APPLIED MEDICAL STATISTICS

TOOLS FOR CLINICAL AND RESEARCH PRACTICE

T. DHASARATHARAMAN,
STATISTICIAN,
KAUVERY HOSPITALS, INDIA

Applied Medical Statistics: Tools for Clinical and Research Practice

T. Dhasaratharaman

Statistician

Kauvery Hospitals, India

Chief Editor

Dr. Venkita S Suresh

Group Medical Director

Kauvery Hospitals, India

Production Team

Dr. Prabhaharan Renganathan, PhD, Copy Editor

Mr. Chandru, Clinical Pharmacist

Ms. Suryaprabha, Assistant Manager-Clinical Research

Kauvery Hospitals, India

Acknowledgment

"I would like to extend my deepest gratitude to all those who contributed to the development of this book. First and foremost, I want to thank my mentor, Dr. Venkita S Suresh, and colleagues, Mr. P Vairamuthu, Ms. B Suja, Ms. D Suryaprabha, Ms. D Ramya and Mr. Chandru, for their expertise and insightful contributions throughout this project. Your shared knowledge of medical statistics and research practices is invaluable.

We would like to thank our editorial team, particularly Dr. Prabhaharan Renganathan, PhD, for their guidance and meticulous attention to ensure that this book serves as a comprehensive and practical tool for clinicians and researchers alike.

I would also like to acknowledge Kauvery Hosptials, India, for providing access to the datasets and resources that were instrumental in illustrating the statistical methods covered in this book.

Finally, I am grateful to my family and friends for their unwavering support and patience during the long hours of my writing and research. This book is a result of your encouragement and belief in my work."

Mr. T. Dhasaratharaman,

Assistant Manager – Statistician,

Kauvery Hospitals, India

Foreword

The management guru Peter Drucker had famously said "What gets measured gets managed".

The basic tenet is that if you're measuring something then you would want to act on it.

But do you not wish to know whether what you have measured is significant?

Medical statistics were designed fundamentally around that question

In health care we care for the health of patients

While caring for them we create enormous medical data

Studying and acting on the information, knowledge and wisdom gained from them make us better, safer and more economical providers of health care.

It is vital that everyone involved in health care constantly harvest, organize, analyze and present their data

Medical Statistics examine that data and guide us on their real significance.

There is a method to it.

Dhasaratharaman's series of articles on the methods, published serially in KAUVERIAN, is now going to be either in your handheld as a first Kauverian study material PDF *Applied Medical Statistics: Tools for Clinical and Research Practice*. .

Each chapter is brief, to the point, practical and helpful. I hope that it helps you to be more confident, competent and content in creation of your scientific papers, towards presentation or publication.

Dr. Venkita S. Suresh,

Group Medical Director,

Kauvery Hospitals, India

Managing Director's Message

Dhasaratharaman has been the statistical back bone of Kauvery Hospitals

He has provided staunch support to the creation of all institutional statistics, not only on our performance but also to support our massive endeavor to seek and achieve the appropriate level of NABH accreditation for our hospitals and also in the formation of our ethical committees that examine, evaluate, and approve institutional research.

His statistical analysis enabled us to understand both the quality and quantity of the sterling service we as a group provided during the terrible pandemic years. We served 18000 patients in our ICUs and Isolation wards and sent 95 % safely home. We sent 98 % of the doctors, who were practicing COVID care in the community and got seriously infected, safely home. 100 % of our healthcare professionals who served COVID patients emerged safe from the pandemic.

Dhasaratharaman is the sole and complete supporter of the entire statistical analysis being undertaken by our increasing number of post graduates towards their dissertations and theses.

The KAUVERIAN, the premier scientific journal being published by the Kauvery Group, edited by Dr. Venkita Suresh, ably supported by his team, has been publishing Dhasaratharaman's articles on Medical Statistics on every issue and this book, *"Applied Medical Statistics: Tools for Clinical and Research Practice"*, is a careful compilation of those articles.

I hope this shall serve as a very useful Handbook for both Post Graduates as well as consultants, other doctors, nurses, physician assistants and all Kauverians who wish to harvest, organize, analyze, present and publish their original work in service of our patients.

I congratulate Dhasaratharaman, the author, Dr. Venkita Suresh, our Group Medical Director who is the spirit behind this unique enterprise, and Dr. Prabhaharan Renagnathan and Chandru, the editorial team who worked hard behind this laudable venture.

Dr. S. Manivannan,

Managing Director,

Kauvery Hospitals, India

Preface

Medical research and clinical practice increasingly rely on the use of statistical tools to analyze data, draw meaningful conclusions, and improve patient outcomes. However, navigating the world of applied statistics can often seem daunting for healthcare professionals, researchers, and students. Understanding the correct application and interpretation of statistical methods is crucial to ensure that clinical studies are both robust and reliable. "*Applied Medical Statistics: Tools for Clinical and Research Practice*" has been written with this goal in mind: to bridge the gap between complex statistical concepts and practical implementation in real-world medical research and practice.

This book is designed to provide readers with a comprehensive yet accessible guide to key statistical techniques that are frequently used in medical and clinical research. From basic descriptive statistics to advanced modeling techniques, each chapter is tailored to equip the reader with a solid understanding of statistical tools in the context of healthcare. This book is unique in its emphasis on real-world applications, supported by a wealth of clinical case studies and examples, ensuring that readers not only grasp the theoretical foundations but also appreciate their practical relevance in healthcare settings. The content is suitable for a wide range of professionals, including clinicians, biomedical researchers, and graduate students in health sciences. Each chapter provides step-by-step instructions and guidance on how to use statistical software, enabling readers to confidently perform their own analyses. In addition, it emphasizes the importance of understanding the underlying assumptions of statistical methods, avoiding common pitfalls, and ensuring that data interpretation is scientifically sound and ethically responsible.

We hope that "*Applied Medical Statistics: Tools for Clinical and Research Practice*" will serve as both a valuable reference and practical guide for healthcare professionals seeking to integrate sound statistical practices into their clinical and research endeavors. By demystifying statistical processes and providing user-friendly insights, we aim to empower the medical community to make data-driven decisions that ultimately enhance patient care and advance medical knowledge.

We extend our heartfelt thanks to countless healthcare professionals and researchers whose work inspires innovation and progress in the medical field. This book is dedicated to tireless research efforts.

Dr. Prabhaharan Renganathan, PhD,

Kauvery Hospitals India,

Contents

Chapter 1. Fundamentals of Statistics

Measures of Central Tendency

(1). *Mean, Arithmetic Mean ($\bar{x}$ or M)*

Sum of the scores in a distribution divided by the number of scores in the distribution. It is the most commonly used measure of the central tendency. It is often reported with its companion statistic, the standard deviation, which shows how far things vary from the average.

(2). *Median (Mdn)*

The midpoint or number in a distribution having 50% of the scores above it and 50% of the scores below it. If there are an odd number of scores, the median is the middle score.

(3). *Mode (Mo)*

The number that occurs most frequently in a distribution of scores or numbers. In some fields, notably education, sample data are often called scores, and the sample mode is known as the modal score.

Measures of Variability

(1). *Range (Ra)*

The difference between the highest and lowest scores in a distribution; a measure of variability.

(2). *Standard deviation (SD)*

The most stable measure of variability, it takes into account each and every score in a normal distribution. This descriptive statistic assesses how far individual scores vary in standard unit lengths from its midpoint of 0. For all normal distributions, 95% of the area is within 1.96 standard deviations of the mean.

(3). *Variance (SD^2)*

A measure of the dispersion of a set of data points around their mean value. It is a mathematical expectation of the average squared deviations from the mean.

Inferential Statistical Tests

Tests concerned with using selected sample data compared with population data in a variety of ways are *called inferential statistical tests***.** There are two main bodies of these tests. The first and most

frequently used are called *parametric statistical tests*. The second are called *nonparametric tests*. For each parametric test, there may be a comparable nonparametric test, sometimes even two or three.

(1). *Parametric tests*

Parametric tests are tests of significance appropriate when the data represent an interval or ratio scale of measurement and other specific assumptions have been met, specifically, that the sample statistics relate to the population parameters, that the variance of the sample relates to the variance of the population, that the population has normality, and that the data are statistically independent.

(2). *Nonparametric tests*

Nonparametric tests are statistical tests used when the data represent a nominal or ordinal level scale or when assumptions required for parametric tests cannot be met, specifically, small sample sizes, biased samples, an inability to determine the relationship between sample and population, and unequal variances between the sample and population. These are a class of tests that do not hold the assumptions of normality.

In the list of statistical terms below, when the test is a parametric test, the designation of will be used at the end of the definition. Conversely, when the test is a nonparametric test, the designation of will be used at the end of the definition.

Statistical Terms

(1). *Alpha coefficient (a):* See *Cronbach's alpha coefficient.*

(2). *Analysis of covariance (ANCOVA)*

A statistical technique for equating groups on one or more variables when testing for statistical significance using the *F*-test statistic. It adjusts scores on a dependent variable for initial differences on other variables, such as pre-test performance or IQ.

(3). *Analysis of variance (ANOVA)*

A statistical technique for determining the statistical significance of differences among means; it can be used with two or more groups and uses the *F*-test statistic.

(4). *Binomial test*

An exact test of the statistical significances of derivations from a theoretically expected distribution of observations into two categories.

(5). *Chi-square (χ^2)*

A nonparametric test of statistical significance appropriate when the data are in the form of frequency counts; it compares frequencies actually observed in a study with expected frequencies to see if they are significantly different.

(6). *Cochran's Q*

Used to evaluate the relation between two variables that are measured on a nominal scale. One of the variables may even be dichotomous or consisting of only two possible values.

(7). *Coefficient of determination (r^2)*

The square of the correlation coefficient (*r*), it indicates the degree of relationship strength by potentially explained variance between two variables.

(8). *Cohen's (d)*

A standardized way of measuring the effect size or difference by comparing two means by a simple math formula. It can be used to accompany the reporting of a *t*-test or ANOVA result and is often used in meta-analysis.

(9). *Cohen's kappa (K)*

A statistical measure of interrater agreement for qualitative (categorical) items. Scores range from –1.0 to 1.0.

(10). *Confidence interval (CI)*

Quantifies the uncertainty in measurement. It is usually reported as a 95% CI, which is the range of values within which it can be 95% certain that the true value for the whole population lies.

(11). *Correlation coefficient (r)*

A decimal number between 0.00 and ±1.00 that indicates the degree to which two quantitative variables are related. The most common one used is the Pearson Product Moment correlation coefficient or just the Pearson coefficient.

(12). *Cumulative frequency distribution*

A graphic depiction of how many times groups of scores appear in a sample.

(13). *Dependent t-test*

A data analysis procedure that assesses whether the means of two related groups are statistically different from each other, for example, one group's mean score (time one) compared with the same group's mean score (time two). It is also called the *paired samples* t-*test.*

(14). *Effect size (θ)*

Any measure of the strength of a relationship between two variables. Effect size statistics are used to assess comparisons between correlations, percentages, mean differences, probabilities, and so on.

(15). *Eta (η)*

An index that indicates the degree of a curvilinear relationship.

