Appendix A: Review of Basic Mathematical Concepts

A

It is essential to know basic mathematical concepts to understand the statistical ideas and methods presented in this book. Appendix A reviews some basic mathematical concepts to help some students understand the use of mathematical concepts utilized in discussing statistical methods. The material presented here follows the order that topics are encountered in the text chapters.

CHAPTER 3

3.1 The Logic of Logarithms

To understand what we mean by *logarithm*, consider some positive number *y*. The base 10 logarithm of *x* is *y*, where *y* satisfies the relation that 10^y is equal to *x*. For example, the base 10 logarithm of 10, often written as $\log_{10}(10)$ or abbreviated as $\log(x)$, is 1 because 10^1 is 10. The value of $\log_{10}(100)$ is 2 because 10^2 equals 100. The value of $\log_{10}(1000)$ is 3 because 10^3, equal to $10 * 10 * 10$, is 1000. Therefore, base 10 logarithms of numbers between 10 and 100 will be between 1 and 2, base 10 logarithms of numbers between 100 and 1000 will be between 2 and 3. In the same way, the logarithms of numbers between 1 and 10 will be between 0 and 1. Numbers less than 1 have a negative logarithm. For example, the base 10 logarithm of 0.1 ($= 1 / 10 = 10^{-1}$) is −1. By expressing the numbers 1, 10, 100, and 0.1 as 10^0, 10^1, 10^2, and 10^{-1}, we get an idea about why the number system we use is base 10 and why the base 10 logarithms are referred to as the *common* logarithms.

3.2 Properties of Logarithms

Given positive numbers *x* and *y*, the number *n*, and that *a* is any positive real number with the exception that $a \neq 1$, then the following properties of logarithms hold:

(1) $\log_a(xy) = \log_a x + \log_a y$
(2) $\log_a(x/y) = \log_a x - \log_a y$
(3) $\log_a x^n = n \log_a x$

3.3 Natural Logarithms

Logarithms with a base of e, where the number e is an irrational number approximately equal to 2.71828, are called the *natural logarithms* and are written in an abbreviated notation as $\ln(x)$ which is equivalent to $\log_e(x)$. Therefore $\log_e(7)$ is equivalent to writing $\ln(7)$.

3.4 Conversion between Bases

Now that you are familiar with logarithms having a base of 10, consider the following general expression, $\log_a(x)$, which is read as the base a logarithm of x where x is any positive number and a is a positive real number where $a \neq 1$. The following theorem can be found in most algebra books:

> If x is any positive number and if a and b are positive real numbers where $a \neq 1$ and $b \neq 1$, then
>
> $$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}.$$

We can use this result to evaluate logarithms with different bases using the base 10 logarithm that is available on most calculators or computer software programs. As an example, we write the following expression $\log_2(x)$ as $\frac{\log_{10}(x)}{\log_{10}(2)}$. Other examples include:

(1) $\log_2(5) = \frac{\log_{10}(5)}{\log_{10}(2)}$

(2) $\log_3(8) = \frac{\log_{10}(8)}{\log_{10}(3)}$

(3) $\log_e(7) = \frac{\log_{10}(7)}{\log_{10}(e)}$.

3.5 Exponential Function

An exponential function has the form $y = a^x$, where $a > 0$. Consider the expression a^n where $a \neq 0$ then $a^n = \{a \cdot a \cdot a \cdot \text{K} \cdot a\}$ so that a is multiplied by itself $n - 1$ times. In this case, n is called the exponent and a is referred to as the base of the exponent. Exponents have the following properties:

If we consider m and n to be integers and the real numbers $a \neq 0$ and $b \neq 0$, then

(1) $a^m a^n = a^{m+n}$
(2) $a^m/a^n = a^{m-n}$
(3) $(a^m)^n = a^{m\,n}$
(4) $(ab)^m = a^m b^m$
(5) $(a/b)^m = a^m/b^m$
(6) $a^{-n} = (1/a)^n$, this can be inferred from (2) if $m = 0$ resulting in $a^m = 1$.

