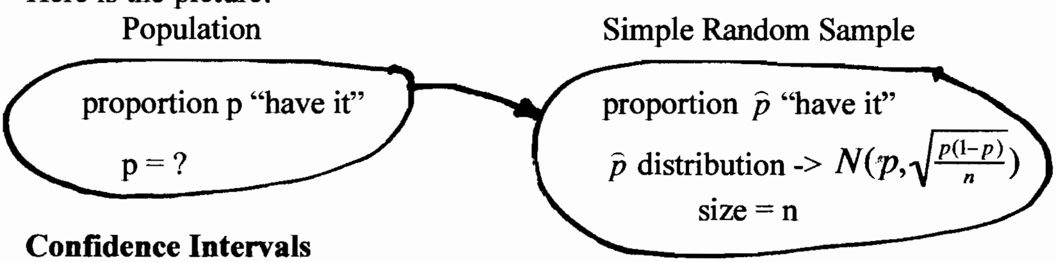
Situation 3. A proportion p of a population has some particular outcome of interest. From this population we draw a Simple Random Sample of size n. The proportion, p, of the population is unknown. The proportion of the sample with the outcome of interest is computed as $\hat{p} = \dfrac{\textit{number of successes in the sample}}{n}$. 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:



Confidence Intervals

We can estimate p with

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

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:

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

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

situation 3 with $p = \frac{1}{3}$ and $n = 40$.

$\hat{p}$ distribution is $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

$\mu = \frac{1}{3}$ $\quad \sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{1}{3}\frac{2}{3}}{40}} = \frac{1}{3\sqrt{20}} = .0745$

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

---

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

situation 3 with $n = 200$ and 140 successes.

$\hat{p} = \text{successes}/n = \frac{140}{200}$

$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{\frac{140}{200} \cdot \frac{60}{200}}{200}} = \frac{1}{200}\sqrt{\frac{140 \cdot 60}{200}} = \frac{1}{200}\sqrt{42} = .0324$

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

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.

Not violated since the population size (2000) is twenty times the sample size (100).

- [x] B. n is so large that both the count of successes np and the count of failures n(1 - p) are 15 or more.

This is violated since we only had 4 "successes".
[To fix this we could use the plus 4 rule]

- [ ] C. There appear to be no violations.

See B above

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

Situation 3 with 80 successes, $n=100$, $c=90\%$

$$p = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{80}{100} \pm 1.645 \sqrt{\frac{\frac{80}{100}\cdot\frac{20}{100}}{100}}$$

$$= .8 \pm \frac{1.645}{10}\sqrt{.8*.2} = .8 \pm .0658$$

A. 0.8 ± .040.

→ B. 0.8 ± .066.

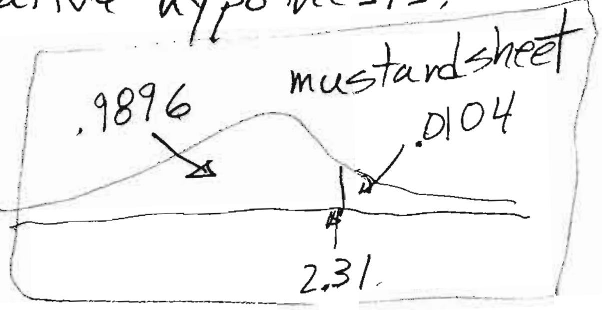
C. 0.8 ± .078.

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 random 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

Situation 3 with 120 successes, $n=400$, $p_0=\frac{1}{4}$,
and a one-sided alternative hypothesis.

$\hat{p} = \frac{120}{400}$, $p_0 = \frac{1}{4}$, $n=400$

$$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{120}{400}-\frac{1}{4}}{\sqrt{\frac{\frac{1}{4}\cdot\frac{3}{4}}{400}}} = 2.31$$



→ A. 0.0104.

B. 0.0146.

C. 0.9896.

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 in 120 cases. Let p represent the probability that the psychic correctly identified the symbols on the cards in a random 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.

Situation 3 with 120 successes, $n = 400$, $C = 95\%$

margin of error < 0.01

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

$$z^* \sqrt{\frac{p^*(1-p^*)}{n}} \le 0.01$$

$$1.960 \sqrt{\frac{.25 * .75}{n}} \le 0.01 \Rightarrow \frac{1.96}{.01}\sqrt{\frac{1}{4}\cdot\frac{3}{4}} \le \sqrt{n}$$

$$196^2 \frac{3}{16} \le n \Rightarrow 7203 \le n$$

- A. n = 1351.
- → B. n = 7203.
- C. n = 9604.

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.

Situation 3 with 46 successes, $n = 50$, $C = 95\%$. Since we only have 4 failures (<15 see #3 above) we use the "plus four" rule and add two successes and two failures to our data. So now we have 48 successes and $n = 54$.

$$p = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{48}{54} \pm 1.960 \sqrt{\frac{\frac{48}{54} \cdot \frac{6}{54}}{54}}$$

$$= \frac{8}{9} \pm .0838 \qquad \left(\frac{8}{9} - .0838, \frac{8}{9} + .0838\right)$$

$$(.8051, .9727)$$

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

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

Situation 3 with $\hat{p} = \frac{1}{2}$ and $n = 1030$

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{\frac{1}{2} \cdot \frac{1}{2}}{1030}} = .01558$$

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

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

A. I will reject the null hypothesis of no difference at 10%.

If the P-value is < 10%. But we don't know that. It could be 20% or 6%.

B. I will not reject the null hypothesis of no difference at 10%.

If the P-value is > 10%. But we don't know that. It could be 20% or 6%

C. There is not enough information given.

We need more information about the P-value

10. I want to take a survey of students at my university to find out [redacted] 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?

Situation 3 with $p \approx .75$, n to be determined, $C = 99\%$, margin of error ≤ 0.02

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

margin of error

$$z^* \sqrt{\frac{p^*(1-p^*)}{n}} < 0{,}02$$

$$2{,}576 \sqrt{\frac{.75 * .25}{n}} < 0{,}02$$

$$\frac{2{,}576}{0{,}02} \sqrt{.75 * .25} < \sqrt{n}$$

$$\left(\frac{2{,}576}{0{,}02}\right)^2 .75 * .25 < n$$

$$3110.52 < n$$

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