(16). *F-test (F)*

A parametric statistical test of the equality of the means of two or more samples. It compares the means and variances between and within groups over time. It is also called *analysis of variance (ANOVA).*

(17). *Factor analysis*

A statistical method for reducing a set of variables to a smaller number of factors or basic components in a scale or instrument being analyzed. Two main forms are *exploratory (EFA)* and *confirmatory factor analysis (CFA).*

(18). *Fisher's exact test*

A nonparametric statistical significance test used in the analysis of contingency tables where sample sizes are small. The test is useful for categorical data that result from classifying objects in two different ways; it is used to examine the significance of the association (contingency) between two kinds of classifications.

(19). *Friedman two-way analysis of variance*

A nonparametric inferential statistic used to compare two or more groups by ranks that are not independent.

(20). *G2*

This is a more conservative goodness-of-fit statistic than the χ^2 and is used when comparing hierarchical models in a categorical contingency (two-by-two) table.

(21). *Independent t-test*

A statistical procedure for comparing measurements of mean scores in two different groups or samples. It is also called the *independent samples* t-*test.*

(22). *Kolmogorav-Smirnov (K-S) test*

A nonparametric goodness of- fit test used to decide if a sample comes from a population with a specific distribution. The test is based on the empirical distribution function (ECDF).

(23). *Kruskal-Wallis one-way analysis of variance*

A nonparametric inferential statistic used to compare two or more independent groups for statistical significance of differences.

(24). *Mann-Whitney U-test (U)*

A nonparametric inferential statistic used to determine whether two uncorrelated groups differ significantly.

(25). *Median test*

A nonparametric test that tests the null hypothesis that the medians of the populations from which two samples are drawn are identical.

(26). *Multiple correlation (R)*

A numerical index describing the relationship between predicted and actual scores using multiple regression. The correlation between a criterion and the best combination of predictors.

(27). *Multivariate analysis of covariance (MANCOVA)*

An extension of ANOVA that incorporates two or more dependent variables in the same analysis. It is an extension of MANOVA where artificial dependent variables (DVs) are initially adjusted for differences in one or more covariates. It computes the multivariate *F* statistic.

(28). *Multivariate analysis of variance (MANOVA)*

It is an ANOVA with several dependent variables.

(29). *One-way analysis of variance (ANOVA)*

An extension of the independent group *t*-test where you have more than two groups. It computes the difference in means both between and within groups and compares variability between groups and variables. Its parametric test statistic is the *F*-test.

(30). *Pearson correlation coefficient (r)*

This is a measure of the correlation or linear relationship between two variables x and y, giving a value between +1 and −1 inclusive. It is widely used in the sciences as a measure of the strength of linear dependence between two variables.

(31). *Pooled point estimate*

An approximation of a point, usually a mean or variance, that combines information from two or more independent samples believed to have the same characteristics. It is used to assess the effects of treatment samples versus comparative samples.

(32). *Post hoc test*

A *post hoc* test (or *post hoc* comparison test) is used at the second stage of the analysis of variance (ANOVA) or multiple analyses of variance (MANOVA) if the null hypothesis is rejected.

(33). *Runs test*

Where measurements are made according to some well-defined ordering, in either time or space. A frequent question is whether or not the average value of the measurement is different at different points in the sequence. This nonparametric test provides a means for this.

(34). *Siegel-Tukey test*

A nonparametric test named after Sidney Siegel and John Tukey, which tests for differences in scale between two groups. Data measured must at least be ordinal.

(35). *Sign test*

A test that can be used whenever an experiment is conducted to compare a treatment with a control on a number of matched pairs, provided the two treatments are assigned to the members of each pair at random.

(36). *Spearman's rank order correlation (ϱ)*

A nonparametric test used to measure the relationship between two rank ordered scales. Data are in ordinal form.

(37). *Standard error of the mean (SEM)*

An estimate of the amount by which an obtained mean may be expected to differ by chance from the true mean. It is an indication of how well the mean of a sample estimates the mean of a population.

(38). *Statistical power*

The capability of a test to detect a significant effect or how often a correct interpretation can be reached about the effect if it were possible to repeat the test many times.

(39). *Student t-test (t)*

Any statistical hypothesis test in which the test statistic follows a Student's *t* distribution if the null hypothesis is true, for example, a *t* test for paired or independent samples.

(40). *t-distribution*

A statistical distribution describing the means of samples taken from a population with an unknown variance.

(41). *T-score*

A standard score derived from a z-score by multiplying the z-score by 10 and adding 50. It is useful in comparing various test scores to each other as it is a standard metric that reflects the cumulative frequency distribution of the raw scores.

(42). *t-test for correlated means*

A parametric test of statistical significance used to determine whether there is a statistically significant difference between the means of two matched, or non-independent, samples. It is also used for pre–post comparisons.

(43). *t-test for correlated proportions*

A parametric test of statistical significance used to determine whether there is a statistically significant difference between two proportions based on the same sample or otherwise non-independent groups.

(44). *t-test for independent means*

A parametric test of significance used to determine whether there is a statistically significant difference between the means of two independent samples.

(45). *t-test for independent proportions*

A parametric test of statistical significance used to determine whether there is a statistically significant difference between two independent proportions.

(46). *Tukey's test of significance*

A single-step multiple comparison procedure and statistical test generally used in conjunction with an ANOVA to find which means are significantly different from one another. Named after John Tukey, it compares all possible pairs of means and is based on a studentized range distribution *q* (this distribution is similar to the distribution of *t* from the *t*-test).

(47). *Wilcoxon sign rank test (W+)*

A nonparametric statistical hypothesis test for the case of two related samples or repeated measurements on a single sample. It can be used as an alternative to the paired Student's *t*-test when the population cannot be assumed to be normally distributed.

(48). *Z-score*

A score expressed in units of standard deviations from the mean. It is also known as a *standard score.*

(49). *Z-test*

A test of any of a number of hypotheses in inferential statistics that has validity if sample sizes are sufficiently large and the underlying data are normally distributed.

Mean

It is also known as an arithmetic mean, or an average. Is very commonly used in papers, so it is important to have an understanding of how it is calculated. It is one of the simplest statistical concepts to grasp.

It is used when the spread of the data is fairly similar on each side of the mid-point, for example when the data are "*normally distributed*".

The "*normal distribution*" is referred to a lot in statistics. It's the symmetrical, *bell-shaped distribution of data* (Fig. 1).

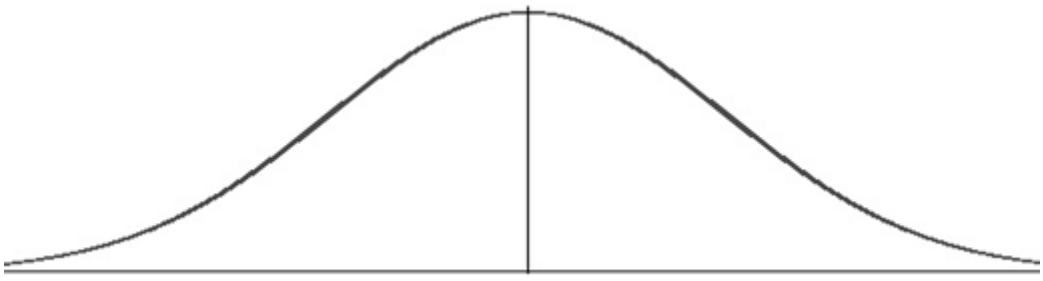


Fig 1. The normal distribution. The centre line shows the mean of the data.

How is it calculated?

The mean is the sum of all the values, divided by the number of values.

Example

Five women in a study on lipid-lowering agents are aged 45, 46, 52, 58 and 59 years.

Add these ages together: 45 + 46 + 52 + 58 + 59 = 260

Now divide by the number of women: 260/5 = 52

So, the mean age is 52 years.

Median

It is also known as Mid-Point. Used in many research papers.

It is used to represent the average when the data are not symmetrical, for instance the *"skewed" distribution (not normal distribution)* (Fig. 2).

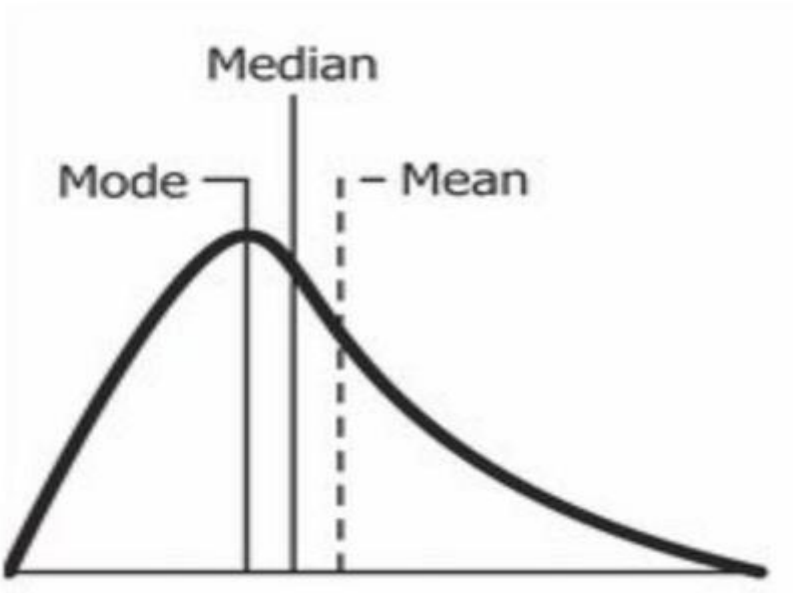


Fig. 2. A skewed distribution. The dotted line shows the median.

How is it calculated?

It is the point, which has half the values above, and half below.

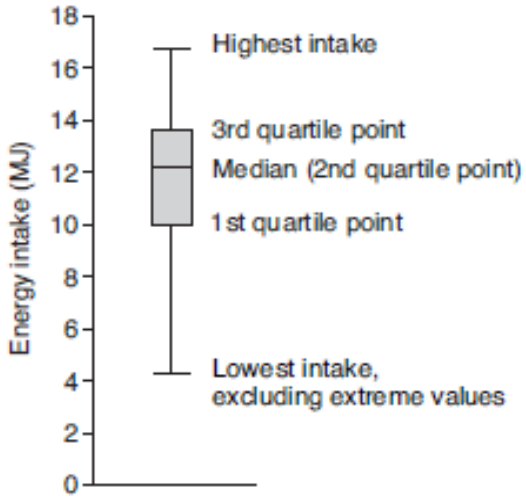
Example 1:

Consider the same example from the one used for mean calculation. There are five patients aged 52, 55, 56, 58 and 59, the median age is 56, the same as the mean – half the women are older, half are younger.

However, in the second example with six patients aged 52, 55, 56, 58, 59 and 92 years, there are two "middle" ages, 56 and 58. The median is half- way between these, i.e., 57 years. This gives a better idea of the mid-point of this skewed data than the mean of 62.

Example 2:

A dietician measured the energy intake over 24 hours of 50 patients on a variety of wards. One ward had two patients that were "nil by mouth". The median was 12.2 megajoules, IQR 9.9 to 13.6. The lowest intake was 0, the highest was 16.7. This distribution is represented by the box and whisker plot.



Box and whisker plot of energy intake of 50 patients over 24 hours. The ends of the whiskers represent the maximum and minimum values, excluding extreme results like those of the two "nil by mouth" patients.

Standard Deviation

This is very important concept, *Standard deviation (SD)* is used for data which are *"normally distributed"***,** to provide information on how much the data vary around their mean.

