Module - 4

Statistical Inference

Q. Differenciate between population and sampling

:- Characteristics of a large group of individuals is studied in any stastical investigation. This group is reffered to as population under consideration.

A finite set of statistical individuals in a population is called a sample.

The process of making inference about the population based on the samples taken from which is known as statistical inference.

Q. Differenciate btw parameter and statistics.

Consider a population of objects, each of which has a numerical value associated with it any quantity that can be computed from this statistical population are called population parameters.

Eg: population mean μ, population variance σ^2 are parameters.

Any Quantity that can be computed from a sample is called a statistic.

Eg: Sample mean $\bar{x}$, sample variance s^2.

Q. Define sampling distribution:-

Let $x_1, x_2 \ldots x_n$ be a sample of size 'n' taken from a population. compute a statistics $t = t(x_1, x_2 \ldots x_n)$

$\because$ $x_1, x_2 \ldots x_n$ are random variables

t is also a random variable. Then the probability distribution of t is called the sampling distribution of statistic t.

The S.D of the sampling distribution is called the standard error of the sampling distribution.

4) Sampling Distribution of Mean

Consider a random sample $x_1, x_2, \ldots x_n$ of size n from a population with mean μ and variance σ^2.

The sample mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$\because$ x_i are random variables

$\bar{x}$ is also a random variable

The ~~Sampling~~ distribution of $\bar{x}$ is called Sampling distribution of mean.

We have the following result

Sampling Distribution of the sample mean $\bar{x}$ of a sample of size n taken from a normal population with mean μ and variance σ^2 is a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

Note:- ~~Var~~ The standard error for the sampling distribution of mean $= \frac{\sigma}{\sqrt{n}}$.

Q. The weights of students of a college are distributed with mean 63 kg and S.D is 8.2 kg. 14 students are choosen at random from the college. what is the probability that their mean weight is less than 65 kg.

∴ Let $\bar{x}$ be the sample mean. we know that $\bar{x}$ follows normal distribution with mean $\mu = 63$ and S.D $\frac{\sigma}{\sqrt{n}} = \frac{8.2}{\sqrt{14}}$ (Standard error)

$= 1.296$

put $z = \frac{\bar{x}-\mu}{\sigma} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{x}-63}{1.296} \sim (0,1)$

when $\bar{x} = 65$

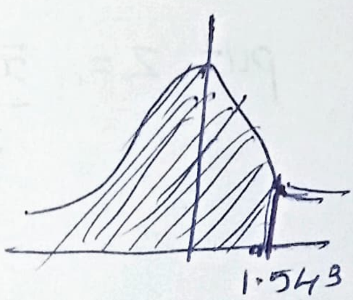
$z = \frac{65-63}{1.296} = \frac{2}{1.296} = \underline{\underline{1.543}}$

∴ $p(\bar{x} < 65) = p(z < 1.543)$

$= 0.5 + p(0 < z < 1.543)$

$= 0.5 + 0.4382$

$= \underline{\underline{0.9382}}$



Q The weights of students of a school are normally distributed with mean 55 kg and SD 2.2 kg. What is the probability that 9 students choosen at random from the school will have mean weight between 53 kg and 57 kg.

: Let x be the mean weight of the student then we know that $\bar{x}$ follows normal distribution with mean $\mu = 55$ and $\sigma (S.E) =$

$$\frac{\sigma}{\sqrt{n}} = \frac{2.2}{\sqrt{9}} = \frac{2.2}{3} = 0.733$$

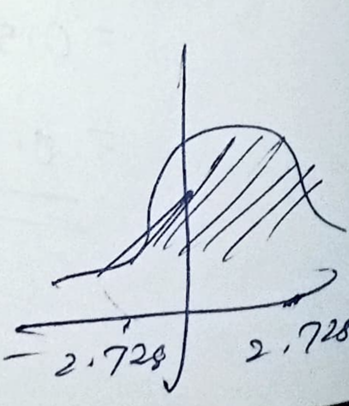
$P(53 < \bar{x} < 57)$

put $z = \frac{\bar{x} - \mu}{\frac{2.2}{3}}$

put $z = \frac{\bar{x} - \mu}{\frac{2.2}{3}} = \frac{\bar{x} - 55}{0.733}$

$\bar{x} = 53 \quad = \frac{53 - 55}{0.733} = -2.728$

$\bar{x} = 57 \quad = \frac{57 - 55}{0.733} = 2.728$



$P(53 < \bar{x} < 57) = P(-2.728 < Z < 2.728)$

$= 2P(0 < Z < 2.728) = 2 \times 0.4967$

$=$

Sampling Distribution of proportion.

Consider a population, the elements of which may be classified into two categories, those possessing a particular characteristic and those not possessing it.

Let p denotes the proportion of elements of the population which possess the characteristic (~~population~~ proportion of success in the population) and $q = 1-p$ be the proportion of those not possessing the characteristic.

