Proportion Inference Overview

Situation 3. A proportion p of a population has some particular outcome of interest. From this population we draw a Simple Random Sample of size n. The proportion, p, of the population is unknown. The proportion of the sample with the outcome of interest is computed as $\hat{p} = \dfrac{number\ of\ successes\ in\ the\ sample}{n}$. For a sufficiently large sample size the $\hat{p}$ statistic has the distribution $N(p, \sqrt{\frac{p(1-p)}{n}})$.

Here is the picture:

Population	Simple Random Sample
proportion p "have it"	proportion $\hat{p}$ "have it"
p = ?	$\hat{p}$ distribution -> $N(p, \sqrt{\frac{p(1-p)}{n}})$
	size = n

Confidence Intervals

We can estimate p with

margin of error

$$p = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

standard error

Where z^* is determined by the confidence level C. The table below gives z^* for some common values of C:

C	90%	95%	99%
z^*	1.645	1.960	2.576

Hypothesis Testing.

The Null Hypothesis:
H_0: $p = p_0$

The Alternative Hypothesis:
H_a: $p > p_0$

First compute the z statistic using the formula:

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

Then for significance level alpha determine z^*. Then if $z > z^*$ we reject the null hypothesis and say that we have statistically significant evidence for the alternative hypothesis. For the other alternative hypotheses, H_a: $p > p_0$ or H_a: $p \neq p_0$, we adjust the test appropriately.

1. Suppose we have a panel of forty expert wine testers. We want to see if they can distinguish between a 1970 and 1971 Chardonnay. We present them each with three glasses of wine. The three glasses are either two glasses of the 1970 and one glass of the 1971 in a random order or two glasses of the 1971 and one glass of the 1970 in a random order. The experts are supposed to identify which of the three glasses contains a different wine. If each of the forty experts is just guessing and X is the number who guess correctly, then the sample proportion $\hat{p}$ = X/40 who guess correctly has a sampling distribution with a mean μ and standard deviation σ of

- ○ A. μ = 1/3 and σ = .0056.
- ○ B. μ = 1/2 and σ = .0063.
- ◉ C. μ = 1/3 and σ = .0745.

Since the "odd" wine can be in any of the 3 positions the random chance of success is $p = \frac{1}{3}$. The sample size is large enough to postulate a Normal distribution for $\hat{p}$.

so the $\hat{p}$ distribution is given by

$$\hat{p} = N(\mu, \sigma) = N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$= N\left(\frac{1}{3}, \sqrt{\frac{\frac{1}{3}(1-\frac{1}{3})}{40}}\right)$$

$$= N\left(\frac{1}{3}, .0745\right)$$

Even if the distribution is not Normal the mean and standard deviation are the same. Refer to page 493.

2. A radio talk show host is interested in the proportion p of adults in his listening area who think the drinking age should be lowered to eighteen. To make this determination, he poses the following question to his listeners: "Do you think that the drinking age should be reduced to eighteen in light of the fact that eighteen-year-olds are eligible for military service?" He asks listeners to phone in and vote "yes" if they agree the drinking age should be lowered and "no" if not. Of the 200 people who phoned in, 140 answered "yes." The standard error for the proportion $\hat{p}$ of those who phoned in and answered "yes" is

- A. 0.46.
- B. 0.032.
- C. 0.00105.

$$\text{standard error} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

see pg 496

$$= \sqrt{\frac{\frac{140}{200}\left(1-\frac{140}{200}\right)}{200}} = .0324$$

3. An inspector inspects large truckloads of potatoes to determine the proportion p in the shipment with major defects prior to using the potatoes in making potato chips. She intends to compute a 95% confidence interval for p To do so, she selects an SRS of 100 potatoes from the over 2000 potatoes on the truck. Suppose that only 4 of the potatoes sampled are found to have major defects. Which of the following assumptions for inference about a proportion using a confidence interval are violated?

- A. The population is at least ten times as large as the sample.
- B. n is so large that both the count of successes np and the count of failures n(1 - p) are 15 or more.
- C. There appear to be no violations.

The n is not large enough. See page 497 of the text.