Interpretation

SD indicates how much a set of values is spread around the average.

A range of one SD above and below the mean (abbreviated to ± 1 SD) includes 68.2% of the values.

±2 SD includes 95.4% of the data.

±3 SD includes 99.7%.

Example 1:

Let us say that a group of patients enrolling for a trial had a normal distribution for weight. The mean weight of the patients was 80 kg. For this group, the SD was calculated to be 5 kg.

1 SD below the average is 80-5 = 75kg.

1 SD above the average is 80+5 = 85kg.

±1 SD will include 68.2% of the subjects, so 68.2% of patients will weigh between 75 and 85 kg.

95.4% will weigh between 70 and 90 kg (±2 SD).

99.7% of patients will weigh between 65 and 95 kg (±3 SD).

This data correlates to the following graph (Fig. 3).

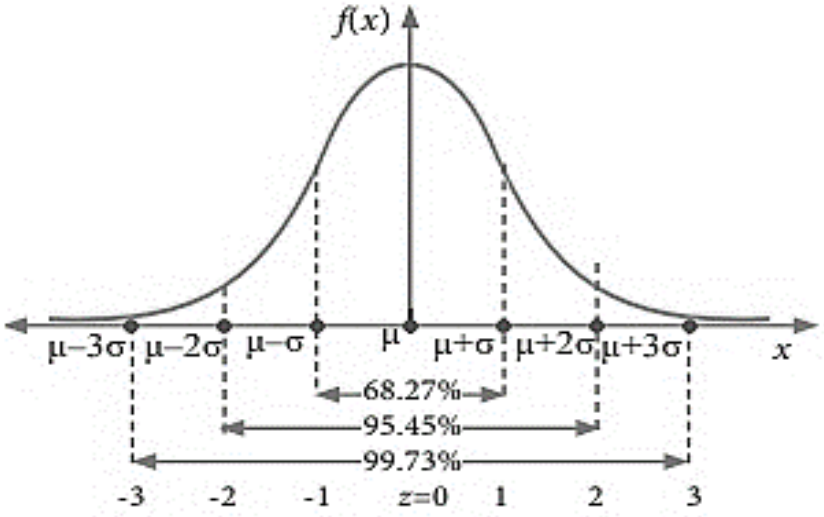


Fig. 3. Graph showing normal distribution of weights of patients enrolling in a trial with mean 80 kg, SD 5 kg.

If we have two sets of data with the same mean but different SDs, then the data set with the larger SD has a wider spread than the data set with the smaller SD.

For example, if another group of patients enrolling for the trial has the same mean weight of 80 kg but an SD of only 3, ±1 SD will include 68.2% of the subjects, so 68.2% of patients will weigh between 77 and 83 kg (Fig. 4).

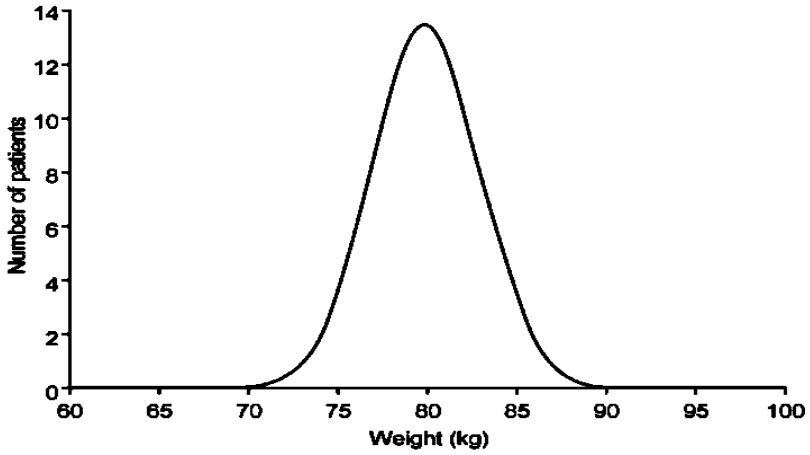


Fig. 4. Graph showing normal distribution of weights of patients enrolling in a trial with mean 80 kg, SD 3 kg.

Example 2:

SD should only be used when the data have a normal distribution. However, means and SDs are often wrongly used for data which are not normally distributed.

A simple check for a normal distribution is to see if 2 SDs away from the mean are still within the possible range for the variable. For example, if we have some length of hospital stay data with a mean stay of 10 days and a SD of 8 days then:

Mean - (2 × SD) = 10 - (2 × 8) = 10 - 16 = -6 days

This is clearly an impossible value for length of stay, so the data cannot be normally distributed. The mean and SDs are therefore not appropriate measures to use.

***p*-value**

The *p* (probability) value is used when we wish to see how likely it is that a hypothesis is true. The hypothesis is usually that there is no difference between two treatments, known as the *"null hypothesis"***.**

Interpretation

The *p*-value gives the probability of any observed difference having happened by chance.

p = 0.5 means that the probability of the difference having happened by chance is 0.5 in 1, or 50:50.

p = 0.05 means that the probability of the difference having happened by chance is 0.05 in 1, i.e., 1 in 20.

It is the figure frequently quoted as being *"statistically significant"*, i.e., unlikely to have happened by chance and therefore important. However, this is an arbitrary figure.

If we look at 20 studies, even if none of the treatments work, one of the studies is likely to have a P value of 0.05 and so appear significant!

The lower the *p*-value, the less likely it is that the difference happened by chance and so the higher the significance of the finding.

p = 0.01 is often considered to be "highly significant". It means that the difference will only have happened by chance 1 in 100 times. This is unlikely, but still possible.

p = 0.001 means the difference will have happened by chance 1 in 1000 times, even less likely, but still just possible. It is usually considered to be "very highly significant".

Example 1:

Out of 50 new babies on average 25 will be girls, sometimes more, sometimes less.

Say there is a new fertility treatment and we want to know whether it affects the chance of having a boy or a girl. Therefore, we set up a null hypothesis - that the treatment does not alter the chance of having a girl. Out of the first 50 babies resulting from the treatment, 15 are girls. We then need to know the probability that this just happened by chance, i.e., did this happen by chance or has the treatment had an effect on the sex of the babies?

The *p*-value gives the probability that the null hypothesis is true.

The *p*-value in this example is 0.007. Do not worry about how it was calculated, concentrate on what it means. It means the result would only have happened by chance in 0.007 in 1 (or 1 in 140) times if the treatment did not actually affect the sex of the baby. This is highly unlikely, so we can reject our hypothesis and conclude that the treatment probably does alter the chance of having a girl.

Example 2:

Patients with minor illnesses were randomized to see either Dr XXXX ended up seeing 176 patients in the study whereas Dr. YYYY saw 200 patients (Table 1**).**

Table 1. Number of patients with minor illnesses seen by two groups

Doctor	**Dr. XXXX (*n* = 200)**	**Dr. YYYY (*n* = 176)**	***p*-value**
Patients satisfied with consultation (%)	186 (93)	168 (95)	0.4
Mean (SD) consultation length (min)	16 (3.1)	6 (2.8)	< 0.001
Patients getting a prescription (%)	58 (29)	76 (43)	0.3
Mean (SD) number of days off work	3.5 (1.3)	3.6 (1.3)	0.8
Patients needing a follow-up appointment (%)	46 (23)	72 (41)	0.05

Caution

The "null hypothesis" is a concept that underlies this and other statistical tests.

The test method assumes (hypothesizes) that there is no (null) difference between the groups. The result of the test either supports or rejects that hypothesis.

The null hypothesis is generally the opposite of what we are actually interested in finding out. If we are interested if there is a difference between two treatments, then the null hypothesis would be that there is no difference and we would try to disprove this.

Try not to confuse statistical significance with clinical relevance. If a study is too small, the results are unlikely to be statistically significant even if the intervention actually works. Conversely a large study may find a statistically significant difference that is too small to have any clinical relevance.

Chi Squared test (χ^2)

Usually written as χ^2 (for the test) or X^2 (for its value); *Chi is pronounced as in sky without the s.*

It is a measure of the difference between actual and expected frequencies.

Interpretation

The *"expected frequency"* is that there is no difference between the sets of results (the null hypothesis). In that case, the X^2 value would be zero.

The larger the actual difference between the sets of results, the greater the X^2 value. However, it is difficult to interpret the X^2 value by itself as it depends on the number of factors studied.

Example

A group of patients with bronchopneumonia were treated with either amoxicillin or erythromycin. The results are shown in Table 2.

Table 2. Type of antibiotic given

Tablet	Amoxicillin (%)	Erythromycin (%)	Total (%)
Improvement at 5 Days	150 (60%)	162 (66%)	314 (63%)
No improvement at 5 Days	102 (40%)	82 (34%)	184 (37%)
Over-all	252 (100%)	244 (100%)	496 (100%)

First, look at the table to get an idea of the differences between the effects of the two treatments. Remember, do not worry about the X^2 value itself, but see whether it is significant. In this case P is 0.14, so the difference in treatments is not statistically significant.

Caution

Instead of the χ^2 test, "*Fisher's exact test*" is sometimes used. Fisher's test is the best choice as it always gives the exact *p*-value, particularly where the numbers are small.

The χ^2 test is simpler for statisticians to calculate but gives only an approximate *p*-value and is inappropriate for small samples. Statisticians may apply "Yates' continuity correction" or other adjustments to the χ^2 test to improve the accuracy of the *p*-value.

Chapter 3. Understanding and use of sensitivity, specificity, and predictive values

Background

We shall discuss the basic knowledge to calculate Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV). We shall also discuss the advantage and limitations of these measures and how we should use these measures in our day-to-day clinical practice. I illustrate below how to calculate sensitivity and specificity, combining two tests and how to apply these results to our day-to-day practice.

Interpretation

Think of any screening test for a disease. For each patient:

(a) the disease itself may be present or absent
(b) the test result may be positive or negative

We need to know how useful the test is.

Table 1 shows 2 × 2 (two-by-two) table

Disease vs Test result	Present	Absent
Positive	A (True Positive)	B (False Positive)
Negative	C (False Negative)	D (True Negative)

Sensitivity

If a patient has the disease, we need to know how often the test will be positive, i.e., "positive in disease".

This is calculated from:

$$\boldsymbol{S = \frac{TP}{TP + FN}}\ \textbf{Or A/(A+C)}$$

This is the rate of pick-up of the disease in a test, and is called the Sensitivity.

Specificity

If the patient is in fact healthy, we want to know how often the test will be negative, i.e., "negative in health".

This is given by,

$$\boldsymbol{S = \frac{TN}{TN + FP}}\ \textbf{Or D/(B+D)}$$

A test can exclude the possibility of the disease, and is known as the Specificity at this rate.

Positive Predictive Value

If the test result is positive, what is the likelihood that the patient will have the condition?

Look at:

$$\boldsymbol{PPV = \frac{TP}{TP + FP}}\ \textbf{Or A/(A+B)}$$

This is known as the Positive Predictive Value (PPV).

Negative Predictive Value

If the test result is negative, what is the likelihood that the patient will be healthy?

Here we use:

$$\boldsymbol{NPV = \frac{TN}{TN + FN}}\ \textbf{Or D/(C+D)}$$

This is known as the Negative Predictive Value (NPV).

In a perfect test, the sensitivity, specificity, PPV and NPV would each have a value of 1. The lower the value (the nearer to zero), the less useful the test is in that respect.

