## 4.1 THE BINOMIAL TEST

### 4.1.1 Function and Rationale

There are many populations which are conceived as consisting of only two classes. Examples of such classes are: male and female, literate and illiterate, member and nonmember, married and single, institutionalized and ambulatory. For such cases, all of the possible observations from the population will fall into one of two discrete categories. Such a population is usually called a *binary population* or a dichotomous population.

Suppose a population consists of only two categories or classes. Then each observation ($X$) sampled from the population may take on one of two values, depending on the category sampled. We could denote the possible values of the random variable by using any pair of values, but it is most convenient to denote



each outcome as either 1 or 0. We shall assume further that the probability of sampling an object from the first category is $p$, and the probability of sampling an object from the other category is $q = 1 - p$. That is,

$$P[X = 1] = p \qquad \text{and} \qquad P[X = 0] = 1 - p = q$$

It is also assumed that each probability is constant regardless of the number of objects sampled or observed.

Although the value of $p$ may vary from population to population, it is fixed for any one population. However, even if we know (or assume) the value of $p$ for some population, we cannot expect that a random sample of observations from the population will contain exactly the proportions $p$ and $1 - p$ for each of the two categories. Random sampling will usually prevent the sample from duplicating precisely the population values of $p$ and $q$. For example, we may know from the official records that the voters in a certain county are evenly split between the Republican and Democratic parties in registration. But a random sample of the registered voters in that county might contain 47 percent Democrats and 53 percent Republicans, or even 56 percent Democrats and 44 percent Republicans. Such differences between the observed and the population values arise because of "chance" or random fluctuations in the observations. We should not be surprised by small deviations from the population values; however, large deviations—although possible—are unlikely.

The *binomial distribution* is used to determine the probabilities of the possible outcomes we might observe if we sampled from a binomial population. If our hypothesis is $H_0$: $p = p_0$, we can calculate the probabilities of the various outcomes when we assume that $H_0$ is true. The test will tell us whether it is reasonable to believe that the proportions (or frequencies) of the two categories in our sample could have been drawn from a population with the hypothesized values of $p_0$ and $1 - p_0$. For convenience in discussing the binomial distribution, we shall denote the outcome $X = 1$ as "success" and the outcome $X = 0$ as "failure." In addition, in a series of $N$ observations,

$$Y = \sum_{i=1}^{N} X_i$$

is the number of "successes" or the number of outcomes of the type $X = 1$.

### 4.1.2 Method

In a sample of size $N$, the probability of obtaining $k$ objects in one category and $N - k$ objects in the other category is given by

$$P[Y = k] = \binom{N}{k} p^k q^{N-k} \qquad k = 0, 1, \ldots, N \tag{4.1}$$

where $p$ = the proportion of observations expected where $X = 1$
$q$ = the proportion of observations expected where $X = 0$

and
$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \quad \text{(see Footnote 1)}$$

Appendix Table E has values of $P[Y = k]$ for various values of $N$ and $p$.

A simple illustration will clarify Eq. (4.1). Suppose a fair die is rolled five times. What is the probability that exactly two of the rolls will show "six"? In this case, $Y$ is the random variable (the outcome of five tosses of the die), $N$ = the number of rolls (5), $k$ = the observed number of sixes (2), $p$ = the expected proportion of sixes ($\frac{1}{6}$), and $q = \frac{5}{6}$. The probability that exactly two of the five rolls will show six is given by Eq. (4.1):

$$P[Y = k] = \binom{N}{k} p^k (1-p)^{N-k}$$

$$P[Y = 2] = \frac{5!}{2!3!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = .16$$

The application of the formula to the problem shows us that the probability of obtaining exactly two "sixes" when rolling a fair die five times is $p = .16$.

Now when we test hypotheses, the question is usually *not*, "What is the probability of obtaining *exactly* the values which were observed?" Rather, we usually ask, "What is the probability of obtaining values *as extreme or more extreme than* the observed value when we assume the data are generated by a particular process?" To answer questions of this type, the probability desired is

$$P[Y \geq k] = \sum_{i=k}^{N} \binom{N}{i} p^i q^{N-i} \tag{4.2}$$

In other words, we sum the probability of the observed outcome with the probabilities of outcomes which are even more extreme.

