Chapter 4a
Probability Models

4a.1 Introduction

Chapters 2 and 3 are concerned with data description (descriptive statistics) where a sample of values of the variable X is obtained and the distribution of the observed values of X within the sample is examined. Chapter 4 is concerned with methods of obtaining a sample, by sampling or experimentation, which will allow us to use the information in the sample to make inferences about a population. The majority of the remainder of this book is devoted to methods for using a sample to make inferences about a population (inferential statistics). The basic idea of inferential statistics is to use the distribution of X in a suitably chosen sample to make inferences about the distribution of X in the population (the population distribution of X). In other words, the goal of inferential statistics is to obtain a sample of values of X and use these sample values to make inferences about the process which generates them by making inferences about the (theoretical) probabilities which these values must obey.

To make inferences about the population distribution of X we need to formulate a suitable probability model for the distribution of X. This probability model may be only partially specified with one or more unknown parameters, in which case we will need to estimate this parameter or these parameters in order to make inferences about the population distribution. This chapter provides a general discussion of probability models and presents some specific probability models.

4a.2 Probability models for a variable with a finite number of values

Let $x_1, x_2, \ldots, x_k$ denote the k distinct possible values of the variable X. A probability model for the distribution of X is an assignment of k probabilities $p_1, p_2, \ldots, p_k$ to the k possible values of X. For $i = 1, 2, \ldots, k$, p_i denotes the probability that X will take on the value x_i, in symbols we write $P(X = x_i) = p_i$. We can think of these probabilities as the theoretical relative frequencies with which X will take on these values, according to this probability model. Notice that any collection of k probabilities, each of which is between 0 and 1, which sum to 1 defines a potential probability model for the distribution of X.

We can think of a probability model for the distribution of X in terms of a box model. Imagine a box of balls where each ball is labeled with one of the k possible values of X and where the proportion of balls in the box labeled x_i is p_i, for $i = 1, 2, \ldots, k$. According to the probability model, measuring or observing the value of X is equivalent to selecting a ball at random from this box and observing the label on the ball. If we are formulating a model

for sampling from a physical population of units, then a ball represents a population unit and the balls in the box constitute the population. If X represents the outcome of some process of measurement or experimentation, then a ball represents a particular outcome and the box of balls represents the population of possible outcomes or values for X.

Tabular and graphical representations of probability models or population distributions are analogous to relative frequency distributions and bar graphs or histograms. That is, a probability distribution for a variable with k possible values is a table listing the possible values and the associated probabilities. If X is qualitative, so that the possible values $x_1, \ldots, x_k$ are simply names for k possible categories, then we can use a bar graph with k bars to give a graphical representation of the probability distribution. If X is quantitative, so that the possible values $x_1, \ldots, x_k$ are meaningful numerical values, then we can use a probability histogram to represent the distribution graphically. Notice that a graphical representation of a probability distribution uses area to represent probability.

The probability model (probability distribution) for a dichotomous (two–valued) variable is the Bernoulli model with success probability p. It is conventional to refer to one of the two possible values (outcomes) as a "success" and the other as a "failure." These generic labels are not meant to imply that observing a success is good. Rather, we can think of choosing one of the two possible outcomes and asking, "Did we observe the chosen outcome?", with the two possibilities being yes (a success) and no (a failure). The Bernoulli model with success probability p has two probabilities $P(\text{success}) = p$ and $P(\text{failure}) = q = 1 - p$, where p is between zero and one.

The two examples below provide simple examples of Bernoulli distributions with success probability $p = 2/3$. The first indicates how this distribution applies to selecting a unit at random from a physical population. The second indicates the application to observing the outcome of an experimental trial.

Example. A box containing balls of two types. Consider a box containing balls of which 2/3 are red and 1/3 are green. Define observing a red ball as a success. If we mix the balls thoroughly and then select one ball at random, the probability that we will obtain a red ball is 2/3 and the probability that we will obtain a green ball is 1/3. The corresponding probability distribution specifies the probability of observing a red ball as $P(\text{red}) = 2/3$ and the probability of observing a green ball as $P(\text{green}) = 1/3$. Thus, with success corresponding to red, the color of the ball selected follows the Bernoulli distribution with success probability $p = 2/3$.

Example. Tossing a fair die once. Suppose that a fair (balanced) die is tossed once and the number of dots on the upturned face is observed. Define a success to be the occurrence of a 1, 2, 3, or 4. Since the die is fair, the probability of a success on a single trial is $p = 4/6 = 2/3$. Therefore, with success defined as above, tossing the fair die once yields a Bernoulli variable with success probability $p = 2/3$.

The next example provides an instance where theoretical considerations in the form of a simple Mendelian inheritance model lead to a Bernoulli distribution.

Example. Inheritance in peas (flower color). In his investigations, during the years 1856 to 1868, of the chromosomal theory of inheritance Gregor Mendel performed a series of experiments on ordinary garden peas. One characteristic of garden peas that Mendel studied was the color of the flowers (red or white). When Mendel crossed a plant with red flowers with a plant with white flowers, the resulting offspring all had red flowers. But when he crossed two of these first generation plants, he observed plants with white as well as red flowers.

The gene which determines the color of the flower occurs in two forms (alleles). Let R denote the allele for red flowers (which is dominant) and r denote the allele for white flowers (which is recessive). When two plants are crossed the offspring receives one allele from each parent, thus there are four possible genotypes (combinations) RR, Rr, rR, and rr. The three genotypes RR, Rr, and rR, which include the dominant R allele, will yield red flowers while the fourth genotype rr will yield white flowers. If a red flowered RR genotype parent is crossed with a white flowered rr genotype parent, then all of the offspring will have genotype Rr and will produce red flowers. The basic Mendelian inheritance model assumes that a pair of alleles is formed by randomly choosing one allele from each parent. Under this model, if two of these first generation Rr genotype plants are crossed, each of the four possible genotypes RR, Rr, rR, and rr is equally likely and plants with white as well as red flowers will occur. Under this simple model, with each of the four genotypes having the same probability of occurring, the probability that a plant will have red flowers is $P(\text{red}) = 3/4$ and the probability that a plant will have white flowers is $P(\text{white}) = 1/4$. This Bernoulli distribution for flower color is summarized in Table 1.