Example 1

Imagine a blood test for gastric cancer, tried out on 100 patients admitted with haematemesis. The actual presence or absence of gastric cancers was diagnosed from endoscopic findings and biopsy. The results are shown.

Disease vs Test Result	Present	Absent
Positive	**A** (True Positive) 20	**B** (False Positive) 30
Negative	**C** (False Negative) 5	**D** (True Negative) 45

Sensitivity = 20/(20+5) = 20/25 = 0.8

If the gastric cancer is present, there is an 80% (0.8) chance of the test picking it up.

Specificity = 45/ (30+45) = 45/75 = 0.6

If there is no gastric cancer, there is a 60% (0.6) chance of the test being negative – but 40% will have a false positive result.

PPV = 20/ (20+30) = 20/50 =0.4

There is a 40% (0.4) chance, if the test is positive, that the patient actually has gastric cancer.

NPV = 45/ (45+5) = 45/50 = 0.9

There is a 90% (0.9) chance, if the test is negative, that the patient does not have gastric cancer. However, there is still a 10% chance of a false negative, i.e. that the patient does have gastric cancer.

Caution

The "Likelihood Ratio" (LR) is the likelihood that the test result would be expected in a patient with the condition compared to the likelihood that that same result would be expected in a patient without the condition.

To calculate the LR, divide the sensitivity by (1 – specificity).

Try using the example above to calculate the LR for a positive result.

LR = Sensitivity/(1-Specificity) = 0.8/(1 – 0.6) = 0.8/0.4 = 2

In this example, LR for a positive result = 2. This means that if the test is positive in a patient, that patient is twice as likely to have gastric cancer than not have it.

One thing you may notice is that in a rare condition, even a diagnostic test with a very high sensitivity may result in a low PPV.

Chapter 4. The importance of statistics in healthcare

The field of **statistics** is concerned with collecting, analysing, interpreting, and presenting data.

In the field of healthcare, statistics is important for the following reasons:

Reason 1: Statistics allows healthcare professionals to monitor the health of individuals using descriptive statistics.

Example: Descriptive statistics are used to *describe* data.

Healthcare professionals often calculate the following descriptive statistics for a given individual:

(a). Mean resting heart rate.
(b). Mean blood pressure.
(c). Fluctuation in weight during a certain time period.

Using these metrics, healthcare professionals can gain a better understanding of the overall health of individuals.

They can then use these metrics to inform individuals on ways they can improve their health or even prescribe specific medications based on the health of the individual.

Reason 2: Statistics allows healthcare professionals to quantify the relationship between variables using regression models.

Example: Another way that statistics is used in healthcare is in the form of regression models.

These are models that allow healthcare professionals to quantify the relationship between one or more predictor variables and a response variable.

For example, a healthcare professional may have access to data on total hours spent exercising per day, total time spent sitting per day, and overall weight of individuals.

They might then build the following multiple linear regression model:

Weight = 124.33 – 15.33(hours spent exercising per day) + 1.04(hours spent sitting per day)

Here's how to interpret the regression coefficients in this model:

(a). For each additional hour spent exercising per day, total weight decreases by an average of 15.33 pounds (assuming hours spent sitting is held constant).
(b). For each additional hour spent sitting per day, total weight increases by an average of 1.04 pounds (assuming hours spent exercising is held constant).

Using this model, a healthcare professional can quickly understand that more time spent exercising is associated with lower weight and more time spent sitting is associated with higher weight.

They can also quantify exactly how much exercise and sitting affect weight.

Reason 3: Statistics allows healthcare professionals to compare the effectiveness of different medical procedures using hypothesis tests.

Example: Another way that statistics is used in healthcare is in the form of hypothesis tests.

These are tests that healthcare professionals can use to determine if there is a statistical significance between different medical procedures or treatments.

For example, suppose a doctor believes that a new drug is able to reduce blood pressure in obese patients. To test this, he may measure the blood pressure of 40 patients before and after using the new drug for one month.

He then performs a paired samples t- test using the following hypotheses:

(a). $\mathbf{H_0}$**:** $\mu_{after} = \mu_{before}$ (the mean blood pressure is the same before and after using the drug)
(b). $\mathbf{H_A}$**:** $\mu_{after} < \mu_{before}$ (the mean blood pressure is less after using the drug)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then he can reject the null hypothesis and conclude that the new drug leads to reduced blood pressure.

Note: This is just one example of a hypothesis test that is used in healthcare. Other common tests include a one sample t-test, two sample t-test, one-way ANOVA, and two-way ANOVA.

Reason 4: Statistics allows healthcare professionals to understand the effect of lifestyle choices on health using incidence rate ratio.

Example: An **incidence rate ratio** allows healthcare professionals to compare the incident rate between two different groups.

For example, suppose it's known that people who smoke develop lung cancer at a rate of 7 per 100 person-years.

Conversely, suppose it's known that people who do not smoke develop lung cancer at a rate of 1.5 per 100 person-years.

We would calculate the incidence rate ratio (often abbreviated IRR) as:

(a). IRR = Incidence rate among smokers / Incidence rate among non-smokers
(b). IRR = (7/100) / (1.5/100)
(c). IRR = **4.67**

Here is how a healthcare professional would interpret this value: The lung cancer rate among smokers is 4.67 times as high as the rate among non-smokers.

Using this simple calculation, healthcare professionals can gain a good understanding of how different lifestyle choices (like smoking) affect health in individuals.

Chapter 5. Statistical type I and type II errors

In statistics, a **Type I error** is a false positive conclusion, while a **Type II error** is a false negative conclusion.

Making a statistical decision always involves uncertainties, so the risks of making these errors are unavoidable in hypothesis testing.

The probability of making a Type I error is the significance level, or alpha (*a*), while the probability of making a Type II error is beta (*β*). These risks can be minimized through careful planning in your study design.

Type I error

Consider we are testing two brands of paracetamol to evaluate if Brand 1 is better in curing subject's suffering from fever as compared to Brand 2. As both brands contain paracetamol, it is expected that the effect of both brands is similar. Let us try to build a statistical hypothesis around this

Null hypothesis (H_0): Brand 1 is equal to Brand 2

Alternate Hypothesis (H_1): Brand 1 is better than Brand 2

Let us try to evaluate error that can occur.

Error 1: Based on analysis it is concluded that Brand 1 is better than Brand 2, basically we reject H_0. Knowing that Brand 1 is equal to Brand 2 (H_0), we are making an error here by rejecting H_0. This is called as Type I error. Statistically it is defined as

Type I error = (Reject H_0/H_0 is true).

Probability of Type I error is called as level of significance and is denoted as α.

Example: Statistical significance and Type I error

In your clinical study, you compare the symptoms of patients who received the new drug intervention or a control treatment. Using a *t* test, you obtain a *p* value of 0.035. This *p* value is

lower than your alpha of 0.05, so you consider your results statistically significant and reject the null hypothesis.

However, the *p* value means that there is a 3.5% chance of your results occurring if the null hypothesis is true. Therefore, there is still a risk of making a Type I error.

Type II error

Consider we are testing paracetamol against placebo to evaluate if paracetamol is better in curing subject's suffering from fever as compared to placebo. (It is expected that the effect of paracetamol is better than placebo). Let us try to build a statistical hypothesis around this

Null hypothesis (H_0): paracetamol is equal to placebo

Alternate hypothesis (H_1): paracetamol is better than placebo

Let us try to evaluate the error that can happen.

Error 2: If analysis concludes that paracetamol is equal to placebo, we accept H_0. Knowing that paracetamol is better than placebo (H_1) we are making an error here by accepting H_0. This is called as Type II error. Statistically it is defined as

Type II error = (Accept H_0/H_1 is true).

Probability of type II error is denoted as β.

In above case if analysis concludes that paracetamol is better than placebo, we reject H_0, which would be correct decision. Probability of such a decision taking place is called as "Power".

Power = Probability (Reject H_0/H_1 is true) which is actually 1-β.

We can tabulate type I and type II error as

Decision Take/Actual Fact	H_o is True	H_1 is True
Reject H_o	Type I Error	No Error
Accept H_o	No Error	Type II Error

Example: Statistical power and Type II error

When preparing your clinical study, you complete a power analysis and determine that with your sample size, you have an 80% chance of detecting an effect size of 20% or greater. An effect size of 20% means that the drug intervention reduces symptoms by 20% more than the control treatment.

However, a Type II may occur if an effect that's smaller than this size. A smaller effect size is unlikely to be detected in your study due to inadequate statistical power.

Chapter 6. Statistics – data collection – case study

Case study research is a qualitative research method that is used to examine contemporary real-life situations and apply the findings of the case to the problem under study. Case studies involve a detailed contextual analysis of a limited number of events or conditions and their relationships. It provides the basis for the application of ideas and extension of methods. It helps a researcher to understand a complex issue or object and add strength to what is already known through previous research.

Steps of Case Study Method

In order to ensure objectivity and clarity, a researcher should adopt a methodical approach to case studies research. The following steps can be followed:

(1) **Identify and define the research questions**: The researcher starts with establishing the focus of the study by identifying the research object and the problem surrounding it. The research object would be a person, a program, an event or an entity.

(2) **Select the cases**: In this step the researcher decides on the number of cases to choose (single or multiple), the type of cases to choose (unique or typical) and the approach to collect, store and analyze the data. This is the design phase of the case study method.

(3) **Collect the data**: The researcher now collects the data with the objective of gathering multiple sources of evidence with reference to the problem under study. This evidence is stored comprehensively and systematically in a format that can be referenced and sorted easily so that converging lines of inquiry and patterns can be uncovered.

(4) **Evaluate and analyze the data**: In this step the researcher makes use of varied methods to analyze qualitative as well as quantitative data. The data is categorized, tabulated and cross checked to address the initial propositions or purpose of the study. Graphic techniques like placing information into arrays, creating matrices of categories, creating flow charts etc. are used to help the investigators to approach the data from different ways and thus avoid making premature conclusions. Multiple investigators may also be used to examine the data so that a wide variety of insights to the available data can be developed.

(5) **Presentation of Results**: The results are presented in a manner that allows the reader to evaluate the findings in the light of the evidence presented in the report. The results are corroborated with sufficient evidence showing that all aspects of the problem have been

adequately explored. The newer insights gained and the conflicting propositions that have emerged are suitably highlighted in the report.

Chapter 7. Statistical significance

Statistical Significance signifies that result of a statistical experiment or testing is not occurring randomly and is attributable to certain cause. Statistical significance of a result could be strong or weak and it is very important for sectors, which are heavily dependent on research works like insurance, pharma, finance, physics and so.

Statistical Significance helps in choosing the sample data so that one can judge the result or outcome of testing to be realistic and not be caused by a random cause.

Statisticians generally formulates the degree of statistical significance by sampling error. Generally, sampling error of 5% is acceptable. Sample size is also important, as it should be representative sample instead of very large sample considering the fact that large samples are prone to errors.

Significance Level

A level at which an event is considered to be statistical significant is termed as significance level. Statisticians uses a test statistic called p-value to get the statistical significance. If p-value of an event falls below a particular level then the event is considered as statistical significant. P-value is function of standard deviations and means of data samples. P-value is the probability of an event, which certifies that result of statistical testing is occurring by chance, or due to some sampling error. In other words, it is the risk of failure of a statistical test. Opposite of p-value is confidence level, which is 1 - p-value.

