametric tests and their non-parametric counterparts.

Parameter	Parametric Test	Non-parametric Test
One mean	"t-test" for μ	"Sign Test"
Two means, independent samples	Two sample "t-test"	"Wilcoxon Rank Sum test"
Two means, dependent samples	"Paired t-test"	Wilcoxon signed Rank test
Two variances	"F-test"	"Levene's Test"
Several means, independent samples	One-way ANOVA	"Kruskal–Wallis H Test"
Linear regression coefficient	"t–test" for β	"Non-parametric regression" or "Spearman Rank Correlation Coefficient".

Source: McClave, 1997.

analysis problem is given by Khamis (1992). However, it is emphasized in that article that each of the statistical procedures carries with it certain assumptions that must be verified before applying the procedure.

Many of the parametric procedures listed earlier are robust against violations of the assumptions (i.e. even if a given assumption is violated, the procedure provides approximately accurate results). For instance, the analysis of variance is robust against modest violations of the normality and equal variance assumptions. So, generally, one need not resort to the non-parametric procedure except when there are severe violations of these assumptions. If you are uncertain about which procedure to use in a given case (i.e., you are uncertain about the extent of the violation of a given assumption), you might try applying both the parametric (one-way analysis of variance) and the non-parametric (Kruskal-Wallis) tests to the data set. If the results agree, or approximately agree, then the conclusion can be made with no inconsistencies. If the results clearly disagree, further analysis of the data is required, to discover the reason for the inconsistency (Zar, 1996).

It may seem that the basic assumptions attached to the various statistical procedures are prohibitive (i.e. very few data sets would qualify for the given procedure) or artificial. However, in practice, the standard statistical procedures cover an enormous number of data analytic problems, especially given that they are applicable even when violations of the assumptions occur. Only when the violations are severe does one need to resort to alternative approaches. For instance, in the case of the normality assumption, as long as the sample data are unimodal with no extreme outliers, the parametric procedure is probably adequate (Prabhakara, 2006).

Many of the tests listed earlier involve a cut-off of 30 for the sample size. That is, the nature of the assumptions in these instances depends on whether the sample size falls below 30. This cut -off value is determined by the central limit Theorem as a good criterion for these tests. Although many data sets in biomedical research tend to be well behaved distribution ally, it is still important to check the validity of many assumptions associated with the statistical test procedure to be applied. By doing so, and modifying the approach to the data analysis accordingly, reliable and accurate rate conclusions can be made from your data, and misleading or incorrect results can be minimized.

Whenever working with research data, it is a good idea to consult with an applied statistician or biostatistician to ensure that the handling of the data and experimental design is done properly.

REFERENCES

Birkes D, Dodge Y. 1993: *Alternative Method of Regression*. John Wiley and Sons: New York.

Denise FP, Bernadette PH. 1995. *Nursing Research Principles and Methods* (5th edn). J. B. Lippincott Company, 151; 453– 459.

Eze JI, Obiegbu ME, Jude NE. 2005. *Methods for Construction and Business Managers*. The Nigerian Institute of Building; 191–207.

Fienberg SE. 1970. Randomization and social affairs. *The 1979 Draft Lottery Science*, **171**: 255-261.

Gupta SP. 2001. *Statistical Methods*. Sultan Chand and Sons; 80-82.

Himes JH. 1991. Lack of correspondence among measures identifying the obese. *Prevent. Med.*, **7**:107–114.

Khamis HJ. 1987. Statistics refresher: Testing hypothesis and diagnostic test evaluation. *JDMS*, **3**: 123–129.

Khamis, H. 1988. Statistics II: choice of sample size. *JDMS*, **4**:176-183.

Khamis HJ, Warner R. 1997. Statistical methods in obstetric ultrasound. In *Diagnostic Medical Sonography: A Guide to Clinical Practice, Obstetrics and Gynaecology* (2nd edn), Berman MC, Cohen HL (eds). Lyppincott: New York; 639–650.

Khamis HJ. 1992. Deciding on the correct statistical technique. *JDMS*, **8**: 193–198.

McClave JT, Dietrich FH, Sincich T. 1997. *Statistics* (7th edn). Upper Saddle River N. J. Prentice Hall.

Nwabuokei PO. 1989. *Fundamentals of Statistics* (2nd edn). Koruma Books; 117-127.

Oyeka CA. 1996. *An Introduction to Applied Statistical Methods in the Sciences* (7th

edn). Nobern Avocation Publishing Company; 348 –359, 377- 406.

Parzen E. 1960. *Modern Probability Theory and its Applications*. John Wiley and Sons, Inc.; 430-433.

Pearl R. 1929. Cancer and tuberculosis. *A. M. J. Epidemiol.,* **9**: 99–159.

Prabhakara GN. 2006. *Biostatistics* (3rd edn). Jaypee Brother.

Rastogi VB. 2007. *Fundamentals of Biostatistics*. Are Books India.

Schefler WC. 1984. *Statistics for Health Professionals* Reading M.A. Addison Wesley Publishing; 143 –149.

Snedecov GW, Cochran WG. 1980. *Statistical Methods* (7th edn). The IOWA State University Press: Ames, IA.

Ugwu AC. 2007. Sonographic assessment of gallbladder sizes and contraction index in an adult Nigerian population. Completed M.Sc dissertation, University of Nigeria.

Zar JH. 1996. *Biostatistical Analysis* (3rd edn). Upper saddle River N. J. Prentice Hall.