Table 1. Pea plant flower color distribution.

flower color	probability
red	3/4
white	1/4

Example. Inheritance in peas (seed shape and color). We will now consider the Mendelian inheritance model for two independently inherited characteristics. In particular we will consider the characteristics seed shape, with possible shapes of round (R, dominant) and wrinkled (r, recessive), and seed color, with possible colors of yellow (Y, dominant) and green (y, recessive). If an $RRYY$ genotype plant with round yellow seeds is crossed with an $rryy$ genotype plant with wrinkled green seeds, the offspring will all have round yellow seeds and genotype $RrYy$. If two of the resulting $RrYy$ genotype plants with round yellow seeds are crossed, there are 16 equally likely possible genotypes. The nine genotypes $RRYY, RRYy, RRyY, RrYY, RrYy, RryY, rRYY, rRYy, rRyY$ yield

round yellow seeds; the three genotypes $rrYY, rrYy, rryY$ yield wrinkled yellow seeds; the three genotypes $RRyy, Rryy, rRyy$ yield round green seeds; and, the single genotype $rryy$ yields wrinkled green seeds. The fact that these 16 possible genotypes are equally likely yields the probability distribution summarized in Table 2.

Table 2. Pea plant seed shape/color distribution.

shape/color	probability
round yellow	9/16
wrinkled yellow	3/16
round green	3/16
wrinkled green	1/16

4a.3 Probability models for discrete quantitative variables

The remainder of our discussion of probability models is restricted to probability models for quantitative variables. Let us begin by considering certain parameters associated with the probability model for a quantitative variable X with possible numerical values $x_1, x_2, \ldots, x_k$ and corresponding probabilities $p_1, p_2, \ldots, p_k$. In this case the possible values are meaningful numerical values, such as counts or measurements of a physical quantity, and can be viewed as points on the number line. Thus we can discuss the shape of a distribution, the locations of representative values such as the population median or quartiles, and quantities which describe the location of and variability in the distribution.

The population mean $\mu = \mu_X$ of X (population mean of the distribution of X) is the location of the balancing point of the probability histogram or, algebraically, the weighted average of the possible values of X with weights given by the corresponding probabilities. More formally, we have

$$\mu_X = x_1 p_1 + \cdots + x_k p_k = \sum_{i=1}^{k} x_i p_i.$$

The population mean is the long run average value of X in the sense that if we observed values of X repeatedly and if these observations occurred with the specified probabilities, then in the long run the average of these values would be equal to μ_X, since x_1 would occur $p_1 \times 100\%$ of the time, x_2 would occur $p_2 \times 100\%$ of the time, and so on. For this reason the population mean of X is often called the expected value of X and denoted $\mu_X = E(X)$ with the understanding that expected value is short for long run average expected value.

The population variance $\sigma^2 = \sigma_X^2$ of X (population variance of the distribution of X) is defined as the weighted average of the squared deviations of the possible values of X

from the mean μ_X with weights given by the corresponding probabilities. More formally, we have

$$\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + \cdots + (x_k - \mu_X)^2 p_k = \sum_{i=1}^{k}(x_i - \mu_X)^2 p_i.$$

The population standard deviation $\sigma = \sigma_X$ of X is the (positive) square root of the population variance.

We can also define population quantiles or percentiles, such as the population median and the population quartiles, as the points on the number line where the probabilities (areas) in the probability histogram to the left and right of the point are equal to the appropriate values.

Example. Tossing a fair die. Suppose we toss a fair (balanced) die once and let X denote the number on the upturned face with possible values of $1, \ldots, 6$. Since the die is balanced each of the six possible values of X is equally likely to appear when we toss the die and the appropriate probability model is the uniform distribution on the set $\{1, \ldots, 6\}$ which assigns probability $1/6$ to each possible outcome. The corresponding population mean and variance are $\mu = 7/2$ and $\sigma^2 = 35/12$; the population median is also $7/2$, since this distribution is symmetric.

In general, the discrete uniform distribution on the integers $1, \ldots, k$ assigns probability $1/k$ to each of the k integers in this set of possible values. The population mean (and median) of this uniform distribution is $\mu = (k + 1)/2$ and the population variance is $\sigma^2 = (k - 1)(k + 1)/12$.

Example. Tossing a fair die, revisited. Now suppose we toss a fair (balanced) die twice and let X denote the sum of the two numbers we obtain. Taking order into account, there are 36 possible outcomes (36 combinations of the two numbers obtained) and, since the die is fair, each of these outcomes is equally likely. Since these 36 combinations are mutually exclusive we can add the probabilities corresponding to each distinct value of X yielding the distribution given in Table 3 and Figure 1. This distribution is symmetric so the population mean (and median) of the sum is $\mu = 7$. It can be shown that the population variance of the sum is $\sigma^2 = 35/6$.

Figure 1. Probability histogram for the sum when a fair die is tossed twice.
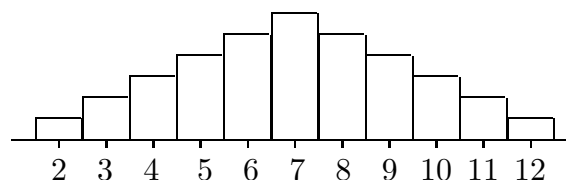
Table 3. Probability distribution for the sum when a fair die is tossed twice.

sum	probability
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

4a.4 Probability models for counts

We will now consider three probability models for counts. The first two models, the binomial and the hypergeometric, can be motivated by considering the distribution of the number of successes in a sample, selected with or without replacement, from a dichotomous population. A dichotomous population is a population of units which can be partitioned into two distinct groups; the population success group and the population failure group.

