Frequently Asked Questions

1. Explain when do we go for the estimation of population proportions?

Answer:

A qualitative characteristic which cannot be measured quantitatively is known as an attribute. For example, honesty, beauty, intelligence etc. Quite often we come across the situations where it may not be possible to measure the characteristic under study but it may be possible to classify the whole population into various classes with respect to the attributes under study.

We consider the cases where the population is divided into two classes only say C and C' with respect to an attribute. Such a classification is termed as dichotomous classification. Hence, any sampling unit in the population may be placed in class C or C' respectively according as it possess or does not possess the given attribute. In the study of attributes we are interested in the estimate of population proportion. Also when the means are not available best measure to estimate is proportion

2. Give some examples of the estimation of population proportions.

Answer:

1) Proportion of defective items in a large consignment of such items
2) Proportion of the literates or the bread winners in a town
3) What percentage of sales is accounted for by a particular product?
4) Proportion of viewing public that a particular program
5) Proportion of trees with a diameter of 11 inches or more

3. Under random sampling for attributes, show that population mean is equal to population proportion and the sample mean is the sample proportion.

Answer:

Let us suppose that a population with N units Y_1, Y_2, …., Y_N is classified into two disjoint and exhaustive classes C and C' respectively with respect to a given attribute. Let the number of individuals in the classes C and C' be X and X' respectively such that X+X' = N

Then,

P = The Proportion of units possessing the given attribute = X/N
Q= The Proportion of units does not possess the given attribute = X'/N = 1-P.
In statistical language P and Q are the proportion of successes and failures respectively in the population.
Let us consider random sample of size n from this population. Let 'x' is the number of units in the sample possessing the given attribute
Then p = Proportion of sampled units possessing the given attribute = x/n and q= Proportion of sampled units which do not possess the given attribute = 1 – p
With the i^{th} sampling unit, let us associate a variate Y_i (i=1,2,…,N) defined as follows

Y_i = 1, if it belongs to the class C i.e., if it possess the given attribute
And Y_i = 0, if it belongs to the class C' i.e., if it does not possess the given attribute.
Similarly let us associate a variable y_i (i=1,2,….,n) with the i^{th} sample unit defined as follows.

y_i = 1 , if the i^{th} sampled unit possess the given attribute
And Y_i = 0 , if the i^{th} sampled unit does not possess the given attribute.
Then $\sum Y_i$ = X, the number of units in the population possessing the given attribute

$\sum y_i$ = x, the number of sample units possessing the given attribute

Thus, $\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{X}{N} = P$ and $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{x}{n} = p$

4. When we draw a random sample of size n from a population then show that sample proportion can be used as an point estimate of population proportion.

Answer:

We know that in random sampling the sample mean provides an unbiased estimate of the population mean. That is

$$E(\overline{y}) = \overline{Y}$$

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{X}{N} = P \quad \text{-- -----(1)}$$

and $$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{x}{n} = p \quad \text{-----------------(2)}$$

But from Equations (1) and (2) we have
$\overline{Y} = P$ and $\overline{y}$ = p
Hence, E (p) = P
Since a sample proportion p is an unbiased estimator of a population proportion P we can use sample proportion as a point estimate of population proportion.

5. What do you mean by interval estimation of a population proportion?

Answer:

Interval estimation for a population proportion is nothing but finding an interval for a given level of confidence (1- α) in which a population proportion is expected to lie. A confidence interval for the population proportion P gives an estimated range of values which is likely to include an unknown population parameter, in this case P, the population proportion, the estimated range being calculated from a given set of sample data.

6. Derive Confidence Interval for population proportion based on a large sample i.e. when n tends to infinity (n is a sample size).

Answer:

We have to find 100 (1- α) % C.I for the population proportion. Let X be a number of units possessing an attribute (in the population) . Let P be the population proportion

Let X~B(n,P) Then E(X) = nP and V(X) = nPQ=nP(1-P)

Let us take a sample of size n from the above population.

