

BIostatISTICS LECTURES

AS A PREFACE

More than a century ago, a French mathematician, Pierre Louis developed and advocated his "numerical method" for the appraisal of therapy. However, he was opposed by most of the leading clinicians of the day. Claude Bernard, arguably the father of modern experimental medicine, was critical of the application of mathematics in medicine; he declared: mathematicians average too much and reason about phenomena as they construct them in their minds, but not as they exist in nature". He went on to urge doctors "to reject statistics as a foundation for experimental therapeutic and pathological science". Ironically, almost 100 years later, his disciples have abandoned his words almost completely. Statistics has become an indispensable part in medical research. Almost every medical publication nowadays includes "statistical method" section as to show its credence.

However, in recent years, the popular use of statistics has become the popular abuse of statistics. The medical literature is filled with an immense statistical information and endless contradictory findings. There are good evidences suggesting that the inappropriate use and wild manipulation of statistics have contributed significantly to this tragedy of confusion of knowledge. People are writing books and papers based on inappropriate application of statistics. Some of these authors are very popular because they are not afraid to provide solutions to problems that have not yet been solved. However, some investigators do not realise that they have made statistical errors due to either ignorance or lack of statistical knowledge. Whatever the reason, to rely on the analysis of data, the nature of which one does not understand, is the first step in losing intellectual honesty. Therefore, understanding the principle behind a statistical analysis is critical to the interpretation of data.

The following fifteen topics of biostatistics represent a collection of elementary statistical principles, theorems, axioms and definitions which are to be presented in a rather non-technical format. The topics are organised into two parts: part one deals with probability and statistical concepts and part two deals with applied statistics, in which statistical analysis of data and inferences are discussed. As the intended audience of this

course are researchers with little or no knowledge of statistics, therefore, all of the statistical statements in this course are presented without mathematical proofs.

Any learning will be incomplete without some assimilation of the subject's principles. To illustrate the ideas and principles behind each of the topics, I also present a collection of exercises and problems for either solution or discussion. Some questions can be classified as "exercises" which are aimed at illustrating the basic principle; other questions can be classified as "problems" which would normally require several skills to solve them. Most of these questions were extracted from medical journals and real-life experience. If you are unable to solve a question, do not despair. The Socratic method of teaching does not aim at drilling people in giving quick answers, but to educate by means of questions. If repeated efforts have been unsuccessful, you can contact me and we will, hopefully, work it out. Remember that the solution to any worthwhile problem rarely comes to us easily without hard work; it is rather the result of intellectual effort of days, or weeks, or months, even years. Well, none of the questions in this course would take you years or months, but may take hours to work out.

The advent of electronic computers has certainly revolutionised the statistical practice. Computer has replaced pencil in solving complex equations in data analysis. But, statistics is not just a collection of theorems or formulae; it is a style of thinking. Computing is also a style of thinking. I, therefore, do not believe that statistics can be reduced to button pushing on a computer and still retains its style of thinking, although there are people who claim to have done so. With this belief in mind, most of the exercises in this course are meant to be solved manually with the help of a calculator or any general spreadsheet package - not by a statistical software. I always believe that in doing exercises, a manual working solution is more enjoyable and fulfilling than an "automatic" solution.

This course is the result of one inexperienced man's effort. It is hardly perfect and leaves a lot of open problems. Furthermore, the materials have been written in a very short time and thus mistakes are inevitable. If you find one, please let me know (and keep it secret).

Good luck.

