Johns Hopkins University
Department of Biostatistics
Math Review for Introductory Courses

Rationale

Biostatistics courses will rely on some fundamental mathematical relationships, functions and notation. The purpose of this Math Review is to provide an opportunity for you to review some basic definitions, relationships and perform some hand calculations. Biostatistics 610 will require use of a hand-held calculator for homework and examinations. In the Biostatistics 621-624 series, the statistical analysis package Stata will also be used.

A. Definitions
Some basic definitions and nomenclature are useful.

1) Integers are whole numbers, both negative and positive:
(..., -3, -2, -1, 0. 1, 2, 3 ...)

2) Positive integers are integers greater than 0:
(1, 2, 3, 4, ...)

3) Real numbers are all numbers, both integers and nonintegers such as fractions.

4) $|a|$ is called the <u>absolute value</u> of the number a:

If $a \geq 0$, then $|a| = a$

If $a \leq 0$, then $|a| = -a$

For example, $|3| = 3$ and $|-3| = 3$.

5) If b is any real number and n is a positive integer, then b^n is defined as b multiplied by itself n times. In other words, $b^n = b \cdot b \cdot b \cdot b \ldots \cdot b$.

b multiplied by itself n times

Here b is referred to as the <u>base</u> and n as the <u>exponent</u>. We can write

$b^0 = 1$ (by convention)

$b^{-n} = \dfrac{1}{b^n}$ if n is a positive integer

$\sqrt[n]{b} = b^{1/n}$ (the nth root of b).

6) The <u>logarithm</u> to base b of a positive number x is that number y which satisfies the equation $b^y = x$ and we write $\log_b x = y$. We have

$$10^{-2} = \frac{1}{10^2} = 0.01 \quad ; \quad \log_{10}(0.01) = -2$$

$$e^3 = e \cdot e \cdot e \quad ; \quad \log_e(e^3) = 3$$

$$2^{10} = 1{,}024 \quad ; \quad \log_2(1{,}024) = 10$$

7) The <u>summation sign</u> is a useful notation to indicate the sum of the values of a variable for observations 1 through n. Let x represent some variable such as age. We can let x_i indicate the value of age for individual i, where i takes on values from 1 to n, in a group of n individuals.
The sum of the ages of all n individuals can be written as:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

8) The <u>product sign</u> is a useful notation to indicate the product of the values of a variable for observations 1 through n. Using the previous examples, the product of the ages of all n individuals can be written as:

$$\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n$$

B. Examples of Using Summation Signs

Suppose we have 3 observations, each described by a variable x and a variable y.

Let $\quad x_1 = 5 \qquad y_1 = 9 \qquad$ and $n = 3$
$\qquad\quad x_2 = 6 \qquad y_2 = 12$
$\qquad\quad x_3 = 2 \qquad y_3 = 4$

For the values of x and y given above, compute the following by pencil and paper just for the mental exercise, <u>not by calculator</u>. (You can later verify your results using your calculator.)

$$\sum_{i=1}^{3} x_i =$$

$$\sum_{i=1}^{3} x_i^2 =$$

$$\sum_{i=1}^{3} y_i =$$

$$\sum_{i=1}^{3} \frac{x_i}{y_i} =$$

$$\sum_{i=1}^{3} (x_i + y_i) =$$

$$\sum_{i=1}^{3} (y_i - 2) =$$

$$\sum_{i=1}^{3} x_i + \sum_{i=1}^{3} y_i =$$

$$\left(\sum_{i=1}^{3} y_i\right) - 2 =$$

$$\sum_{i=1}^{3} x_i y_i =$$

$$\sum_{i=1}^{3} (y_i - 2)^2 =$$

$$\sum_{i=1}^{3} x_i \cdot \sum_{i=1}^{3} y_i =$$

$$\left(\sum_{i=1}^{3} y_i^2\right) - 2 =$$

$$\sum_{i=1}^{3} (x_i - y_i) =$$

$$\sum_{i=1}^{3} (x_i - y_i)^2 =$$

$$x_1^{-1} =$$

$$\sum_{i=1}^{3} |x_i| =$$

$$x_3^{-3} =$$

$$\sum_{i=1}^{3} |x_i - y_i| =$$

$$y_3^{1/2} =$$

$$x_2^0 =$$

$$\log_{10}(5 \cdot x_1 y_3) =$$

$$\log_2 y_3 =$$

$$\prod_{i=1}^{3} x_i =$$

$$\frac{\displaystyle\prod_{i=1}^{3} x_i}{\displaystyle\prod_{i=1}^{3} y_i} =$$

C. More Examples

Compute the following commonly used equations by hand:

