STATISTICS IN ANAESTHESIA: PART ONE

ANAESTHESIA TUTORIAL OF THE WEEK 302

20TH JANUARY 2014

Sameer Hanna-Juma
Northampton Hospital NHS Trust, UK
Correspondence to Samhanna76@hotmail.com

QUESTIONS

Before continuing, try to answer the following questions. The answers can be found at the end of the article.

1. Which of the following statements is correct?
    a. Type I error occurs when there is a difference that is only due to chance and it is mainly due to statistical errors in the study design.
    b. Histograms are useful to demonstrate the mode of distribution of a data set.
    c. Type II error occur when there is real difference that is not shown statistically.
    d. Mean, Median and Mode are equal in normal distribution.
    e. A Box and Whisker plot is usually used to demonstrate normally distributed data.

2. Which of the following is correct?
    a. Standard error of mean (SEM) is a measure of spread.
    b. NNT (number needed to treat) is the reciprocal of relative risk reduction.
    c. The standard error of mean (SEM) decreases as the sample size increases.
    d. Absolute risk reduction is the percentage in the intervention group minus percentage in the control group.
    e. Sample size does not affect the confidence intervals.

INTRODUCTION

Statistics is the practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole (population) from those in a representative sample (Oxford Dictionary). This tutorial will cover the basic knowledge that underpins statistics in clinical practice and will cover the following topics:

- Types of Data
- Data Collection
- Data Presentation
- Inferential Statistics

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

TYPES OF DATA

Data are observations collected together from measuring a set of variables. Variables are elements, features or factors that are liable to vary or change. The first stage in the correct handling of data is to recognize what type of data it is. This determines how it is described and is the first step in deciding the type of statistical test to use. The two main types of data are qualitative and quantitative, and most studies will have a combination of both. While quantitative data are easy to analyse and fairly reliable, qualitative data provide more in depth description of the sample.

Qualitative Data

These are variables that do not have a numerical value. They usually describe a meaning and give a name or label to variables. Note that the label given to a variable may be a number like ASA-1, but this label does not have a true numerical value. Qualitative data may be ordinal or nominal, which are described in the box below.

Nominal (or Categorical or Attribute) data	Ordinal (or Ordered categorical or Ranked) data
• Categories do not overlap • Have no numerical value • Called binary if only 2 possible categories like male or female • Commonly used to describe the sample characteristics • Examples include blood group and operation type	• Categories ordered in meaningful way • There is intrinsic magnitude that cannot be measured • commonly used in surveys and to analyse non-numerical variables • Examples include ASA score and Categorical pain scores

Note that in categorical pain scores where: none=0, mild=1, moderate=2, severe=3, the moderate pain is more than mild pain but not twice as much.

Quantitative Data

These are variables that are truly numerical. There are various types of quantitative data, but this usually does not affect the choice of statistical test. Quantitative data may be discrete or continuous as below.

Discrete	Continuous
• Data that vary by finite steps • Example include numerical pain scale	• Data measured along a continnum • Examples include visual analoge pain scale

Continuous data can be ratio or non-ratio data. On a ratio scale the variable of zero truly represents the absence of quantity or characteristic and a single unit change has the same interpretation at any part of the scale. However, on a non-ratio scale the variable changes can be non-linear throughout different parts of the scale.

DATA COLLECTION

There are various methods of collecting data. The more common types of data sampling include simple random sampling and stratified sampling.

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Simple Random Sample

A truly random sample is one where every member of the population has an equal chance of being included in the study. This is the most realistic representation of the whole population, but it is difficult to achieve a true random sample in medical research. For example if you are studying the effect of haemodynamic monitoring on the outcome of emergency laparotomy you should be able to include every single patient having an emergency laparotomy if selected.

