Sampling

In statistical investigation the interest usually lies in the assessment of variation of one or more characteristics of objects belonging to a group. This group of objects under study is called a population . Thus in statistics population is an aggregation of objects. animate or inanimate, under study.

The population may be finite or infinite. Examining the entire population for sake of assessment of variation may be difficult or even impossible to do. In such situations we consider a sample.

A finite subset of statistical individuals in a population is called a sample and the number of individuals in a sample is called the sample size.

The process of obtaining a sample is called a sampling. Sampling in which each member may be chosen more than once is called sampling with replacement. While if each member cannot be chosen more than once is called sampling without replacement.

Random sampling: If the sample units are selected at random then it is called random sampling. In this case each unit of population has an equal chance of being included in the sample.

Let X be random variable contain probability distribution with mean μ and variance σ^2. Let X_1, X_2 , ----- X_n be n independent random variables each having the same distribution of X. Then (X_1, X_2 , ----- X_n) is called a random sample of size n from X. μ is called population mean. σ^2 is called population variance.

- Let (X_1, X_2 , ----- X_n) be a random sample of size n from X .

  The statistic $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ is called sample mean and $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ is called sample variance.

- Let X be a random variable with expectation E(X)=μ and variance V(X)= σ^2 then E($\bar{X}$) $= \mu$ and V($\bar{X}$) = $\frac{\sigma^2}{n}$ where n sample size.
- Let (X_1, X_2 , ----- X_n) be a random sample of size n≥ 2 from a distribution N(μ, σ^2) then mean sample $\bar{X} \sim$N($\mu, \frac{\sigma^2}{n}$) and $\frac{ns^2}{\sigma^2} \sim \chi^2 \ (n-1)$.

Problem

1. Let $\bar{X}$ be the mean of a sample of size 5 from a normal distribution with mean $\mu = 0$, variance 125 find C so that P($\bar{X} < c) = 0.9$.

   Sol: n=5, X$\sim N(0,125)$, $\bar{X} \sim$N(0,$\frac{125}{5}$)

   Pr($\bar{X} < c$) = 0.9

   Pr($Z < \frac{c}{5}$) = 1.28 $\Rightarrow$ c= 6.40

2. If $\bar{X}$ is the mean of a random sample size n from aa normal distribution with mean μ and variance 100 find n so that Pr $\{\mu - 5 < \bar{X} < \mu + 5\} = 0.954$.

   Sol: , X$\sim N(\mu, 100)$, $\bar{X} \sim$N($\mu, \frac{100}{n}$)

   Pr $\left\{\frac{\mu-5-\mu}{\sigma} < \bar{X} < \frac{\mu+5-\mu}{\sigma}\right\} = 0.954$

   Pr $\left\{\frac{-5}{\sigma} < \bar{X} < \frac{5}{\sigma}\right\} = 0.954 \Rightarrow 2\phi\left(\frac{5}{\sigma}\right) = 1.954 \Rightarrow$2.5= $\sqrt{\frac{100}{n}}$ $\Rightarrow$ n=16

3. Let S^2 be the variance of a random sample of size 6 from the N $(\mu, 12)$ then find Pr $\{2.3 < S^2 < 22.2\}$

Sol: $\bar{X} \sim \mathrm{N}\left(\mu, \frac{12}{6}\right)$, $\frac{ns^2}{\sigma^2} \sim \chi^2\ (n-1)$

$\Rightarrow \frac{6s^2}{12} \sim \chi^2\ (5)$

Pr $\{2.3 < S^2 < 22.2\} \Rightarrow$ Pr $\{1.15 < S^2/2 < 11.1\}$

$\Rightarrow \phi(11.1) - \phi(1.15) = 0.90$

CENTRAL LIMIT THEOREM

Let X_1, X_2, ----- X_n denote a random sample of size n from a distribution that has mean μ and variance (σ^2. Then random variable $Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ has a limiting distribution N(0,1).

Proof : Let $S_n = X_1 + X_2 + \text{-----} + X_n$ where $E(X_i) = \mu,\ V(X_i) = \sigma^2$

$E(S_n) = E(\mathrm{X1} + \mathrm{X2} + ------ + \mathrm{Xn}) = \mathrm{n}\mu$

$V(S_n) = V(\mathrm{X1} + \mathrm{X2} + ------ + \mathrm{Xn}) = \mathrm{n}\sigma^2$

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

$M_{Y_n}(t) = E(e^{ty_n})$ Since the random variables are independent we get

$$M_{Y_n}(t) = \prod_{i=1}^{n} E(e^{t\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right)}$$

$$= \left[E(e^{t\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right)}\right]^n$$

