17.1 Random Samples and Sampling Distributions *P.462*

Population
(N, μ, σ^2)

Each population is associated with a probability distribution e.g. Poisson / Normal *P.462*

Random Sampling *P.463*
1. Objective
2. Replacement / without replacement

Sample
(n, $\bar{x}$, s^2)

Sample distribution *P.464*

Ex 17.1
Construction of the sampling distribution

Calculation

Sample Statistics
($\bar{x}$, s^2)

Inference

17.2 Relationship between Sample mean and Population mean *P.469*
17.3 Relationship between Sample variance and Population variance *P.483*

Inference (i.e. known sample → unknown population) *Ex 17.3*

1. Point estimation - use sample data to calculate a single no. that can be used as an estimate of the population.
   (a) For many many sample
      (i) Sample mean ($\bar{x}$)

      $E(\bar{x}) = \mu$, *P.469* $\quad Var(\bar{x}) = \dfrac{\sigma^2}{n}$ *P.470*

      $\Rightarrow$ unbiased estimate(*P.470*) of the population mean.
      Ex 17.2

      (ii) Sample variance (s^2)

      $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$ *P.483*

      $\Rightarrow$ unbiased estimate of the population variance.
   (b) For one sample, n must be large (n $\to \infty$) *P.486*
2. Model Fitting - *Ch.18*

Exercise 17.3 Q2

Solution

2. unbiased estimation of
   (i) population mean = mean of sample ($\bar{x}$)
   (ii) population variance = sample variance σ_{n-1}

$$\mu = \frac{1}{10}(280+250+300+320+280+280+270+290+300+300)$$
$$= \$287_{\#}$$

$$\sigma^2 = \frac{1}{10-1}\sum_{i=1}^{n}(x_i - 287)^2$$
$$= 378.9\ (\$^2)_{\#}$$

Appendix 1 (ESTIMATION OF POPULATION PARAMETERS)

Statistical inference is the process of inferring information about a population from the data of samples drawn from it. There are two main lines of attack: estimation and hypotheses testing.

In estimation of population parameter, we can seek for a point estimation or an interval estimation.

(A) Point Estimation

A sample statistics used to provide an estimate of the corresponding population parameter is called a point estimator e.g. the sample mean $\bar{x}$ may be used as a point estimator of the population mean μ. In choosing an estimator which is to be considered as the 'best', it must be: unbiased, efficient and consistent.

(a) Unbiased estimator: if $\hat{\theta}$ is an estimate for the unknown parameter θ associated with the distribution of a random variable X, $\hat{\theta}$ is an unbiased estimator for θ if $E(\hat{\theta}) = \theta$. Otherwise, it is biased.

(b) Efficient estimator: for all unbiased estimators of the population parameter, the one with the smallest variance is called the most efficient estimator. Thus, if $\hat{\theta}_1$, $\hat{\theta}_2$ are unbiased estimators of θ, $\hat{\theta}_1$ is more efficient if $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$.

(c) Consistent estimator: let $\hat{\theta}$ be an estimate of θ based on a sample of size n. If $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ and $\lim_{n\to\infty} V(\hat{\theta}) = 0$ then $\hat{\theta}$ is a consistent estimate of θ.

[Note: (a) Because a sample is only a portion of the population — often a minute portion — the value of a sample statistic inevitably provides an imperfect estimate of population parameter of interest.

(b) Different sample statistics may be used to provide an estimate of a population parameter. For example, having select a random sample of size n, we may use the sample mean $\bar{x}$, the sample median, or the sample mode as a point estimate of the population mean μ. It can be shown that all the three are unbiased estimators of μ, and $\bar{x}$ is the most efficient estimator.]

Example (1) Show that the mean $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ is an unbiased and consistent estimator of the population mean μ.

(2) Show that $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ is an unbiased estimator of the population variance σ^2.

Solution (1) $E(\bar{x}) = \mu$, $\text{var}(\bar{x}) = \dfrac{\sigma^2}{n}$

$\therefore \lim_{n\to\infty} \text{var}(\bar{x}) = 0$

$\therefore \bar{x}$ is an unbiased and consistent estimator of μ.

(2)
$$E(s^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$
$$= \frac{1}{n-1}E[\sum_{i=1}^{n}\{(x_i-\mu)-(\bar{x}-\mu)\}^2]$$
$$= \frac{1}{n-1}[\sum_{i=1}^{n}E\{(x_i-\mu)^2\} - nE\{(\bar{x}-\mu)^2\}]$$

since $2\sum_{i=1}^{n}E[(x_i-\mu)(\bar{x}-\mu)] = 2(\bar{x}-\mu)\sum_{i=1}^{n}E(x_i-\mu) = 0$

Now $E[(x_i-\mu)^2] = \sigma^2$, $E[(\bar{x}-\mu)^2] = \dfrac{\sigma^2}{n}$

$\therefore E(s^2) = \dfrac{1}{n-1}[n\sigma^2 - \sigma^2] = \sigma^2$

i.e. s^2 is an unbiased estimator of σ^2.

(B) Interval estimation

In the previous section we have use a random sample to give an estimate of a population parameter. These estimates are, of course, just numbers and are called point estimates.

It is often preferable to give an interval estimate within which we have a reasonable degree of confidence that the parameter will lie.

For an interval estimation of the population parameter θ, we seek a pair of statistics, L and U, such that θ would be within the interval [L,U] with some preassigned probability. The interval [L,U] is called a confidence interval, and L, U are called the confidence limits.

Consider a population parameter θ, if the sample statistic used to estimate θ is distributed with mean $\hat{\theta}$ and variance s^2 then $[\hat{\theta} - s, \hat{\theta} + s]$ is called the 68.27% confidence interval for estimating θ, and $[\hat{\theta} - z_c s, \hat{\theta} + z_c s]$ is the confidence interval with confidence coefficient (or critical value) z_c.