8 Tests of Hypotheses Based on a Single Sample

8.3 Tests Concerning a Population Proportion

Tests Concerning a Population Proportion

Let p denote the proportion of individuals or objects in a population who possess a specified property (e.g., cars with manual transmissions or smokers who smoke a filter cigarette).

If an individual or object with the property is labeled a success (S), then p is the population proportion of successes.

Tests concerning p will be based on a random sample of size n from the population. Provided that n is small relative to the population size, X (the number of S's in the sample) has (approximately) a binomial distribution.

3

Tests Concerning a Population Proportion

Furthermore, if n itself is large [$np \geq 10$ and $n(1 - p) \geq 10$], both X and the estimator $\hat{p} = X/n$ are approximately normally distributed.

We first consider large-sample tests based on this latter fact and then turn to the small sample case that directly uses the binomial distribution.

4

Large-Sample Tests

Large-Sample Tests

Large-sample tests concerning p are a special case of the more general large-sample procedures for a parameter θ.

Let $\hat{\theta}$ be an estimator of θ that is (at least approximately) unbiased and has approximately a normal distribution.

The null hypothesis has the form H_0: $\theta = \theta_0$ where θ_0 denotes a number (the null value) appropriate to the problem context.

Large-Sample Tests

Suppose that when H_0 is true, the standard deviation of $\hat{\theta}$, $\sigma_{\hat{\theta}}$, involves no unknown parameters.

For example, if $\theta = \mu$ and $\hat{\theta} = \overline{X}$, $\sigma_{\hat{\theta}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$, which involves no unknown parameters only if the value of σ is known.

Large-Sample Tests

A large-sample test statistic results from standardizing $\hat{\theta}$ under the assumption that H_0 is true (so that $E(\hat{\theta}) = \theta_0$):

$$\text{Test statistic: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

If the alternative hypothesis is H_a: $\theta > \theta_0$, an upper-tailed test whose significance level is approximately α is specified by the rejection region $z \geq z_\alpha$.

The other two alternatives, H_a: $\theta < \theta_0$ and H_a: $\theta \neq \theta_0$, are tested using a lower-tailed z test and a two-tailed z test, respectively.

Large-Sample Tests

In the case $\theta = p$, $\sigma_{\hat{\theta}}$ will not involve any unknown parameters when H_0 is true, but this is atypical.

When $\sigma_{\hat{\theta}}$ does involve unknown parameters, it is often possible to use an estimated standard deviation $S_{\hat{\theta}}$ in place of $\sigma_{\hat{\theta}}$ and still have Z approximately normally distributed when H_0 is true (because when n is large, $s_{\hat{\theta}} \approx \sigma_{\hat{\theta}}$ for most samples).

The large-sample test we have seen earlier furnishes an example of this: Because σ is usually unknown, we use $s_{\hat{\theta}} = s_{\overline{X}} = s/\sqrt{n}$ in place of $\sigma/\sqrt{n}$ in the denominator of z.

9

Large-Sample Tests

The estimator $\hat{p} = X/n$ is unbiased $(E(\hat{p}) = p)$, has approximately a normal distribution, and its standard deviation is $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$.

When H_0 is true, $E(\hat{p}) = p_0$ and $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/n}$, so $\sigma_{\hat{p}}$ does not involve any unknown parameters. It then follows that when n is large and H_0 is true, the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

has approximately a standard normal distribution.

10

Large-Sample Tests

If the alternative hypothesis is H_a: $p > p_0$ and the upper-tailed rejection region $z \geq z_\alpha$ is used, then

P(type I error) = P(H_0 is rejected when it is true)

= $P(Z \geq z_\alpha$ when Z has approximately a standard normal distribution) $\approx \alpha$

Thus the desired level of significance α is attained by using the critical value that captures area α in the upper tail of the z curve.

11

Large-Sample Tests

Rejection regions for the other two alternative hypotheses, lower-tailed for H_a: $p < p_0$ and two-tailed for H_a: $p \neq p_0$, are justified in an analogous manner.

Null hypothesis: H_0: $p = p_0$

Test statistic value: $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

12

Large-Sample Tests

Alternative Hypothesis	Rejection Region
H_a: $p > p_0$	$z \geq z_\alpha$ (upper-tailed)
H_a: $p < p_0$	$z \leq -z_\alpha$ (lower-tailed)
H_a: $p \neq p_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed)

These test procedures are valid provided that $np_0 \geq 10$ and $n(1 - p_0) \geq 10$.