Let x denote the number of successes out of n samples.

Then the sample proportion $\bar{p} = \frac{x}{n}$

The distribution of $\bar{p}$ is called sampling distribution of proportion.

Theorem

Consider a binomial population in which the proportion of success is p. Let $\bar{p}$ denote the proportion of success in a random sample of size n ~~seen~~ drawn from the population. If n is large ($np \geq 10$ and $n(1-p) \geq 10$), then $\bar{p}$ has approximately a normal distribution with mean p and standard deviation $\sqrt{\frac{pq}{n}}$

when $q = 1-p$

Note

The S.D [S.E] ~~of the~~ for the sample proportion is $\sqrt{\frac{pq}{n}}$.

Q. An air line claims that 72% all its planes to a certain region arrive on time. Assuming the air lines plane is true, find the probability ~~of~~ of a sample of 40 arrivals not more than 30 would be on time.

∵ Let $\bar{p}$ be the sample proportion of flights arriving on time.

We know that $\bar{p}$ follows normal distribution with mean $p = \frac{72}{100} = 0.72$ and S.D (S.E) $= \sqrt{\frac{pq}{n}}$

$q = 100 - 0.72$
$= 0.28$

$= \sqrt{\frac{0.72 \times 0.28}{40}}$

$= \sqrt{0.00504} = 0.07\overset{1}{\cancel{38}}$

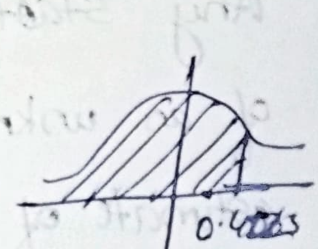
~~Take~~ $p\left(\bar{p} \le \frac{30}{40}\right)$

Take $z = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{\bar{p} - 0.72}{\cancel{0.0738}\ 0.071} \sim N(0,1)$.

when $\bar{p} = \frac{30}{40} \Rightarrow z = \frac{\frac{3}{4} - 0.72}{\cancel{0.0738}\ 0.071} = \cancel{0.4065}\ 0.4225$

$p\left(\bar{p} \le \frac{30}{40}\right) = p(z < 0.\cancel{4065}\,4225)$

$= 0.5 + p(0 \le z \le \cancel{0.4065}\ 0.4225)$



$= 0.5 +$

Statistical Inference is broadly classified into two.

1. Estimation of parameters
2. Testing of hypothesis

Estimation deals with methods of determining numbers which may be taken as the values of the unknown parameters as well as with the determination of intervals which will contain the unknown parameters with a specified probability, based on some samples taken from the population.

Point Estimation

Any statistic suggested as an estimate of an unknown parameter is called a point estimate of that parameter.

Interval Estimation

In interval estimation, we find out two statistics T_1 and T_2 such that the probability that the interval (T_1, T_2) contain the true value of unknown parameter has a preassigned value

$1-\alpha$ called the confidence coefficient or the confidence level of the interval. such an interval is called confidence interval with confidence coefficient $1-\alpha$.

Confidence interval for Mean.

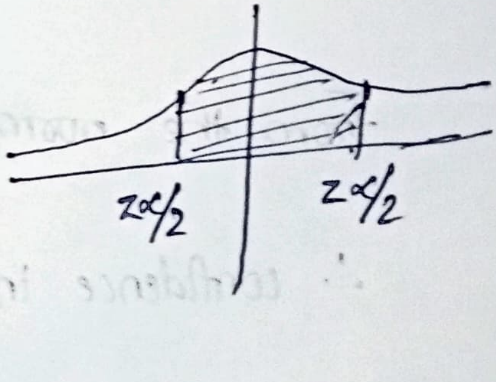
Suppose that a large random sample of size n is drawn from a population unknown mean and known standard deviation σ. if the sample mean $\bar{x}$, then the ~~coffi~~ confidence interval for the population mean with confidence coefficient $1-\alpha$ is given by

$$\cancel{\bar{x}-\frac{\sqrt{\alpha}}{\alpha}} \left(\bar{x}-\frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{x}+\frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

when $z_{\frac{\alpha}{2}}$ is obtained from the normal table such that probability of $\cancel{|z| < z_{\frac{\alpha}{2}}}$

$$\cancel{P(|z| < z_{\alpha/2}) = 1-\alpha}$$

$$P(|z| < z_{\alpha/2}) = 1-\alpha$$



If σ is unknown we use sample standard deviation 's'.

Q. The mean of a random sample size 121 drawn from a population with S.D 11.3 is 105.2. Find a 95% confidence interval for the population mean.

