Lecture - 20

Probability & Statistics.

Use the tools of probability to solve statistics (problems).

The science of statistics deals with drawing conclusions from observed data.

Population: A large collection of items that have measurable values associated with them.

[By suitably sampling from this collection, and then analyzing the sampled items, one hopes to draw some conclusions about the collection as a whole.

Sample/random sample: If $X_1, X_2, \ldots, X_n$ are independent random variables having a common distribution F, then we say that they constitute a sample or random sample from the distribution F.

(1) the population distribution F will not be completely specified and one will attempt to use the data to make inferences about F

——— non parametric ~~interface for~~ inference problem.

(2) Sometimes, F is specified up to some unknown parameters. For example, F is normal distribution having unknown mean and variance.

— parametric inference problem.

Big deal: How to estimate the values of these parameters and how to verify that those are correct estimates.

Statistic: A statistic is a random variable whose value is determined by the sample data.

Two important statistics are **sample mean** and the **sample variance**.

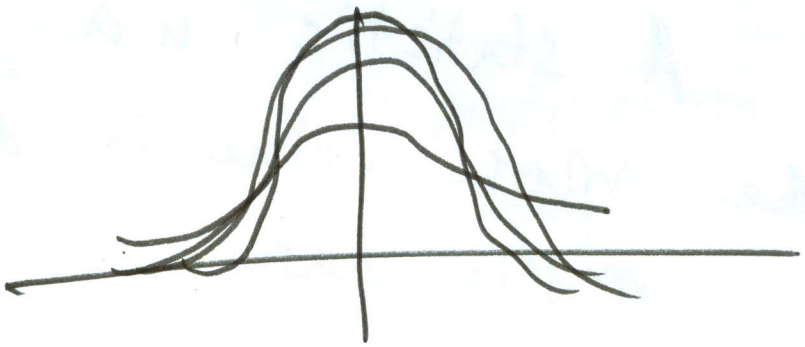
Sample mean: Let $X_1, X_2, \cdots, X_n$ be a sample of values from a population. Then the sample mean is defined by

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Let X_i have mean μ and variance σ^2, $i = 1, 2, \cdots, n$. This μ is called the population mean and σ^2 is called the population variance.

Then $$E[\bar{X}] = \frac{E[X_1] + E[X_2] + \cdots + E[X_n]}{n} = \mu$$

$$\text{var}(\bar{X}) = \frac{1}{n^2}\sum \text{var}(X_i) = \frac{\sigma^2}{n}$$



μ

C.L.T X_i's are i.i.d

$$P\left\{\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} < x\right\} \simeq \Phi(x)$$

as $n \to \infty$

$$\Downarrow$$

$$P\left\{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < x\right\} \simeq \Phi(x)$$

$$\swarrow$$

Here sample mean is μ and variance is $\sigma/\sqrt{n}$.

$\therefore$ $\bar{X}$ will approximately be normal when the sample size n is large.

10. The weights of a population of workers have mean 167 and standard deviation 27.

If a sample of 36 workers is chosen, approximate the probability that the sample mean of their weights lie between 163 and 170.

Soln:

$$P\{163 < \bar{x} < 170\}$$

$$\cong 2P\{N(0,1) < 0.8889\} - 1$$

$$\simeq 0.6259$$

Sample variance.

Let $X_1, \ldots, X_n$ be random sample from a distribution with mean μ and variance σ^2.

Let $\bar{X}$ be the sample mean.

Then the statistic S^2 defined by

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

is called the sample variance and $S = \sqrt{S^2}$ is called the sample standard deviation.

Notation $\sum_{i=1}^{n}(x_i - \bar{x})^2$, $\bar{x} = \frac{\sum x_i}{n}$

$$= \sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right)$$

$$= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

Then it follows that

$$(n-1)S^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

$$\therefore (n-1)E[S^2] = E\left(\sum_{i=1}^{n} X_i^2\right) - nE[\bar{X}^2]$$

$$= nE(X_i^2) - nE[\bar{X}^2]$$

$$= n\,Var(X_1) + n\left(E[X_1]^2\right) - n\,Var(\bar{X}_0) - n\left(E[\bar{X}]\right)^2$$

$$= n\sigma^2 + n\mu^2 - n(\sigma^2/n) - n\mu^2$$

$$= (n-1)\sigma^2$$

$$\therefore E[S^2] = \sigma^2$$

∴ the expected value of the sample variance s^2 is equal to the population variance.