Example 11

Natural cork in wine bottles is subject to deterioration, and as a result wine in such bottles may experience contamination.

The article "Effects of Bottle Closure Type on Consumer Perceptions of Wine Quality" (*Amer. J. of Enology and Viticulture*, 2007: 182–191) reported that, in a tasting of commercial chardonnays, 16 of 91 bottles were considered spoiled to some extent by cork-associated characteristics.

Does this data provide strong evidence for concluding that more than 15% of all such bottles are contaminated in this way?

Example 11

cont'd

Let's carry out a test of hypotheses using a significance level of .10.

1. p = the true proportion of all commercial chardonnay bottles considered spoiled to some extent by cork-associated characteristics.

2. The null hypothesis is H_0: $p = .15$.

3. The alternative hypothesis is H_a: $p > .15$, the assertion that the population percentage exceeds 15%.

Example 11

cont'd

4. Since $np_0 = 91(.15) = 13.65 > 10$ and $nq_0 = 91(.85) = 77.35 > 10$, the large-sample z test can be used. The test statistic value is

$$z = (\hat{p} - .15)/\sqrt{(.15)(.85)/n}.$$

5. The form of H_a implies that an upper-tailed test is appropriate: Reject H_0 if $z \geq z_{.10} = 1.28$.

6. $\hat{p} = 16/91 = .1758$, from which

$$z = (.1758 - .15)/\sqrt{(.15)(.85)/91} = .0258/.0374 = .69$$

Example 11

cont'd

7. Since .69 < 1.28, z is not in the rejection region. At significance level .10, the null hypothesis cannot be rejected.

Although the percentage of contaminated bottles in the sample somewhat exceeds 15%, the sample percentage is not large enough to conclude that the population percentage exceeds 15%.

The difference between the sample proportion .1758 and the null value .15 can adequately be explained by sampling variability.

17

Large-Sample Tests

β and Sample Size Determination When H_0 is true, the test statistic Z has approximately a standard normal distribution.

Now suppose that H_0 is *not* true and that $p = p'$. Then Z still has approximately a normal distribution (because it is a linear function of $\hat{p}$), but its mean value and variance are no longer 0 and 1, respectively. Instead,

$$E(Z) = \frac{p' - p_0}{\sqrt{p_0(1-p_0)/n}} \qquad V(Z) = \frac{p'(1-p')/n}{p_0(1-p_0)/n}$$

The probability of a type II error for an upper-tailed test is $\beta(p') = P(Z > z_\alpha$ when $p = p')$.

18

Large-Sample Tests

This can be computed by using the given mean and variance to standardize and then referring to the standard normal cdf.

In addition, if it is desired that the level α test also have $\beta(p') = \beta$ for a specified value of β, this equation can be solved for the necessary n.

19

Large-Sample Tests

General expressions for $\beta(p')$ and n are given in the accompanying box.

Alternative Hypothesis	**$\beta(p')$**
H_a: $p > p_0$	$\Phi\left[\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
H_a: $p < p_0$	$1 - \Phi\left[\frac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
H_a: $p \neq p_0$	$\Phi\left[\frac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$
	$-\Phi\left[\frac{p_0 - p' - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

20

Large-Sample Tests

The sample size n for which the level α test also satisfies $\beta(p') = \beta$ is

$$n = \begin{cases} \left[\dfrac{z_\alpha \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p'(1-p')}}{p' - p_0} \right]^2 & \text{one-tailed test} \\ \left[\dfrac{z_{\alpha/2} \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p'(1-p')}}{p' - p_0} \right]^2 & \text{two-tailed test (an approximate solution)} \end{cases}$$

Example 12

A package-delivery service advertises that at least 90% of all packages brought to its office by 9 A.M. for delivery in the same city are delivered by noon that day.

Let p denote the true proportion of such packages that are delivered as advertised and consider the hypotheses H_0: $p = .9$ versus H_a: $p < .9$.

If only 80% of the packages are delivered as advertised, how likely is it that a level .01 test based on $n = 225$ packages will detect such a departure from H_0? What should the sample size be to ensure that $\beta(.8) = .01$?