Suppose now that we want to know the probability of obtaining two *or fewer* sixes when a fair die is rolled five times. Here again $N = 5$, $k = 2$, $p = \frac{1}{6}$, and $q = \frac{5}{6}$. Now the probability of obtaining 2 or fewer sixes is denoted $P[Y \leq 2]$. From Eq. (4.1) the probability of obtaining 0 sixes is $P[Y = 0]$, the probability of obtaining one six is $P[Y = 1]$, etc. Using Eq. (4.2) we have

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

[1] $N!$ is "*N factorial*," which is defined as

$$N! = N(N-1)(N-2)\cdots(2)(1).$$

For example, $4! = (4)(3)(2)(1) = 24$ and $5! = 120$. By definition, $0! = 1$. Appendix Table W gives factorials for values of $N$ through 20. Appendix Table X gives binomial coefficients

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

for values of $N$ through 20.



That is, the probability of obtaining two or fewer sixes is the sum of three probabilities. If we use Eq. (4.1) to determine the three probabilities, we have

$$P[Y = 0] = \frac{5!}{0!5!}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^5 = .40$$

$$P[Y = 1] = \frac{5!}{1!4!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 = .40$$

$$P[Y = 2] = \frac{5!}{2!3!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = .16$$

and thus

$$\begin{aligned} P[Y \leq 2] &= P[Y = 0] + P[Y = 1] + P[Y = 2] \\ &= .40 + .40 + .16 \\ &= .96 \end{aligned}$$

We have determined that the probability under $H_0$ (the assumption of a fair die) of obtaining two or fewer sixes when a die is rolled five times is $p = .96$.

**SMALL SAMPLES.** In the one-sample case, when binary categories are used, a common hypothesis is $H_0$: $p = \frac{1}{2}$. Table D of the Appendix gives the one-tailed probabilities associated with the occurrence of various values as extreme as $k$ under the null hypothesis $H_0$: $p = \frac{1}{2}$. When referring to Appendix Table D, let $k$ equal the smaller of the observed frequencies. This table is useful when $N \leq 35$. Although Eq. (4.2) could be used, the table is more convenient. Table D gives the probabilities associated with the occurrence of various values as small as $k$ for various $N$'s. For example, suppose we observe seven successes and three failures. Here $N = 10$ and $k = 7$. Table D shows that the one-tailed probability of occurrence under $H_0$: $p = \frac{1}{2}$ for $Y \leq 3$ when $N = 10$ to be .172. Because of the symmetry of the binomial distribution when $p = \frac{1}{2}$, $P[Y \geq k] = P[Y \leq N - k]$. Thus, in this example, $P[Y \leq 3] = P[Y \geq 7] = .172$.

The probabilities given in Table D are one-tailed. A one-tailed test is used when we have predicted in advance which of the two categories should contain the smaller number of cases. When the prediction is simply that the two frequencies will differ, a two-tailed test would be used. For a two-tailed test, the probability values in Appendix Table D would be doubled. Thus for $N = 10$ and $k = 7$, the two-tailed probability associated with the occurrence under $H_0$ is .344.

The following example illustrates the use of the binomial test in a study in which $H_0$: $p = \frac{1}{2}$.

**Example 4.1.** In a study of the effects of stress,[2] an experimenter taught 18 college students 2 different methods of tying the same knot. Half of the subjects (randomly selected from

[2] Barthol, R. P., and Ku, N. D. (1953). Specific regression under a non-related stress situation. *American Psychologist*, **10**, 482.

**TABLE 4.1**
**Knot-tying method chosen under stress**

| | Method chosen | | |
|---|---|---|---|
| | First-learned | Second-learned | Total |
| Frequency | 16 | 2 | 18 |

the group of 18) learned method $A$ first, and half learned method $B$ first. Later—at midnight, after a 4-hour final examination—each subject was asked to tie the knot. The prediction was that stress would induce regression, i.e., that the subjects would revert to the first-learned method of tying the knot. Each subject was categorized according to whether the subject used the knot-tying method learned first or the one learned second when asked to tie the knot under stress.

