3.1

SOME TESTS BASED ON THE BINOMIAL DISTRIBUTION

THE BINOMIAL TEST AND TIMATION OF p

One example of the binomial test has already been presented. In Example 2.3.1 the binomial test was applied to a quality control problem. This entire chapter (Chapter 3) is little more than an elaboration of Example 2.3.1, showing the many uses and amazing versatility of that simple little binomial test. With a little ingenuity the binomial test may be adapted to test almost any hypothesis, with almost any type of data amenable to statistical analysis. In some situations the binomial test is the most powerful test; in those situations the test is claimed by both parametric and nonparametric statistics. In other situations more powerful tests are available, and the binomial test is claimed only by nonparametric statistics. However, even in situations where more powerful tests are available, the binomial test is sometimes preferred because it is usually simple to perform, simple to explain, and sometimes powerful enough to reject the null hypothesis when it should be rejected.

We will now formally present the binomial test and, at the same time, introduce the format for presenting tests. We feel that there is a need for some format in presenting tests both for the convenience of the reader and for ready review by the users of nonparametric techniques.

▶ The Binomial Test

Data The sample consists of the outcomes of n independent trials. Each outcome is in either "class 1" or "class 2," but not both. The number of observations in class 1 is O_1 and the number of observations in class 2 is $O_2 = n - O_1$.

Assumptions

1. The n trials are mutually independent.
2. Each trial has probability p of resulting in the outcome "class 1," where p is the same for all n trials.

Test Statistic Since we are concerned with the probability of the outcome "class 1," we will let the test statistic T be the number of times the outcome is "class 1." That is,

$$T = O_1 \tag{1}$$

Null Distribution Let p^* be the probability specified in the null hypothesis. The null distribution of T is the binomial distribution with parameters $p = p^*$ and n = the sample size. The null distribution of T is tabulated in Table A3 for $n \le 20$ and selected values of p.

3.1 Binomial Test

①α

$\boxed{n \leq 20}$

$H_0 : p \geq p^*$	$H_0 : p \leq p^*$	$H_0 : p = p^*$
$H_a : p < p^*$	$H_a : p > p^*$	$H_a : p \neq p^*$
$c = \{T : T \leq t\}$	$c = \{T : T > t\}$	$c = \{T : T \leq t_1 \text{ or } T > t_2\}$
find t such	find t such	find t_1 & t_2 such th
$P(y \leq t) = \alpha$	$P(y \leq t) = 1 - \alpha$	$P(y \leq t_1) = \alpha_1$
		$P(y \leq t_2) = 1 - \alpha_2$

or

using P.Value

$P.V = P(y \leq T)$	$P.V = P(y \geq T)$	$P.V = 2P(y \leq T)$ or
		$P.V = 2P(y \geq T)$
		* the smaller

Cont. 3.1 : Binomial Test

$\boxed{n > 20}$

$H_0: p \geq p^*$ $H_a: p < p^*$ $C = \{T : T \leq t\}$ where $t = np^* + z_\alpha \sqrt{np^*(1-p^*)}$	$H_0: p \leq p^*$ $H_a: p > p^*$ $C = \{T : T > t\}$ where $t = np^* + z_{1-\alpha} \sqrt{np^*(1-p^*)}$	$H_0: p = p^*$ $H_a: p \neq p^*$ $C = \{T : T \leq t_1 \text{ or } T > t_2\}$ wher $t_1 = np^* + z_{\alpha/2} \sqrt{np^*(1-p^*)}$ $t_2 = np^* + z_{1-\alpha/2} \sqrt{np^*(1-p^*)}$
	or using P. Value	
$P(y \leq T) = P\left(z \leq \frac{T - np^* + .5}{\sqrt{n.p^*(1-p^*)}}\right)$ 1	$P(y \geq T) = 1 - P\left(z \leq \frac{T - np^* - .5}{\sqrt{np^*(1-p^*)}}\right)$ 2	use 1 & 2 – and double them – the smaller

A method for finding a confidence interval for p, the unknown probability of any particular event occurring, is closely related to the binomial test.

Confidence Interval for a Probability or Population Proportion

Data A sample consisting of observations on n independent trials is examined, and the number Y of times the specified event occurs is noted.

Assumptions

1. The n trials are mutually independent.
2. The probability p of the specified event occurring remains constant from one trial to the next.

$n \le 30$

Method A For n less than or equal to 30, and confidence coefficients of 0.90, 0.95, or 0.99, use Table A4. Simply enter the table with sample size n and the observed Y. Reading across gives the exact lower and upper bounds in the columns for the desired confidence interval.

Method B For n greater than 30, or confidence coefficients not covered in Table A4, use the normal approximation. $n > 30$

$$L = \frac{Y}{n} - z_{1-\alpha/2}\sqrt{Y(n-Y)/n^3} \tag{14}$$

and

$$U = \frac{Y}{n} + z_{1-\alpha/2}\sqrt{Y(n-Y)/n^3} \tag{15}$$

where $z_{1-\alpha/2}$ is the quantile of a normally distributed random variable, obtained from Table A1. The confidence coefficient is approximately $1 - \alpha$.

Computer Assistance Computer packages that find confidence intervals for the binomial parameter p, or population proportion p, include *Minitab*, *S-Plus*, and *StatXact*.

For the sake of illustration, both methods of computing confidence intervals are used in the following example.

EXAMPLE 3

In a certain state 20 high schools were selected at random to see if they met the standards of excellence proposed by a national committee on education. It was found that 7 schools did qualify and accordingly were designated "excellent."

Y