Confidence Intervals for Proportions

7.3

Today’s topics

Review:

- Confidence Intervals for mean and variance

New:

- Confidence Interval for proportions

General Form of CI for mean (review)

Estimate $\pm$ (Critical Value * SE of estimate)

e.g. if σ is known:

$$\bar{x} \;\pm\; z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

if σ is unknown:

$$\bar{x} \;\pm\; t_{n-1,\alpha/2} * \frac{s}{\sqrt{n}}$$

Confidence Interval for σ^2 (review)

Confidence Interval for σ^2 :

$$\left(\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}} \right)$$

Confidence Interval for Proportions

Goal: Create a (1-α)100% confidence interval for true proportion, p, based on a sample.

e.g. What is the true proportion of people who do not have severe symptoms due to COVID-19.

We can think of a sample of size n as observations coming from n Bernoulli trials with probability p.

Notes

Estimating p in the Bernoulli Distr. ($\hat{p}$)

Let $X_1, X_2, \ldots, X_n$ be iid ~ Bernoulli(p). $f(x) = p^x(1-p)^{1-x}$

Y = the number of total successes (Y = $\sum X_i$)

What we (should) know:

- Y ~ Binom(n,p)
- MLE of p, $\hat{p}$ = Y/n $L(p) = p^y(1-p)^{n-y}$, etc…
- $\hat{p}$ is an unbiased point estimator for p
  - E[$\hat{p}$] = p

Estimating $\hat{p}$ in the Bernoulli Distribution

More things you (should) know:

If $X_1, X_2, \ldots X_n \sim$ Bern(p), then for all X, $\mu_X = p$ and σ_X^2 = p(1-p)

Since, $\hat{p} = \frac{Y}{n} = \frac{\sum X_i}{n}$, by the Central limit theorem,

$$\frac{\hat{p} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} \sim N(0,1) \text{ when } n \to \infty$$

(or approximately Normal(0,1) when n is large enough)

Estimating $\hat{p}$ in the Bernoulli Distribution

$\dfrac{\hat{p} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} = \dfrac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim$ N(0,1) approximately when n is large enough

Interesting side notes:

The variance of a Bernoulli is $p(1-p)$. If we sum n independent Bernoulli R.V.s, we get a $Binom(n, p)$ distribution, and its variance is $np(1-p)$.

$\hat{p} = \frac{1}{n}\sum X_i$,

so Var[$\hat{p}$] $= \frac{1}{n^2}$Var[$\sum X_i$] $= \frac{1}{n^2}$[np(1-p)] $= \dfrac{\text{p(1−p)}}{n}$

Notes

Creating a CI for p

Now that we know $\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ is approx $\sim$ N(0,1),

For a given confidence coefficient, 1-α, we can find $z_{\alpha/2}$ such that

$$P\left[-z_{\alpha/2} \leq \frac{(Y/n) - p}{\sqrt{p(1-p)/n}} \leq z_{\alpha/2}\right] \approx 1 - \alpha.$$

Creating a CI for p

$$P\left[-z_{\alpha/2} \le \frac{(Y/n) - p}{\sqrt{p(1-p)/n}} \le z_{\alpha/2}\right] \approx 1 - \alpha.$$

$$P\left[-\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le -p \le -\frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] \approx 1 - \alpha.$$

multiply by (-1)

$$P\left[\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] \approx 1 - \alpha.$$

Creating a CI for p

$$P\left[\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] \approx 1-\alpha.$$

We'll need to use another approximation to get p out of the endpoints: Just use $\hat{p}$ ($\hat{p}$ = Y/n)

$$P\left[\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{(Y/n)(1-Y/n)}{n}} \le p \le \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{(Y/n)(1-Y/n)}{n}}\right] \approx 1-\alpha.$$

Creating a CI for p

Previous slide:

$$P\left[\frac{Y}{n} - z_{\alpha/2}\sqrt{\frac{(Y/n)(1-Y/n)}{n}} \le p \le \frac{Y}{n} + z_{\alpha/2}\sqrt{\frac{(Y/n)(1-Y/n)}{n}}\right] \approx 1-\alpha.$$

For large n, and observed $Y = y$, the following interval is a 100(1-α)% Confidence interval for p:

$$\boxed{\left[\frac{y}{n} - z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}, \frac{y}{n} + z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}\right]}$$

or

$$\boxed{\frac{y}{n} \pm z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}}$$