*i. Null hypothesis.* $H_0$: $p = q = \frac{1}{2}$, that is, there is no difference between the probability of using the first-learned method under stress ($p$) and the probability of using the second-learned method under stress ($q$). Any difference between the frequencies which may be observed is of such magnitude that it might be expected in a sample from the population of possible results under $H_0$. $H_1$: $p > q$, that is, when under stress, the probability of using the first-learned method is greater than the probability of using the second-learned method.

*ii. Statistical test.* The binomial test is chosen because the data are in two discrete categories and the design is of the one-sample type. Since methods $A$ and $B$ were randomly assigned to being first-learned and second-learned, there is no reason to think that the first-learned method would be preferred to the second-learned under $H_0$, and thus $p = q = \frac{1}{2}$.

*iii. Significance level.* Let $\alpha = .01$ and $N$ is the number of cases = 18.

*iv. Sampling distribution.* The sampling distribution is given in Eq. (4.2) above. However, when $N \leq 35$, and when $p = q = \frac{1}{2}$, Table D gives the probabilities associated with the occurrence under $H_0$ of observed value as small as $k$, and thus it is not necessary to calculate the sampling distribution directly in this example.

*v. Rejection region.* The region of rejection consists of all values of $Y$ (where $Y$ is the number of subjects who used the second-learned method under stress), which are so small that the probability associated with their occurrence under $H_0$ is equal to or less than $\alpha = .01$. Since the direction of the difference was predicted in advance, the region of rejection is one-tailed.

*vi. Decision.* In the experiment, all but two of the subjects used the first-learned method when asked to tie the knot under stress (late at night after a long, final examination). These data are shown in Table 4.1. In this case, $N$ is the number of independent observations = 18. $k$ is the smaller frequency = 2. Appendix Table D shows that for $N = 18$, the probability associated with $k \leq 2$ is .001. Inasmuch as this probability is smaller than $\alpha = .01$, the decision is to reject $H_0$ in favor of $H_1$. Thus we conclude that $p > q$, that is, that people under stress revert to the first-learned of two methods.

**LARGE SAMPLES.** Appendix Table D cannot be used when $N$ is larger than 35. However it can be shown that, as $N$ increases, the binomial distribution tends toward the normal distribution. More precisely, as $N$ increases, the distribution of the variable $Y$ approaches a normal distribution. The tendency is rapid when $p$ is close to $\frac{1}{2}$, but is slower when $p$ is close to 0 or 1. That is, the greater the disparity between $p$ and $q$, the larger must be $N$ before the approximation is usefully close to the normal distribution. When $p$ is near $\frac{1}{2}$, the approximation may be used for a statistical test when $N > 25$. When $p$ is near 0 or 1, a rule of thumb is that $Npq$ should be greater than 9 before the statistical test based on the normal approximation is sufficiently accurate to use. Within these limitations, the sampling distribution of $Y$ is approximately normal, with mean $Np$ and variance $Npq$, and, therefore, $H_0$ may be tested by



$$z = \frac{x - \mu_x}{\sigma_x} = \frac{Y - Np}{\sqrt{Npq}} \tag{4.3}$$

where $z$ is approximately normally distributed with mean 0 and standard deviation 1.

The approximation to the normal distribution becomes better if a correction for "continuity" is used. The correction is necessary because the normal distribution is continuous while the binomial distribution involves discrete variables. To correct for continuity, we regard the observed frequency $Y$ of Eq. (4.3) as occupying an interval, the lower limit of which is one-half unit below the observed frequency while the upper limit is one-half unit above the observed frequency. The correction for continuity consists of reducing, by .5, the difference between the observed value of $Y$ and its expected value $\mu_Y = Np$. Therefore, when $Y < \mu_Y$ we add .5 to $Y$, and when $Y > \mu_Y$ we subtract .5 from $Y$. That is, the observed difference is reduced by .5. Thus $z$ becomes

$$z = \frac{(Y \pm .5) - Np}{\sqrt{Npq}} \tag{4.4}$$

where $Y + .5$ is used when $Y < Np$, and $Y - .5$ is used when $Y > Np$. The value of $z$ obtained by the application of Eq. (4.4) is asymptotically normally distributed with mean 0 and variance 1. Therefore, the significance of an obtained $z$ may be determined by reference to Table A of the Appendix. Table A gives the one-tailed probability associated with the occurrence under $H_0$ of values as extreme as an observed $z$. (If a two-tailed test is required, the probability yielded by Appendix Table A must be doubled.)

