17 Point Estimate

A **point estimate** is a single number, calculated from the obtained sample data, which is used to estimate the value of an unknown population parameter.

17.1 Point Estimate of the Population Mean

Very often, the mean of a population μ is an **unknown**. In this course, we assume that the population is of infinite size (or very large). One has to estimate the mean by conducting some survey for n random sample $\{X_1, X_2, \ldots, X_n\}$ where X_i are **independent** and are **identically distributed**.

• Then we shall estimate it by a point estimate which is defined as a "single value estimate" for our captured population parameter.

• A very good point estimate of the population mean is the **sample mean**

$$\bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right).$$

Remark 1. We have

$$E(\bar{X}) = \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i)\right) = \frac{1}{n}(n\mu) = \mu.$$

Since the expected value of $\bar{X}$ is equal to the unknown mean μ, it is also called an **unbiased estimator** of μ.

Remark 2. We also note that

$$Var(\bar{X}) = \frac{1}{n^2}\left(\sum_{i=1}^{n} Var(X_i)\right) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Thus we know that if σ^2 is finite then

$$\lim_{n\to\infty} Var(\bar{X}) = 0.$$

17.2 Point Estimate of Population Variance*

This subsection is for your information only. To estimate the population variance, there are two cases to study.

• In the first case, we assume that the **population mean μ is known**. In this case a good point estimate of the population variance is

$$W^2 = \frac{1}{n}\left(\sum_{i=1}^{n}(X_i - \mu)^2\right).$$

Remark 3. We have

$$E(W^2) = \frac{1}{n}\left(\sum_{i=1}^{n}E(X_i - \mu)^2\right) = \frac{1}{n}(n\sigma^2) = \sigma^2.$$

The expected value of W^2 is equal to σ^2, so it is an **unbiased estimator** of σ^2.

Remark 4. We note that

$$Var(W^2) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} (X_i - \mu)^2\right).$$

Therefore we have

$$Var(W^2) = \frac{n}{n^2}\left(Var\left((X - \mu)^2\right)\right)$$

where X has the same distribution of X_i. Because

$$Var\left(\sum_{i=1}^{n} (X_i - \mu)^2\right) = \sum_{i=1}^{n} Var\left((X_i - \mu)^2\right)$$

as all X_i are independent and

$$Var\left((X_1 - \mu)^2\right) = Var\left((X_2 - \mu)^2\right) = \cdots = Var\left((X_n - \mu)^2\right).$$

Now we have

$$\begin{aligned} Var((X-\mu)^2) &= E((X-\mu)^{2\times 2}) - E((X-\mu)^2)^2 \\ &= E((X-\mu)^4) - \sigma^4 \end{aligned}$$

by using the identity

$$Var(Y) = E(Y^2) - E(Y)^2$$

with $Y = (X-\mu)^2$. Finally we have

$$Var(W^2) = \frac{\big(E((x-\mu)^4) - \sigma^4\big)}{n}.$$

Thus we know that if σ^4 and $E(X-\mu)^4$ are finite then

$$\lim_{n\to\infty} Var(W^2) = 0.$$

• In the second case, we assume that the **population mean μ is unknown**. In this case a good point estimate of the population variance is

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) \quad \text{where} \quad \bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right).$$

Remark 5. We have

$$\begin{aligned}
E(S^2) &= \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) \\
&= \frac{1}{(n-1)}\left(\sum_{i=1}^{n} E\left(\frac{X_1 + X_2 + \cdots + (1-n)X_i + \cdots + X_n}{n}\right)^2\right) \\
&= \frac{n}{(n-1)n^2}\left(E((X_1 + X_2 + \cdots + (1-n)X_i + \cdots X_n)^2)\right) \quad i \text{ can be } 1, 2, \cdots \\
&= \frac{n}{(n-1)n^2}(Var(X_1 + X_2 + \cdots + (1-n)X_i + \cdots X_n) \\
&\quad +(E(X_1 + X_2 + \cdots + (1-n)X_i + \cdots X_n))^2) \\
&= \frac{(n-1)\sigma^2 + (n-1)^2\sigma^2 + 0^2}{(n-1)n} = \sigma^2.
\end{aligned}$$

The expected value of S^2 is equal to σ^2, it is an **unbiased estimator** of σ^2.

18 Interval Estimate

We should note that $\bar{X}$ is a random variable and in the following we are going to introduce the concept of **confidence interval** for the mean μ.

18.1 Case of Infinite Population and Known Variance σ^2

• Suppose a random sample $X_1, X_2, \cdots, X_n$ of size n is drawn from a population of mean μ (unknown) and **known variance** σ^2. Then according to the **central limit theorem**, for large n, the random variable

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{18.1}$$

follows the standard normal distribution $N(0, 1)$ where

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{18.2}$$

is the sample mean.