Stratified Random Sample

The random sample is divided into subgroups called strata. This is to try and make the sample more representative of the population and reduce the effect of confounding factors. For example in the above study the sample group is stratified into 2 subgroups – smokers and non-smokers first. You will then randomly admit patients to a control group without a haemodynamic monitoring device, and a treatment group with a haemodynamic monitoring device.This will eliminate the confounding effect of smoking on outcome following emergency laparotomy as equal numbers will be included in both groups.

Other types of samples include cluster sampling, multistage sampling, contact sampling and multiphase sampling.

DATA PRESENTATION

All studies have a large number of raw data that are collected during the study period. These raw data can be presented in three different ways; tabular, graphic (chart), or numerical (descriptive statistics) forms. Each mode of presentation has its uses, advantages and disadvantages.

Tabular presentation

Frequency and cumulative frequency tables are common ways of presenting data in clinical research. These methods can be used to present all different types of variables including nominal, ordinal and quantitative data. In order to present continuous data by this mode it needs to be arranged into groups (intervals) first. Examples include the ASA score and age groups of your study population.

ASA Category	Frequency (number of patients) $n = 390$
ASA-1	60
ASA-2	230
ASA-3	65
ASA-4	32
ASA-5	3

Age Category	Frequency (number of patients) $n = 390$	Cumulative frequency (number of patients)
20-40 Years	24	24
41-60 Years	210	234
61-80 Years	126	360
81-100 Years	30	390

A relative frequency table is another way of presenting the same data above. In this table the numbers will be replaced by percentages of the total sample number.

ASA Category	Relative frequency (% of patients in each category) $n = 390$
ASA-1	15.3
ASA-2	58.9
ASA-3	16.6
ASA-4	8.2
ASA-5	0.77

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Cross tabulation and ranking of data is another use for frequency tables.

Graphic (Chart) Presentation

Charts and diagrams are very important, especially when presenting a large pool of data. They are powerful visual tools which highlight important relationships between different variables. In clinical statistics the following charts are frequently used:

Pie Charts

Pie charts are useful to show the proportion of different groups that constitute the total sample of the study. The whole pie represents the total sample while the size occupied by each group will be proportional to their number. Pie charts are used for ordinal and nominal data. They can be useful to highlight potential imbalance of the study sample or potential confounding factors.
The example in figure 1 shows the different surgical groups in a study of supra-glottic airway devices used in clinical practice.

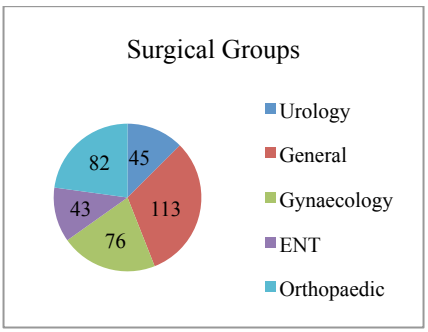

Figure 1 Pie chart

Bar Charts

Bar charts are used to compare different classes of data. The *x* axis is usually dimensionless while the *y* axis represents the frequency of each class. Each class could represent a single group, or be further divided into subgroups. The example in figure 2 shows the same data above presented as a bar chart.
Note that in each class gender subgroups are presented on this chart.

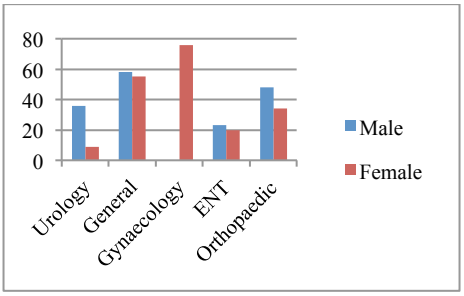

Figure 2 Bar chart

Histograms

This is a specialized bar chart used to give a visual presentation of interval data. Quantitative data, and in particular continuous data, are divided into intervals in order to be integrated into frequency tables. Histograms are useful to show the mode of distribution of the data. This will have a significant implication on the choice of statistical tests as we will explore later in this article. We can clearly see in the example in figure 3 that the data do not follow a normal distribution; otherwise the middle of each bar would be lying on the normal distribution curve. The statistical analysis here is different from data that are normally distributed. Histograms are also useful to demonstrate descriptive statistics like mean, mode and standard deviation.