If p-value of a result is 10% then that means CL of the result is 90% Z-value = 1.645

If p-value of a result is 5% then that means CL of the result is 95% Z-value = 1.960

If p-value of a result is 2% then that means CL of the result is 98% Z-value = 2.33

If p-value of a result is 1% then that means CL of the result is 99% Z-value = 2.58

An example of findings reported with p values are below:

Statement: Drug 24 reduced patients' symptoms compared to Drug 23. Patients who received Drug 24 (n=100) were 2.1 times less likely than patients who received Drug 23 (n = 100) to experience symptoms of Disease A, $p<0.05$.

Or

Statement: Individuals who were prescribed Drug 24 experienced fewer symptoms (M = 1.4, SD = 0.8) compared to individuals who were prescribed Drug 23 (M = 5.4, SD = 1.9). This finding was statistically significant, $p= 0.02$.

Chapter 8. Statistical risk ratio data analysis

The relative risk (or risk ratio) is an intuitive way to compare the risks for the two groups. Simply divide the cumulative incidence in exposed group by the cumulative incidence in the unexposed group:

$$\textbf{Risk Ratio} = \frac{CI_e}{CI_u}$$

where, CI_e is the cumulative incidence in the 'exposed' group and CI_u is the cumulative incidence in the 'unexposed' group.

Risk is a relatively intuitive concept that we encounter every day, but interpretation of risk (especially low risk) is often inconsistent.

The risk of death while travelling to the shops to buy a lotto ticket can be higher than the risk of winning the jackpot!

Relative risk is used in "cohort studies", prospective studies that follow a group (cohort) over a period of time and investigate the effect of a treatment or risk factor.

Interpretation

Risk is the probability that an event will happen. It is calculated by dividing the number of events by the number of people at risk.

One boy is born for every two births, so the probability (risk) of giving birth to a boy is

1/2 = 0.5

If one in every 100 patients suffers a side-effect from a treatment, the risk is

1/100 = 0.01

Risk ratios are calculated by dividing the risk in the treated or exposed group by the risk in the control or unexposed group.

A risk ratio of one indicates no difference in risk between the groups.

If the risk ratio of an event is >1, the rate of that event is increased compared to controls.

If < 1, the rate of that event is reduced.

Risk ratios are frequently given with their 95% CIs – if the CI for a risk ratio does not include one (no difference in risk), it is statistically significant.

Risk ratio < 1

It is also possible for the risk ratio to be less than 1; this would suggest that the exposure being considered is associated with a reduction in risk. A randomized clinical trial was begun in order to test whether low-dose aspirin was beneficial in reducing myocardial infarctions (heart attacks). The study population consisted of over 22,000 male physicians who were randomly assigned to either low-dose aspirin or a placebo (an identical looking pill that was inert). They followed these physicians for about five years. Some of the data is summarized in the 2 × 2 table shown below.

Treatment	Myocardial Infarction	No Infarction	Total	Cumulative Incidence
Aspirin	139	10,898	11,037	139/11,037 = 0.0126
Placebo	239	10,795	11,034	239/11,034 = 0.0217

$$\text{Risk Ratio} = \frac{0.0126}{0.0217} = 0.58$$

Note that the "exposure" of interest was low-dose aspirin, and the aspirin group is summarized in the top row. The group assigned to take aspirin had an incidence of 1.26%, while the placebo (unexposed) group had an incidence of about 2.17%. The cumulative incidence in the aspirin group was divided by the cumulative incidence in the placebo group, and RR = 0.58. An appropriate interpretation of this would be:

Those who take low dose aspirin regularly have ***0.58 times the risk*** of myocardial infarction compared to those who do not take aspirin.

Note also that the unexposed (comparison, reference) group must be specified. For example, if we simply said, "Those who take low dose aspirin regularly have *0.58 times the risk* of myocardial infarction", the question is "compared to what?" Is it those who didn't take any aspirin, those who took low-dose aspirin but used it irregularly, those who took high dose aspirin, those who took acetaminophen...?

In general:

(1). If the risk ratio is 1 (or close to 1), it suggests no difference or little difference in risk (incidence in each group is the same).

(2). A risk ratio > 1 suggests an increased risk of that outcome in the exposed group.

(3). A risk ratio < 1 suggests a reduced risk in the exposed group.

Chapter 9. Significance testing of correlation coefficient

Where there is a linear relationship between two variables there is said to be a correlation between them. Examples are height and weight in children, or socio-economic class and mortality.

The strength of that relationship is given by the "*correlation coefficient*".

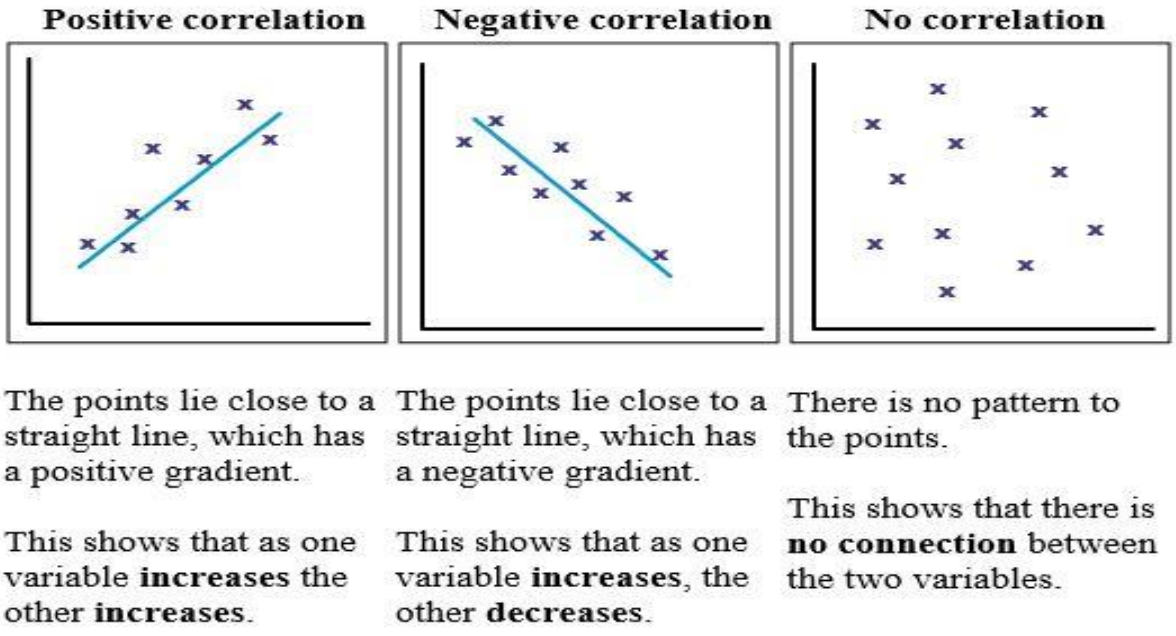
Interpretation

The correlation coefficient is usually denoted by the letter **"*r*"** for example $r = 0.8$.

SPSS 26.0 Software – Analyze – Correlate – Bivariate.

A **positive correlation** coefficient means that as one variable is increasing the value for the other variable is also increasing – the line on the graph slopes up from left to right. Height and weight have a positive correlation: children get heavier as they grow taller.

A **negative correlation** coefficient means that as the value of one variable goes up the value for the other variable goes down – the graph slopes down from left to right. Higher socio-economic class is associated with a lower mortality, giving a negative correlation between the two variables.



If there is a perfect relationship between the two variables, then $r = 1$ (if a positive correlation) or $r = -1$ (if a negative correlation).

If there is no correlation at all (the points on the graph are completely randomly scattered) then *r* = 0.

The following is a good rule of thumb when considering the size of a correlation:

***r* = 0–0.2:** very low and probably meaningless.

***r* = 0.2–0.4:** a low correlation that might warrant further investigation.

***r* = 0.4–0.6:** a reasonable correlation.

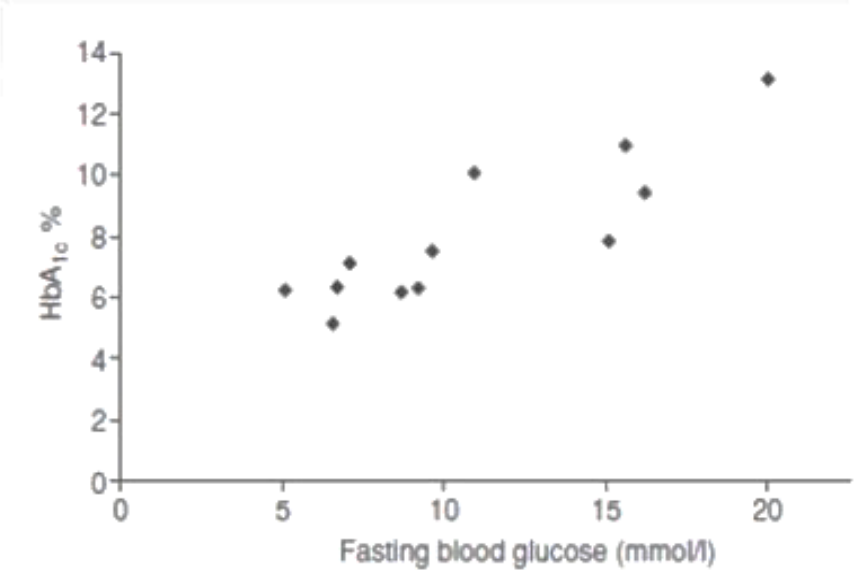
***r* = 0.6–0.8:** a high correlation.

***r* = 0.8–1:** a very high correlation. Possibly too high! Check for errors or other reasons for such a high correlation.

The same applies to negative correlations too.

Example 1

A nurse wanted to be able to predict the laboratory HbA1c result (a measure of blood glucose control) from the fasting blood glucoses which she measured in her clinic. On 12 consecutive diabetic patients she noted the fasting glucose and simultaneously drew blood for HbA1c. She compared the pairs of measurements and drew the graph.

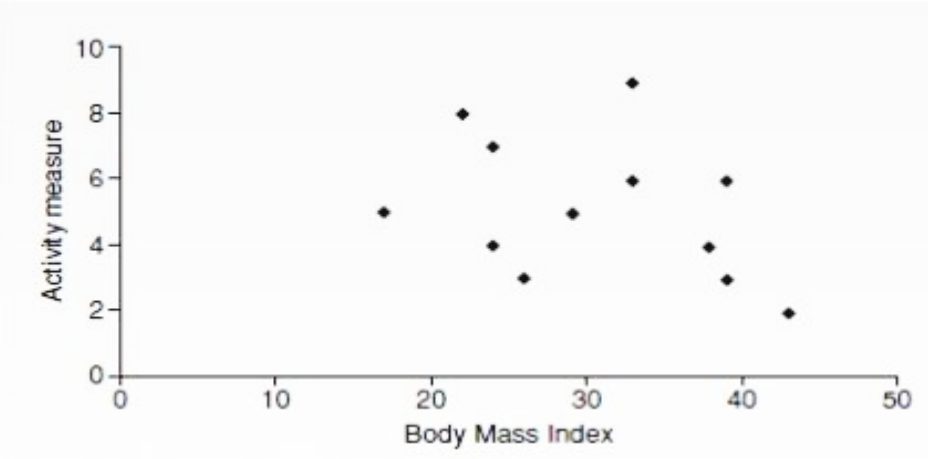


For these results *r* = 0.88, showing a very high correlation.