BIostatistics

Topic 1: Review of Mathematical Notations

Even a man of slow intellect who is trained and exercised in arithmetic, if he gets nothing else from it, will at least improve and become sharper than before

Plato

A wise man, Bertrand Russell, recently wrote in one of his books: "there was a footpath leading across fields to New Southgate, and I used to go there alone to watch the sunset and contemplate suicide. I did not, however, commit suicide, because I wished to learn more about mathematics". I am sure most of us would not deny the power of mathematics, but there may be more than one of us who has distaste for the discipline. If we think seriously about mathematics, we find that it is the basis for everything we do nowadays. It is no wonder that mathematics is regarded as *the* prince of science. One of my colleagues here at the Garvan describes mathematical ideas as "eternal truth". I like the description. In most sciences, particularly biological science, the speed of "discoveries" makes one dizzy; unfortunately, these so-called discoveries often tear down ideas that others have built. Even worse, what one has established another undoes. In mathematics alone each generation adds a new story to the old (permanent) structure, because the idea is simply timeless.

Statistics is usually defined as a branch of applied mathematics, which is in turn a modern discipline of modern mathematics. In practice, modern mathematics is one of the principal tools of statistics. Therefore, to understand statistics, it is a must to have some knowledge of elementary mathematics. In this topic, we will survey some basic ideas of modern mathematics such as the concept of functions, equations, operation of summations, etc. before we venture into the statistical discussion.

I. THE REAL NUMBER SYSTEMS

In this section, we will review briefly the structure of the real number system.

- (a) The number 1, 2, 3, and so on are called **natural numbers**. If we add or multiply any two natural numbers, the result is always a natural number. However, if we subtract or divide two natural numbers, the results are not *always* a natural number. For example $8 - 5 = 3$ and $8/2 = 4$ are natural numbers, but $5 - 8 = -3$ and $2/7$ do not result in natural numbers. Thus, within a natural number system, we can add and multiply, but can not always subtract or divide.
- (b) To overcome the limitation of subtraction, we extend the natural number system to the system of **integers**. We do by including, together with all the natural numbers, all of their negatives and the number zero (0). Thus, we can represent the system of integers in the form:

$$\dots -3, -2, -1, 0, 1, 2, 3, \dots$$

- (c) But we still can not always divide any two integers. For example $8/(-2) = -4$ is an integer, but $8/3$ is not an integer. To overcome this problem, we extend the system of integers to the system of **rational numbers**. We define a number as rational if it can be expressed as a ratio of two integers. Thus, all four basic arithmetic operations (addition, subtraction, multiplication and division) are all possible in the rational number system.
- (d) There also exists some numbers in everyday use that are not rational number; that is, they can not be expressed as a ratio of two integers. For example $\sqrt{2}$, $\sqrt{3}$, π , etc. are not rational numbers; such numbers are called **irrational numbers**.
- (e) The term **real number** is used to describe a number that is either rational or irrational. To give a complete definition of real numbers would involve the introduction of a number of new ideas, and we shall not embark on this task now. However, it is possible to get a good idea of what a real number is by thinking it in terms of decimals. **Any rational number can be expressed as a decimal** simply by dividing the denominator into the numerator by long division. It is found that in every

case the decimal either terminates or develops a pattern that repeats indefinitely. For example $1/4 = 0.25$ but $1/6 = 0.1616161\dots$

- (f) The square root of any real number (positive, zero or negative) is always nonnegative. Thus $\sqrt{-2}$ and $\sqrt{-2/5}$ etc are not real numbers. As long as we consider only real numbers, no meaning can be attached to the square root of a negative number. These numbers however do exist in a certain sense and are called **complex numbers**. We shall not consider such numbers at all in this course.

II. INTEGRAL INDICES AND EXPONENTS

If x and y are any real numbers, m and n are any natural numbers, then the following relations are true:

$$(a) x^m \times x^n = x^{m+n}$$

$$(b) x^{-m} = \frac{1}{x^m}$$

$$(c) \frac{x^m}{x^n} = x^{m-n}$$

$$(d) (x^m)^n = x^{mn}$$

$$(e) (x^m)^n = x^{mn}$$

$$(f) (xy)^m = x^m y^m$$

$$(g) x^{m/n} = \sqrt[n]{x^m}$$

It follows from (g) that:

$$(h) \sqrt{x} = x^{1/2}$$

and

$$(i) \sqrt{x^2} = \pm x = |x|$$

While we are on this topic, it may be useful to remind that there is more than a fundamental difference between $x^2 + y^2$ and $(x + y)^2$ (verify it for yourself). This difference will be explored later in this series, however, a number of binomial identities may be helpful here:

$$(x \pm y)^2 = x^2 \pm 2xy + y^2$$

$$(x \pm y)^3 = x^3 \pm 3x^2y + 3xy^2 \pm y^3$$

etc.