4. One hundred rats whose mothers were exposed to high levels of tobacco smoke during pregnancy were put through a simple maze. The maze required the rats to make a choice between going left or right at the outset. Eighty of the rats went right when running the maze for the first time. Assume that the 100 rats can be considered an SRS from the population all rats born to mothers exposed to high levels of tobacco smoke during pregnancy. (Note that this assumption may or may not be reasonable, but researchers often assume lab rats are representative of such larger populations because they are often bred to have very uniform characteristics.) Let p be the proportion of rats in this population that would go right when running the maze for the first time. A 90% confidence interval for p is

- ○ A. 0.8 ± .040.
- ✓ B. 0.8 ± .066.
- ○ C. 0.8 ± .078.

$$\hat{p} \pm z^* \sqrt{\frac{p(1-p)}{n}}$$

where $\hat{p} = \frac{80}{100}$

$z^* = 1.645$

$n = 100$

$$\frac{80}{100} \pm 1.645 \sqrt{\frac{\frac{80}{100}\left(1-\frac{80}{100}\right)}{100}}$$

$0.8 \pm .066$

5. A noted psychic was tested for ESP. The psychic was presented with 400 cards face down and asked to determine if each of the cards was one of four symbols: a star, cross, circle, or square. The psychic was correct in 120 cases. Let p represent the probability that the psychic correctly identifies the symbols on the cards in a randc trial. Suppose you wish to see if there is evidence that the psychic was doing better than just guessing. To make this determination you test the hypotheses $H_0: p = 0.25$ versus $H_a: p > 0.25$. The P-value of your test is

- ✓ A. 0.0104.
- ○ B. 0.0146.
- ○ C. 0.9896.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{120}{400} - .25}{\sqrt{\frac{.25 * .75}{400}}} = 2.31$$

The p-value comes from the area on the mustard sheet $= 1 - .9896 = .0104$

6. A noted psychic was tested for ESP. The psychic was presented with 400 cards face down and asked to determine if each of the cards was one of four symbols: a star, cross, circle, or square. The psychic was correct ir 120 cases. Let p represent the probability that the psychic correctly identified the symbols on the cards in a randc trial. How large a sample n would you need to estimate p with margin of error 0.01 and 95% confidence? Use the guess p* = 0.25 as the value for p.

- ○ A. n = 1351.
- ✔ B. n = 7203.
- ○ C. n = 9604.

Margin of error = $z^* \sqrt{\frac{p(1-p)}{n}}$

at 95% confidence $z^* = 1.960$

so

$m = z^* \sqrt{\frac{p(1-p)}{n}} = 1.960 \sqrt{\frac{.25 * .75}{n}} < .01$

solving for n we get $7203 < n$

7. Drug sniffing dogs must be 95% accurate in their responses, since we don't want them to miss drugs and also don't want false positives. A new dog is being tested and is right in 46 of 50 trials. Find a 95% confidence interval for the proportion of times the dog will be correct.

- ✔ A. (0.845, 0.995)
- ○ B. (0.805, 0.973)
- ○ C. (0.819, 0.959)

Confidence interval is $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

For 95% confidence $z^* = 1.960$

$\hat{p} = \frac{46}{50}$, $n = 50$

So the interval is

$$\frac{46}{50} \pm 1.96 \sqrt{\frac{\frac{46}{50}\left(1-\frac{46}{50}\right)}{50}} = .92 \pm .075$$

$(.92 - .075, .92 + .075) = (.845, .995)$

8. A poll finds that 54% of the 600 people polled favor the incumbent. Shortly after the poll is taken, it is disclosed that he had an extramarital affair. A new poll finds that 50% of the 1030 polled now favor the incumbent. The standard error for a confidence interval for the candidate's latest support level is

- A. 0.016.
- B. 0.020.
- C. 0.00025.

standard error $= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\hat{p} = .50 \qquad n = 1030$

$\sqrt{\frac{.5 * .5}{1030}} = .016$

9. I read an article about a new drug which stated that "the incidence of side effects was similar to placebo, P-val > 0.05." I want to know if the results are significant at α = 10%. With the information given,

- A. I will reject the null hypothesis of no difference at 10%.
- B. I will not reject the null hypothesis of no difference at 10%.
- C. There is not enough information given.

They were not able to reject the null hypothesis because the P-value was larger than their $\alpha = .05$. Changing the α to 10% may change the conclusion. However without knowing the actual P-value we cannot tell.

10. I want to take a survey of students at my university to find out how what proportion like the new bus service on campus. How many will I need to survey if I want to estimate with 99% confidence the true proportion to within 2% if I believe that 75% of students like the bus service?

- A. 3111
- B. 1801
- C. 25

$$\text{Margin of error} = z^* \sqrt{\frac{p(1-p)}{n}}$$

For Confidence level 99%, $z^* = 2.576$

$p = .75$ so

$$m = z^* \sqrt{\frac{p(1-p)}{n}} = 2.576 \sqrt{\frac{.75 * .25}{n}} < .02$$

Solving for n we get

$$3110.52 < n$$