PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE19CS203 – STATISTICS FOR DATA SCIENCE

Unit - 3 - Probability Distributions

Question Bank - SOLVED

Confidence Intervals for Proportions

Exercises for Section 5.2

1. A soft-drink manufacturer purchases aluminum cans from an outside vendor. A random sample of 70 cans is selected from a large shipment, and each is tested for strength by applying an increasing load to the side of the can until it punctures. Of the 70 cans, 52 meet the specification for puncture resistance.

   a) Find a 95% confidence interval for the proportion of cans in the shipment that meet the specification.
   b) Find a 90% confidence interval for the proportion of cans in the shipment that meet the specification.
   c) Find the sample size needed for a 95% confidence interval to specify the proportion to within ± 0.05.
   d) Find the sample size needed for a 90% confidence interval to specify the proportion to within ± 0.05.
   e) If a 90% confidence interval is computed each day for 300 days, what is the probability that more than 280 of the confidence intervals cover the true proportions?

   [Text Book Exercise – Section 5.2 – Q. No. 3 – Pg. No. 342]
   Solution:

   (a) X = 52, n = 70,
   ˜p = (52+2)/(70+4) = 0.72973, z.025 = 1.96.
   The confidence interval is $0.72973 \pm 1.96\sqrt{0.72973(1-0.72973)/(70+4)}$, or (0.629, 0.831).

   (b) X = 52, n = 70, ˜p = (52+2)/(70+4) = 0.72973, z.05 = 1.645.

The confidence interval is $0.72973 \pm 1.645\sqrt{0.72973(1-0.72973)/(70+4)}$ or (0.645, 0.815).

(c) Let n be the required sample size.
Then n satisfies the equation $0.05 = 1.96\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$
Replacing $\tilde{p}$ with 0.72973 and solving for n yields n = 300.

(d) Let n be the required sample size.
Then n satisfies the equation $0.05 = 1.645\sqrt{\tilde{p}(1-\tilde{p})/(n+4)}$
Replacing $\tilde{p}$ with 0.72973 and solving for n yields n = 210.

(e) Let X be the number of 90% confidence intervals that cover the true proportion.
Then $X \sim Bin(300, 0.90)$,
X is approximately normal with mean $\mu_X = 300(0.90) = 270$ and standard deviation $\sqrt{300(0.90)(0.10)}$
= 5.196152.
To find $P(X > 280)$, use the continuity correction and find the z-score of 280.5.
The z-score of 280.5 is $(280.5-270)/5.196152 = 2.02$.
The area to the right of $z = 2.02$ is $1-0.9783 = 0.0217$.
$P(X > 280) = 0.0217$

2. In a random sample of 150 customers of a high-speed internet provider, 63 said that their service had been interrupted one or more times in the past month.
   a) Find a 95% confidence interval for the proportion of customers whose service was interrupted one or more times in the past month.
   b) Find a 99% confidence interval for the proportion of customers whose service was interrupted one or more times in the past month.
   c) Find the sample size needed for a 95% confidence interval to specify the proportion to within ± 0.05.
   d) Find the sample size needed for a 99% confidence interval to specify the proportion to within ± 0.05.

[Text Book Exercise – Section 5.2 – Q. No. 12 – Pg. No. 343]
Solution:

(a) $X = 63$, $n = 150$, $\tilde{p} = (63+2)/(150+4) = 0.42208$, $z_{.025} = 1.96$.
The confidence interval is
$0.42208 \pm 1.96\sqrt{0.42208(1-0.42208)/(150+4)}$, or (0.344, 0.500).

(b) $X = 63$, $n = 150$, $\tilde{p} = (63+2)/(150+4) = 0.42208$, $z_{.005} = 2.58$.

The confidence interval is

$0.42208 \pm 2.58\sqrt{0.42208(1-0.42208)/(150+4)}$, or (0.319, 0.525).

(c) Let n be the required sample size.

Then n satisfies the equation 0.05 = $1.96\sqrt{\tilde{p}(1-\tilde{p}/(n+4)}$

Replacing $\tilde{p}$ with 0.42208 and solving for n yields n = 371.

Therefore the level 100(1−α)% is closest to 80%.