:- $n = 121$

$\sigma = 11.3$

$\bar{x} = 105.2$

$1-\alpha = 95\% = 0.95$

$\alpha = 1 - 0.95$

$= 0.05$

$\therefore \frac{\alpha}{2} = \frac{0.05}{2} = \cancel{0.020}\ 0.025$

The confidence interval $= \left(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right)$

From the normal table $Z_{\frac{\alpha}{2}} = \cancel{1.960}\ Z_{0.025} = 1.96$

$\therefore$ confidence interval $= \left(105.2 - \frac{11.3}{\sqrt{121}} \times 1.96,\right.$

$\left. 105.2 + \frac{11.3}{\sqrt{121}} \times 1.96\right)$

$$= \left(105.2 - \frac{11.3}{11} \times 1.96,\ 105.2 + \frac{11.2}{11} \times 1.96\right)$$

$$= (105.2 - 2.01,\ 105.2 + 2.01)$$

$$= \underline{\underline{(103.1865,\ 107.2134)}}$$

Q. Suppose the following 10 values represent random observations from a normal population

2, 6, 7, 9, 5, 1, 0, 3, 5, 4.

Construct a 99% confidence interval for the population mean.

∴ The confidence interval for mean =

$$\left(\bar{x} - \frac{s}{\sqrt{n}} Z_{\alpha/2},\ \bar{x} + \frac{s}{\sqrt{n}} Z_{\alpha/2}\right)$$

$n = 10$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{2+6+7+9+5+1+0+3+5+4}{10}$$

$$= \frac{42}{10} = \underline{\underline{4.2}}$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{1}{9}\left[\cancel{79.2}\ 66.72\right] = \underline{\underline{\cancel{8.8}\ \cancel{7.68}\ 7.73}}$$

$x-\bar{x}$	-2.2	$\cancel{2.2}$ 1.8	$\cancel{3.2}$ 2.8	5.2	1.2	-3.2	-4.2	-1.2	1.2	0.2
$(x-\bar{x})^2$	4.84	4.84	10.24	27.04	1.44	10.24	17.64	1.44	1.44	0.04

$\cancel{S = \sqrt{8.8}}$: $S = \sqrt{7.68} = 2.772$.

$1-\alpha = 0.99$ ⟶ $99\% = \frac{99}{100} = 0.99$

$\alpha = 1 - 0.99$

$= 0.01$

$\alpha/2 = 0.005$

$Z_{\alpha/2} = 2.576$

∴ confidence interval for mean =

$$\left(4.2 - \frac{2.772}{\sqrt{10}} \times 2.576,\ 4.2 + \frac{2.772}{\sqrt{10}} \times 2.576\right).$$

Q. Find the least sample size required if the length of the 95% confidence interval for the mean of a normal population with SD = 8 should be less than 10.

∴ confidence interval for mean =

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2},\ \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

length = b-a

$$= \left(\cancel{\bar{x}} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) - \left(\cancel{\bar{x}} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

$$= 2 \cdot \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$$

$\sigma = 8.$

$Z_{\alpha/2} = 1-\alpha = 1-95\%$

$= 1-0.95$

$=$

$\alpha/2 = 0.025$

$z = 1.960$

$2 \cdot \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < 10$

$\frac{2 \times 8}{\sqrt{n}} \times 1.960 < 10$

$\frac{2 \times 8 \times 1.960}{10} < \sqrt{n}$

$3.136 < \sqrt{n}$

squaring both sides.

$9.834 < n \Rightarrow n = 10, 11, 12, 13 \ldots$

So least sample size $= 10$

Confidence Interval for proportion

The confidence interval for the proportion p is given by $\left(\bar{p} - \sqrt{\frac{\bar{p}\bar{q}}{n}} \times Z_{\alpha/2},\ \bar{p} + Z_{\alpha/2}\sqrt{\frac{\bar{p}\bar{q}}{n}}\right)$

where $Z_{\alpha/2}$ is obtained from normal table satisfying

$$P\left(|z| < Z_{\alpha/2}\right) = 1-\alpha.$$

Q. In a survey conducted among adults in a city, 42 out of 900 adults reported themselves to be a vegetarians. Construct a 95% confidence interval for the ~~popu~~ proportion of all adults in the city who are vegetarians.

:-

$n = 900$

$x = 42.$

Let 'p' denote proportion of vegetarians.

Then the confidence interval for p is

$$\left(\bar{p} - \sqrt{\frac{\bar{p}\bar{q}}{n}} \times Z_{\alpha/2},\ \bar{p} + \sqrt{\frac{\bar{p}\bar{q}}{n}} \times Z_{\alpha/2}\right)$$

$$\bar{p} = \frac{x}{n} = \frac{42}{900}$$

$$\bar{q} = 1-\bar{p} = 1-\frac{42}{900} = \frac{858}{900}$$

$$1-\alpha = 0.95$$

$$\alpha = 1-0.95$$

$$= 0.05 \longrightarrow Z_{\alpha/2} = 0.05/2 = 0.025$$

$$Z_{\alpha/2} = \underline{1.960}$$

$\therefore$ Confidence interval =

$$\left(\frac{42}{900} - \sqrt{\frac{\frac{42}{900} \cdot \frac{858}{900}}{900}} \cdot 1.960,\ \frac{42}{900} + \sqrt{\frac{\frac{42}{900} \cdot \frac{858}{900}}{900}} \times 1.960\right)$$

$= (0.0328,\ 0.0604)$

below = 30
Small

Confidence interval for mean when the sample size is small.