A graph like this is known as a **"scatter plot"**.

Example 2

An occupational therapist developed a scale for measuring physical activity and wondered how much it correlated to Body Mass Index (BMI) in 12 of her adult patients.



In this example, *r* = –0.34, indicating a low correlation.

The fact that the *r* value is negative shows that the correlation is negative, indicating that patients with a higher level of physical activity tended to have a lower BMI.

Caution

Correlation tells us how strong the association between the variables is, but does not tell us about cause and effect in that relationship.

The **"Pearson correlation coefficient"**, Pearson's *r*, is used if the values are sampled from "normal" populations. Otherwise the **"Spearman correlation coefficient"** is used. However, the interpretation of the two is the same.

Where the author shows the graph, you can get a good idea from the scatter as to how strong the relationship is without needing to know the *r* value.

It is very easy for authors to compare a large number of variables using correlation and only present the ones that happen to be significant. So, check to make sure there is a plausible explanation for any significant correlations.

Also bear in mind that a correlation only tells us about linear (straight line) relationships between variables. Two variables may be strongly related but not in a straight line, giving a low correlation coefficient.

Chapter 10. Statistical regression analysis

Definition

The Regression Analysis is a technique of studying the dependence of one variable (called dependent variable), on one or more variables (called explanatory variable), with a view to estimate or predict the average value of the dependent variable in terms of the known or fixed values of the independent variables.

Regression analysis is used to find how one set of data relates to another.

This can be particularly helpful where we want to use one measure as a proxy for another – for example, a near-patient test as a proxy for a lab test.

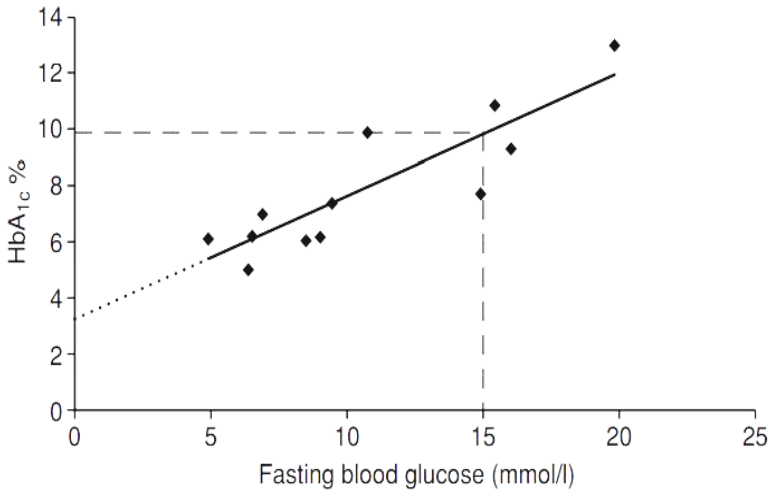
Interpretation

A regression line is the "best fit" line through the data points on a graph.

The regression coefficient gives the "slope" of the graph, in that it gives the change in value of one outcome, per unit change in the other.

Example

Consider the graph shown in previous section. A statistician calculated the line that gave the "best fit" through the scatter of points



The line is called a "regression line".

To predict the HbA1c for a given blood glucose the nurse could simply plot it on the graph, as here where a fasting glucose of 15 predicts an HbA1c of 9.95.

This can also be done mathematically. The slope and position of the regression line can be represented by the "regression equation":

HbA1c = 3.2 + (0.45 × blood glucose).

The 0.45 figure gives the slope of the graph and is called the "regression coefficient".

The "regression constant" that gives the position of the line on the graph is 3.2: it is the point where the line crosses the vertical axis.

Try this with a glucose of 15:

HbA1c =3.2+(0.45x15) = 3.2+6.75=9.95

This regression equation can be applied to any regression line. It is represented by:

y = a + bx

To predict the value y (value on the vertical axis of the graph) from the value x (on the horizontal axis), b is the regression coefficient and a is the constant.

Other types of regression

The example above is a **"linear regression",** as the line that best fits the points is straight. Other forms of regression include:

Logistic regression: This is used where each case in the sample can only belong to one of two groups (e.g. having disease or not) with the outcome as the probability that a case belongs to one group rather than the other.

Poisson Regression: It is mainly used to study waiting times or time between rare events.

Cox proportional hazards regression model: It is used in survival analysis where the outcome is time until a certain event.

Caution

Regression should not be used to make predictions outside of the range of the original data. In the example above, we can only make predictions from blood glucoses which are between 5 and 20.

Regression vs correlation

Regression and correlation are easily confused.

Correlation measures the strength of the association between variables.

Regression quantifies the association. It should only be used if one of the variables is thought to precede or cause the other.

Chapter 11. Statistical hypothesis - using the t-test

The t-test is a statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis.

A t-test is most commonly applied when the test statistic follows a normal distribution if the value of a scaling term in the test statistic is known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics (under certain conditions) follow a Student's t distribution.

The t-test can be used, for example, to determine if the means of two sets of data are significantly different from each other.

Among the most frequently used *t*-tests are:

(1) A **one-sample** location test of whether the mean of a population has a value specified in a null hypothesis.
(2) A **two-sample** location test of the null hypothesis such that the means of two populations are equal. All such tests are usually called Student's t-tests, though strictly speaking that name should only be used if the variances of the two populations are also assumed to be equal; the form of the test used when this assumption is dropped is sometimes called Welch's t-test. These tests are often referred to as **unpaired or independent samples t-tests,** as they are typically applied when the statistical units underlying the two samples being compared are non-overlapping.

Parametric statistics are used to compare samples of "normally distributed" data. If the data do not follow a normal distribution, these tests should not be used.

Interpretation

A parametric test is any test which requires the data to follow a specific distribution, usually a normal distribution. Common parametric tests we come across are the t-test and the χ^2 test.

Analysis of variance (ANOVA)

This is a group of statistical techniques used to compare the means of two or more samples to see whether they come from the same population - the "null hypothesis". These techniques can also allow for independent variables which may have an effect on the outcome.

t-test (also known as Student's t)

t-tests are typically used to compare just two samples. They test the probability that the samples come from a population with the same mean value.

χ^2 test

A frequently used parametric test is the χ^2 test. It is discussed later.

Example

Two hundred adults seeing an asthma nurse specialist were randomly assigned to either a new type of bronchodilator or placebo.

After 3 months the peak flow rates in the treatment group had increased by a mean of 96 l/min (SD 58), and in the placebo group by 70 l/min (SD 52). The null hypothesis is that there is no difference between the bronchodilator and the placebo.

The t statistic is 11.14, resulting in a P value of 0.001. It is therefore very unlikely (1 in 1000 chance) that the null hypothesis is correct so we reject the hypothesis and conclude that the new bronchodilator is significantly better than the placebo.

(Here we are not bothered how we get these values. It is beyond the scope of this app!)

Caution

Parametric tests should only be used when the data follow a “normal” distribution. You may find reference to the “Kolmogorov Smirnov” test. This tests the hypothesis that the collected data are from a normal distribution and therefore assesses whether parametric statistics can be used.

Sometimes authors will say that they have "transformed" data and then analyzed it with a parametric test. This is quite legitimate – it is not cheating! For example, a skewed distribution might become normally distributed if the logarithm of the values is used.

Chapter 12. Statistics – Mcnemar test

In statistics, McNemar's test is a statistical test used on paired nominal data. It is applied to 2 × 2 contingency tables with a dichotomous trait, with matched pairs of subjects, to determine whether the row and column marginal frequencies are equal. It is named after Quinn McNemar, who introduced it in 1947.

McNemar test is utilized for two related examples as a part of circumstances where the states of mind of individuals are noted previously, then after the fact treatment to test the essentialness of progress in sentiment if any.

The McNemar test is especially helpful when the information speaks the truth two related samples. For the most part this information is utilized as a part of circumstances where the states of mind of individuals are noted before overseeing the treatment and are then contrasted and investigations in the wake of managing the treatment. It can along these lines be said that utilizing McNemar test we can judge if there is any adjustment in the demeanours or supposition of individuals subsequent to regulating the treatment with the utilization of table as demonstrated as follows:

Before Treatment	**After Treatment**	
Favour	A	B
Do Not Favour	C	D

As can be seen C and B don't change their supposition and show 'Do Not Favour' and 'Favour' individually even after the treatment has been administered. However, A which was good before treatment demonstrates a 'Do Not Favour' reaction after treatment and vice versa for D. It can hence be said that A+DA+D shows change in individuals' reaction.

The null hypothesis for McNemar test is that (A+D)2(A+D)2 cases change in one direction and the same proportion of change takes place in other direction.

McNemar test statistic uses a transformed test model as follows:

$x^2=(|A-D|-1)^2/A+D)$ (Degree of freedom = 1.)

Acceptance Criteria: If the calculated value is less than the table value, accept null hypothesis.

Rejection Criteria: If the calculated value is more than table value then null hypothesis is rejected.

Illustration

In a before and after experiment the responses obtained from 300 respondents were classified as follows:

Before Treatment	After Treatment	
Favour	A = 60	B = 90
Do Not Favour	C = 120	D = 30

Test at 5% significance level, using McNemar test if there is any significant difference in the opinion of people after the treatment.

Solution:

HoHo: There is no difference in the opinion of people even after the experiment.

The test statistic is calculated using the formula:

$x^2=(|A-D|-1)^2(A+D)$

$(|60-30|-1)^2(60+30) = 9.34 \times 2$

$(|A-D|-1)^2/(A+D) = (|60-30|-1)^2/(60+30) = (29)^2/90 = 841/90 = 9.34$

The value of test at 5% significance level for 1 D.F. is 3.84. Since the test is greater than the table value, the null hypothesis is rejected i.e. the opinion of people has changed after the treatment.

Chapter 13. Analysis of variance Two-Way ANOVA

The two-way ANOVA is an extension of the one-way ANOVA. The "two-way" comes because each item is classified in two ways, as opposed to one way. For example, one-way classifications might be gender, political party, religion, or race. Two-way classifications might be by gender and political party, gender, and race, or religion and race.

Each classification variable is called a factor and so there are two factors, each having several levels within that factor. The factors are called the "row factor" and the "column factor" because the data is usually arranged in table format. Each combination of a row-level and a column-level is called a treatment.

The two-way ANOVA that we are going to discuss requires a balanced design. The balanced design is where each treatment has the same sample size.

Example 1

A pharmaceutical company is testing a new drug to see if it helps reduce the time to recover from a fever. They decide to test the drug on three different races (Caucasian, African American, and Hispanic) and both genders (male and female). This makes six treatments (3 races × 2 genders = 6 treatments). They randomly select five test subjects from each of those six treatments, so all together, they have 3 × 2 × 5 = 30 test subjects. The response variable is the time in minutes after taking the medicine before the fever is reduced. The data might look something like this.

Data

Category	Male	Female
Caucasian	54, 49, 59, 39, 55	25, 29, 47, 26, 28
African American	53, 72, 43, 56, 52	46, 51, 33, 47, 41
Hispanic	33, 30, 26, 25, 29	18, 21, 34, 40, 24

Example 2

The following data represent clotting times (mins) of plasma from eight subjects treated in four different ways. The eight subjects (blocks) were allocated at random to each of the four treatment groups.