To show how good this approximation is when $p = \frac{1}{2}$ even for $N < 25$, we can apply it to the knot-tying data discussed earlier. In that case, $N = 18$, $Y = 2$, and $p = q = \frac{1}{2}$. For these data, $Y < Np$, that is, $2 < 9$, and by Eq. (4.4),

$$\begin{aligned} z &= \frac{(2 + .5) - (18)(1/2)}{\sqrt{(18)(1/2)(1/2)}} \\ &= -3.06 \end{aligned}$$

Appendix Table A shows that a value of $z$ as extreme as $-3.06$ has a one-tailed probability associated with its occurrence under $H_0$ of .0011. This is essentially the same probability we found by the other analysis, which used a table of exact probabilities. However, remember that in this example $p = \frac{1}{2}$, so the approximation should do well.

### 4.1.3 Summary of Procedure

In brief, these are the steps in the use of the binomial test of $H_0$: $p = \frac{1}{2}$:

1. Determine $N$ = the total number of cases observed.
2. Determine the frequencies of the observed occurrences in each of the two categories.
3. The method of finding the probability of occurrence under $H_0$ of the observed values, or values even more extreme, depends upon the sample size:
   (*a*) If $N \leq 35$, Appendix Table D gives the one-tailed probabilities under $H_0$ of various values as small as an observed $Y$. Specify $H_1$, and determine whether the test should be one-tailed or two-tailed.
   (*b*) If $N > 35$, test $H_0$ by using Eq. (4.4). Appendix Table A gives the probability associated with the occurrence under $H_0$ of values as large as an observed $z$. Table A gives one-tailed probabilities; for a two-tailed test, double the obtained probability.
4. If the probability associated with the observed value of $Y$ or an even more extreme value is equal to or less than $\alpha$, reject $H_0$. Otherwise, do not reject $H_0$.

### 4.1.4 Power-Efficiency

Since there is no parametric technique applicable to data measured as a dichotomous variable, it is not meaningful to inquire about the power-efficiency of the binomial test when used with such data.

If a continuous variable is dichotomized and the binomial test is used on the resulting data, the test may be wasteful of data. In such cases, the binomial test has power-efficiency (in the sense as defined in Chap. 3) of 95 percent for $N = 6$, decreasing to an asymptotic efficiency of $2/\pi = 63$ percent as $N$ increases. However, if the data are basically dichotomous, even though the variable has an underlying continuous distribution, the binomial test may have no more powerful and practicable alternative.

### 4.1.5 References

For other discussions of binomial distribution and its applications, see Hays (1981) or Bailey (1971).

**TABLE D**

**Table of probabilities associated with values as small as (or smaller than) observed values of $k$ in the binomial test**

Given in the body of the table are one-tailed probabilities under $H_0$ for the binomial test when $p = q = \frac{1}{2}$.

Entries are $P[Y \leq k]$. Note that entries may also be read as $P[Y \geq N - k]$