One-sided CIs for p

Upper bound

$$\left[0, \frac{y}{n} + z_{\alpha}\sqrt{\frac{(y/n)[1-(y/n)]}{n}}\right]$$

Lower bound

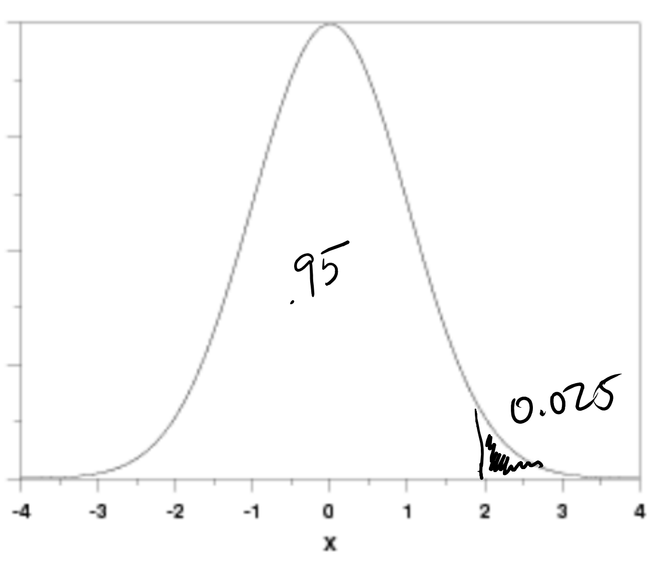
$$\left[\frac{y}{n} - z_{\alpha}\sqrt{\frac{(y/n)[1-(y/n)]}{n}}, 1\right]$$

Example

In a certain political campaign, Captain America has a poll taken at random among the voting population. The results show that y = 185 out of n = 351 voters favor him. Create a 95% Confidence interval for the true proportion of voters, p, who favor Captain America.

$$0.527 \pm 1.96\sqrt{\frac{(0.527)(0.473)}{351}}$$

(0.475, 0.579)

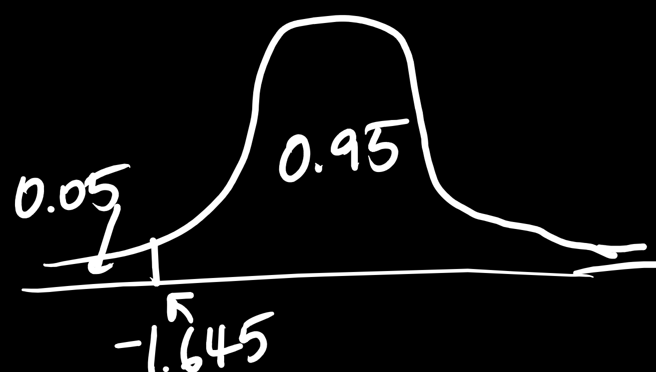


In a certain political campaign, Captain America has a poll taken at random among the voting population. The results show that y = 185 out of n = 351 voters favor him.

Create a **95%** **<u>1 sided lower bound</u>** **Confidence interval** for the true proportion of voters, p, who favor Captain America.

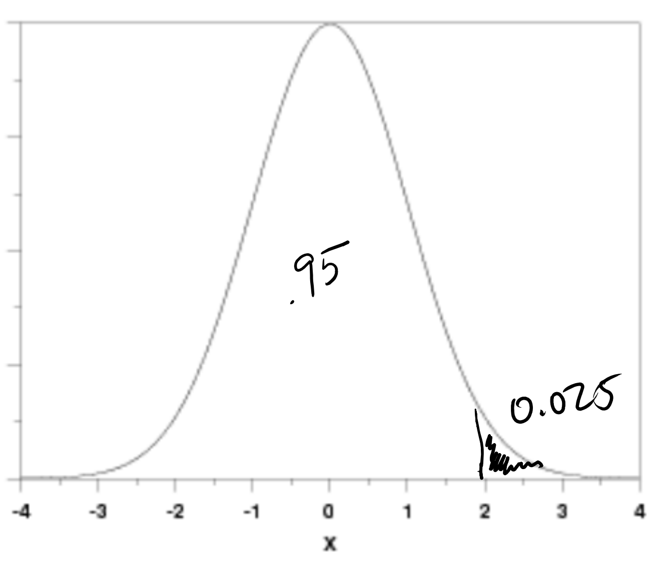
$$\left(0.527 - 1.645\sqrt{\frac{(0.527)(0.473)}{351}}, 1\right)$$

$$= (0.483, 1)$$



Example

In Covidland, out of a random sample of size 100, 60 people prefer funnels to straws. Based on this information, construct a 95% confidence interval for the true proportion of people who prefer funnels.



Notes