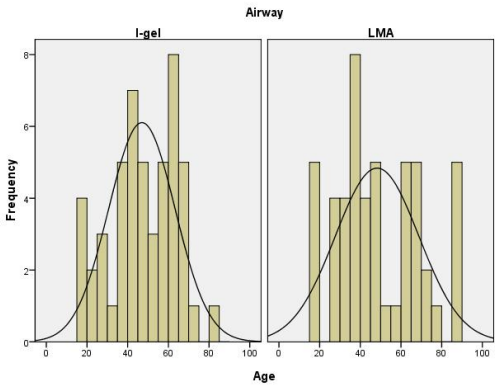

Figure 3 Histogram

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Frequency Curves

These are very similar to histograms, but without the bars. They serve a similar use to histograms, but one advantage is that they can be used to compare the distribution of 2 or more groups on the same chart. In figure 4 we can compare the trends in blood pressure between the control group and the study group. There is a decrease in the number of patients with blood pressure above 140 mmHg in the study group compared to the control group.

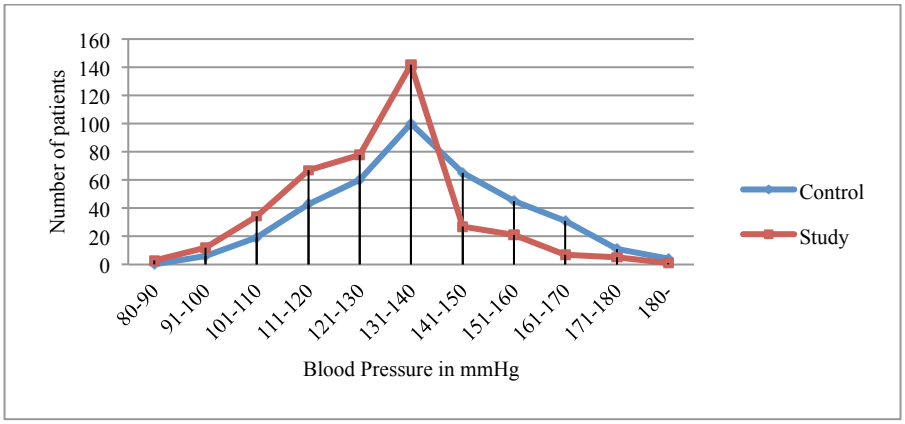

Figure 4 Frequency curve

Cumulative Frequency Curves (Ogive)

This is a graph of a cumulative distribution, which shows data values on the horizontal axis and either the cumulative frequencies, the cumulative relative frequencies or cumulative percent frequencies on the vertical axis. This type of graph is useful to identify the proportion of a sample that falls below or above certain limit.

Scatterplots

These are used to determine if there is any relationship between two sample variables. They can also be used to statistically calculate the strength of the relationship using a correlation coefficient. The example in figure 5 demonstrates the relationship between the dose of new muscle relaxant (MR) in mg/Kg body weight and the time required to regain a train of four (TOF) in a sample study. The data show a direct relationship between an increased dose and time to TOF recovery.

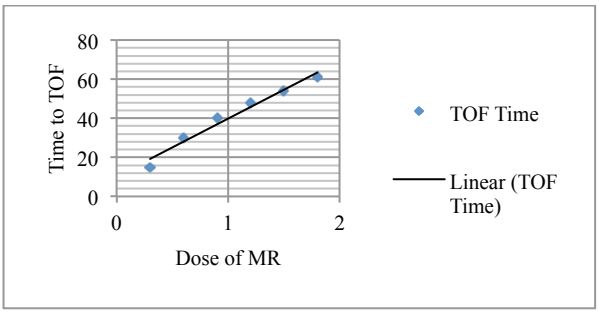

Figure 5 Scatter plot

- Histograms are commonly used to determine mode of distribution of data to aid the choice of appropriate statistical tests.
- A scatterplots is an important visual aid to demonstrate the degree of correlation in multivariate analysis.