This kind of expansion will be dealt with in a probability topic.

III. LOGARITHMS

DEFINITION: If $\log_a x = y$ then $x = a^y$. The values of a and x must be greater than zero.

It follows from this definition that $\log 1 = 0$ and $\log_a a = 1$.

It is customary to write $\log_e x$ as $\ln x$, and $\log_{10} x$ as $\log x$.

Some important properties of logarithm are:

(a) $\log(xy) = \log x + \log y$

(b) $\log\left(\frac{x}{y}\right) = \log x - \log y$

(c) $\log x^n = n \log x$

(d) Change of base (from a to b) law: $\log_a x = \frac{\log_b x}{\log_b a}$

IV. FUNCTIONS

(a) GENERAL FUNCTIONS

The idea of function is one of the most fundamental concepts in modern mathematics. A function expresses the hypothesis of one quantity depending on (or being determined by) another quantity. For example:

- (i) bone mass is dependent on age of subject;
- (ii) height is dependent on races etc.

If a function f assigns a value y in the range to a certain x in the domain, then we write:

$$y = f(x)$$

where, in modern statistics, x is called "**independent**" variable and y is "**dependent**" variable, although this terminology is sometimes controversial. In formal mathematics, we refer to the possible value of x as **domain**, and the possible value of y as **range**.

For example:

$$f(x) = 3x^2 - 7x + 2$$

$$\text{if } x = -3, \text{ then: } f(-3) = 3(-3)^2 - 7(-3) + 2 = 50.$$

However, there are certain functions in which there are restrictions in the domain. For instance, for the function $y = \sqrt{x}$, the restriction is that $x \geq 0$ since we can not take root of a negative number; for $y = \log_a x$, the value of x must be greater than zero because any value of $x \leq 0$, y is not defined. Similarly for a function of the form $y = \frac{1}{x}$, then x must be different from 0.

(b) LINEAR FUNCTIONS

A linear function is usually of the form:

$$y = a + bx \quad [1]$$

where a is called the **intercept** (when $y = 0$) and b is called the **slope** which represents the rate of change in y with respect to the change in x by one unit. In a two dimensional space x - y (Figure 1A), for any two given points (x_1, y_1) and (x_2, y_2) , the slope can be determined by the relation:

$$b = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\text{change in } y}{\text{change in } x}$$

However, for a series of n points of x and y , we could extend this formula into a series of n simultaneous equations and estimate a and b by the method of **least squares** which is readily available in several statistical softwares.

If there are two lines, say, $y = a_1 + b_1x$ and $y = a_2 + b_2x$, then we can make a number of observations:

- (a) The two lines are parallel if and only if its slopes are equal, i.e. $b_1 = b_2$ (Figure 1B);
- (b) On the other hand, if $b_1 \neq b_2$, then the lines are not parallel. In statistics, we call this phenomenon "**interaction**" (Figure 1C).