Suppose that a simple random sample of size n is selected with replacement from a dichotomous population with population success proportion p. We can view the results of this sampling process as forming a sequence of n trials, where a trial is the selection of a unit from the population. These trials possess two important properties.

1. For each trial the probability of success is p.

2. The outcomes of the trials are independent.

A sequence of trials with these properties is known as a sequence of independent Bernoulli trials with success probability p.

The probability of observing a specified sequence of successes and failures as the outcome of a sequence of independent Bernoulli trials is equal to the product of the probabilities of success (S) and failure (F) (as appropriate) on each trial. For example, with three Bernoulli trials the probability of observing the sequence SFS, denoted $P(SFS)$, is equal to the product of the probabilities of the three outcomes, *i.e.*, $P(SFS) = P(S)P(F)P(S)$. In the present context on each trial $P(S) = p$ and $P(F) = 1 - p$ and we have $P(SFS) = p(1-p)p = p^2(1-p)$.

To determine the appropriate probability model for the number of successes X in a sequence of n Bernoulli trials we need to determine the number of sequences of $S's$ and $F's$ which give each possible value of X. First consider the case with $n = 3$ trials. The

eight possible sequences of $S's$ and $F's$ and the associated probabilities and values of X are given in Table 4.

Since the eight sequences in Table 4 are mutually exclusive, we can add the probabilities corresponding to each distinct value of X to get the probability distribution in Table 5. This distribution of the number of successes in three Bernoulli trials is the binomial distribution with parameters $n = 3$ and p.

Table 4. Outcomes of a simple random sample of size 3 selected with replacement.

sequence	probability	X
FFF	$(1-p)^3$	0
SFF	$p(1-p)^2$	1
FSF	$p(1-p)^2$	1
FFS	$p(1-p)^2$	1
SSF	$p^2(1-p)$	2
SFS	$p^2(1-p)$	2
FSS	$p^2(1-p)$	2
SSS	p^3	3

Table 5. Probability distribution for the number of successes in 3 Bernoulli trials. (Binomial distribution)

sum X	probability P(X)
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	p^3

Notice that the probabilities in Table 5 are of the form $cp^x(1-p)^{n-x}$, where x is the number of successes, $n-x$ is the number of failures, and c is the number of ways to choose locations for the x $S's$ in the sequence of n $S's$ and $F's$. The number c, usually denoted by $\binom{n}{x}$ or C_x^n, is the binomial coefficient giving the number of combinations of n things taken x at a time. In general, for $x = 0, 1, \ldots, n$, the probability of observing $X = x$ successes in a sequence of n Bernoulli trials with success probability p is

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

This probability function determines the binomial distribution with parameters n and p. Most statistical software programs and some calculators will compute these probabilities
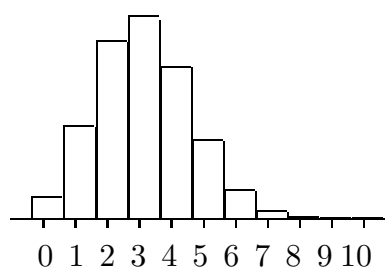
or the cumulative probabilities $P(X \leq x)$. It can be shown that the mean of this binomial distribution is $\mu = np$ and the variance is $\sigma^2 = npq = np(1 - p)$.

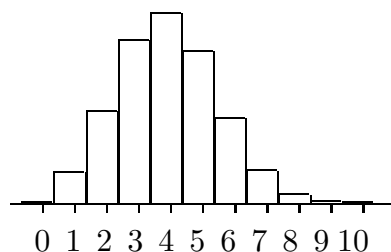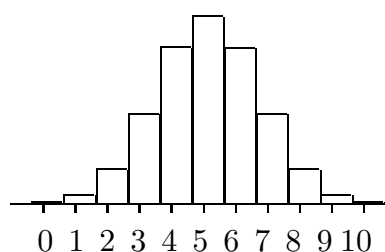**Figure 2. Probability histograms for the number of successes.**
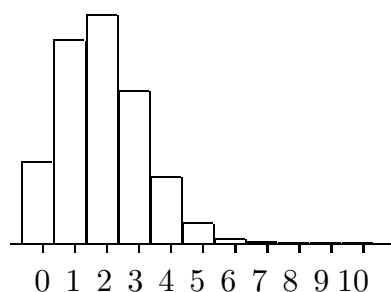
binomial $(n = 10, p = .2, \mu = 2, \sigma^2 = 1.6)$ **binomial** $(n = 10, p = .3, \mu = 3, \sigma^2 = 2.1)$

binomial $(n = 10, p = .4, \mu = 4, \sigma^2 = 2.4)$ **binomial** $(n = 10, p = .5, \mu = 5, \sigma^2 = 2.5)$

Several binomial distributions are presented in Figure 2. Notice that the binomial distribution is skewed right when p is small (near zero) and that, for $p < .5$, there is less skewness as p increases with the distribution becoming symmetric when $p = .5$. If we examined binomial distributions with $p > .5$ we would observe the analogous pattern with skewness to the left and with less skewness for p near .5.

Now suppose that a simple random sample of size n is selected without replacement from a dichotomous population with population success proportion p. As before we can view the results of this sampling process as forming a sequence of n trials, where a trial is the selection of a unit from the population. However, it is clear that neither of the two properties of independent Bernoulli trials is satisfied when we sample without replacement.

To motivate the appropriate probability model for the number of successes X in a sample chosen without replacement first consider the case with $n = 3$ trials. In this situation the distribution of X depends on the number of population units which are classified as successes M and the total number of units in the population N. Suppose that there are $N = 20$ population units of which $M = 5$ are successes. As before the probability of observing a specified sequence of $S's$ and $F's$ is equal to the product of the probabilities of the three outcomes; however, the probabilities of success and failure on a particular trial depend on what has happened on the previous trials. For example, for the sequence SFS

we write $P(SFS) = P(S)P(F|S)P(S|SF)$, where $P(F|S)$ is the conditional probability of a failure given that we have observed a success and $P(S|SF)$ is the conditional probability of a success given that we have observed a success and a failure. With $M = 5$ population success units and $N - M = 15$ population failure units we find that $P(S) = 5/20$, $P(F|S) = 15/19$, and $P(S|SF) = 4/18$, so that $P(SFS) = (5/20)(15/19)(4/18)$. The eight possible sequences of $S's$ and $F's$ and the associated probabilities and values of X are given in Table 6.