The following notation is used
for the <u>sample mean</u>:

$$\frac{\displaystyle\sum_{i=1}^{3} x_i}{n} = \bar{x} =$$

$$\left(\sum_{i=1}^{3} x_i\right)^2 =$$

The two equations below are algebraically equivalent formulations of the <u>sample variance</u>:

$$\frac{\displaystyle\sum_{i=1}^{3}(x_i - \bar{x})^2}{n-1} =$$

$$\frac{\left(\displaystyle\sum_{i=1}^{3} x_i^2\right) - n\bar{x}^2}{n-1} =$$

D. More on Logarithms

1) Write the following equations in terms of logarithms:

$$2^3 = 8$$

$$10^2 = 100$$

$$10^{-3} = 0.001$$

2) Write the following equations in terms of exponents:

$$\log_2 128 = 7$$

$$\log_5 125 = 3$$

$$\log_{1/2} \left(\frac{1}{16} \right) = 4$$

$$\log_{10} (0.01) = -2$$

Common logarithm =   a logarithm with base 10 such that $\log_{10} y = x$ implies $10^x = y$.
Often $\log_{10} y$ is written as $\log y$.

Natural logarithm =   a logarithm with base e such that $\log_e y = x$ implies $e^x = y$. Often $\log_e y$ is written as lny.

Note: e or Euler's constant (2.71828 ...) is important in describing biological relationships and is useful in many statistical applications.

Properties of logarithms:

$$\log_b(xy) = \log_b x + \log_b y$$
$$\log_b(x/y) = \log_b x - \log_b y$$
$$\log_b(x^r) = r \cdot \log_b x$$

E. Scientific Notation: Expressing a number as a product of a number N between 1 and 10 and an integral power of ten in order to simplify notation of calculations: $(N) \cdot (10)^k$

(e.g., $390.672 = 3.90672 \times 10^2$
$0.0001576 = 1.576 \times 10^{-4}$)

Rules

1) The exponent of 10 is determined by counting the number of places that the decimal point was moved when going from the original number to the number between 1 and 10.

2) The exponent is

    a) negative if the original number is less than 1
    b) positive if the original number is greater than 10
    c) 0 if the original number is between 1 and 10

Express the following in scientific notation:

    2.14
   31.79
  412.9
 8,000,000
    0.14
   0.0379

0.00000049

F. Significant Figures: Any digits in a number which contribute to the specification of its magnitude apart from zeroes that determine the position of the decimal point.

HINT: It helps to first write the number in scientific notation in order to determine the number of significant figures.

e.g., $92,800,000 = 9.28 \times 10^7$ has **3** significant figures

$0.0909 = 9.09 \times 10^{-2}$ has **3** significant figures

$0.00052 = 5.2 \times 10^{-4}$ has **2** significant figures

a) Specify the number of significant figures corresponding to each number:

0.045
4.5
4.05
4.502
20.04

G. Rounding Correct to *n* Decimal Places: the process of rounding a number to *n* decimal places.

Rules for rounding

1) Round to the nearest number

2) If the number determining the rounding is 5, set a policy to always round to the even number (or always to the odd number) to minimize overestimation or underestimation. This is important for hand calculations. If you always round "up", your calculations may tend to be overestimates. By choosing to round to the even number, we would round 2.45 to 2.4, 3.35 to 3.4, 4.25 to 4.2, 5.75 to 5.8. The rounded values are correct to 1 decimal place. In this way, 1/2 of the time we round "up" and 1/2 of the time we round "down".

(eg, 3.14159 is 3.14 correct to 2 decimal places; 17.45 is 17.4 correct to 1 decimal place)

a) Correct the following numbers to two decimal places:
7.865
7.847
7.853
7.875

H. Equation for a Straight Line:

Suppose you've collected independent pairs of data (X_i, Y_i), $i = 1$ to n, for n observations. Suppose we let Y_i represent our outcome of interest, and X_i some fixed continuous covariate. Is there a perfect linear relationship between Y_i and X_i; that is, do the points fall exactly on a straight line? If so, we could write:

$$Y = \beta X + \alpha$$
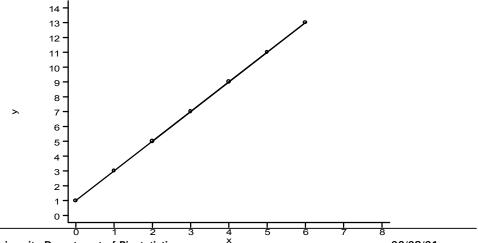
where α is the y-intercept and β is the slope of the line (the change in X for each unit change in Y).

Suppose we have:

Observation i	X_i	Y_i
1.	0	1
2.	1	3
3.	3	7
4.	4	9
5.	6	13

Plot each set of points and connect the points by a straight line. We would see the following straight line (linear) relationship:

From the graph, we can determine both the y-intercept and the slope.

The <u>y-intercept</u>, α, is the value of Y when X=0. Here, $\alpha = 1$.

