Statistical Inference

By
Dr. Asaad Al-Yassen
Assist. Professor of dermato-epidemiology
& occupational health
Department of Community Medicine
University of Basrah

Is the procedure followed in drawing calculations from a sample value for the population values, i.e. how to generalize information collected from a single sample to the entire population.

What is meant by population and sample?

Population: is an aggregate inmate or inanimate objects.
A sample: is a subset of the population.

Statistical Inference of Quantitative data:

- We can observe a sample mean ($\bar{X}$) and then use it as an estimate of the un known population mean (**μ**).
- The sample mean (**X**) and standard deviation (**Sd**) are used to estimate the mean of the population (**μ**).
- The sample mean is unlikely to be exactly equal to the population mean.
- A different sample would give a different, and this difference is due to sampling variation.
- If we take **100** samples, the means are $\bar{X}1, \bar{X}2, \bar{X}3, \ldots \bar{X}n$, and the mean of the population is **μ**. The difference between these mean and **μ** is called **standard error** of the sample mean (**SE**(**x**)).
- The standard error of the mean can be calculated as follow:

$$SE(x) = \frac{\sigma}{\sqrt{n}}$$

σ → it is the standard deviation of the population.(we seldom know the population standard deviation, so we can use the sample Sd to estimate the standard error.

n → is the total number of the sample.

- The size of the standard error depends on:

1. How much variation there is in the population.

2. Size of the sample, the larger the sample, the smaller standard error, so there is inverse relation between sample size and standard error of the mean.

Uses of Statistical Inference:

a. Estimation of population mean (**μ**):

***** when a sample is taken, we can use it's mean to estimate the population mean.

***** If repeated samples of size **n** were taken, the sample means would follow a normal distribution with mean (**μ**) and the standard deviation equal to **SE** (**x**).

***** So that, the characteristics of the normal distribution could be applied to the distribution of the sample means.

- So the Confidence Interval (**CI**) for population mean (**μ**)

At 95% CI for μ = $\bar{X} \pm [\, 1.96 \times SE(x) \,]$

i.e. **95% CI for μ** =

$$\underbrace{\bar{X} - [\, 1.96 \times SE(x) \,]}_{\text{lower limit}} \text{ ---- } \underbrace{\bar{X} + [\, 1.96 \times SE(x) \,]}_{\text{upper limit}}$$

At 99.7% CI for μ = $\bar{X} \pm [\, 2.58 \times SE(x) \,]$

i.e. **99.7% CI for μ** =

$$\underbrace{\bar{X} - [\, 2.58 \times SE(x) \,]}_{\text{lower limit}} \text{ ---- } \underbrace{\bar{X} + [\, 2.58 \times SE(x) \,]}_{\text{upper limit}}$$

Example: If a random sample has a mean of 5.11 with standard error 0.41, calculate the 95% confidence interval for population mean.

95% CI for μ = $\bar{X}$ ± [1.96 x SE (x)]

lower limit = 5.11 - [1.96 x 0.41]
= 4.31

upper limit = 5.11 + [1.96 x 0.41]
= 5.91

4.31 – 5.91 include the mean of the population from which the sample was taken.

b. Test whether sample belong to the population or not:

At 95% Confidence Interval

X = μ - 1.96 SE(x) ------------ μ + 1.96 SE(x)

At 99% Confidence Interval

X = μ - 2.58 SE(x) ------------ μ + 2.58 SE(x)

Statistical Inference of Qualitative data:

Proportion

- The proportion (**p**) in the sample with the characteristic is:

$$P = \frac{r}{n}$$

p = proportion of successful trial
r = number of successful trials
n = number of trials

Estimation of a population proportion (π):

- We can observe a sample proportion (**p)** and then use it to estimate the un known population proportion **(π).**

At 95% Confidence Interval for

$$\pi = p \pm [\ 1.96 \times SE(p)\ ]$$

At 99% CI for $\pi = p \pm [\ 2.58 \times SE(p)\ ]$

$$SE(p) = \sqrt{\frac{P(1-p)}{n}}$$