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Numerical Presentation (Descriptive Statistics)

Even though descriptive statistics refers to data presentation in tabular, chart and numerical forms, in medical research they mainly refer to numerical presentation of data. The main aim of descriptive statistics is to present a meaningful summary of the sample data rather than drawing conclusions about the whole population. This is essential before deciding on the appropriate statistical tests for inferential analysis. When using descriptive statistics there is a risk of losing important details, despite the fact that it provides a powerful summary that may enable comparisons across variables. Descriptive statistics can be used for both Univariate (single variable) analysis and Multivariate analysis. Three key characteristics for Univariate analysis are distribution, central tendency and measurements of spread.

Distribution

In healthcare statistical analysis the normal (Gaussian) distribution has central importance, and is the most common distribution of biological data (height, weight, blood pressure) in healthy individuals. Visual aids such as histograms can be used to determine the distribution of data, but there are other formal statistical tests to determine if the data are normally distributed, such as the Shapiro-Wilkes test and D'Agostino-Peasrson omnibus test. The normal (parametric) distribution is characterized by a single peak (unimodal) and a symmetrical spread of variables on either side. All central tendency measures (mean, mode and median) are equal in a normal distribution and they are represented by the point of maximum frequency. The spread of data is equal on either side, which represents standard deviation (SD). Two parameters (mean and SD) can fully describe the shape of the curve in Figure 6.

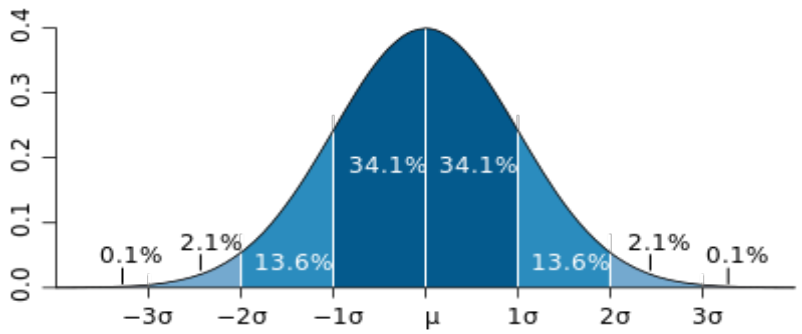

Figure 6 Normal distribution curve. Sixty-eight percent of the population will be within 1 SD of the mean, 95% will be within 2 SD of the mean, and 99.7% will be within 3 SD of the mean. Most medical probability studies will investigate mean ± 2SD (95% of the population)

In non-parametric data the variables are not equally distributed around the central tendency point. The data could be clustered around one side and sparse on the other, which is called skewness (Fig 7). Data can also have more than one peak (multimodal). Kurtosis is another term which describes the peak of the curve, with a normal distribution having a kurtosis of zero. A curve with a sharper peak and longer tail is termed *positive*, while *negative* kurtosis indicates a wider, flattened distribution. In these situations statistical analysis should be performed using non-parametric statistical tests. In situations when mode of distribution is not clear it is safer to use non-parametric statistical tests.