Table 6. Outcomes of a simple random sample of size 3 selected without replacement, when there are M=5 population success units and N=20 population units.

sequence	probability	X
FFF	$(15 \cdot 14 \cdot 13)/(20 \cdot 19 \cdot 18)$	0
SFF	$(5 \cdot 15 \cdot 14)/(20 \cdot 19 \cdot 18)$	1
FSF	$(15 \cdot 5 \cdot 14)/(20 \cdot 19 \cdot 18)$	1
FFS	$(15 \cdot 14 \cdot 5)/(20 \cdot 19 \cdot 18)$	1
SSF	$(5 \cdot 4 \cdot 15)/(20 \cdot 19 \cdot 18)$	2
SFS	$(5 \cdot 15 \cdot 4)/(20 \cdot 19 \cdot 18)$	2
FSS	$(15 \cdot 5 \cdot 4)/(20 \cdot 19 \cdot 18)$	2
SSS	$(5 \cdot 4 \cdot 3)/(20 \cdot 19 \cdot 18)$	3

Table 7. Probability distribution for the number of successes in a simple random sample of size 3 selected without replacement, when there are M=5 population success units and N=20 population units.

sum X	probability P(X)
0	$(15 \cdot 14 \cdot 13)/(20 \cdot 19 \cdot 18) = .3991$
1	$3(5 \cdot 15 \cdot 14)/(20 \cdot 19 \cdot 18) = .4605$
2	$3(5 \cdot 4 \cdot 15)/(20 \cdot 19 \cdot 18) = .1316$
3	$(5 \cdot 4 \cdot 3)/(20 \cdot 19 \cdot 18) = .0088$

Since the eight sequences in Table 6 are mutually exclusive, we can add the probabilities corresponding to each distinct value of X to get the probability distribution in Table 7. This distribution of the number of successes in a simple random sample of size 3 selected without replacement from a dichotomous population of $N = 20$ population units with $M = 5$ population success units is the hypergeometric distribution with parameters $M = 5$, $N = 20$, and $n = 3$. Since only M population units are classified as successes and only $N - M$ are classified as failures, we must have $x \le M$ and $n - x \le N - M$. Subject to

these restrictions the hypergeometric probabilities can be computed using the probability function

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}.$$

Most statistical software programs and some calculators will compute these probabilities or the cumulative probabilities $P(X \leq x)$. The hypergeometric distribution with parameters M, N, and n has mean
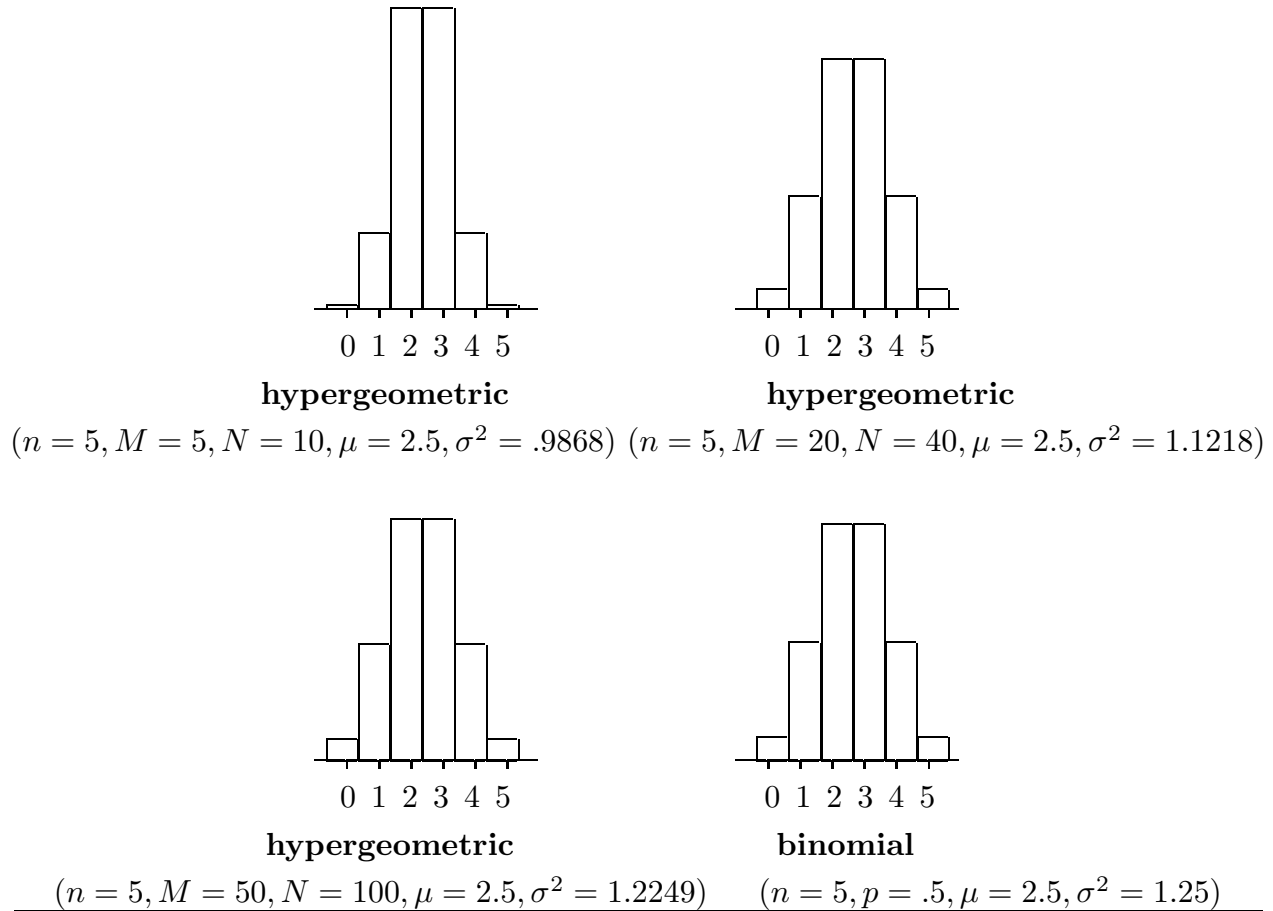
$$\mu = np = n\left(\frac{M}{N}\right)$$

and variance

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right) = npq\left(\frac{N-n}{N-1}\right),$$

where $p = M/N$ is the population success proportion.