| N | k = 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 062 | 312 | 688 | 938 | 1.0 | | | | | | | | | | | | | |
| 5 | 031 | 188 | 500 | 812 | 969 | 1.0 | | | | | | | | | | | | |
| 6 | 016 | 109 | 344 | 656 | 891 | 984 | 1.0 | | | | | | | | | | | |
| 7 | 008 | 062 | 227 | 500 | 773 | 938 | 992 | 1.0 | | | | | | | | | | |
| 8 | 004 | 035 | 145 | 363 | 637 | 855 | 965 | 996 | 1.0 | | | | | | | | | |
| 9 | 002 | 020 | 090 | 254 | 500 | 746 | 910 | 980 | 998 | 1.0 | | | | | | | | |
| 10 | 001 | 011 | 055 | 172 | 377 | 623 | 828 | 945 | 989 | 999 | 1.0 | | | | | | | |
| 11 | | 006 | 033 | 113 | 274 | 500 | 726 | 887 | 967 | 994 | 999+ | 1.0 | | | | | | |
| 12 | | 003 | 019 | 073 | 194 | 387 | 613 | 806 | 927 | 981 | 997 | 999+ | 1.0 | | | | | |
| 13 | | 002 | 011 | 046 | 133 | 291 | 500 | 709 | 867 | 954 | 989 | 998 | 999+ | 1.0 | | | | |
| 14 | | 001 | 006 | 029 | 090 | 212 | 395 | 605 | 788 | 910 | 971 | 994 | 999 | 999+ | 1.0 | | | |
| 15 | | | 004 | 018 | 059 | 151 | 304 | 500 | 696 | 849 | 941 | 982 | 996 | 999+ | 999+ | 1.0 | | |
| 16 | | | 002 | 011 | 038 | 105 | 227 | 402 | 598 | 773 | 895 | 962 | 989 | 998 | 999+ | 999+ | 1.0 | |
| 17 | | | 001 | 006 | 025 | 072 | 166 | 315 | 500 | 685 | 834 | 928 | 975 | 994 | 999 | 999+ | 999+ | 1.0 |
| 18 | | | 001 | 004 | 015 | 048 | 119 | 240 | 407 | 593 | 760 | 881 | 952 | 985 | 996 | 999 | 999+ | 999+ |
| 19 | | | | 002 | 010 | 032 | 084 | 180 | 324 | 500 | 676 | 820 | 916 | 968 | 990 | 998 | 999+ | 999+ |
| 20 | | | | 001 | 006 | 021 | 058 | 132 | 252 | 412 | 588 | 748 | 868 | 942 | 979 | 994 | 999 | 999+ |

*Note:* Decimal points omitted, and values less than .0005 are omitted.

**TABLE D (continued)**

| N | k = 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 21 | | | | 001 | 004 | 013 | 039 | 095 | 192 | 332 | 500 | 668 | 808 | 905 | 961 | 987 | 996 | 999 |
| 22 | | | | | 002 | 008 | 026 | 067 | 143 | 262 | 416 | 584 | 738 | 857 | 933 | 974 | 992 | 998 |
| 23 | | | | | 001 | 005 | 017 | 047 | 105 | 202 | 339 | 500 | 661 | 798 | 895 | 953 | 983 | 995 |
| 24 | | | | | 001 | 003 | 011 | 032 | 076 | 154 | 271 | 419 | 581 | 729 | 846 | 924 | 968 | 989 |
| 25 | | | | | | 002 | 007 | 022 | 054 | 115 | 212 | 345 | 500 | 655 | 788 | 885 | 946 | 978 |
| 26 | | | | | | 001 | 005 | 014 | 038 | 084 | 163 | 279 | 423 | 577 | 721 | 837 | 916 | 962 |
| 27 | | | | | | 001 | 003 | 010 | 026 | 061 | 124 | 221 | 351 | 500 | 649 | 779 | 876 | 939 |
| 28 | | | | | | | 002 | 006 | 018 | 044 | 092 | 172 | 286 | 425 | 575 | 714 | 828 | 908 |
| 29 | | | | | | | 001 | 004 | 012 | 031 | 068 | 132 | 229 | 356 | 500 | 644 | 771 | 868 |
| 30 | | | | | | | 001 | 003 | 008 | 021 | 049 | 100 | 181 | 292 | 428 | 572 | 708 | 819 |
| 31 | | | | | | | | 002 | 005 | 015 | 035 | 075 | 141 | 237 | 360 | 500 | 640 | 763 |
| 32 | | | | | | | | 001 | 004 | 010 | 025 | 055 | 108 | 189 | 298 | 430 | 570 | 702 |
| 33 | | | | | | | | 001 | 002 | 007 | 018 | 040 | 081 | 148 | 243 | 364 | 500 | 636 |
| 34 | | | | | | | | | 001 | 005 | 012 | 029 | 061 | 115 | 196 | 304 | 432 | 568 |
| 35 | | | | | | | | | 001 | 003 | 008 | 020 | 045 | 088 | 155 | 250 | 368 | 500 |

*Note:* Decimal points omitted, and values less than .0005 are omitted.