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

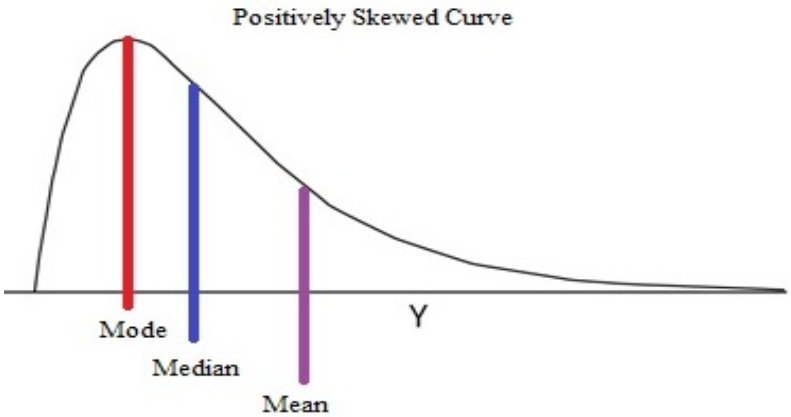

Figure 7 Positively skewed data. Notice the different positions for mode, median and mean.

Central tendency

The central tendency of a distribution is an estimate of the "center" of a distribution of values. The three central tendency measures are mean, median and mode:

- Mean is simply the total sum of the values divided by the number of variables (arithmetic mean). It is used as a measure of central tendency for parametric data and should not be used to report central tendency of ordinal or nominal data.
- Median is the middle value when all the data are arranged in numerical order. This means that 50% of the data are below and 50% above that value. This is preferable to measuring central tendency in non-parametric data since it is less affected by outliers than the mean.
- Mode is the most frequently occurring observation in a set of data. It is not a good indicator of central tendency but it is the only way for measuring central tendency in nominal and ordinal data.

Measures of Spread

Range is the simplest measure of spread, but with limited practical use. It is the difference between the maximum and the minimum value in a data set. Variance and standard deviation (SD) are the main measures of parametric data spread. These measures are more accurate because they include the total sample values in the calculation. Variance is calculated from the sum of the square difference of each value from the mean divided by the total study population. Values are squared otherwise they will cancel each other in normal distribution because equal values are negative and positive. The SD is the square root of the variance. Percentiles (Quartiles) are the main measures of non-parametric data spread. Quartiles are self-explanatory: the 1st quartile has 25% of the data below it, the 2nd quartile corresponds to the median and has 50% of data below it, and the 3rd quartile has 75% of data below it.

Percentiles are usually demonstrated in a Box and Whisker plot (figure 8).

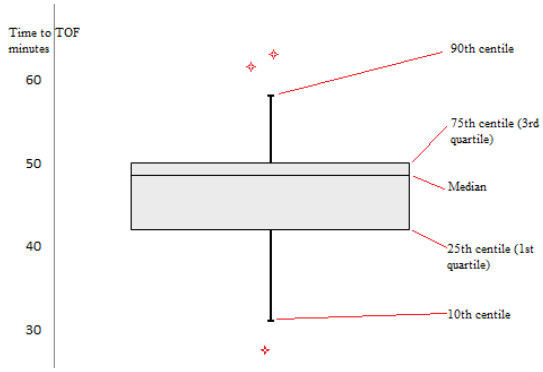

Figure 8 Box and whisker plot

The box in this example represents 50% of the sample (50% had their train of four return between 42 to 50 minutes following administration of muscle relaxant). The whiskers represent the 10th and 90th centile, and the stars are outliers (10% of population had their TOF return after 31 minutes while 90% had their TOF return after 58 minutes). Note that the median is not in the middle of the Box, which demonstrates the non-normal distribution of the data (an equal number had their TOF return between 42-48 minutes and 48-50 minutes).

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

INFERENTIAL STATISTICS

Inference is the process of deriving logical conclusions from premises known or assumed to be true, and it presupposes that the whole population is represented by the study sample. It is important to understand there is a degree of assumption and this will lead on to the concept of probability. Inferential statistics can be descriptive (the sample mean represents the population mean) or analytical (the study of relationships between different variables in a sample that can be generalized to the population of interest). Hence two basic components of inferential statistics are sample and probability.