Notice that the mean number of successes $\mu = np$ is the same whether we sample with replacement (the binomial (n, p) mean) or without replacement (the hypergeometric (M, N, n) mean with $p = M/N$). However, the variance of the hypergeometric (without replacement) distribution is smaller by a factor of $f = (N - n)/(N - 1)$, i.e., the binomial variance is npq and the hypergeometric variance is $fnpq$. The factor f is known as the finite population correction factor and its effect is most noticeable when N is small relative to n. If N is very large relative to n, then $f \approx 1$ and the two variances are essentially equal. Actually, if N is very large relative to n, then the binomial and hypergeometric distributions are essentially the same. The difference between the binomial and the hypergeometric distributions is illustrated, for $n = 5$ and $p = .5$, by the probability histograms in Figure 3. All of the distributions of Figure 3 are symmetric with mean $\mu = 2.5$, since we have $n = 5$ trials and the population success proportion is $p = .5$. Notice that as the size of the population increases the hypergeometric distributions become more similar to the binomial distribution and, in particular, there is very little difference between the hypergeometric distribution with $M = 50$ and $N = 100$ and the binomial distribution.

If X denotes the number of successes in a simple random sample of size n selected from a dichotomous population with population success proportion p, then $\hat{p} = X/n$, the proportion of successes in the sample, provides an obvious estimator of the population success proportion p. The distribution of $\hat{p}$ describes the variability (from sample to sample) in $\hat{p}$ as an estimator of p. The distribution of $\hat{p}$ is also known as the sampling distribution of $\hat{p}$. The binomial and hypergeometric distributions can be used to determine the respective distributions of $\hat{p}$ when sampling with replacement and sampling without replacement.

Figure 3. Hypergeometric and binomial probability histograms, with p=.5.

hypergeometric
$(n = 5, M = 5, N = 10, \mu = 2.5, \sigma^2 = .9868)$

hypergeometric
$(n = 5, M = 20, N = 40, \mu = 2.5, \sigma^2 = 1.1218)$

hypergeometric
$(n = 5, M = 50, N = 100, \mu = 2.5, \sigma^2 = 1.2249)$

binomial
$(n = 5, p = .5, \mu = 2.5, \sigma^2 = 1.25)$

If X denotes the number of successes in a simple random sample of size n selected with replacement from a dichotomous population with population success proportion p, then the possible values of X are $x = 0, 1, \ldots, n$ and the corresponding possible values of $\hat{p}$ are $x/n = 0, 1/n, 2/n, \ldots, 1$. Thus, when the sample is selected with replacement, the probability that $X = x$ is equal to the probability that $\hat{p} = x/n$, *i.e.*, for $x = 0, 1, \ldots, n$,

$$P\left(\hat{p} = \frac{x}{n}\right) = P(X = x) = \binom{n}{x}p^x(1-p)^{n-x}.$$

In this case, the mean of the sampling distribution of $\hat{p}$ is $E(\hat{p}) = p$ $(E(\hat{p}) = E(X)/n)$ and the variance of the sampling distribution of $\hat{p}$ is $\text{var}(\hat{p}) = pq/n$ $(\text{var}(\hat{p}) = \text{var}(X)/n^2)$. Notice that this sampling distribution does not depend on the size of the population.

If X denotes the number of successes in a simple random sample of size n selected without replacement from a dichotomous population with population success proportion p, then the sampling distribution of $\hat{p}$ depends on the size of the population. Let N denote the size of the dichotomous population being sampled and let M denote the number of

population units classified as successes. When the sample is selected without replacement, for $x = 0, 1, \ldots, n$, subject to the restrictions $x \leq M$ and $n - x \leq N - M$,

$$P\left(\hat{p} = \frac{x}{n}\right) = P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}.$$

In this case, the mean and variance of the sampling distribution of $\hat{p}$ are

$$E(\hat{p}) = p = \left(\frac{M}{N}\right)$$

and

$$\text{var}(\hat{p}) = \left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{n(N-1)}\right) = \frac{pq}{n}\left(\frac{N-n}{N-1}\right),$$

where $p = M/N$ is the population success proportion. Note that, as before, $E(\hat{p}) = E(X)/n$ and $\text{var}(\hat{p}) = \text{var}(X)/n^2$.

We will now consider the Poisson distribution which provides a realistic model for counts of "rare events" in many practical settings. Consider a sequence of events occurring randomly in time or space and a count such as the number of radioactive particle emissions per unit time, the number of meteorites that collide with a satellite during a single orbit, the number of defects per unit length of some material, or the number of weed seeds per unit volume in a large batch of wheat seeds. We can picture the time (or location) of each occurrence as a point on the positive part of the number line. Consider the following assumptions about the times (locations) of these occurrences:

  1) The probability of exactly one occurrence in a small interval of length t is approximately νt, where $\nu > 0$ is the mean rate at which events occur per unit time (the mean rate of occurrence).

  2) The probability of more than one occurrence in a small interval of length t is negligible compared to the probability of exactly one occurrence in a small interval of length t.

  3) The numbers of occurrences in non–overlapping intervals are independent in the sense that information concerning the number of events in one interval reveals nothing about the number of events in the other interval.

