

## Review of Last Class: Testing a Proportion

Suppose the data are a sample proportion  $\hat{p}$  from the sample of size  $n$  where the true population proportion is an unknown quantity  $p$ . The *null hypothesis* is

$$H_0 : p = p_0$$

where  $p_0$  is some given proportion.

The *alternative hypothesis* is almost always one of

$$H_a : p > p_0, \quad \text{or} \quad (1)$$

$$H_a : p < p_0, \quad \text{or} \quad (2)$$

$$H_a : p \neq p_0. \quad (3)$$

where the choice among (1)–(3) depends on the context of the problem.

## Test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}. \quad (4)$$

If  $H_0$  is true, then  $z$  has a standard normal distribution with mean 0 and standard deviation 1.

Side comment:

The formula for the standard error is

$$\sqrt{\frac{p_0(1 - p_0)}{n}}.$$

This is different from the confidence interval calculation where it's

$$\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

The difference is that in a hypothesis testing problem, we already have a specific value  $p_0$  that we're testing against. Therefore, it makes sense to use that as the basis for calculating a standard error.

## Computing the P-value

The calculation of the P-value depends on the alternative hypothesis.

For (1), compute  $\Pr\{Z > z\}$  where  $Z$  is a standard normal random variable and  $z$  is the number computed in (4).

For (2), compute  $\Pr\{Z < z\}$ .

For (3), compute  $\Pr\{Z < -|z|\} + \Pr\{Z > |z|\}$  where  $|z|$  means the magnitude of  $z$  (ignoring the sign). In practice, this almost always results in twice the P-value computed for (1) or (2).

## Another side comment:

In the example with  $n = 31, p_0 = 0.067, X = 6$ , the usual condition  $np_0 > 15, n(1 - p_0) > 15$  is not satisfied (because  $np_0 = 2.077$ ). We could do an exact calculation, as follows.

The problem is to calculate  $\Pr\{X \geq 6\}$  when  $X$  has a binomial distribution with  $n = 31, p = .067$ . This is  $1 - \Pr\{X \leq 5\}$ .

1. In Excel, go to Formulas → More Functions → Statistical → BINOMDIST
2. Enter “Number” =5, “Trials” =31, “Probability” =.067, “Cumulative” =TRUE.
3. The answer comes up as 0.9844. Therefore,  $1 - .9844 = .0156$  is the required (one-sided) P-value.

## Interpreting the P-value

1. It's usual to interpret a P-value less than .05 as “significant” . However if it's very important to make sure we don't get a spurious result, we may adopt a more stringent criterion, e.g.  $P < .01$ .
2. In presenting research results, a common practice is just to ignore results for which  $P > .05$ , but when  $P < .05$ , state the exact P-value. That way, the reader can judge for herself just how strong the result it. This is what they did in the paper about skin cancer in marathon runners.

## Testing the mean of a quantitative variable

*An example:* The mean height of male students at the University of Georgia (page 64) was 71 inches.

In a sample of 14 male students in this class, the mean  $\bar{x}$  was 72.57 inches and the standard deviation  $s$  was 3.131 inches.

*Is this a statistically significant difference?*

Let's set this up as a formal hypothesis test.

Assume that the students in this class are a random sample of the population of all male students at UNC, and let  $\mu$  be the mean height of that population. Also write  $\mu_0 = 71$ . The null hypothesis is

$$H_0 : \mu = \mu_0.$$

For the alternative hypothesis, we again have three possibilities analogous to (1)–(3), namely

$$H_a : \mu > \mu_0, \tag{5}$$

$$H_a : \mu < \mu_0, \tag{6}$$

$$H_a : \mu \neq \mu_0. \tag{7}$$

In this case, there is no *a priori* reason to think that the students at UNC are either taller or shorter than the students at the University of Georgia, so it is most logical to choose (7) as our alternative hypothesis.



We are given  $\bar{x} = 72.57$ , and its standard error is  $\frac{s}{\sqrt{n}} = \frac{3.131}{\sqrt{14}} = 0.8368$ . The  $t$  statistic is

$$t = \frac{\bar{x} - \mu}{S.E.} = \frac{72.57 - 71}{0.8368} = 1.876.$$

If  $H_0$  is true, then  $t$  has a  $t$  distribution with degrees of freedom  $df = n - 1 = 13$ , by the same theory as used for confidence intervals in Chapter 8. Therefore the P-value is the probability that a random variable with the  $t_{13}$  distribution is greater than 1.876, multiplied by 2 (because it is a two-tailed test — therefore  $t < -1.876$  has the same meaning as  $t > 1.876$ ).

At this point, I'm going to differ a little from what the book tells you to do (on page 431). They recommend the use of the Minitab software package to find the exact P-value. I'm going to show you how to get the approximate value from the table on page A3.

df	Confidence level					
	80%	90%	95%	98%	99%	99.8%
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
Inf	1.282	1.645	1.960	2.327	2.576	3.091

In this table, corresponding to  $df = 13$ , we find  $t$  values of 1.350, 1.771, 2.160, 2.650, 3.012, 3.852 corresponding to confidence levels of 80%, 90%, 95%, 98%, 99%, 99.8%.