Sample

This is the portion of recruits in a study. For example, in a study of the effect of perioperative haemodynamic monitoring on outcome of patients presenting for emergency laparotomy, the patients that are enrolled during the study period will be the sample, and all patients that require emergency laparotomy will be the population. It is very important to calculate the size of the sample needed before commencing any trial This is related to the study *power*.

Probability

For each and every event there are a number of possible outcomes. The chance of any outcome ranges from 0 (never) to 1 (always). In clinical research, probability (or *P-value*) is an essential part of presenting any type of inferential data. This is to reassure the reader that the outcome was secondary to the effect of the studied variable and has not occurred purely by chance. So when a study presents a difference between the blood transfusion requirements of the control group and a group that received tranexamic acid with a *P-value* of 0.01, it means that the probability of this happening purely by chance and not due to the use of tranexamic acid is very small (but not impossible). Other common terms used in medical statistics to describe the probability of an event happening include odds ratio and risk ratio.

Odds ratio

Odds can be defined as the probability of an event happening. For example in a group of 160 women in labour, 110 have a normal delivery and 50 have a caesarean section. Hence the odds of having caesarean section are 50:110 equal 0.45. A more important and commonly applied concept in clinical research is the odds ratio. This is used to measure the effect of certain intervention on the probability of an event happening. An example is the odds ratio for normal delivery in women in labour who are having epidural compared to a control group. In a group of parturients with an epidural, 90 have a normal delivery while another 60 have a caesarean section. The odds are 60:90 equal 0.66. The odds ratio for having a caesarean section with epidural analgesia compared to the control group is therefore 0.66:0.45 equal 1.46. An odds ratio of 1 mean there is no significant difference between the 2 groups, while in this case the odds ratio indicates that the probability of caesarean section is more likely when an epidural is used for analgesia.

Risk ratio

Risk ratio is commonly used in epidemiological studies and is very similar to odds ratio. It is important to understand in risk ratio calculations that the denominator is the total population. For example the risk of having a caesarean section in the above groups is 50:160 equal 0.312 compare to the odds of 0.45. In large scale epidemiological studies risk ratio, absolute risk reduction and relative risk reduction are important statistics that are used to decide about the effectiveness of a particular intervention, as well as the financial implications of introducing the treatment on a large population scale (figure 9). The risk ratio can't be calculated for case control studies; however the odds ratio gives an approximate estimate of the risk ratio.

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Absolute Risk Reduction	• difference between the risk of event happening in the treatment group compared to the control • calculated by percentage in the intervention group - percentage in the control group
Relative Risk Reduction	• proportion by which an intervention reduces the risk of an event happening • calculated by percentage in the intervention group divided by percentage in the control group
Number Needed to Treat	• number of patients need to receive treatment for one to benefit (the smaller the more effective is the treatment), it s calculate as the reciprocal of the Absolute Risk Reduction

Figure 9 the difference between absolute risk reduction, relative risk reduction and number needed to treat

The significance of the above values needs to be evaluated within the context of the study. In particular studies, a large sample population may give misleading information if only relative risk reduction is calculated.

Example: the risk of PONV following administration of new antiemetic drug compared to the gold standard.

- Group A (control): 10 patients had vomiting out of 1000
- Group B (new treatment): 5 patients had vomiting out of 1000

Absolute risk reduction equal 99.5% – 99% = 0.5%, while Relative risk reduction = 0.005:0.01 = 0.5 or 50%.

Some authors will use relative risk reduction to highlight the 50% risk reduction of PONV. However, if you calculate the Number Needed to Treat (NNT) = 100:0.5 = 200 patients need to receive the treatment to reduce the risk of PONV in one patient compared to the gold standard treatment. This might change your mind about using this treatment if you know that it could be associated with significant cost or side effects such as dysrhythmias.