Let p=x/n, be the sample proportion where x is the number of units possessing an attribute in the sample and n is the sample size

E(p)= E(x/n)=[1/n] E(x) = [1/n]nP = P

$V(p)=V(x/n) = [1/n^2]V(x)= [1/n^2]nPQ=PQ/n$

As the sample size is large (as n tends to infinity) Binomial distribution tends to Normal distribution

Therefore, p ~ N(P,PQ/n)

$$Z= \frac{p-P}{\sqrt{\frac{PQ}{n}}} \sim N(0,1)$$

For the given (1-α) we can always find two quantities $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ from Standard Normal variate tables such as

$$P[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}] = 1-\alpha$$

$$P\left[-Z_{\alpha/2} \leq \frac{p-P}{\sqrt{\frac{PQ}{n}}} \leq Z_{\alpha/2}\right] = 1-\alpha$$

$$P\left[-Z_{\alpha/2}\sqrt{\frac{PQ}{n}} \leq p - P \leq Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right] = 1-\alpha$$

$$P\left[p-Z_{\alpha/2}\sqrt{\frac{PQ}{n}} \leq P \leq p + Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\right] = 1-\alpha$$

Therefore, 100 (1- α) % C.I for the population proportion P is given by

$$[p-Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\ ,\ p\ +Z_{\alpha/2}\sqrt{\frac{PQ}{n}}\ ]$$

7. Suppose the value of P is unknown, then how do you get an interval estimate of population proportion?

Answer:

If P is not known then we use the estimate of P

The estimate of P is p

Therefore 100 (1- α) % C.I for the population proportion P is given by

$$[p-Z_{\alpha/2}\sqrt{\frac{pq}{n}}\ ,\ p\ +Z_{\alpha/2}\sqrt{\frac{pq}{n}}\ ]$$

8. What do you mean by Confidence Interval for the difference of proportions?

Answer:

- Confidence Interval for the difference of proportions is centred at the difference of the two sample proportions. This interval gives an estimated range of values which is likely to include the difference in the population proportions P1- P2 between two groups. The estimated range being calculated from a given sets of sample data.

- We use the two sample proportions to construct a confidence interval for the difference in population proportions
- Examples:

    - The difference in the proportion of males and females who will vote for Congress.
    - The difference in the proportion of Republicans and democrats who will vote for Gore.
    - The difference in the proportion of smokers and non-smokers in a district.

9. Derive a Confidence Interval for the difference of two population proportions of two independent populations.

Answer:

On the basis of independent random samples of sizes n_1 and n_2 from two Binomial populations it is desirable to estimate the difference between the parameters P1 and $P_{2.}$

For example: If we wish to estimate the difference between the proportion of voters in two different constitutes that favour a candidate X for a particular election.

We have to find 100 (1- α) % C.I for the difference of population proportions. Let X_1 be a number of units possessing an attribute in the first population. Let P_1 be the population proportion of that population

Let X_{1i}~$B(n_1,P_1)$ Then $E(X_1) = n_1P_1$ and $V(X_1) = n_1P_1Q_1=n_1P_1(1-P_1)$

Let us take a sample of size n_1 from the above population. Let $p_1=x_1/n_1$, be the sample proportion where x_1 is the number of units possessing an attribute in the sample of the first population and n_1 is the sample size

$E(p_1)= E(x_1/n_1)=[1/n_1]\ E(x_1) = [1/n_1]n_1P_1 = P_1$

$V(p_1)=V(x_1/n_1) = [1/_{n1}{}^2]V(x_1)= [1/_{n1}{}^2]n_1P_1Q_1=P_1Q_1/n_1$

Let X_2 be a number of units possessing an attribute in the second population . Let P_2 be the population proportion of that population

Let X_{2i}~$B(n_2,P_2)$ Then $E(X_2) = n_2P_2$ and $V(X_2) = n_2P_2Q_2=n_2P_2(1-P_2)$

Let us take a sample of size n_2 from the above population. Let $p_2=x_2/n_2$, be the sample proportion where x_2 is the number of units possessing an attribute in the sample of the second population and n_2 is the sample size