Treatment 1	Treatment 2	Treatment 3	Treatment 4
8.4	9.4	9.8	12.2
12.8	15.2	12.9	14.4
9.6	9.1	11.2	9.8
9.8	8.8	9.9	12.0
8.4	8.2	8.5	8.5
8.6	9.9	9.8	10.9
8.9	9.0	9.2	10.4
7.9	8.1	8.2	10.0

Solution

Variables: Treatment 1, Treatment 2, Treatment 3, Treatment 4

Source of Variation	Sum Squares	Mean Square
Between blocks (rows)	78.98	11.28
Between treatments (columns)	13.01	4.33
Residual (error)	13.77	0.655
Corrected total	105.77	

F (VR between blocks) = 17.20 P < 0.0001

F (VR between treatments) = 6.615 P = 0.0025

Here we can see that there was a statistically highly significant difference between mean clotting times across the groups. The difference between subjects is of no particular interest here

Chapter 14. Statistical non-parametric Mann -Whitney U test

Non-parametric statistics are used when the data are **not normally distributed** and so are not appropriate for "parametric" tests.

Interpretation

Rather than comparing the values of the raw data, statisticians "rank" the data and compare the ranks.

Example

Mann–Whitney U Test

The Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed.

A GP introduced a nurse triage system into her practice. She was interested in finding out whether the age of the patients attending for triage appointments was different to that of patients who made emergency appointments with the GP.

Six hundred and forty-six patients saw the triage nurse and 532 patients saw the GP. The median age of the triaged patients was 50 years (1st quartile 40 years, 3rd quartile 54), for the GP it was 46 (22, 58).

Note how the quartiles show an uneven distribution around the median, so the data cannot be normally distributed and a non-parametric test is appropriate.

The statistician used a "Mann–Whitney U test" to test the hypothesis that there is no difference between the ages of the two groups.

This gave a Mann-Whitney U value of 133 Wilcoxon W 200 with a P value of < 0.001. Ignore the actual U value but concentrate on the P value, which in this case suggests that the triage nurse's patients were very highly significantly older than those who saw the GP.

Caution

The "Wilcoxon signed rank test", "Kruskal Wallis" and "Friedman" tests are other non-parametric tests. Do not be put off by the names – go straight to the P value.

Chapter 15. Definitions of probability

There is a number of definitions of probability proposed by many authors.

However, the majority of definitions can be divided into three groups:

(1) Definitions of probability as a quantitative measure of the **'degree of certainty'** of the observer of the experiment.
(2) Definitions that reduce the concept of probability to the more primitive notion of **'equal likelihood'**.
(3) Definitions that take as their point of departure the **'relative frequency'** of occurrence of the event in a large number of trials.

The classical definition of probability

Uses the sample space to determine the numerical probability that an event will happen, also called theoretical probability. The probability P(A) of an event A is equal to the number of possible simple events (outcomes) favourable to A divided by the total number of possible simple events of the experiment, i.e., P(A) = $\frac{m}{n}$

where, m = number of simple events into which the event A can be decomposed.

The classical definition of probability reduces the concept of probability to the concept of equiprobability of events, which is regarded as a primitive concept and hence not subject to the formal definition.

Statistical definition of probability

First of all, the question arises in a majority of cases, as to a reasonable way of selecting the 'equally likely cases'.

Lengthy observations as to the occurrence or non-occurrence of an event A in a large number of repeated trials under the same set of conditions show that for a wide class of phenomena, the number of occurrences or non-occurrences of the event A is subject to a stable law; then it turns out that for sufficiently large N the ratio m/N in most of such series of observations, assumes an almost constant value. Since this constant is an objective numerical characteristic of the

phenomena, it is natural to call it the statistical probability of the random event A under investigation.

"The probability of an event A can be approximated by the proportion of times that A occurs when the experiment is repeated a very large number of times."

Chapter 16. Conditional probability

It is the probability of some event A, given the occurrence of some other event B. Conditional probability is written P(A|B), and is read as "the probability of A, given B".

The probability that event B occurs, given that event A has already occurred is

P(B|A) = P(A and B)/P(A).

This formula comes from the general multiplication principle and a little bit of algebra.

Since we are given that event A has occurred, we have a reduced sample space. Instead of the entire sample space S, we now have a sample space of A since we know A has occurred. If you then divided the numerator and denominator of the right-hand side by the number in the sample space S, then you have the probability of A and B divided by the probability of A.

Example 1

The question, "Do you smoke?" was asked of 100 people. Results are shown in the table.

	Yes	**No**	**Total**
Male	19	41	**60**
Female	12	28	**40**
Total	**31**	**69**	**100**

What is the probability of a randomly selected individual being a male who smokes? This is just a joint probability.

The number of "Male and Smoke" divided by the total = 19/100 = 0.19

What is the probability of a randomly selected individual being a male? This is the total for male divided by the total = 60/100 = 0.60. Since no mention is made of smoking or not smoking, it includes all the cases.

What is the probability of a randomly selected individual smoking? Again, since no mention is made of gender, this is a marginal probability, the total who smoke divided by the total = 31/100 = 0.31.

What is the probability of a randomly selected male smoking? This time, you're told that you have a male - think of stratified sampling. What is the probability that the male smokes? Well, 19 males smoke out of 60 males, so 19/60= 0.31666.

What is the probability that a randomly selected smoker is male? This time, you're told that you have a smoker and asked to find the probability that the smoker is also male. There are 19 male smokers out of 31 total smokers, so 19/60= 0.6129.

Chapter 17. Probability distribution of Bernoulli trials

A Bernoulli experiment is a random experiment, the outcome of which can be classified in one of two mutually exclusive and exhaustive ways, mainly, **Success or Failure** (e.g., female or male, life or death, non-defective or defective).

So, it is also called a Binomial trial!

A sequence of Bernoulli trials occurs when a Bernoulli experiment is performed several independent times so that the probability of success, say, p, remains the same from trial to trial. That is, in such a sequence we let p denote the probability of success on each trial. In addition, frequently $q=1-p$ denote the probability of failure; that is, we shall use q and $1-p$ interchangeably.

Let X be a random variable associated with Bernoulli trial by defining it as follows:

X (success) = 1 and X (failure) = 0. That is the two outcomes, success and failure, is denoted by one and zero, respectively. It can be written as $f(x) = p^x(1-p)^{1-x}$ and we say that X has a Bernoulli distribution.

Properties

(1) A Bernoulli (Success-Failure) experiment is performed n times.
(2) The trials are independent.
(3) The probability of success on each trial is a constant p; the probability of failure is $q=1-p$.
(4) The random variable X counts the number of successes in the n trials.

A binomial distribution will be denoted by the symbol b(n,p) and we say that the distribution of X is b(n,p). The constants n and p are called the parameters of the binomial distribution, they correspond to the number n of independent trials and the probability p of success on each trial.

A random variable X will have Bernoulli distribution with probability p if its probability distribution is

$P(X = x) = p^x (1 - p)^{1-x}$, for x = 0, 1 and $P(X = x) = 0$ for other values of x.

Here, 0 is failure and 1 is the success.

Example 1: In a hospital, a particular type of operation has 95% chance of success. On a given day, 10 operations were performed. This can be viewed as 10 Bernoulli trials, with each trial having a probability of success $p=0.95$, and a probability of failure $1-p=0.05$.

Chapter 18. Statistical Independent Events and Probability

Definition

The independent t-test, also called the two-sample t-test, independent-samples t-test or student's t-test, is an inferential statistical test that determines whether there is a statistically significant difference between the means in two unrelated groups.

In probability, two events are said to be independent if the probability of one is not affected by the occurrence or non-occurrence of the other. This definition requires further explanation, so consider the following example.

Earlier in this module we considered data from a population of **N=100** men who had both a PSA test and a biopsy for prostate cancer. Suppose we have a different test for prostate cancer. This prostate test produces a numerical risk that classifies a man as at low, moderate or high risk for prostate cancer. A sample of 100 men underwent the new test and had a biopsy. The data from the biopsy results are summarized below.

Prostate Test Risk	Prostate Cancer	No Prostate Cancer	Total
Low	10	50	60
Moderate	6	30	36
High	4	20	24
Over All	20	100	120

1. The probability that a man has prostate cancer given he has a low risk is P (Prostate Cancer | Low Risk) = 10/60 = 0.167.
2. The probability that a man has prostate cancer given he has a moderate risk is P (Prostate Cancer | Moderate Risk) = 6/36 = 0.167.
3. The probability that a man has prostate cancer given he has a high risk is P (Prostate Cancer | High Risk) = 4/24 = 0.167.

Note that regardless of whether the hypothetical Prostate Test was low, moderate, or high, the probability that a subject had cancer was 0.167. In other words, knowing a man's prostate test result does not affect the likelihood that he has prostate cancer in this example. In this case, the probability that a man has prostate cancer is independent of his prostate test result.

Chapter 19. Types of sampling methods in statistics

Probability sampling strategies typically use a random or chance process, although there are important exceptions to this rule. Random sampling is a strategy for selecting study participants in which each and every person has an equal and independent chance of being selected. What does it mean to be independent? The researchers select each person for the study separately.

The types of sampling roughly fall into two main categories - probability sampling and judgment sampling. Most sampling theory has been developed for probability sampling and we will later consider the detailed theory of some of these methods.

Simple Random Sampling

Simple random sampling is the most straightforward of the random sampling strategies. We use this strategy when we believe that the population is relatively homogeneous for the characteristic of interest.

For example, let's say you were surveying first-time parents about their attitudes toward mandatory seat belt laws. You might expect that their status as new parents might lead to similar concerns about safety. On campus, those who share a major might also have similar interests and values; we might expect psychology majors to share concerns about access to mental health services on campus.

Stratified Random Sampling

Stratified random sampling is used when we have subgroups in our population that are likely to differ substantially in their responses or behaviour. This sampling technique treats the population as though it were two or more separate populations and then randomly samples within each.

For example, you are interested in visual-spatial reasoning and previous research suggests that men and women will perform differently on these types of task. Therefore, you divide your sample into male and female members and randomly select equal numbers within each subgroup (or "stratum"). With this technique, you are guaranteed to have enough of each subgroup for meaningful analysis.

Cluster Sampling

Cluster sampling is useful when it would be impossible or impractical to identify every person in the sample. Suppose a hospital does not print a staff directory. It would be most practical in this instance to sample staff from hospital. Rather than randomly sample 10% of staff from each hospital, which would be a difficult task, randomly sampling every staff in 10% of the hospital would be easier.

For example, suppose an organization wishes to find out **Hospital Anniversary Year 11 staff are participating in across Trichy.** It would be too costly and take too long to survey every staff, or even some staff from every hospital. Instead, 100 staff are randomly selected from all over Trichy.

These Hospitals are considered clusters. Then, every Year 11 staff in these 100 hospitals is surveyed. In effect, staff in the sample of 100 hospital **represent all Year 11 staff in Trichy.**

Multistage Sampling

Our final strategy within the broader category of probability sampling is multistage sampling. This is our most sophisticated sampling strategy and it is often used in large epidemiological studies. To obtain a representative national sample, researchers may select zip codes at random from each state. Within these zip codes, streets are randomly selected. Within each street, addresses are randomly selected. While each zip code constitutes a cluster, which may not be as accurate as other probability sampling strategies, it still can be very accurate.

