Confidence interval for proportion

Thursday, March 11, 2021 5:03 PM

7.6. Confidence interval for a proportion.

Estimate the proportion of people who are left handed.

$$p = \frac{20}{100} = 20\%$$

n: size of the sample.

$$p_s = \left(\frac{x}{n}\right) \quad (*)$$

For Binomial distributions-

$X \sim B(n, p)$

→ Normal distribution $\xrightarrow{n>5}$ $N(np, npq)$

where, $q = 1-p$.

$$N\left(\frac{1}{n}E(X), \frac{1}{n^2}\text{var}(X)\right)$$

$$= N\left(\frac{1}{n}\cdot np, \frac{1}{n^2}\cdot npq\right)$$

$$= N\left(p, \frac{pq}{n}\right), \quad q = 1-p.$$

$Z = aX + bY$

$E(Z) = aE(X) + bE(Y)$

$\text{var}(Z) = a^2\text{var}(X) + b^2\text{var}(Y)$

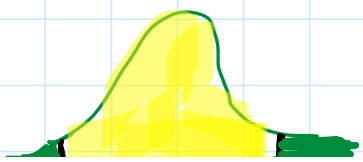
$Z = \frac{1}{a}X$

$E(Z) = \frac{1}{a}E(X)$

95% confidence interval

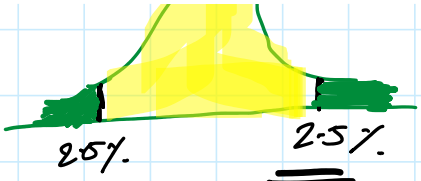
C.I.:

$$\left(p_s - 1.96\sqrt{\frac{pq}{n}}, \; p_s + 1.96\sqrt{\frac{pq}{n}}\right)$$



$$\left(p_s - 1.96\sqrt{\frac{pq}{n}},\ p_s + 1.96\sqrt{\frac{pq}{n}}\right)$$

$p_s \leftarrow$ sample proportion.



0.025 $\qquad$ ± 1.96

$$z_{score} = -1.96$$

{90% confidence interval} $z_{score} = 1.645$

{99% confidence interval} $z_{score} = 2.576$

A large urban area is considering introducing making two of the five lanes on one of the main routes into the area into high-occupancy vehicle lanes at peak travel times to try to ease traffic congestion. Planning officials commission a study to estimate the proportion of vehicles currently carrying only the driver.

The study recorded 653 vehicles entering the area on that route between 8.10 am and 8.15 am on a randomly selected ordinary working day. Analysis of the video recording suggested that 362 of these vehicles carried only the driver.

i) Calculate a 95% confidence interval for the proportion of vehicles currently carrying only the driver on that route at peak times.

ii) If the planning officials want to collect more data, suggest how they might do it in order to get the best information possible for their situation.

$n = 653$, $p_s = \frac{362}{653}$, $q_s = 1 - \frac{362}{653} = \frac{291}{653}$

So 95% confidence interval is given by

$$\left(\frac{362}{653} - 1.96\sqrt{\frac{\frac{362}{653}\times\frac{291}{653}}{653}},\ \frac{362}{653} + 1.96\sqrt{\frac{362\times 291}{653^3}}\right)$$

$= (0.516, 0.592)$

1. The sample size, n, and number of people, s, in the sample satisfying a particular criterion is given below for a number of surveys. Assuming these are random samples, calculate

i) a 95% confidence interval for the population proportion (0.239, 0.378)

ii) a 90% confidence interval for the population proportion satisfying the criterion in each survey.

a) $n = 120; s = 37$

$$p_s = \frac{s}{n} = \frac{37}{120}$$

$$q_s = 1 - p_s = 1 - \frac{37}{120} = \frac{83}{120}$$

$$N\left(p_s, \frac{p_s q_s}{n}\right)$$

$$\sqrt{\frac{p_s q_s}{n}} = \frac{37 \times 83}{120^3} = \sqrt{0.0018}$$

95% Confidence interval is

$$= \left(\frac{37}{120} - 1.96\sqrt{0.0018}, \frac{37}{120} - 1.96\sqrt{0.0018}\right)$$

$$= (0.226, 0.391)$$

2. A random sample of 100 bolts was measured and 12 of them were found to lie outside the production tolerance limits. Find a 95% confidence interval for the proportion of bolts which lie outside the production tolerance levels.

$$p_s = \frac{12}{100}, \quad q_s = \frac{88}{100}$$

95% CI $\Rightarrow$ (0.0563, 0.184)

4. During the morning 300 cars were observed on a busy road and 143 of them carried no passengers.

143 of them carried no passengers.

i) Find a 90% confidence interval for the proportion of cars on the road which carry no passengers. (0.429, 0.524)

ii) **State** any assumptions that you have had to make in constructing the confidence interval.

iii) An α% confidence interval is constructed using the same sample. The interval has width 0.15. Find the value of α.

width of the interval = 0.15

$$\left(p_s - k\sqrt{\frac{p_s q_s}{n}},\ p_s + k\sqrt{\frac{p_s q_s}{n}}\right) = (a, b)$$

width = (b − a)

$$p_s + k\sqrt{\frac{p_s q_s}{n}} - p_s + k\sqrt{\frac{p_s q_s}{n}}$$

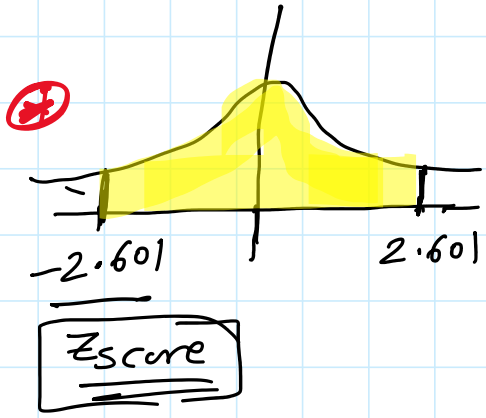
$$0.15 = 2k\sqrt{\frac{p_s q_s}{n}}$$

$$0.15 = 2k\sqrt{\frac{143 \times 157}{300^3}}$$

$$\Rightarrow k = 2.601$$

$$\Phi(2.601) = 0.991$$

$$\alpha = 99\%$$



5. A random sample of 250 items for which the prices were compared in two large shops found that Easipay were cheaper on 112 items than Valustore.

i) Find a 95% confidence interval for the proportion of items for which Easipay are cheaper than Valustore. $(0.386, 0.510)$

ii) Estimate the size of the sample needed for an approximate 95% confidence interval to have width 0.05.

$$\text{width} = 2 \cdot (1.96)\sqrt{\frac{0.247}{n}} < 0.05$$

$$\Rightarrow \frac{0.247}{n} < \left(\frac{0.05}{2 \times 1.96}\right)^2$$

$$\Rightarrow n > \left(\frac{2 \times 1.96}{0.05}\right)^2 \times 0.247$$

$$\Rightarrow n > 1520.02$$

$$\sqrt{n} > \frac{2 \times 1.96 \times \sqrt{0.448 \times 0.552}}{0.05}$$

$n > 1520.02$

$n = 1521$ ← size of sample