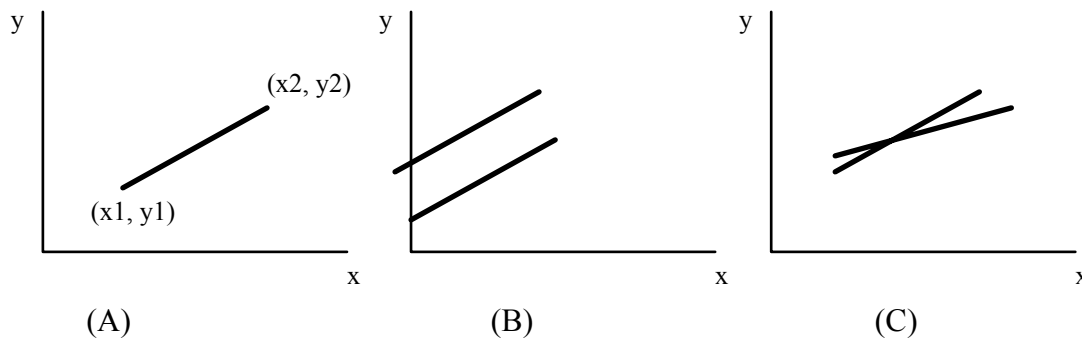


Figure 1: (A) A linear equation of the form $y = a + bx$; (B) parallel lines; (C) interaction.

Equation [1] could be expanded further to include more than one x variable. For instance, bone mineral density (BMD) is strongly dependent on age and weight, we may write this statement as:

$$BMD = a + b(\text{AGE}) + c(\text{WEIGHT})$$

where a , b and c are estimated constants. Thus, for every value of AGE and WEIGHT, a BMD could be estimated.

We will examine this function in the context of **regression analysis** later in this series.

(c) QUADRATIC FUNCTIONS

We often come across situations where the functional relationship between the dependent variable (y) and independent variable (x) is not linear, but a curved one. One of the popular functions is the quadratic function, which is of the form:

$$y = f(x) = ax^2 + bx + c \quad [2]$$

where a , b and c are constants. When $y = 0$, the values of x are: $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. The minimum/maximum of $f(x)$ is defined when $x = \frac{-b}{2a}$ i.e. $\max/\min = f\left(\frac{-b}{2a}\right)$.

For example, the relationship between BMD and age (among elderly subjects) has been suggested to follow a quadratic relationship: $BMD = a(AGE)^2 + b(AGE) + c$.

(d) INEQUALITIES

A statement such as $y = ax + b$ is called an equation, whereas $y < ax + b$ is called an inequality or inequation. The following symbols of inequality are commonly used:

- $>$ and $<$ "greater than" and "less than", respectively.
 \geq and \leq "greater than or equal to" and "less than or equal to", respectively.

A number of results of inequalities are as follows:

- (i) if $y > x$ then $y \pm a > x \pm a$
(ii) if $y > x$ then $-y < -x$
(iii) if $y > x$ then $y/a > x/a$ if $a > 0$
and $y/a < x/a$ if $a < 0$
(iv) if $y > x > 0$ then $x^n > y^n$ for all n belongs to N .

(e) LINEARISATION

It is probably needless to say that a linear function is easy to interpret than a non-linear function. Therefore, attempts have been made to transform non-linear functions to a linear function. For instance if $y = ab^x$, we could use logarithm to linearise it as: $\log y = \log(ab^x) = \log a + \log(b^x) = \log a + x \log b$. The last expression is obviously a linear function of $\log y$ with respect to x .

It should be noticed here that, formally, there is a difference between a linear function and a linear model. For instance $y = a + bx$, $y = a + bx + cx^2$ etc. are all linear model but the latter is non-linear function as oppose to the former - a linear function. We say "linear" because the parameters a , b , and c are linear. However, $y = a^2x + b$ or $y = ab^x$ are both "non-linear" models and non-linear function because both the parameters and the curve are not linear.

V. THE SUMMATION NOTATION Σ

Many statistical computations involved this symbol. Before introducing statistics, we need to be familiar with the sigma symbol Σ , which is a short writing for summation.

For example: instead of writing $x_1 + x_2 + x_3 + x_4 + x_5$ we could write as $\sum_{i=1}^5 x_i$ (reads: sum of x sub. i where i runs from 1 to 5), or, $x_0 + x_2 + x_4 + \dots + x_{10} = \sum_{i=0}^5 x_{2i}$.