Systematic Sampling

Systematic sampling yields a probability sample but it is not a random sampling strategy (it is one of our exceptions). Systematic sampling strategies take every nth person from the sampling frame. For example, you choose a random start page and take every 45th name in the directory until you have the desired sample size. Its major advantage is that it is much less cumbersome to use than the procedures outlined for simple random sampling. The appropriate sampling interval, I, is then calculated by dividing population size, N, by required sample size, n, as follows: $I = N/n$

For example, if a systematic sample of 500 staff were to be carried out in a hospital with an enrolled population of 10,000, the sampling interval would be:

$$I = N/n = 10000/500 = 20$$

Proportionate Sampling

Proportionate sampling is a variation of stratified random sampling. We use this technique when our subgroups vary dramatically in size in our population. For example, we are interested in risk taking among college students and suspect that risk taking might differ between smokers and non-smokers. Given increasing societal pressures against smoking, there are many fewer smokers on campus than non-smokers. Rather than take equal numbers of smokers and non-smokers, we want each group represented in their proportions in the population.

Chapter 20. Measuring association in case-control studies

The cohort studies or clinical trials in which we compared either **cumulative incidence** or **incidence rates** among two or more exposure groups. However, in a true case-control study we don't measure and compare incidence. There is no "follow-up" period in case-control studies.

Variables	Diseased	Non-diseased	Total
Exposed	7	1,000	1,007
Non-exposed	6	5,634	5,640

This view of the population is hypothetical because it shows us the exposure status of all subjects in the population. We therefore know the total number of exposed and non-exposed people (in the "Total" column). If we know all this, we could compute the incidence in each group (the incidence in the exposed individuals would be 7/1007 = 0.70%, and the incidence in the non-exposed individuals would be 6/5640 = 0.11%), and we could compute the risk ratio (RR = 6.53). All of our computations involved the "Diseased" column and the "Total" column.

Another way of looking at this association is to consider that the "Diseased" column tells us the relative exposure status in people who developed the outcome (7/6 = 1.16667), and the "Total" column tells us the relative exposure status of the entire source population (1007/5640 = 0.1785). The ratio of these two distributions (7/6)/(1007/5640) = 6.53, because it is just an algebraic rearrangement of the same four numbers we used to compute the cumulative incidences and the risk ratio. Note also that the relative exposure distribution in the "Total" population is very similar the relative exposure distribution in the "Non-diseased" portion of the source population, because the disease is rare. Consequently, in order to estimate the risk ratio we could use the relative distribution of exposure in the "Non-diseased" subjects - OR, to be more efficient, we could just take a sample of non-diseased subjects in order to estimate their exposure distribution. We could for example, just sample 1% of the non-diseased people and I then determine their exposure status. The data might look something like this:

Variables	Diseased	Non-diseased	Total
Exposed	7	10	unknown
Non-exposed	6	56	unknown

The Odds Ratio

The relative exposure distributions (7/6) and (10/56) are really ***odds***, i.e. the odds of exposure among cases and non-diseased controls. If we compute the ratio of these two odds we would get:

$$OR = \frac{7/6}{10/56} = \mathbf{6.53}$$

i.e., almost identical to the risk ratio we calculated when we had all the information for the source group. Note that we would get the same answer if we computed the odds ratio by dividing the odds of disease in the exposed (7/10) by the the odds of disease in the non-exposed group (6/56).

$$OR = \frac{7/10}{6/56} = \mathbf{6.53}$$

Because this expression is just an algebraic rearrangement of the previous equation.

Chapter 21. A randomized controlled trial: Different types of blinding

Background

Blinding means withholding which group each participant has been assigned to. Studies may use **Single-, Double- or Triple-blinding.**

Single-blinding occurs in many different kinds of studies, but double- and triple-blinding are mainly used in medical research.

A trial can be "open labeled" or "blinded". By the process of blinding, we make the participant and/or assessing physician unaware of the treatment he/she is going to receive. Thus, the element of bias which can creep in owing to personal preference or subjective component to the assessment of outcome (e.g., a tool like physician global score is used to assess the outcome) can be eliminated. The process has now been further extended to include the statistician analyzing the result to make it fool proof. Thus, blinding is helpful in eliminating intentional or unintentional bias, increasing the objectivity of results, and ensuring the credibility of study conclusions.

Types of Blinding

Open-labeled or unblinded: All parties involved in a study are aware of the treatment the participants are receiving. Although blinding is desirable, sometimes it may not be possible or feasible. This type of study design suffers from low credibility but may be acceptable if endpoints are indisputably objective (e.g., survival or death)

Example: You are studying the impact of a new school instruction program aimed at improving students' reading comprehension skills.

You randomly assign some students to the new program (the treatment group), while others are instructed with a standard program (the control group). You use single blinding: you do not inform students whether they are receiving the new instruction program or the standard one.

If students become aware of which program they have been assigned to – for example, by talking to previous students about the content of the program – they may change their behavior. Students in the control group might work harder on their reading skills to make up for not receiving the

new program, or conversely to put in less effort instead since they might believe the other students will do better than them anyway.

Thus, the results of your study could be invalid unless you prevent any unblinding.

Single-blind: The participants in a study might drop out from study or might give false assessment if they come to know that they are receiving "no treatment." In addition, they might develop a placebo effect, if they know they are receiving "new treatment." All these biases can be eliminated by single-blinding. In this, a group of individuals (usually the participants) do not know the intervention he or she is going to receive. Conventionally, it refers to participant-blinded but logically the group of individuals blinded can also be the outcome assessor. Thus, a single-blind trial can be either participant-blind or assessor-blind, and it is better to specify who is blinded, instead of saying single-blind

Example: You have developed a new flu vaccine. In order to test the effectiveness of your new treatment, you run an experiment, giving half of your participants the flu vaccine and the other half a fake vaccine that will have no effect (to control for the placebo effect).

If participants in the control group realize they have received a fake vaccine and are not protected against the flu, they might modify their behavior in ways that lower their chances of becoming sick – frequently washing their hands, avoiding crowded areas, etc. This behavior could narrow the gap in sickness rates between the control group and the treatment group, thus making the vaccine seem less effective than it really is.

To prevent such an outcome, in a single-blind study, you hide from the participants which vaccine – real or fake – each of them received.

Double-blind: Like participants, the investigator/observer may influence the results of the study, if they are aware if a group of individuals are receiving a particular treatment. For example, if the endpoint is subjective (e.g. physician global scale), they might record a more favorable response for treatment of their preference. In addition, they might influence participants' assessment of a particular treatment during follow-up meetings. In double-blinding, neither the participant nor the investigator/observer/outcome assessor is aware of the treatment allotted. The investigator is the person carrying out the research. The observer or the outcome assessor is the person who assesses the parameters of the study.

Example: In the flu vaccine study that you are running, you have recruited several experimenters to administer your vaccine and measure the outcomes of your participants.

If these experimenters knew which vaccines were real and which were fake, they might accidentally reveal this information to the participants, thus influencing their behavior and indirectly the results.

They could even directly influence the results. For instance, if experimenters expect the vaccine to result in lower levels of flu symptoms, they might accidentally measure symptoms incorrectly, thus making the vaccine appear more effective than it really is.

To avoid this, you hide group assignments from both the participants and the experimenters giving the vaccines – a double-blind study.

Triple-blind: Triple-blinding is done to eliminate the bias of data analysts. In triple blinding, the participant, investigator, and the data analyst are unaware of the treatment given.

Example: In your vaccine study, you have also recruited assistants to analyze the data you gathered on flu infection rates. You decide to hide the group assignments from the participants, the people administering the experiment, and the people analyzing the data – a triple-blind study.

To achieve triple blinding, you assign each participant to group 1 or group 2, but do not inform the data analysts which number represents which group.

However, instead of expressing whether the trial is single, double, or triple blinded, it is more pertinent to specify who exactly is going to be blinded.

Chapter 22. Statistics – Black-Scholes model

Background

The Black Scholes model is a mathematical model to check price variation over time of financial instruments such as stocks, which can be used to compute the price of the European call option. This model assumes that the price of assets, which are heavily traded, follows a geometric Brownian motion having a constant drift and volatility. In the case of a stock option, the Black Scholes model incorporates the constant price variation of the underlying stock, the time value of money, the strike price of the option and it is time to expiry.

The Black Scholes Model was developed in 1973 by Fisher Black, Robert Merton and Myron Scholes and it is still widely used in European financial markets. It provides one of the best ways to determine fair prices of options.

Input

The Black Scholes model requires five inputs.

1) Strike price of an option
2) Current stock price
3) Time to expiry
4) Risk-free rate
5) Volatility

Assumption

The Black Scholes model assumes the following points.

1) Stock prices follow a lognormal distribution.
2) Asset prices cannot be negative.
3) No transaction cost or tax.
4) Risk-free interest rate is constant for all maturities.
5) Short selling of securities with the use of proceeds is permitted.
6) No riskless arbitrage opportunity present.

Formula: C= SN (D1) – Ke-rtN(D2)

Here,

1) d1=lnStK+r+22tt,
2) d2 = d1 – t,
3) C - call option price,
4) S= current stock price or price of the underlying security,
5) K= strike price,
6) r = risk-free interest rate,
7) t = time to maturity,
8) N= normal distribution,

Limitation

The Black Scholes model has the following limitations.

1) Only applicable to Indian options as American options could be exercised before their expiry.
2) Constant dividends and constant risk-free rates may not be realistic.
3) Volatility may fluctuate with the level of supply and demand of option thus being constant may not be true

Example

Step I: To determine the historical volatility, daily log returns have been calculated by using the moving average method.

Daily return = ln (today's closing price/yesterday's closing price)

Daily Standard deviation (SD) = (Variance of daily returns) 0.5

Historical volatility = Daily SD × (250) 0.5

(250 trading days in a year is taken for the above calculation purpose)

Step II: To derive the fair value of call and put options of single strike prices, first we collect all required data in the Black formula from the NSE and then apply them in the BSOPM. The next action is to determine the variations between the model value and the actual market prices.

Step III: The last step is the comparison of the fair option premium with the actual price of the option premium.

Monthly log-returns of the corresponding scrips have been used to find out the historical volatility:

Monthly return = ln [(this month's closing price)/(last month's closing price)]

Volatility = Standard deviation of the monthly returns

Of note, 7.4 per cent is the risk-free rate of return, which has been used in this study. This Rf is the current yield on 10-year government bonds issued by the Indian Government. The time to maturity is calculated as the fractional value of the number of days remaining to the maturity date. NSC and BSE websites are referred for collecting the data, i.e. spot prices of the different stocks. BSOPM has been used then to determine the call and put option fair price using the single strike price of all the stocks. The following hypothesis has been framed and a paired sample test has been conducted to derive whether there is a significant difference between BSOPM price and real market price.

Null hypothesis (H_0): There is no significant difference between BSOPM prices and actual market prices.

Alternate hypothesis (H_a): There is a significant difference between BSOPM prices and actual market prices at a 95 per cent level of confidence.

If *P*-value> 0.05, then the null hypothesis is accepted.

Conclusion

A total of 60 hypotheses are framed (30 for call option contracts and 30 for put option contracts) and tested using the paired sample *t*-test. The paired *t*-test compares the means and standard deviations of the two series of numbers and determines if there is any significant difference between the two series of numbers. The following stocks are chosen for the analysis.