Study	OR	95% CI	Inference
A	0.9	0.6–1.0	Useless
B	1.0	0.9–1.4	
C	1.2	1.1–1.5	Less useful
D	1.4	1.1–1.6	Useful
E	1.5	1.2–1.8	Most useful

TESTS OF SIGNIFICANCE

Overview

Test	Parametric	Nonparametric
Distribution	Normal	Nonnormal
Type of data	Quantitative	Quantitative
Sample size	Large	Small
Compares	Mean (SD)	%, proportions and fraction

1. Categorical vs Categorical Data

- **2 variables unrelated**
  - Chi-square
  - Fisher's exact
- **2 variables related**
  - McNemar
- **> 2 variables unrelated**
  - Chi-square
  - Fisher's exact

2. Categorical vs Quantitative

a. *2 variables*

- **Parametric unrelated**
  - Unpaired t test
- **Parametric related**
  - Paired t test
- **Nonparametric unrelated**
  - Mann Whitney/Wilcoxon rank sum test
- **Nonparametric related**
  - Wilcoxon signed rank test

b. *More than 2 variables*

- **Parametric unrelated**
  - ANOVA
- **Parametric related**
  - MANOVA
- **Nonparametric unrelated**
  - Kruskal-Wallis
- **Nonparametric related**
  - Friedman's test

Step 8: t test is applied

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SED}}$$

$$t = \frac{6-8}{1.8} = 1.1 \text{ (ignoring the sign)}$$

Step 9: Degree of freedom is calculated as explained earlier
$\text{df} = n_1 + n_2 - 2 = 9 + 9 - 2 = 16$

Step 10: t value table is referred

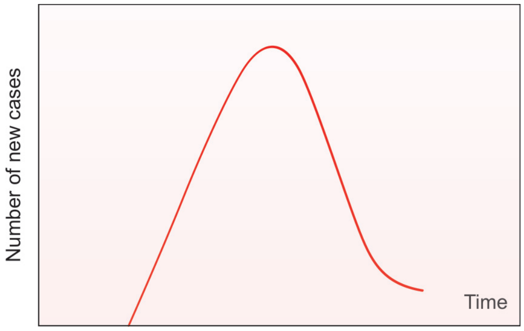
Interpretation: Calculated t-value is less than the table value. Hence, do not reject H_0. Hence, it can be concluded that the difference is not significant.

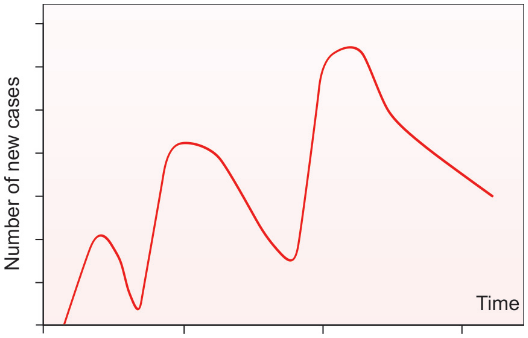
EPIDEMIOLOGY

EPIDEMIC CURVES

Theoretical Overview

Common exposure point source	Propagated
• Rapid rise and rapid fall (explosive)	• Slow rise and slow fall (only when Number of susceptible is depleted/ there is no more exposure)
• No secondary wave	• Secondary wave present
• All cases in one incubation period (IP) as there is brief exposure	• More than one IP
• Clustering of cases	• Herd immunity plays an important role in such type of epidemic
• For example, Bhopal gas tragedy, Minamata disease, Chernobyl gas disaster, food poisoning	• For example, hepatitis A, polio





Practical Overview 8

1. List the primary sampling units (column A) and their population sizes (column B). Each cluster has its own *cluster population size (a)*.
2. Calculate the cumulative sum of the population sizes (column C). The *Total population (b)* will be the last figure in column C.
3. Determine the *number of clusters (d)* that will be sampled in each strata randomly.
4. Determine the *Number of Individuals to be sampled from each cluster (c)*. In order to ensure that all individuals in the population have the same probability of selection irrespective of the size of their cluster, the same number of individuals has to be sampled from each cluster.
5. Divide the total population by the number of clusters to be sampled, to get the *sampling interval* (SI).
6. Choose a random number between 1 and the SI. This is the *random start* (RS). The first cluster sampled contains this cumulative population (column D).
7. Calculate the following series: $RS; RS + SI; RS + 2SI; \ldots. RS + (d - 1) \times SI.$
8. The clusters selected are those for which the cumulative population (column C) contains one of the serial numbers calculated in item 7. Depending on the population size of the cluster, it is possible that big clusters will be sampled more than once. Mark the sampled clusters in another column (column D).
9. Calculate for each of the sampled clusters the *probability of each cluster sampled (Prob 1)* (column E).

   Prob 1 = $(a \times d) \div b$

   a = Cluster population

   b = Total population

   d = Number of clusters
10. Calculate for each of the sampled clusters the *probability of each individual being sampled in each cluster (Prob 2)* (column G).

    Prob 2 = c/a

    a = Cluster population

    c = Number of individuals to be sampled in each cluster
11. Calculate the overall basic weight of an individual being sampled in the population. The basic weight (BW) is the inverse of the probability of selection.

$$BW = 1/(\text{prob } 1 \times \text{prob } 2)$$

Check for correct calculation: BW obtained from all the individuals being sampled in the population should be approximately same.

Practical Example

Q 7. Calculate sample size for a descriptive study to find the prevalence of anaemia in school children 5 to 15 years for a 95% power with 5% precision. Earlier studies have indicated that prevalence is approximately 40%. In addition, if the non-response rate is 10%, then what is the final sample size?

Answer:

For proportions

$$n = \frac{Z^2pq}{d^2}$$

where, n = sample size

Z^2 = abscissa of normal curve (1 – desired confidence interval)

As CI is 95%, so Z^2 is 4

p = estimated proportion of an attribute that is present in the population = 0.4

q = complement of $p = 1 - p = 0.6$

d = absolute precision = 0.05

$$n = \frac{4 \times 0.4 \times 0.6}{0.05 \times 0.05} = 384$$

If there is 10% nonresponse rate, then the final sample size will be

$0.9\ n = 384$

Final $n = 426.66$ (approximately 427)

- Common error done by students in this step

  $n = 384$

  $1.1\ n = 422.4$ (approximately 423)

Second approach is wrong as if you back calculate with second approach, then you will attain a sample size of 380.

Practical Example

Q 8. Suppose we require a 95% confidence interval for the mean of a continuous variable with a standard deviation of 15 to be no wider than 10 (i.e. *d* = 5%).

Answer:

For mean

$$n = \frac{Z^2\sigma^2}{d^2} = \frac{4 \times 15 \times 15}{5 \times 5} = 36$$

where, n = sample size

Z^2 = abscissa of normal curve (1 – desired confidence interval)

σ^2 = variance (square of standard deviation) of an attribute in population

d = absolute precision

Inference: In order to estimate the mean of a continuous variable (SD = 15) with 95% confidence interval no wider than 10, 36 participants would be required.