If we let X denote the number of occurrences in a period of length t, then these three assumptions imply that X follows the Poisson distribution with parameter $\lambda = \nu t$. The possible values of X are $0, 1, \ldots$, with no theoretical upper bound on the value, and for $\lambda > 0$ the Poisson probabilities can be computed using the probability function

$$P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda},$$

where $e \approx 2.718$ is the base of the natural logarithm and $x! = x(x-1)\cdots 1$ is x factorial. The mean and variance of the Poisson distribution with parameter λ are both equal to λ.

Example. Radioactive disintegrations. This example is taken from Feller (1957), p. 149 and Cramér (1946) p. 436. In a famous experiment by Rutherford, Chadwick, and Ellis (*Radiations from Radioactive Substances*, Cambridge, 1920) a radioactive substance was observed during 2608 consecutive time intervals of length $t = 7.5$ seconds each. The number of particles reaching a counter was recorded for each period. The results are summarized in Figure 4 and Table 8. (In Table 8 the observations greater than or equal to 10 are grouped together. The data actually contain 10 tens, 4 elevens, and 2 twelves.) The last column of Table 8 contains expected relative frequencies (probabilities) computed using a Poisson model with λ estimated from these data. These Poisson probabilities appear to match the observed relative frequencies fairly well. A formal test of the goodness of fit of this Poisson model to these data, which is discussed in Chapter 11, indicates that the model does fit well ($\chi^2 = 12.885$, 9 d.f., P–value .17).

Table 8. Relative frequency distribution for radioactive disintegrations.

number	observed frequency	observed relative frequency	expected relative frequency
0	57	.0219	.0209
1	203	.0778	.0807
2	383	.1469	.1562
3	525	.2013	.2015
4	532	.2040	.1949
5	408	.1564	.1509
6	273	.1047	.0973
7	139	.0533	.0538
8	45	.0173	.0260
9	27	.0104	.0112
≥ 10	16	.0051	.0065
total	2608	.9991	.9999

Figure 4. Histogram for radioactive disintegrations (with ≥ 10 expanded).
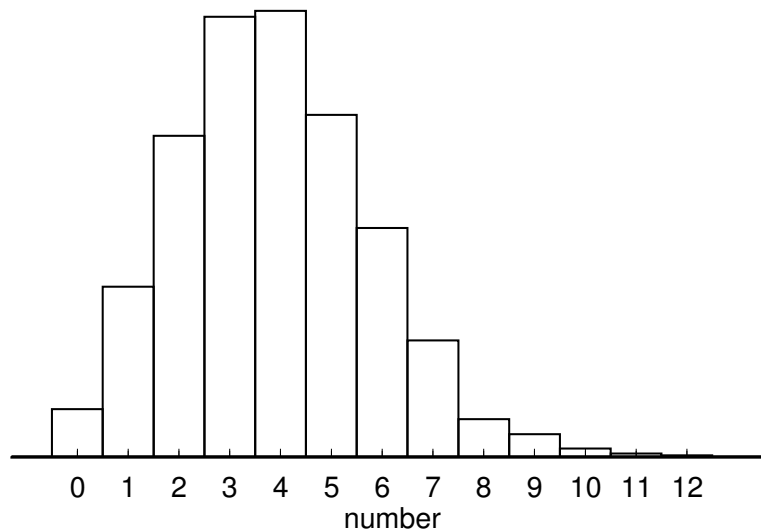
    Example. Bacteria counts. This example is taken from Feller (1957), p.153. The original source is T. Matuszewsky, J. Supinska, and J. Neyman (1936), *Zentralblatt für Bakteriologie, Parasitenkunde und Infektionskrankrankheiten*, II Abt., **95**. A Petri dish with bacteria colonies was examined under a microscope. The dish was divided into small squares and the number of bacteria colonies, visible as dark spots, was recorded for each square. In this example t is the area the square within which the count is determined and we will take this area to be one. If the bacteria colonies were randomly distributed over the Petri dish, without being clustered together, then the Poisson model should hold. The results for one of several experiments are summarized in Figure 5 and Table 9. The last column of Table 9 contains expected relative frequencies (probabilities) computed using a Poisson model, with λ estimated from these data. In this example the observed relative frequency in the "≥ 6" line is for "exactly 6", but, the expected relative frequency is for all values greater than or equal to 6. These Poisson probabilities appear to match the observed relative frequencies fairly well. Therefore, the evidence supports the contention that the bacteria colonies are randomly distributed over the Petri dish. A formal test of the goodness of fit of this Poisson model to these data, which is discussed in Chapter 11, indicates that the model does fit well ($\chi^2 = .8386$, 5 d.f., P–value .9745).

Figure 5. Histogram for bacteria counts.
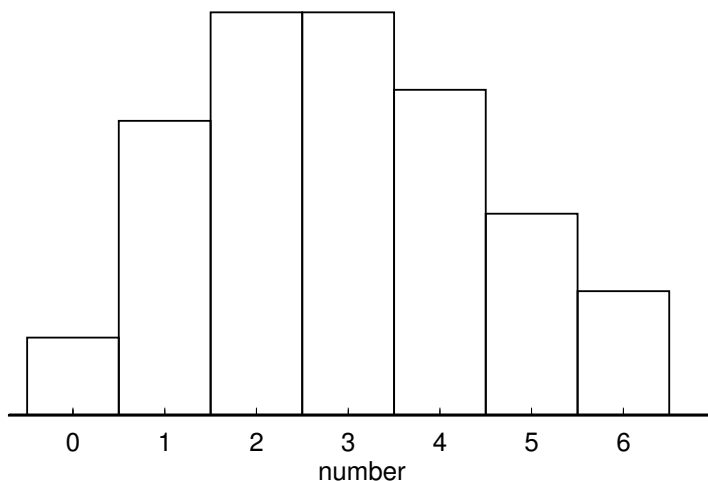
Table 9. Relative frequency distribution for bacteria counts.