The <u>slope of the line</u>, β, is the change in the value of Y for each one unit change in X. The slope can be derived from any two points on the straight line and is equal to $\triangle Y / \triangle X$ or $(Y_2 - Y_1)/(X_2 - X_1)$. Using the points for observations 1 and 2, we can calculate the slope as $(3-1)/(1-0) = 2$.

Then, we can express the linear relationship between X and Y by the equation of the straight line

$$Y = 2X + 1.$$

Math Review for Biostatistics Courses

Answer Key

Page 2 $\sum_{i=1}^{3} x_i = 5 + 6 + 2 = 13$ $\qquad$ $\sum_{i=1}^{3} x_i^2 = 25 + 36 + 4 = 65$

Page 3 $\sum_{i=1}^{3} y_i = 9 + 12 + 4 = 25$ $\qquad$ $\sum_{i=1}^{3} \frac{x_i}{y_i} = \frac{5}{9} + \frac{6}{12} + \frac{2}{4} = 1\frac{5}{9}$

$\sum_{i=1}^{3}(x_i + y_i) = 38$ $\qquad$ $\sum_{i=1}^{3}(y_1 - 2) = 7 + 10 + 2 = 19$

$\sum_{i=1}^{3} x_i + \sum_{i=1}^{3} y_i = 38$ $\qquad$ $\left(\sum_{i=1}^{3} y_i\right) - 2 = 25 - 2 = 23$

$\sum_{i=1}^{3} x_i y_i = 45 + 72 + 8 = 125$ $\qquad$ $\sum_{i=1}^{3}(y_i - 2)^2 = 7^2 + 10^2 + 2^2 = 153$

$\sum_{i=1}^{3} x_i \cdot \sum_{i=1}^{3} y_i = 13 \cdot 25 = 325$ $\qquad$ $\left(\sum_{i=1}^{3} y_i^2\right) - 2 = (81 + 144 + 16) - 2 = 239$

$\sum_{i=1}^{3}(x_i - y_i) = -4 + (-6) + (-2) = -12$ $\qquad$ $\sum_{i=1}^{3}(x_i - y_i)^2 = (-4)^2 + (-6)^2 + (-2)^2 = 56$

$x_1^{-1} = \frac{1}{5}$ $\qquad$ $\sum_{i=1}^{3}|x_i| = 13$

$x_3^{-3} = \frac{1}{2^3} = \frac{1}{8}$ $\qquad$ $\sum_{i=1}^{3}|x_i - y_i| = 12$

Page 4 $y_3^{1/2} = \sqrt{4} = 2$ $\qquad$ $x_2^0 = 1$

$$\log_{10}(5 \cdot x_1 y_3) \quad = \log_{10}(5 \cdot 5 \cdot 4)$$
$$= \log_{10}(100) = 2$$

$$\log_2 y_3 = \log_2 4 = 2$$

$$\prod_{i=1}^{3} x_i = 5 \cdot 6 \cdot 2 = 60$$

$$\frac{\prod_{i=1}^{3} x_i}{\prod_{i=1}^{3} y_i} = \frac{5 \cdot 6 \cdot 2}{9 \cdot 12 \cdot 4} = \frac{5}{36}$$

$$\frac{\sum_{i=1}^{3} x_i}{3} = \bar{x} = \frac{5+6+2}{3} = \frac{13}{3}$$

$$\left(\sum_{i=1}^{3} x_i\right)^2 = (13)^2 = 169$$

$$\frac{\sum_{i=1}^{3} x_i}{3} = \bar{x} = \frac{\left(5-\frac{13}{3}\right)^2 + \left(6-\frac{13}{3}\right)^2 + \left(2-\frac{13}{3}\right)^2}{2} = \frac{13}{3}$$

$$\frac{\left(\sum_{i=1}^{3} x_i^2\right) - n\bar{x}^2}{n-1} = \frac{65 - 3\left(\frac{13}{3}\right)^2}{2} = \frac{13}{3}$$

Note: $n = 3$

Page 5 $2^3 = 8$ $\log_2 8 = 3$

$10^2 = 100$ $\log_{10} 100 = 2$

$10^{-3} = 0.001$ $\log_{10} 0.001 = -3$

$\log_2 128 = 7$ $2^7 = 128$

$\log_5 125 = 3$ $5^3 = 125$

$\log_{1/2}(1/16) = 4$ $\left(\dfrac{1}{2}\right)^4 = \left(\dfrac{1}{16}\right)$

$\log_{10}(0.01) = -2$ $10^{-2} = 0.01$

Page 6   2.14 2.14×10^0

31.79 3.179×10^1

412.9 4.129×10^2

8,000,000 8×10^6

0.14 1.4×10^{-1}

0.379 3.79×10^{-2}

0.49 4.9×10^{-7}

$0.045 = 4.5 \times^{-2}$ 2 significant figures

4.5   2

4.05   3

4.502   4

$20.04 = 2.004 \times 10^1$ 4

Page 7   7.865   7.86

7.847   7.85

7.853   7.85

7.875   7.88