• For a particular given confidential level $\beta = 1 - \alpha$ (β can be $0.90, 0.95$ and 0.99), we can obtain a **Confidence Interval (C.I.)** for μ by considering

$$P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \le Z_{\alpha/2}\right) = \beta,$$

i.e., **the $100\beta\%$ confidence interval for the mean μ**:

$$\left[\bar{X} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}\right].$$

• We note that

$$\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \le Z_{\alpha/2}$$

is equivalent to

$$\bar{X} - Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}.$$

• Here μ is an unknown constant but NOT a random variable. But the end points of the interval are random variables.

Remark 6. For different confident levels, we have
For $\beta = 0.90$, $Z_{0.050} = 1.645$.
For $\beta = 0.95$**,** $Z_{0.025} = 1.960$.
For $\beta = 0.99$, $Z_{0.005} = 2.575$.

Remark 7. The C.I. obtained with confident level β means that it has a probability of β to cover the unknown population μ.

We should NOT say that μ has a probability of β to be contained in the C.I. obtained. Because μ is a fixed unknown and is **NOT** a random variable but the end points of the C.I. are random variables.

Remark 8. Given the sample mean $\bar{X}$, the population variance σ^2 and the confident level (usually it is is taken to be $90\%, 95\%$ or 99%), the larger the sample size n is, the smaller the length of the C.I. will be and the length will tend to zero.

Remark 9. Given the sample mean, the population variance and the sample size, the larger the confident level is, the longer the length of the C.I. will be.

Example 1. The mean demand of certain new product is μ (unknown) in the network of department stores. A survey of 40 random samples yields a mean of 600 per day. From the past experience of new products, the variance σ^2 is known to be 400. Construct a 90% and 95% confidence intervals for the mean μ.

• We have $\bar{X} = 600$, $\sigma = 20$ and $n = 40$, thus we have:
For $\beta = 0.90$, $z_{0.05} = 1.645$ and the C.I. is

$$\left[600 - 1.645 \times \frac{20}{\sqrt{40}},\ \ 600 + 1.645 \times \frac{20}{\sqrt{40}}\right] = [594.8, 605.2].$$

For $\beta = 0.95$, $z_{0.025} = 1.96$ and the C.I. is

$$\left[600 - 1.96 \times \frac{20}{\sqrt{40}},\ \ 600 + 1.96 \times \frac{20}{\sqrt{40}}\right] = [593.8, 606.2].$$

Exercise 1. In a public exam, a sample of 100 scores yields an average of 57.4. From the past experience, the standard deviation is around 10. Construct a 90% and 95% confidence intervals for the average score μ.

18.2 Case of Infinite Population and Unknown Variance σ^2

We know how to construct a C.I. for the population mean μ at a given confident level when the population variance σ^2 is known. The key idea is (18.1), i.e.,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

follows the standard normal distribution approximately.

- If the population variance is **NOT known**, and we get n random samples, $\{X_1, X_2, \cdots, X_n\}$, then we have a new random variable:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \tag{18.3}$$

where

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

and

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

• Here S^2 is called the **sample variance** and it is a point estimate for the population variance σ^2.

• It can be shown that T in (18.3) follows the *Student's t distribution* with degree of freedom $(n-1)$. However, this probability distribution will NOT be covered in this course.

• Thus one can construct C.I.s for different confident levels by using the *Student's t distribution*. However, it can be shown that for sample size $n \geq 30$, the *Student's t distribution* with degree of freedom n can be well approximated by the standard normal distribution $N(0,1)$. But anyway in this course, you may still adopt the standard normal to replace the Student t when $n \geq 30$.

Remark 10. In this course, when the variance is unknown, we shall assume a sample size greater than or equal to 30, and we can apply the normal approximation.

Example 2. We repeat the previous example. The mean demand of certain new product is μ (unknown) in the network of department stores. A survey of 40 samples yields a mean of 600 per day and sample standard deviation is 18. Construct a 90% and 95% confidence intervals for the mean μ.

We have $\bar{X} = 600$, $S = 18$ and $n = 40$, thus we have:

For $\beta = 0.90$, $z_{0.05} = 1.645$ and the C.I. is

$$\left[600 - 1.645 \times \frac{18}{\sqrt{40}}, \quad 600 + 1.645 \times \frac{18}{\sqrt{40}}\right] = [595.3, 604.7].$$

For $\beta = 0.95$, $z_{0.025} = 1.96$ and the C.I. is

$$\left[600 - 1.96 \times \frac{18}{\sqrt{40}}, \quad 600 + 1.96 \times \frac{18}{\sqrt{40}}\right] = [594.2, 605.6].$$

Exercise 2. In a public exam, a sample of 100 scores yields an average of 57.4 and the sample standard deviation is 8. Construct a 90% and 95% confidence intervals for the mean μ.

A Summary

1. Point estimate of the population mean.
2. Interval estimate of the population mean.
3. Sample variance.
4. Confident interval for the population mean μ with known variance.
5. Confident interval for the population mean μ with unknown variance.