number	observed frequency	observed relative frequency	expected relative frequency
0	5	.0424	.0533
1	19	.1610	.1562
2	26	.2203	.2290
3	26	.2203	.2239
4	21	.1780	.1641
5	13	.1102	.0962
≥ 6	8	.0678	.0772
total	118	1.0000	.9999

4a.5 Probability models for continuous quantitative variables

We will now consider probability models for the distribution of a continuous quantitative variable. A probability model for the distribution of a continuous variable X can be represented by a density curve. A **density curve** is a nonnegative curve for which the area under the curve (over the x–axis) is one. We can think of the density curve as a smooth version of a probability histogram with the rectangles of the histogram replaced by a smooth curve indicating where the tops of the rectangles would be. With a continuous variable X it does not make sense to talk about the probability that X would take on a particular value, after all if we defined positive probabilities for the infinite collection (continuum) of possible values of X these probabilities could not add up to one. It does, however, make sense to talk about the probability that X will take on a value in a specified interval or range of values. Given two constants $a < b$ the probability that X takes on a

value in the interval from a to b, denoted by $P(a \le X \le b)$, is equal to the area under the density curve over the interval from a to b on the x–axis. Areas of this sort based on the density curve give the probabilities which a single value of X, chosen at random from the infinite population of possible values of X, will satisfy.

Given a probability model for the distribution of a continuous variable X, *i.e.*, given a density curve for the distribution of the continuous variable X, we can define population parameters which characterize relevant aspects of the distribution. For example, we can define the population mean μ as the balance point of the unit mass bounded by the density curve and the number line. We can also think of the population mean as the weighted average of the infinite collection of possible values of X with weights determined by the density curve. We can similarly define the population median M as the point on the number line where a vertical line would divide the area under the density curve into two equal areas (each of size one–half).

The most widely used continuous probability model is the normal probability model or normal distribution. The normal distribution with mean μ and standard deviation σ can be characterized by its density curve, which is the familiar bell shaped curve. The normal density curve corresponds to the probability density function

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

The standard normal density curve, which has mean $\mu = 0$ and standard deviation $\sigma = 1$, is shown in Figure 6.

Figure 6. The standard normal density curve.
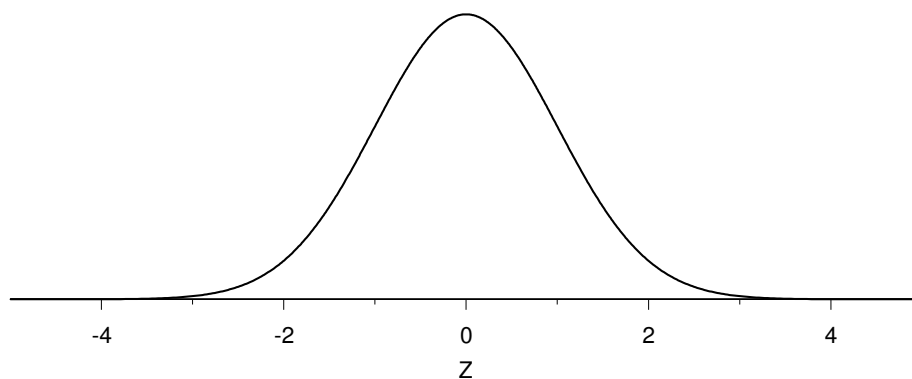
The normal distribution with mean μ and its density curve are symmetric around μ, *i.e.*, if we draw a vertical line through μ, then the two sides of the density curve are mirror images of each other. Therefore the mean of a normal distribution μ is also the median of the normal distribution. The mean μ locates the normal distribution on the number line so that if we hold σ constant and change the mean μ, the normal distribution is simply shifted

along the number line until it is centered at the new mean. In other words, holding σ fixed and changing μ simply relocates the density curve on the number line; it has no effect on the shape of the curve. Figure 7 provides the density curves for normal distributions with respective means $\mu = 0$ and $\mu = 2$ and common standard deviation $\sigma = 1$.

Figure 7. Normal distributions with common standard deviation one and means of zero and two.
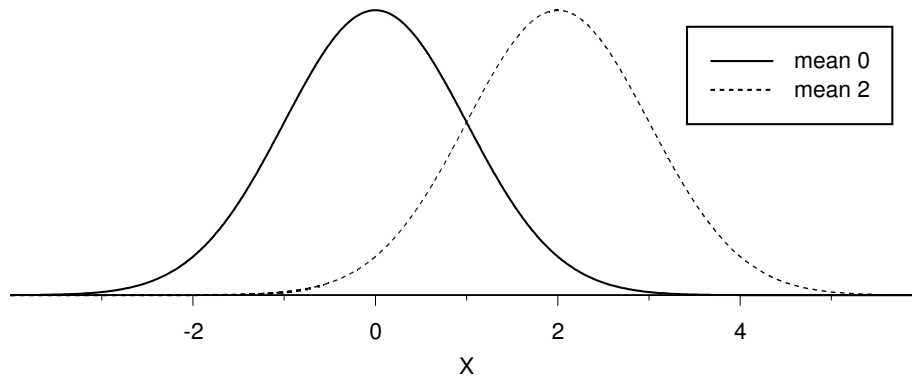
Figure 8. Normal distributions with common mean zero and standard deviations one–half, one, and two.
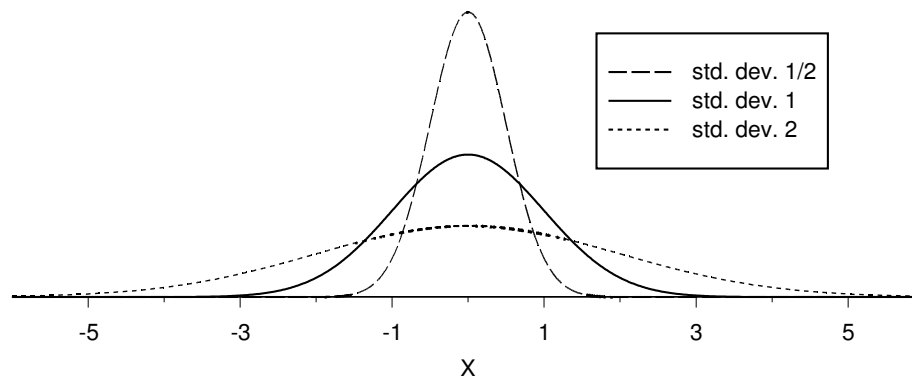
The standard deviation σ indicates the amount of variability in the normal distribution. If we hold μ fixed and increase the value of σ, then the normal density curve becomes flatter, while retaining its bell–shape, indicating that there is more variability in the distribution. Similarly, if we hold μ fixed and decrease the value of σ, then the normal density curve becomes more peaked around the mean μ, while retaining its bell–shape, indicating that there is less variability in the distribution. Normal distributions with mean $\mu = 0$ and respective standard deviations $\sigma = .5$, $\sigma = 1$, and $\sigma = 2$ are plotted in Figure 8.

