sampling distribution of $\hat{p}$

population proportion: p
sample proportion: $\hat{p} = \frac{x}{n}$
x : number of successes in the sample.
$\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
For large n, approximately $\hat{p} \sim$ Normal $(p, \sqrt{\frac{p(1-p)}{n}})$.

CI for p

Sample size needed: $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$. (at least 10 successes and 10 failures in the sample)
Data come from a random sample.

$(1-\alpha)100\%$CI is:
$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Example

When the 2000 GSS asked subjects whether they would be willing to accept cuts in their standard of living to protect the environment, 344 of 1170 subjects said yes. Find a 95% confidence interval for the population proportion who would say yes.
The sample proportion is $\hat{p} = 344/1170 = 0.294$
A 95% CI:
$\hat{p} \pm 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.294 \pm 1.96\sqrt{\frac{0.294*0.706}{1170}} = 0.294 \pm 0.026 = (0.268, 0.320)$.
exercise: 23, page 327.

Exercise

In 2008 a GSS asked 182 people whether they received health insurance from their employers. A total of 60 people said they did. Get a 95% CI for the population proportion.

Solution

$\hat{p} = 60/182 = 0.330.$

95% CI: $0.330 \pm 1.96\sqrt{\frac{0.330 * 0.670}{182}} = 0.330 \pm 0.068 = (0.262, 0.398).$

sample size for desired margin of error

$m = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, and

$$n = \hat{p}(1-\hat{p})\left(\frac{z}{m}\right)^2$$

Use $\hat{p} = 1/2$ if a value for $\hat{p}$ is unavailable.

Example 7.14

Among a sample of 517 music teachers, 403 believe that video games have a positive effect on music education. Estimate the sample size needed so that a 95% CI will have a margin of error of 0.03.
$\hat{p} = \frac{403}{517} = 0.7795$,
$n = \hat{p}(1 - \hat{p})\left(\frac{z}{m}\right)^2 = 0.7795 * (1 - 0.7795)(\frac{1.96}{0.03})^2 \approx 734$.

exercise 30 page 328:
1). A gallup poll estimates the proportion of people who believe the economy is getting better to be 0.33. What sample size is needed so that the 95% CI will have a margin of error 0.03?
2).Estimate sample size needed if no estimate of p is available.

Solutions

a. $n = 0.33 * 0.67(\frac{1.96}{0.03})^2 = 944.$

b. $n = 0.50 * 0.50(\frac{1.96}{0.03})^2 = 1068.$

Exercise

A researcher wants to estimate the population proportion of frogs with fungus infection in a pond. Her sample proportion is $\hat{p}$=0.35 based on a sample of 100.

1). Construct a 95% confidence interval for the population proportion p.

2). Construct a 99% confidence interval for p.

Solutions

1). $0.35 \pm 1.96\sqrt{\frac{0.35*0.65}{100}} = 0.35 \pm 0.09 = (0.26, 0.44)$.

2). $0.35 \pm 2.576\sqrt{\frac{0.35*0.65}{100}} = 0.35 \pm 0.12 = (0.23, 0.47)$.

CIs for small samples

adjusted sample proportion: $\tilde{p} = \frac{x+2}{n+4}$
Replace $\hat{p}$ and n with $\tilde{p}$ and $\tilde{n} = n + 4$ in the CI formula.
$\tilde{p} \pm z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$

Exercise

2. A 2008 GSS survey asked 182 whether they received health insurance from their employers and 60 of them said yes.
Get a 90% CI for the population proportion who receive health insurance from their employers.

$\hat{p} = 60/182 = 0.330.$

90% CI: $0.330 \pm 1.645\sqrt{\frac{0.330*0.670}{182}} = 0.330 \pm 0.057 = (0.273, 0.387)$.

Example

The General Social Survey recently asked 1294 people whether they performed any volunteer work during the past year. A total of 517 people said they did.

1). Find a point estimate for the proportion of people who performed volunteer work during the past year.

2). Construct a 95% confidence interval for the proportion of people who performed volunteer work during the past year.

3). A sociologist states that 50% of Americans perform volunteer work in a given year. Does the confidence interval contradict his statement? Explain.

Solutions

1). $\hat{p} = \frac{517}{1294} = 0.400$.
2). $0.400 \pm 1.96\sqrt{.400 * .600/1294} = 0.400 \pm 0.027 = (0.373, 0.427)$.
3). Yes, the confidence interval contradicts his statement as .50 is not included in the interval.