Null and Alternative Hypothesis

Statistical analysis allows the strength of evidence supporting or refuting a theory to be quantified by analysis of experimental observations. In medical statistics it is assumed there is no difference in a certain variable among the different groups until proven otherwise. This is called the null hypothesis. The alternative hypothesis is the opposite, and is usually the question of greater interest to the researcher. It assumes the difference is secondary to instigation of a particular treatment or intervention. The main aim of most statistical tests is to reject the null hypothesis and approve the alternative hypothesis.

Types of Error

Type I error occurs when inappropriately rejecting the null hypothesis, and is also called α error. This means that a statistically significant difference has been found where no real difference exists. The level of significance is the *P-value*, which represents the smallest value for α for which the null hypothesis would be rejected, usually set at 0.05 (1 in 20 occurrences) in medical statistics.

The *P-value* should be calculated and mentioned in a research paper, but it is not enough to simply say if it is below or above 0.05. For example, if 2 different studies investigating the same intervention had *P-values* of 0.051 and 0.049, in reality they have very similar outcomes. However, this could be misinterpreted if they were reported as having *P-value* < 0.05 and *P-value* > 0.05.

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Type II error occurs when inappropriately accepting the alternative hypothesis, also called β error. This means that no statistical difference has been found where one does in fact exist. The most common cause of this type of error is inadequate sample size; hence it is very important before starting any study to calculate the sample size required to detect a difference, and therefore avoid the risk of type II error. This is called the power of a study.

Study power is the probability that it will detect a statistically significant difference if one exists. It is calculated as (1-β). Most medical research will accept a β value of 0.2, and it is unlikely that a proposal for a study will be approved without a power calculation.

Types of Statistical Tests

It is important to choose the appropriate statistical test for each type of data. There are a few systematic steps that should be followed to establish the appropriate test for your data.

1. Identify whether the data are Qualitative or Quantitative
2. For Quantitative data, determine the type of distribution
3. Decide how many groups are being compared
4. Determine whether the data is paired or not.

The exact nature of the tests will not be discussed in this article, but the process can be represented as in Figure 10.

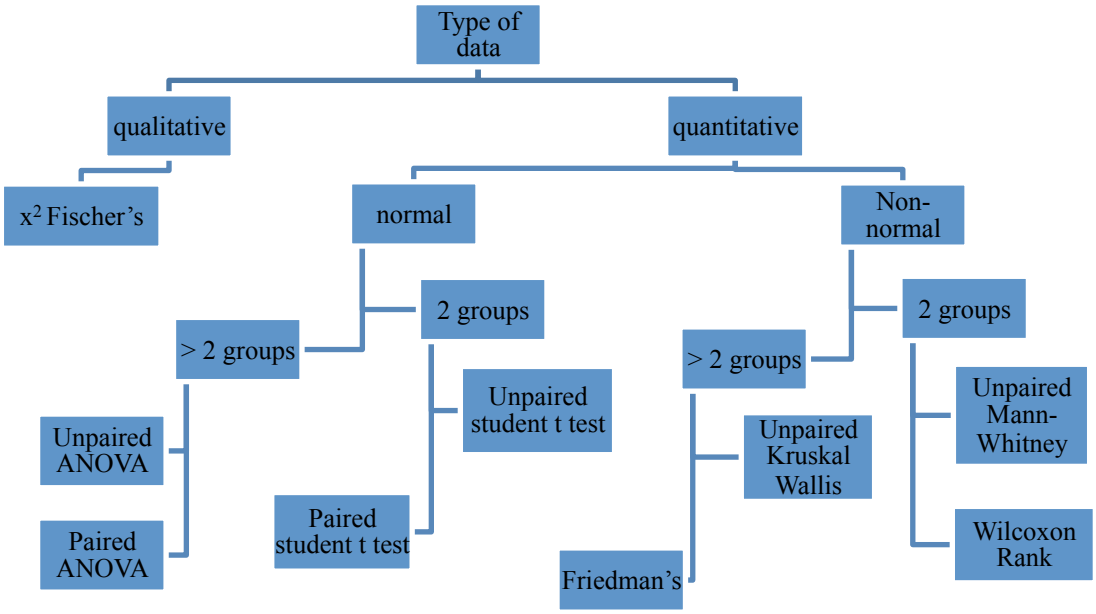

Figure 10 Method for selecting an appropriate statistical test

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

Sample and Population

To appreciate how accurately a study sample represents the population from which it is taken a few other statistical concepts can be used. These include the standard error of mean (SEM) and confidence intervals (CI).