If a is a positive real number and $n \neq 0$, then the n^{th} root of a is denoted as $a^{1/n}$. This can also be expressed as $\sqrt[n]{a}$. However if n is an even number and a is a negative number, then $a^{1/n}$ is not a real number.

Example A.1

Consider the following example to help understand the usefulness of the exponential function. Epidemics are usually characterized by individuals in a population who are susceptible to some kind of infection, such as the flu. Susceptible individuals have never been infected and also have no immunity against infection. Assume that the number of individuals susceptible to flu infection during a flu epidemic decreases exponentially according to $y = y_0 e^{-ct}$, where y_0 is the base population at time 0, c is the rate of infection and t represents time. If a flu epidemic enters a population of 20,000 with an infection rate of 0.01 per day, then the number of individuals susceptible to the flu at time t is given by

$\mathrm{y} = 20{,}000e^{-0.01*t}$, where t is time in days.

(a) Find the number of individuals susceptible at the beginning of the epidemic.

Since the beginning of the epidemic is at time $t = 0$, the number susceptible is

$$\mathrm{y} = 20{,}000e^{-0.01*(0)} = 20{,}000e^{(0)} = 20{,}000.$$

(b) Approximately how many individuals are susceptible after 10 days?

At time $t = 10$, the number susceptible is

$$\mathrm{y} = 20{,}000e^{-0.01*(10)} = 20{,}000e^{(-0.1)} \approx 18{,}096.$$

(c) After how many days will half of the population be infected with the flu?

Half of 20,000 is 10,000, so

$$10{,}000 = 20{,}000e^{-0.01*t}.$$

Dividing both sides by 20,000,

$$10{,}000/20{,}000 = (20{,}000/20{,}000)e^{-0.01*t}$$

and taking the natural logarithm of both sides, we have

$$\ln(0.5) = \ln(e^{-0.01*t}).$$

By evaluating the natural logarithm of the right side, which gives

$$\ln(0.5) = -0.01 * t$$

and finally solving for t, we have

$$t = \ln(0.5)/(-0.01) \approx 69 \text{ days}.$$

CHAPTER 4

4.1 Factorials

We denote the product of $3 \cdot 2 \cdot 1$ by the symbol 3!, which is read "3 factorial." For any natural number n,

$$n! = n(n-1)(n-2)\ldots 1.$$

In general, $n! = n(n-1)!$. Then the following should be true: $1! = 1(0!)$. Since this statement is true if and only if 0! is equal to 1, we define $0! = 1$.

4.2 Permutations

If we choose two of three objects (A B C), we have the following six arrangements: AB, AC, BA, BC, CA, and CB. There are 3 choices for the first position and 2 choices for the second position. Thus, $3 * 2 = 6$. Each arrangement is called a permutation. The number of permutations of 3 objects taken 2 at a time is 6. We denote this as $P(3, 2) = 6$. In general,

$$P(N, n) = N(N-1)(N-2)\cdots(N-n+1).$$

By multiplying and dividing by $(N - n)!$, we get

$$P(N,n) = \frac{N!}{(N-n)}.$$

If $n = N$, we have $P(N, N) = N!$. Note that the order of arrangement is important in permutation (order is not ignored).

4.3 Combinations

If we ignore the order of arrangement in permutation, then any such selection is called a *combination*. The number of combinations of 3 objects taken 2 at a time is 3, including AB, AC, and BC. We denote this by $\binom{3}{2} = 3$. In general, $\binom{N}{n} n! = P(N, n)$. For each combination, there are $P(n, n) = n!$ ways of arranging the objects in that combination. Hence, we have

$$\binom{N}{n} = \frac{P(N,n)}{n!} = \frac{N(N-1)\ldots(N-n+1)}{n!} = \frac{N!}{n!(N-n)!}.$$

CHAPTER 15

15.1 Taylor Series Expansion

The Taylor series expansion has been used in statistics to obtain an approximation to a nonlinear function, and then the variance of the function is based on the Taylor series approximation to the function. Often the approximation provides a reasonable estimate to the function, and sometimes the approximation is even a linear function. This idea of variance estimation has several names in the literature, including the linearization and the delta method.