Similarly, one could represent $x^0 + x^1 + x^2 + \dots + x^n$ by $\sum_{k=0}^n x^k$. Obviously a variety of expressions can be simplified by this operation. I include some interesting identities in this note for your exercise.

VI. THE PRODUCT NOTATION Π

FACTORIAL NOTATION

Suppose that we have n objects, the of arrangements (permutations) of these n objects in a line can be shown to be equal to $n!$ (read n factorial); this symbol is defined as:

$$n! = n(n-1)(n-2)(n-3)\dots 1$$

For example $3! = 3 \times 2 \times 1 = 6$

$$5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$$

As can be seen, for a large n , the value of $n!$ is very very large and may not be calculated by normal hand calculator, an approximation is suggested:

$$n! \approx \sqrt{2\pi n} n^n e^{-n}$$

where $e = 2.71828\dots$ is the base of natural logarithms and $\pi = 3.1416\dots$. This approximation is called the **Stirling's formula** of approximation.

THE PRODUCT NOTATION \prod

This operator is defined as follows:

$$\prod_{i=1}^n x_i = x_1 \times x_2 \times x_3 \times \dots \times x_n$$

It follows that: $\prod_{i=1}^n k = k^n$

and that $\prod_{i=1}^n kx_i = k^n (x_1 \times x_2 \times x_3 \times \dots \times x_n)$

It can also be shown that:

$$\prod_{i=1}^n x_i y_i = \left(\prod_{i=1}^n x_i \right) \left(\prod_{i=1}^n y_i \right)$$

VII. COMMON GREEK CHARACTERS AND MATHEMATICAL SYMBOLS

In statistics, it has become somewhat conventional to denote a **parameter of a population** by a Greek letter and its corresponding **sample statistic** by a normal English letter. The latter is readily understood by all of us, while the former may require a revision. I list in the following table some of the commonest Greek characters that we are going to use or encounter in this course.

UPPERCASE	LOWER CASE	PRONUNCIATION
α	α	alpha
β	β	beta
χ	χ	chi
δ	δ	delta
ε	ε	epsilon
ϕ	ϕ	phi
γ	γ	gamma
μ	μ	mu
ν	ν	nu
π	π	pi
θ	θ	theta
ρ	ρ	rho
σ	σ	sigma
λ	λ	lambda
κ	κ	kappa
ω	ω	omega
κ	κ	sigma
ψ	ψ	psi

COMMON MATHEMATICAL SYMBOLS

And of course, a learning of mathematics can not be complete without being able to communicate in its language. Here are some of the commonly-used symbols in mathematics which you are required to be conversant with:

SYMBOL	MEANING
\in :	Belong to
\notin	Not belong to
\Rightarrow	imply; it follows that
\Leftarrow	Implied by
\Leftrightarrow	Equivalent to; if and only if
\mathbb{R}	Real numbers
\forall	For every
\exists	There exists

VIII. EXERCISES

1. The VDR gene locus has two alleles, say T and t , which gives rise to three genotypes TT , Tt and tt . Notice that the number of T 's alleles in genotypes TT , Tt and tt are 2, 1 and 0, respectively. In a study of 130 women in the USA, the frequency distribution of VDR genotypes were as follows:

TT :	53
Tt :	57
tt :	20

- Use this distribution to find the relative frequency of the allele T . What is the relative frequency of the allele t ?
2. According to the Hardy-Weinberg equilibrium law, if the relative frequencies of the T and t alleles are p and q , respectively ($p + q = 1$), then the frequencies of TT , Tt and tt are given by p^2 , $2pq$ and q^2 , respectively. Use the results of Question 1 (p and q) to work out the expected frequency distribution of VDR genotypes for the 130 women. Are the observed data in Question 1 consistent with the Hardy-Weinberg's equilibrium law (do not need to use statistical test) ?
 3. According to the theory of genetics, the theoretical mean phenotype values of TT , Tt and tt are $(\mu + a)$, $(\mu + d)$ and $(\mu - a)$, respectively, where μ refers to the overall population mean, a refers to the additive genetic effect and d refers to the dominant genetic effect. In a sample of women, the mean BMD for the three genotypes were 1.23, 1.15 and 0.98, respectively. Find the value of a and d .
 3. Consider the following statement: ". . . DNA was extracted from 80 osteoporotic and 85 age-matched control women. PCR was used to amplify the DNA sequence and the Bsm-I restriction enzyme was used to detect VDR alleles. The results show no significant difference in the frequency of the B allele between osteoporotic and controls (0.44 and 0.39), nor in the frequency of the BB genotype (0.20 and 0.21)" (Gallagher et al JBMR 1994).
Assuming that the distribution of genotypes in the osteoporotic and control subjects followed the Hardy-Weinberg equilibrium law, can you estimate (reconstruct) the number of subjects in each genotype for each group ?

5. What value of x which should be excluded from the domain of each of the following functions:

(a) $f(x) = \frac{x+2}{x-1}$ (b) $f(x) = \frac{x}{x^2-4}$ (c) $f(x) = \log(x)$

(d) $g(x) = \sqrt{4-x^2}$ (e) $f(x) = \sqrt{x^2-x}$

6. The rate of change in BMD is sometimes modelled as a quadratic function as follows:

$$\delta y = at^2 + bt + c$$

where δy is the rate of change in BMD, t is time (in years), a , b and c are constant. Use your knowledge of inequality to find out when ($t = ?$) is δy is positive and when δy is negative.

7. Given that $f(x) = \ln\left(\frac{x+1}{x}\right)$, show that $f(1) + f(2) + f(3) = \ln 4$.

8. If $C(n, k) = \frac{n!}{k!(n-k)!}$, find an expression (not calculate) for:

(a) $C(5, 3)$ and (b) $C(n+1, k)$.

9. The Normal (Gaussian) distribution is determined by two parameters: the mean (μ) and the standard deviation (σ), and is given by:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Find the expression for $f(10; 15, 5)$.

10. If $x = \ln 2$, $y = \ln 3$, express the relationship between x and y the following:

(a) $\ln 9$ (b) $\ln \sqrt{2}$ (c) $\ln\left(\frac{3}{2}\right)$ (d) $\ln 3\sqrt{3}$ (e) $\ln 0.25$

(f) $\ln\left(\frac{8}{9}\right)$ (g) $\ln \sqrt{6}$ (h) $\ln 12$ (i) $\ln 4.5$ (j) $\ln 72$

11. (i) If $e^x - e^{-x} + 1 = 0$, show that $x = \ln[\sqrt{5}-1] - \ln 2$.

(ii) If $(e^{-x} + e^x)/(e^{-x} - e^x) = 2$ show that $x = -\frac{1}{2} \log_e 3$

(iii) If $y = \frac{1}{2}(e^x + e^{-x})$ show that $x = \ln\left(y + \sqrt{y^2 + 1}\right)$.

12. Show that the following results are true:

(a) $\log_a e = 1/\log_e a$ (b) $\log_q p \cdot \log_r q \cdot \log_p r = 1$ (c) $\log_{na} x = \log_a x / (1 + \log_a n)$

13. The female of a certain species of insect produces 20 surviving female offspring per generation. There are 4 generations per year. What is the total number of female offspring that descend in one year ?.

14. Suppose that only dried lentils and dried soybeans are available to satisfy a person's daily requirement for protein, which is 75 g. One gram of lentils contains 0.26 g of proteins and 1 g of soybeans contains 0.35 g of proteins. Let his consumption be x g of lentils and y g of soybeans. Express the relation between x and y in the form of a linear function i.e. $y = a + bx$.