Since the random variables are identically distributed

$$M_{Y_n}(t) = E\left[1 + \frac{t\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right)}{1!} + \frac{t^2\left(\frac{X_i - \mu}{\sigma\sqrt{n}}\right)^2}{2!} + o(n^{-3}/2)\right]^n$$

$$M_{Y_n}(t) = \left[1 + \frac{tE(X_i - \mu)}{\sigma\sqrt{n}} + \frac{t^2E(X_i - \mu)^2}{\sigma^2 n2!} + o(n^{-3}/2)\right]^n$$

$$M_{Y_n}(t) = \left[1 + \frac{t^2}{2n} + o(n^{-3}/2)\right]^n$$

For every fixed t. the term $o\left(n^{-3}/2\right) \to 0, n \to \infty$.

$$\lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty}\left[1+\frac{t^2}{2n}+o(n^{-3/2})\right]^n$$

$$= \lim_{n\to\infty}\left[1+\frac{t^2}{2n}\right]^{\frac{2nt^2}{2t^2}}$$

$$M_{Y_n}(t) = e^{t^2/2}$$

Which is mgf of a random variable with N(0,1) ⇒ Y_n has limiting distribution N(0,1).

Problem

1. Compute an approximate probability that the mean sample of size 15 from a distribution having pdf $f(x) = \begin{cases} 3x^2 & 0<x<1 \\ 0 & elsewhere \end{cases}$ is between 3/5 and 4/5.
   Sol: n=15 , $\mu = \frac{3}{4}, \sigma^2 = \frac{3}{80}$
   $$\bar{X} \sim \mathrm{N}\left(\frac{3}{4}, \frac{3}{80}\right)$$
   $$\Pr\left\{\frac{3}{5} < \bar{X} < \frac{4}{5}\right\} = \Pr\{-3 < z < 1\} = \phi(1) - (1-\phi(3)) = 0.840$$

2. Let $\bar{X}$ the mean of a random sample of size 100 from a distribution which is χ^2 (50). Compute an approximate value of $\Pr\{49 < \bar{X} < 51\}$.
   Sol: $X \sim \chi^2$ (50) , $\mu = n = 50$, $\sigma^2 = 2n = 100$, n=50
   $\frac{\bar{X}-\mu}{\sigma} \sim N(\mu, \frac{\sigma^2}{n})$
   $$\Pr\{49 < \bar{X} < 51\} = \Pr\left\{\frac{49-\mu}{\sigma} < \frac{\bar{X}-\mu}{\sigma} < \frac{51-\mu}{\sigma}\right\} = \Pr\left\{\frac{49-\mu}{\sigma} < \frac{\bar{X}-\mu}{\sigma} < \frac{51-\mu}{\sigma}\right\}$$
   $$\Pr\{-1 < Y < 1\} = 2\,\phi(1) - 1 = 0.682$$

3. Let $\bar{X}$ the mean of a random sample of size 128 from a Gamma distribution with r=2, $\alpha = 1/4$. Approximate $\Pr\{7 < \bar{X} < 9\}$.
   Sol: , $\mu = \frac{r}{\alpha} = 8$, $\sigma^2 = \frac{r}{\alpha^2} = 32$
   $$\Pr\{7 < \bar{X} < 9\} = \Pr\left\{\frac{7-8}{0.5} < \frac{\bar{X}-\mu}{\sigma} < \frac{9-8}{0.5}\right\} = 0.954.$$

4. Suppose that X_j j=1,2,-----50 are independent random variable each having a Poisson distribution with $\alpha = 0.03$. Let S= X_1+ X_2+----------+ X_{50} using the central limit theorem evaluate Pr(S≥ 3).
   Sol: S= X_1+ X_2+----------+ X_{50}
   E(S)= 50 × 0.03= 1.5, V(S) = 50 × 0.03= 1.5
   $$Y = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{S - E(S)}{\sqrt{V(S)}} = \frac{S-1.5}{\sqrt{1.5}} \sim N(0,1)$$
   Pr(S≥ 3) = 1- Pr(S$\leq 3) = 1 - \Pr\left(\frac{S-1.5}{\sqrt{1.5}} \leq \frac{3-1.5}{\sqrt{1.5}}\right) = 1 - \phi(\sqrt{1.5}) = 0.1093$

5. A distribution with unknown mean μ has a variance 1.5 find how large a sample should be taken from the distribution in order that the probability will be at least 0.95 that the sample mean will be within 0.5 of the population mean.

Sol: Pr $\{\mu - 0.5 < \bar{X} < \mu + 0.5\} = 0.95$

$$\Pr\left\{\frac{\mu - 0.5 - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\mu + 0.5 - \mu}{\frac{\sigma}{\sqrt{n}}}\right\} = 0.95$$

$$\Pr\left\{\frac{-0.5\sqrt{n}}{\sigma} < Y < \frac{0.5\sqrt{n}}{\sigma}\right\} = 0.95$$

$$2\phi\left(\frac{0.5\sqrt{n}}{\sigma}\right) = 1.95$$

$$\frac{0.5\sqrt{n}}{\sigma} = 1.96$$

$$n = 23.05$$