Case 1. when σ is known

Let $x_1, x_2, \ldots x_n$ be a small sample size 'n' taken from a normal population with mean μ and S.D σ.

Then sampling distribution of mean $\bar{x}$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$ follows normal distribution with mean μ and SD $\frac{\sigma}{\sqrt{n}}$.

Case 2: when σ is unknown

In this case sampling distribution of mean follows Students t-distribution with n-1 degrees of freedom (df)

In case 1 the confidence interval for mean is given by

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2},\ \bar{x} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2}\right)$$ where $Z_{\alpha/2}$ is obtained from the normal table satisfy the condition $P(|Z| < Z_{\alpha/2})$

In case 2 the confidence interval for mean is given by

$$\left(\bar{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right)$$ where

$s^2 = \frac{1}{n-1} \Sigma(x - \bar{x})^2$ and $t_{\alpha/2}$ is obtained from t table with $n-1$ df.

Q. The following gives the values of 10 observations taken from a normal population with SD 1.2 construct a 97% confidence interval for mean.

* 10, 8, 12, 9, 11, 15, 7, 11, 12, 13

$\because n = 10$ (Small)

$\sigma = 1.2$

$\because \sigma$ is known the confidence Interval for mean =

$$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2} \;,\; \bar{x} - \frac{\sigma}{\sqrt{n}} \cdot Z_{\alpha/2}\right)$$

$$\bar{x} = \frac{x_1 + x_2 + \cdots x_n}{n} = \frac{10+8+12+9+11+15+7+11+12+13}{10}$$

$$= 10.8$$

$1 - \alpha = 97\%$

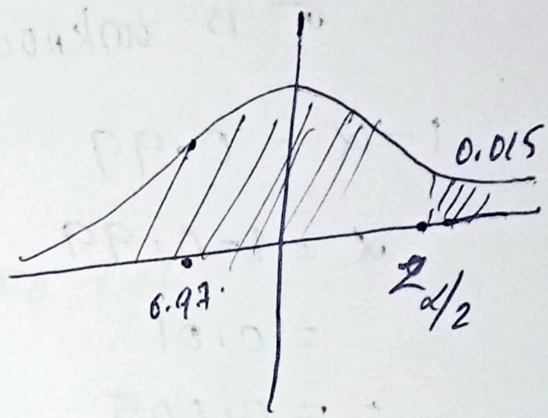
$= 0.97$

$\alpha = 1 - 0.97$

$= 0.03$

$\alpha/2 = 0.015$

$0.5 - 0.015 = 0.485$



From normal table

$Z_{\alpha/2} = 2.17$

$CI = (10.8 - 0.3794 \times 2.17 \;,\; 10.8 + 0.3794 \times 2.17)$

$= (9.977 \;,\; 11.623)$

Q. Suppose the following 10 values represent random observations from a normal population

2, 6, 7, 9, 5, 1, 0, 3, 5, 4.

construct a 99% confidence interval for the mean of the population.

:- Here $n = 10$ (small size)

σ is unknown

$1 - \alpha = 0.99$

$\therefore \alpha = 1 - 0.99$

$= 0.01$

$= 0.005$

confidence interval =

$$\left(\bar{x} - t_{\alpha/2} \times \frac{s}{\sqrt{n}} \,,\; \bar{x} + t_{\alpha/2} \times \frac{s}{\sqrt{n}}\right)$$

$$\bar{x} = \frac{2+6+7+9+5+1+0+3+5+4}{10} = \frac{42}{10} = 4.2$$

$$s^2 = \frac{1}{n-1} \Sigma (x - \bar{x})^2$$

$$= \frac{1}{9} \times (66.72) = \cancel{7.68}\; 7.73$$

$x-\bar{x}$	-2.2	1.8	2.8	5.2	1.2	-3.2	-4.2	-1.2	1.2	0.2
$(x-\bar{x})^2$										

so $S = \cancel{2.7820}.\ 2.780$

$\because \sigma$ is unknown sampling distribution of $\bar{s}$ follow Students t distribution with 9 df.

from t table $t_{\alpha/2} = 3.250$

$\therefore$ confidence level =

$(4.2 - 3.250 \times 0.8791,\ 4.2 + 3.250 \times 0.8791)$

$= \cancel{39.143},\ (1.343,\ 7.057)$