15. The equation of relating the probability of fracture (P) and bone mineral density (BMD) usually follows a logistic function of the form:

$$\frac{P}{1-P} = e^{b(BMD)}$$

where b is a constant. Estimating b of this equation is a bit difficult, however, the estimation would be easier if we express this relation in a linear function by using logarithm. Can you find this linear equation.

16. In an experiment the weight W of the antlers of deer was measured for a number of deers of different ages. The results are given in the following table, W being in kg and the age A in months. Show by equation and graph that the data fit closely with the linear relation

$W = mA + b$, where m and b are slope and intercept, respectively. Find this equation.

A: 20 22 30 34 42 43 46 54 56 68 70

W: 0.08 0.10 0.15 0.20 0.27 0.26 0.31 0.36 0.40 0.49 0.49

17. Shortly after consuming a whisky, the alcohol content of a person's blood rises to a peak value of 0.22 mg/ml, and thereafter slowly decreases. If t is the time in hours after the maximum value is reached and y the blood-alcohol level, the following table gives the experimentally measured values for this subject:

t : 0 0.5 0.75 1.0 1.5 2.0 2.5 3.0

y : 0.22 0.18 0.15 0.13 0.10 0.08 0.06 0.05

The data are believed to follow the relation $y = ba^t$ where b and a are constants. By using logarithm and common statistical software, estimate a and b .

18. Let

$$\begin{aligned} w_1 = 8, \quad w_2 = 11, \quad w_3 = 3, \quad w_4 = 2, \quad w_5 = 17, \quad w_6 = 9, \quad w_7 = 6, \quad w_8 = 11 \\ x_1 = 4.7, \quad x_2 = 3.9, \quad x_3 = 7.2, \quad x_4 = 0.5, \quad x_5 = 2.8, \quad x_6 = 5.1, \quad x_7 = 7.9, \quad x_8 = 4.6 \\ y_1 = 6, \quad y_2 = 1, \quad y_3 = 5, \quad y_4 = 4, \quad y_5 = 4, \quad y_6 = 8, \quad y_7 = 1, \quad y_8 = 7. \\ z_1 = 31, \quad z_2 = 62, \quad z_3 = 7, \quad z_4 = 15, \quad z_5 = 53, \quad z_6 = 16, \quad z_7 = 94, \quad z_8 = 59. \end{aligned}$$

Evaluate

$$(a) \sum_{i=1}^6 w_i \quad (b) \sum_{i=1}^8 z_i \quad (c) \sum_{i=1}^4 y_{2i} \quad (d) \sum_{i=4}^7 x_{2i-8} \quad (e) \sum_{j=3}^7 (y_j - y_{j-2})$$

Questions 19 to 24 are referred to question 18.

19. Let $v_{ij} = w_i y_j$ (e.g. $w_{23} = w_2 y_3 = 11 \times 5 = 55$). Evaluate: (a) $\sum_{i=1}^3 \sum_{j=1}^7 v_{ij}$ (b) $\sum_{i=2}^5 \sum_{j=3}^4 v_{ij}$

20. Show that $\sum_{i=1}^5 3w_i = 3 \sum_{i=1}^5 w_i$

21. Show that $\sum_{j=1}^4 (3w_j - 2y_j) = 3 \sum_{i=1}^4 w_j - 2 \sum_{j=1}^4 y_j$

22. Show that $\sum_{j=1}^4 (z_j - 6) = \sum_{j=1}^4 z_j - 4 \times 6$

23. Show that $\sum_{j=1}^3 (w_j - 3)^2 = \sum_{i=1}^3 w_j^2 - 6 \sum_{j=1}^3 w_j + 3(9)$

24. Show that $\left(\sum_{j=1}^4 w_j \right) \left(\sum_{j=1}^4 y_j \right) \neq \sum_{i=1}^4 w_j y_j$

25. Find the percentage error in Stirling's formula for $n = 10$ and $n = 20$.