$E(p_2)= E(x_2/n_2)=[1/n_2]\ E(x_2) = [1/n_2]n_2P_2 = P_2$

$V(p_2)=V(x_2/n_2) = [1/_{n2}{}^2]V(x_2)= [1/_{n2}{}^2]n_2P_2Q_2=P_2Q_2/n_2$

As the sample size is large (as n tends to infinity) Binomial distribution tends to Normal distribution

Therefore $p_1 \sim N(P_1,P_1Q_1/n_1)$ and $p_2\sim N(P_2,P_2Q_2/n_2)$

$p_1 – p_2 \sim N(P_1 – P_2,\ \ P_1Q_1/n_1 + P_2Q_2/n_2)$

$$Z=\frac{(p_1-p_2)\text{-}(P_1-P_2)}{\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}}\sim N(0,1)$$

For the given (1-α) we can always find two quantities $–Z_{\alpha/2}$ and $Z_{\alpha/2}$ from Standard Normal variate tables such as

$P[-Z_{\alpha/2} \le Z \le Z_{\alpha/2}] = 1- \alpha$

$$P[-Z_{\alpha/2} \le \frac{(p_1-p_2)\text{-}(P_1-P_2)}{\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}} \le Z_{\alpha/2}] = 1-\alpha$$

$$P[-Z_{\alpha/2}\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}} \le (p_1-p_2)\text{-}(P_1-P_2) \le Z_{\alpha/2}\sqrt{\frac{P_1Q_1}{n_1}+\frac{P_2Q_2}{n_2}}] = 1-\alpha$$

P[$(p_1 - p_2)$) -Z $_{\alpha/2}$ $\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$ ≤ $(P_1 - P_2)$ ≤ $(p_1 - p_2)$ + Z $_{\alpha/2}$ $\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$] = 1- α

Therefore, 100 (1- α) % C.I for the difference of population proportions is given by

[($p_1 - p_2$)) -Z $_{\alpha/2}$ $\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$,($p_1 - p_2$) + Z $_{\alpha/2}$ $\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}$]

10. Suppose P1 and P2 are not known, then how do you obtain an interval estimate for the difference in the population proportions?

Answer:

If P_1 and P_2 are not known then we use the estimates of P_1 and P_2 as p_1 and p_2.

Therefore 100 (1- α) % C.I for the difference of two population proportions is given by

[($p_1 - p_2$)) -Z $_{\alpha/2}$ $\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$, ($p_1 - p_2$) + Z $_{\alpha/2}$]

11. What are the steps to be followed for the construction of the C.I for the population proportion?

Answer:

Steps to construct Confidence Interval for the population proportion

- Identify a sample statistic. Use the sample proportion as a sample statistic to estimate population proportion.
- Select a confidence level. The confidence level describes the uncertainty of a sampling method. Often, researchers choose 90%, 95%, or 99% confidence levels; but any percentage can be used.

Find the margin of error= critical value (table value) into standard deviation.= Z α/2 $\sqrt{\frac{pq}{n}}$ where q = 1-p

- Z α/2 is obtained from the table of Normal probabilities for a given level of significance α

- Then 100 (1- α) % C.I for the population proportion P is given by

$$[p - Z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}]$$

12. Narrate the steps to be followed for the construction of the C.I for the difference between two population proportions?

Answer:

Steps to be followed for the construction of the C.I for the difference between two populations proportions

- Identify a sample statistic. Use the difference between samples proportions to estimate the difference between population proportions.
- Select a confidence level. One can choose 90%, 95%, or 99% confidence levels; but any percentage can be used.
- Find the margin of error. = critical value (table value) into standard deviation of ($p1 - p_2$)= $Z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ where $q_1 = 1-p_1$ and $q_2 = 1-p_2$
- $Z_{\alpha/2}$ is obtained from the table of Normal probabilities for a given level of significance α Then 100 (1- α) % C.I for the difference of two population proportions is given by
- $[((p_1 - p_2)) - Z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}},\ (p_1 - p_2) + Z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}]$

13. 1000 randomly selected Americans were asked if they believed the minimum wage should be raised. 600 said yes. Construct a 95% confidence interval for the proportion of Americans who believe that the minimum wage should be raised.

