Nov. 17.

statistical inference → parameter estimation, linear regression
→ hypothesis testing, single sample test

point estimator:

A point estimate of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic

population parameter — sample statistics

Estimation example:

Mean of a single population μ, the estimate is $\hat{\mu} = \bar{x}$, sample mean

Variance of a single population σ^2, the estimate $\hat{\sigma}^2 = s^2$, sample variance

Difference in means of two population $\mu_1 - \mu_2$, estimate $\hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2$

Central limit theorem

The most important result in probability theory.

Fundamental idea:

the sum or average of a large number of Random Variables is approximately Normal.

steps: <1>. Suppose we have a random sample of size n : $X_1, X_2, \cdots, X_n$ taken from a population.

<2>. The population has a mean μ and finite variance σ^2

$$E(X_i) = \mu \quad . \quad V(X_i) = \sigma^2$$

<3>. Sample mean is defined as σ

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_n \qquad \sigma^2$$

std $= \sqrt{var}$

$\sigma = \sqrt{\sigma^2}$

$$E(\bar{X}) = \mu_{\bar{x}} = \frac{\mu + \mu + \cdots + \mu}{n} = \mu$$

Variance $V(\bar{X}) = \sigma^2_{\bar{x}} = V\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}(\sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}$

std $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

<4>. The sample mean is approximatly Normal distributed.

$$\bar{X} \sim \text{Normal}(\mu, \sigma^2/n) \qquad \mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Note: i). It doesn't matter what kind of distribution $X_1, \cdots, X_n$ are

Uniform, binomial, exponential, ...

ii). If $n \geq 30$, the normal approximation of sample mean will be satisfactory regardless of the population shape distribution

If $n < 30$, CLT will work only if the population distribution is not severely nonnormal.

<5>. Normalization.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{is the standard normal distribution}$$

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$N(\mu, \sigma^2)$

$N(\mu, \sigma^2)$

Eg 3.

Manufacturing ropes.

tensile strength has a distribution with mean 75.5 lbs and standard deviation 3.5 lbs.

$E(X) = \mu = 75.5$

$V(X) = \sigma^2 = 3.5^2$

Suppose we take a random sample of 6 ropes what's the prob. that the sample mean $\bar{X}$ 的 strength will exceed 75.75 lbs.

$\left(\frac{3.5}{\sqrt{6}}\right)^2 = (1.429)^2$

$\frac{3.5^2}{6} = 1.429^2$

$\mu_{\bar{x}} = \mu = 75.5$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.5}{\sqrt{6}} = 1.429$

$\bar{X} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2) \sim N(75.5, 1.429^2)$.

$P(\bar{x} \geq 75.75) = P\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \geq \frac{75.75 - 75.5}{1.429}\right)$

$= P(Z \geq 0.175)$

$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

$= 1 - P(Z < 0.175)$

$Z \neq \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}^2}$

$= 1 - 0.56945$

$= 0.43055$

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$