The Taylor series expansion for a function of x variable, $f(x)$, evaluated at the mean or expected value of x, written as $E(x)$, is

$$f(x) = f[E(x)] + f'[E(x)][x - E(x)] + f''[E(x)][x - E(x)]^2/2! + \cdots$$

where f' and f'' are the first and second derivatives of the function. The variance of $f(x)$ is $V[f(x)] = E[f^2(x)] - E^2[f(x)]$ by definition, and using the first order of Taylor series expansion, we have

$$V[f(x)] = \{f'[E(x)]\}^2 V(x) + \cdots$$

The same ideas carry over to functions of more than one random variable. In the case of a function of two variances, the Taylor series expansion yields

$$V[f(x_1, x_2)] \cong \left(\frac{\partial f}{\partial x_1}\right)\left(\frac{\partial f}{\partial x_2}\right) Cov(x_1, x_2).$$

If we define the following new variable at the observational unit level,

$$t_i = \left(\frac{\partial f}{\partial x_1}\right) x_{1i} + \left(\frac{\partial f}{\partial x_2}\right) x_{2i}$$

then we can calculate the variance directly from t_i without calculating the covariance. The new variable is called linearized value of $f(x_1,x_2)$.

Example A.2

Let us apply this idea to a ratio estimate shown in Example 15.5. The ratio estimate of the total number of professional workers with an MPH was 623 using the following estimator:

$$\hat{X} = \left(\frac{\bar{x}}{\bar{y}}\right) Y = \left(\frac{10.5}{19.375}\right) 1150 = 623.$$

The function is the ratio, $\bar{x}/\bar{y}$, and the derivatives with respect to $\bar{x}$ and $\bar{y}$ are

$$\left(\frac{\partial f}{\partial \bar{x}}\right) = \frac{1}{\bar{y}} \quad \text{and} \quad \frac{\partial f}{\partial \bar{y}} = -\frac{\bar{x}}{\bar{y}^2}.$$

The linearized values can then be calculated by

$$t_i = \frac{1}{19.375} x_i - \frac{10.5}{19.375^2} y.$$

For example, the linearized value for the first observation is

$$t_i = \frac{14}{19.375} - \frac{10.5(21)}{19.375^2} = 0.35193.$$

The rest of linearized values are shown here.

Health Department	Number of Professional Workers (y_i)	Number of Workers with MPH (x_i)	Linearized Values (t_i)
1	21	14	0.135193
2	18	8	−0.090572
3	9	3	−0.096899
4	13	6	−0.053944
5	15	8	−0.006660
6	22	13	0.055609
7	30	17	0.038293
8	27	15	0.018980
Mean	19.375	10.5	

The standard error can be calculated by

$$\sqrt{\frac{Y^2\sum(t_i-\overline{t})^2}{n(n-1)}\left(1-\frac{n}{N}\right)}=29.92.$$

We have the same result as shown in Example 15.5.

For a complex survey, the above can be extended to the case of c random variables with the sample weight (w_i). Woodruff (1971) showed that the approximate variance of $\theta = f(x_1,x_2,...,x_c)$ is

$$V(\hat{\theta}) \cong V\left[\sum w_i \sum\left(\frac{\partial f}{\partial y_i}\right)y_{ij}\right].$$

This method of approximation is applied to PSU totals within the stratum. That is, the variance estimate is a weighted combination of the variation across PSUs within the same stratum.

REFERENCE

Woodruff, R. S. "A Simple Method for Approximating the Variance of a Complicated Estimate." *Journal of American Statistical Association* 66:411–414, 1971.