Confidence Intervals for Proportions

7.3

Today’s topics

Review:

- Confidence Intervals for mean and variance

New:

- Confidence Interval for proportions

$Var[\bar{Y}] = \frac{\sigma^2}{n}$ qnorm(0.025) σ $\frac{\sigma}{\sqrt{n}}$

General Form of CI for mean (review)

$-Z_{\alpha/2}$ $Z_{1-\alpha/2}$

- Estimate $\pm$ (Critical Value * $\boxed{\text{SE}}$ of estimate)

  e.g. if σ is known:

95% CI $\alpha = 0.05$ $\alpha/2 = 0.025$

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

right tail prob

  if σ is unknown:

s

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

Confidence Interval for σ^2 (review)

$1-\alpha/2$

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 $(1\text{-}\alpha)100\%$ 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.

P

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

Notes

$\bar{X}$

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 $\boxed{(Y = \sum X_i)}$

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)

$\bar{X}$

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

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

Interesting side notes:

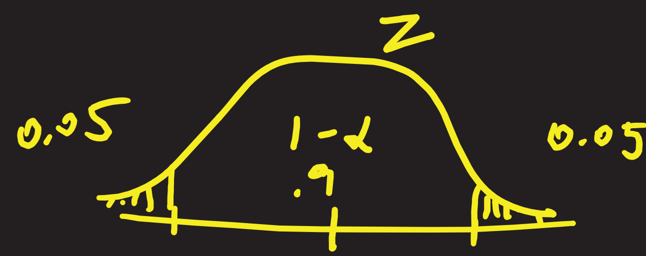
The variance of a Bernoulli is $p(1-p)$. If we sum n independent Bernoulli R.V.s, we get a $Binom(n,p)$ distribution, and its variance is $np(1-p)$.

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

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

Notes

Creating a CI for p



$z_{\alpha/2}$ 1.645

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

90% $\alpha = 0.1$

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

such that .9

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

Creating a CI for p

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

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

multiply by (-1)

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

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)