Computer programs and many calculators can be used to compute normal probabilities or equivalently to compute areas under the normal density curve. These probabilities can also be calculated using tables of standard normal distribution probabilities such as Table

10 at the end of this chapter. Recall that the standard normal distribution is the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The relationship between the standard normal variable Z and the normal variable X, which has mean μ and standard deviation σ, is

$$Z = \frac{X - \mu}{\sigma} \quad \text{or equivalently} \quad X = \mu + Z\sigma.$$

This relationship implies that a probability statement about the normal variable X can be re–expressed as a probability statement about the standard normal variable Z by re–expressing the statement in terms of standard deviation units from the mean. Given two constants $a < b$, observing a value of X between a and b (observing $a \leq X \leq b$) is equivalent to observing a value of $Z = (X - \mu)/\sigma$ between $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$ (observing $(a - \mu)/\sigma \leq (X - \mu)/\sigma \leq (b - \mu)/\sigma$). Furthermore, $Z = (X - \mu)/\sigma$ behaves in accordance with the standard normal distribution so that the probability of observing a value of X between a and b, denoted by $P(a \leq X \leq b)$, is equal to the probability that the standard normal variable Z takes on a value between $(a - \mu)/\sigma$ and $(b - \mu)/\sigma$, which is denoted by $P[(a - \mu)/\sigma < Z < (b - \mu)/\sigma]$, i.e.,

$$P(a < X < b) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right).$$

In terms of areas this probability equality says that the area under the normal density curve with mean μ and standard deviation σ over the interval from a to b is equal to the area under the standard normal density curve over the interval from $(a - \mu)/\sigma$ to $(b - \mu)/\sigma$. Similarly, given constants $c < d$, we have the analogous result that

$$P(c < Z < d) = P(\mu + c\sigma < X < \mu + d\sigma).$$

Most tables of the standard normal distribution and many computer programs provide cumulative standard normal probabilities of the form $P(Z \leq a)$ for selected values of a. To use these cumulative probabilities to compute a probability of the form $P(a \leq Z \leq b)$ note that

$$P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$$

and note that the symmetry of the normal distribution implies that

$$P(Z \leq -a) = P(Z \geq a) = 1 - P(Z \leq a).$$

Calculators will usually provide probabilities of the form $P(a \leq Z \leq b)$ directly.

Example. Heights of adult males. Consider the heights (in inches) of adult males born in the United Kingdom (including the whole of Ireland) which are summarized in the Table 8 of Section 3.3.

These height data provide a good illustration of the fact that normal distributions often provide very good models for a population of physical measurements of individuals, such as heights or weights. Figure 9 provides a histogram for this height distribution and the density curve for a normal distribution chosen to model these data. You can see that the normal distribution provides a very reasonable model for the heights of adult males born in the United Kingdom.

Figure 9. Histogram and normal density curve for the UK height example.

Probability

z'
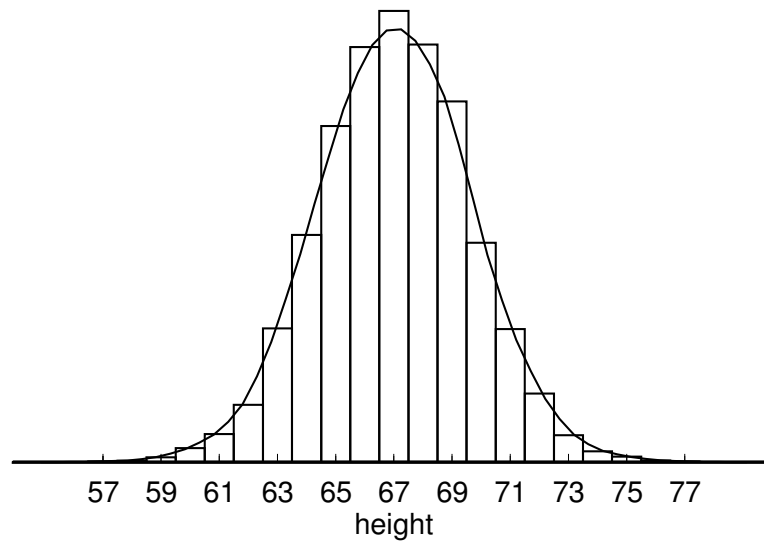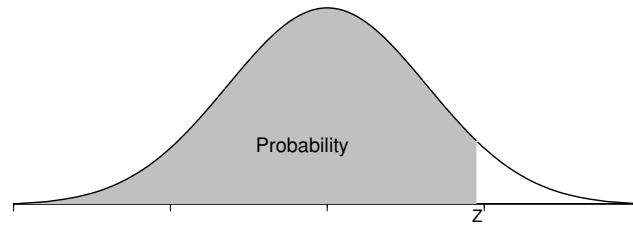
Table 10. Cumulative normal probabilities.
(Areas under the standard normal curve to the left of Z.)

Z	Second decimal place in Z									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

continued on next page

Table 10 (continuation). Cumulative normal probabilities.
(Areas under the standard normal curve to the left of Z.)

Z	Second decimal place in Z									
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999