Solution:

We have, p = 600/1000 =0 .6; q=1-p=0.4

Since 100 (1- α) % =95%; α=0.05. Then from the table of probabilities of Normal distribution $Z_{\alpha/2}$ = 1.96 and n = 1000

Then, 100 (1- α) % C.I for the population proportion P is given by

$$[p - Z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}]$$

$[0.6- 1.96 \sqrt{\frac{(0.6)(0.4)}{1000}}, 0.6+ 1.96 \sqrt{\frac{(0.6)(0.4)}{1000}}]$

=[0.57,0.63]

Hence we can conclude that between 57 and 63 percent of all Americans agree with the proposal. In other words, with a margin of error of .03, 60% agree.

14. Here is the data related to proportion of persons who would vote for a guilty verdict in a particular sexual harassment case from a study by Egbert, Moore, Wuensch, and Castellow. Of 160 mock jurors of both sexes, 105 voted guilty and 55 voted not guilty. Among the 80 male jurors 47 voted guilty. Among the 80 female jurors 58 voted guilty. Construct a 95% confidence interval for the difference between the two proportions based on sex.

Answer:

Proportion of female jurors who voted guilty = p_1= 58/80 = 0.725

Proportion of male jurors who voted guilty = p_2= 47/80 = 0.588

q_1 =1- p_1= 1- 0.725 = 0.275

q_2 =1- p_2= 1- 0.588 = 0.412

Since 100 (1- α) % =95%; α=0.05. Then from the table of probabilities of Normal distribution $Z_{\alpha/2}$ = 1.96

Then 100 (1- α) % C.I for the difference of two proportions in males and females is given by

$$[((p_1 - p_2)) - Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}, (p_1 - p_2) + Z_{\alpha/2} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}]$$

$$[(0.725-0.588)- 1.96 \sqrt{\frac{(0.725)(0.275)}{80} + \frac{(0.588)(0.412)}{80}}, (0.725-0.588)+ 1.96 \sqrt{\frac{(0.725)(0.275)}{80} + \frac{(0.588)90.412)}{80}}]$$

=[-.009, 0.283]

Notice that this confidence interval includes the value of zero which implies there may be negligible difference in the difference of two sexes who voted guilty.

15. A simple random sample of size 100 is drawn from a population of students of size 300 who have two wheelers. Obtain 95% CI for the proportion of students in a college who possess two wheelers?

Answer:

Given n = 100 and N = 300
A fraction of students in a college who have two wheelers
Xi = 0 if she/ he hasn't possess two wheeler
Xi = 1 if she/ he has a two wheeler

Student: 1,2,3, ,4, ………………………….., 99,100
Xi : 0,1,1,………………………………..,1,1,1
∑Xi = 65
Estimate of the proportion of students in a college who have two wheelers

$$p = \frac{\sum_{i=1}^{n} x_i}{100} = \frac{65}{100} = 0.65$$

q=1-p=0.35

Since 100 (1- α) % =95%; α=0.05. Then from the table of probabilities of Normal distribution

$Z_{\alpha/2} = 1.96$

Then 100 (1- α) % C.I for the population proportion P is given by

$$[p - Z_{\alpha/2}\sqrt{\frac{pq}{n}},\ p + Z_{\alpha/2}\sqrt{\frac{pq}{n}}]$$

$$[0.65 - 1.96\sqrt{\frac{(0.65)(0.35)}{100}},\ 0.65 + 1.96\sqrt{\frac{(0.65)(0.35)}{100}}]$$

[0.55651,0.74349]

Hence, we can conclude that between 56 and 74 percent of students in a college possess two wheelers. In other words, with a margin of error of .09, 65% of the students possess two wheelers.