Confidence interval for a population proportion.

We want to measure the proportion of a population that has a given property. For example, what proportion:

(i) of US voters are disgusted with politics?

(ii) of cell phone users develop brain cancer?

(iii) of cross-bred peas will turn out to be green?

We know that a 95% confidence interval will look like

$$\left(\bar{x} - 1.96\frac{s}{\sqrt{n}},\ \bar{x} + 1.96\frac{s}{\sqrt{n}}\right) \qquad (CI_{95})$$

for a random sample of size n, but: what are $\bar{x}$ and s?

Answer: assign a 1 to each member of the population that has the property, and a 0 to each one that doesn't. Suppose the number in the sample who have the property is f. Then the number who don't is n-f. So

$$\bar{x} = \frac{f \cdot 1 + (n-f)\cdot 0}{n} = \frac{f}{n},$$

which is just the sample proportion (the proportion of the sample that has the property). Call this proportion $\hat{p}$. So $\bar{x} = \hat{p}$.

Also,

$$s = \sqrt{\frac{f(1-\hat{p})^2 + (n-f)(0-\hat{p})^2}{n-1}}.$$

If we approximate $n-1$ by n, we get

$$s \approx \sqrt{\frac{f(1-\hat{p})^2 + (n-f)\hat{p}^2}{n}} = \sqrt{\frac{f}{n}\cdot(1-\hat{p})^2 + \frac{(n-f)}{n}\cdot\hat{p}^2}.$$

Now again, $f/n = \hat{p}$, so $\frac{n-f}{n} = 1 - \frac{f}{n} = 1-\hat{p}$. So we get

do the math

$$s \approx \sqrt{\hat{p}(1-\hat{p})^2 + (1-\hat{p})\hat{p}^2} = \sqrt{\hat{p}(1-\hat{p})}.$$

Put this info back into (CI_{95}), to find:

CONCLUSION.

A 95% confidence interval for a population proportion p is given by

$$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right), \quad (*)$$

where $\hat{p}$ is the sample proportion.

Example 1.

Gregor Mendel, in 1856, crossbred pure green pea plants with pure yellow pea plants. Two generations later, he counted 152 green pea plants and 428 yellow ones. Find a 95% confidence interval for the proportion of green plants that would generally result from this process.

Solution.

We have $\hat{p} = 152/(152+428) = 152/(580)$
$= 0.262.$

We compute that

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.262(1-0.262)}{580}} = 0.018,$$

So our 95% confidence interval for p is

$$(0.262 - 1.96 \cdot 0.018,\ 0.262 + 1.96 \cdot 0.018)$$

$$= (0.227,\ 0.297).$$

Note:

98% and 99% confidence intervals for a population proportion look just like (*), but with 2.33 or 2.58, respectively, in place of 1.96.