SEM is not a measure of central tendency or spread, but reflects how close the mean of your sample is to the population mean. It can be calculated by the equation (SEM = SD/$\sqrt{n}$). From this equation it is clear that the larger the sample studied the smaller is the SEM and the better it represents the whole population. This is simply because a larger sample will contain more information about the population parameter of interest, and therefore results in more precise estimations. The CI gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. CI tells us how likely the size of the measured effect in the sample is representative of the population. Factors that affect CI include sample size and SD. The larger the sample size the more representative is the sample of the whole population, and the smaller is the CI. Importantly, a sample with a very wide CI might not be representative of the true population. It is particularly important to note the CI in randomized controlled trials that compare a new treatment with a gold standard. For example if a research paper demonstrated a 20 mmHg increase in mean arterial blood pressure with vasopressin compared to noradrenaline in patients with septic shock, with CI of -5 to +45mmHg, then the conclusion should be that there is no significant difference in effect on the population. This is because based on the CI the predicted change in the population mean arterial pressure could be zero, or even decreased by 5mmHg.

The example in figure 11 demonstrates confidence intervals in studies on the effect of exercise on resting heart rate.

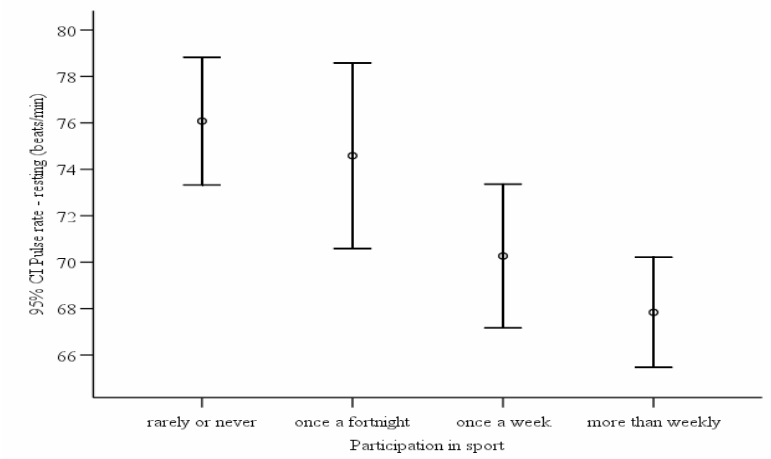

Figure 11 Illustration of confidence intervals in several studies on the effect of exercise on resting heart rate

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.

ANSWERS TO QUESTIONS

1. TTTTF
2. FFTTF

WEBLINKS

http://www.cyclismo.org/tutorial/R/index.html
http://www.stats.gla.ac.uk/steps/glossary/index.html
http://en.wikipedia.org/wiki/Statistics

REFERENCES and FURTHER READING

1. McCluskey A, Lalkhen A. Statistics A publication series from Continuous Education in Anaesthesia, Critical Care and Pain Management, Oxford University Press in 2007
2. Harris M, Taylor G. Medical Statistics Made Easy, 2nd Edition, Scion Publishing Ltd, 2008.
3. Bowers D, Medical Statistics from Scratch an Introduction for Health Professionals, 2nd Edition, John Wiley & Sons Ltd, UK 2008.
4. Spoors C, Kiff K. Oxford Specialty Training: Training in Anaesthesia, 1st Edition, Oxford University Press in 2010

This work is licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/.