The two-sided P-value is one minus the confidence level, expressed as a decimal fraction.

Thus, only a  $t$  value bigger than 2.160 would be considered “significant” with a P-value of .05 or less.

Since in this case  $t = 1.876$ , the P-value is between 0.1 and 0.05.

Since this is bigger than .05, this is *not significant*. We accept the null hypothesis  $\mu = 71$ . The students at the University of North Carolina are not significantly different from those at the University of Georgia (as we would have expected).

## An alternative calculation

Suppose, however, the mean of the UNC students was 73.5 inches. In this case we would calculate  $t = \frac{73.5-71}{0.8368} = 2.988$ . In the table, this comes out between confidence levels 98% and 99% — in other words, the 2-sided P-value is between  $1-0.98=0.02$  and  $1-0.99=0.01$ .

If we had set it up as a 1-sided test in the first place, the P-value would be between 0.005 and 0.01 (half the 2-sided P-value).

Either way, the conclusion would be that there is statistically significant evidence that the UNC students are taller than the Georgia students, though even then, the evidence is not overwhelmingly strong. We would probably not attach too much importance to the conclusion unless it were replicated by further studies.

Suppose, however, the study were not of all male students at UNC and UGa, but specifically of male basketball players. Then, it would not be at all surprising that the UNC players were taller (because the best players want to come to UNC...)

This illustrates the point that the *context* of a study often plays a role in interpreting its outcome — it is more likely to be accepted if it accords with prior expectation. Even then, however, the P-value calculation provides a “reality check” in helping us to identify spurious results.

*Footnote to this section.* Even though we don't have Minitab at our disposal, it is still possible to mimic the book's method of calculation using Excel.

1. In Excel, go to Formulas → More Functions → Statistical → TDIST
2. Enter “X” =1.876, “Deg-freedom” =13, “Tails” — either 1 or 2 for a 1-sided or 2-sided test.
3. The 1-sided P-value is 0.0416 and the 2-sided P-value is 0.0833. Note that this is one instance where the significance (relative to P-value 0.05) of the result depends on whether the test is one-sided or two-sided.

## Summary of method

The method is designed to test whether the mean of some population,  $\mu$  say, is equal to some predetermined value  $\mu_0$ , where  $\mu_0$  could be the result for a larger population, or from an older survey, or could be some target value (e.g. specifying that the mean concentration of an environmental pollutant should be lower than a safe level determined by the EPA).

We assume we have a sample mean  $\bar{x}$ , with a standard deviation  $s$ , calculated from a random sample of size  $n$ .

1. Determine  $H_0 : \mu = \mu_0$  and  $H_a$  (either  $\mu > \mu_0$  or  $\mu < \mu_0$  or  $\mu \neq \mu_0$ )

2. Calculate the test statistic

$$t = \frac{\bar{x} - \mu_0}{S.E.} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

3. Calculate  $df = n - 1$ .

4. Determine the P value — either exactly using Excel or Minitab, or using Table B in the back of the book. Note that if you use Table B, you will not be able to determine P exactly but will be able to identify the nearest value among 0.1, 0.05, 0.025, 0.01, 0.005 and 0.001 (one-sided) or among 0.2, 0.1, 0.05, 0.02, 0.01 and 0.002 (two-sided),

5. Report your conclusions. Usually if  $P > 0.05$  we simply report that the result was not statistically significant, but if  $P \leq 0.05$  we report that fact and our nearest guess for the exact value of P.



## Example Problems

*Question 9.75, page 461.* A bank wants to evaluate which of two credit cards is more attractive to its customers. For a random sample of 100 customers, 40 say they would prefer one that has an annual cost but low interest. Software reports:

Test of  $p=0.50$  vs  $p \text{ not } = 0.50$

X	N	Sample p	95% CI	Z-value	P-value
40	100	0.40000	(0.304,0.496)	-2.00	0.04550

Explain how to interpret these results. What would you report to the company?

*Solution.* For calculating a confidence interval, the sample proportion is  $\hat{p}$  and the standard error is  $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.4 \times 0.6}{100}} = .049$ . Hence the 95% CI is  $0.4 \pm 1.96 \times 0.049 = 0.4 \pm 0.096 = (0.304, 0.496)$ .

For the hypothesis test of  $H_0 : p = p_0 = 0.5$ , we recalculate the standard error based on  $p_0$ , as  $\sqrt{\frac{0.5 \times 0.5}{100}} = 0.05$ . Then the  $z$  statistic is  $\frac{0.4 - 0.5}{0.05} = -2.0$ , and the corresponding *one-sided* P-value is .0228 (from Table A). Therefore, the two-sided P-value is  $2 \times 0.0228 = 0.0456$ . [Slight rounding error here: the exact one-sided P-value is .02275013 and twice that would be .0455, but this makes no difference in practice.]

Report to the company: Less than half the customers favor the new card, and that is a statistically significant difference (though only barely).