Example 12

cont'd

With $\alpha = .01$, $p_0 = .9$, $p' = .8$, and $n = 225$,

$$\beta(.8) = 1 - \Phi\left(\frac{.9 - .8 - 2.33\sqrt{(.9)(.1)/225}}{\sqrt{(.8)(.2)/225}} \right)$$

$$= 1 - \Phi(2.00) = .0228$$

Thus the probability that H_0 will be rejected using the test when $p = .8$ is .9772— roughly 98% of all samples will result in correct rejection of H_0.

Example 12

cont'd

Using $z_\alpha = z_\beta = 2.33$ in the sample size formula yields

$$n = \left[\frac{2.33\sqrt{(.9)(.1)} + 2.33\sqrt{(.8)(.2)}}{.8 - .9} \right]^2 \approx 266$$

Small-Sample Tests

Small-Sample Tests

Test procedures when the sample size n is small are based directly on the binomial distribution rather than the normal approximation. Consider the alternative hypothesis H_a: $p > p_0$ and again let X be the number of successes in the sample.

Then X is the test statistic, and the upper-tailed rejection region has the form $x \geq c$. When H_0 is true, X has a binomial distribution with parameters n and p_0, so

$P(\text{type I error}) = P(H_0 \text{ is rejected when it is true})$

$= P(X \geq c \text{ when } X \sim \text{Bin}(n, p_0))$

Small-Sample Tests

$= 1 - P(X \leq c - 1 \text{ when } X \sim \text{Bin}(n, p_0))$

$= 1 - B(c - 1; n, p_0)$

As the critical value c decreases, more x values are included in the rejection region and P(type I error) increases. Because X has a discrete probability distribution, it is usually not possible to find a value of c for which P(type I error) is exactly the desired significance level α (e.g., .05 or .01).

Instead, the largest rejection region of the form $\{c, c + 1, \ldots , n\}$ satisfying $1 - B(c - 1: n, p_0) \leq \alpha$ is used.

Small-Sample Tests

Let p' denote an alternative value of $p(p' > p_0)$. When $p = p'$, $X \sim \text{Bin}(n, p')$,

so

$\beta(p') = P(\text{type II error when } p = p')$

$= P(X < c \text{ when } X \sim \text{Bin}(n, p'))$

$= B(c - 1; n, p')$

Small-Sample Tests

That is, $\beta(p')$, is the result of a straightforward binomial probability calculation.

The sample size n necessary to ensure that a level α test also has specified β at a particular alternative value p' must be determined by trial and error using the binomial cdf.

Test procedures for H_a: $p < p_0$ and for H_a: $p \neq p_0$ are constructed in a similar manner.

In the former case, the appropriate rejection region has the form $x \leq c$ (a lower-tailed test).

29

Small-Sample Tests

The critical value c is the largest number satisfying $B(c; n, p_0) \leq \alpha$.

The rejection region when the alternative hypothesis is H_a: $p \neq p_0$ consists of both large and small x values.

30

Example 13

A plastics manufacturer has developed a new type of plastic trash can and proposes to sell them with an unconditional 6-year warranty.

To see whether this is economically feasible, 20 prototype cans are subjected to an accelerated life test to simulate 6 years of use.

The proposed warranty will be modified only if the sample data strongly suggests that fewer than 90% of such cans would survive the 6-year period.

31

Example 13

cont'd

Let p denote the proportion of all cans that survive the accelerated test. The relevant hypotheses are H_0: $p = .9$ versus H_a: $p < .9$.

A decision will be based on the test statistic X, the number among the 20 that survive. If the desired significance level is $\alpha = .05$, c must satisfy $B(c; 20, .9) \leq .05$.

From Appendix Table A.1, $B(15; 20, .9) = .043$, whereas $B(16; 20, .9) = .133$. The appropriate rejection region is therefore $x \leq 15$.

32

Example 13

cont'd

If the accelerated test results in $x = 14$, H_0 would be rejected in favor of H_a, necessitating a modification of the proposed warranty.

The probability of a type II error for the alternative value $p' = .8$ is

$$\beta(.8) = P(H_0 \text{ is not rejected when } X \sim \text{Bin}(20, .8))$$

$$= P(X \geq 16 \text{ when } X \sim \text{Bin}(20, .8))$$

$$= 1 - B(15; 20, .8) = 1 - .370 = .630$$

33

Example 13

cont'd

That is, when $p = .8$, 63% of all samples consisting of $n = 20$ cans would result in H_0 being incorrectly not rejected.

This error probability is high because 20 is a small sample size and $p' = .8$ is close to the null value $p_0 = .9$.

34