Chapter 7 – Statistical Inference: Estimation for one sample case

7.1 Statistical inference

Statistical inference (inferential statistics) refers to the methodology used to draw conclusions (expressed in the language of probability) about population parameters on the basis of samples drawn from the population.

Examples

1.) The government of a country wants to estimate the proportion of voters (p) in the country that approve of their economic policies.

2.) A manufacturer of car batteries wishes to estimate the average lifetime (μ) of their batteries.

3.) A paint company is interested in estimating the variability (as measured by the variance, σ^2) in the drying time of their paints.

The quantities p, μ and σ^2 that are to be estimated are called population **parameters**.

A **sample estimate** of a population parameter is called a **statistic.** The table below gives examples of some commonly used parameters toegether with their statistics.

Parameter	Statistic
p	$\hat{p}$
μ	$\overline{X}$
σ^2	S^2

7.2 Point and interval estimation

- A **point estimate** of a parameter is a **single value** (point) that estimates a parameter.

- An **interval estimate** of a parameter is a **range of values** from L (lower value) to U (upper value) that estimate a parameter. Associated with this range of values is a probability or percentage chance that this range of values will contain the parameter that is being estimated.

Examples

Suppose the mean time it takes to serve customers at a supermarket checkout counter is to be estimated.

1) The mean service time of 100 customers of (say) $\bar{x} =$ 2.283 minutes is an example of a point estimate of the parameter μ.

2) If it is stated that the probability is 0.95 (95% chance) that the mean service time will be from 1.637 minutes to 4.009 minutes, the interval of values (1.637, 4.009) is an interval estimate of the parameter μ.

The estimation approaches discussed will focus mainly on the interval estimate approach.

7.3 Confidence intervals terminology

A **confidence interval** is a **range** of values from **L** (lower value) to **U** (upper value) that estimate a population parameter θ with 100(1-α)% confidence.

Θ – pronounced "theta".

L is the **lower confidence limit**.
U is the **upper confidence limit**.

The interval (L, U) is called the **confidence interval**.

1 – α is called the **confidence coefficient.** It is the probability that the confidence interval will contain θ the parameter that is being estimated.
100(1 – α) is called the **confidence percentage**.

Example
Consider example 2 of the previous section.
θ, the parameter that is being estimated, is the population mean μ.

α=0.05
The confidence coefficient is (1–α) = 0.95
The confidence percentage is 100(1–α) = 95.

L = 1.637 U = 4.009
The confidence interval is the interval (1.637, 4.009).

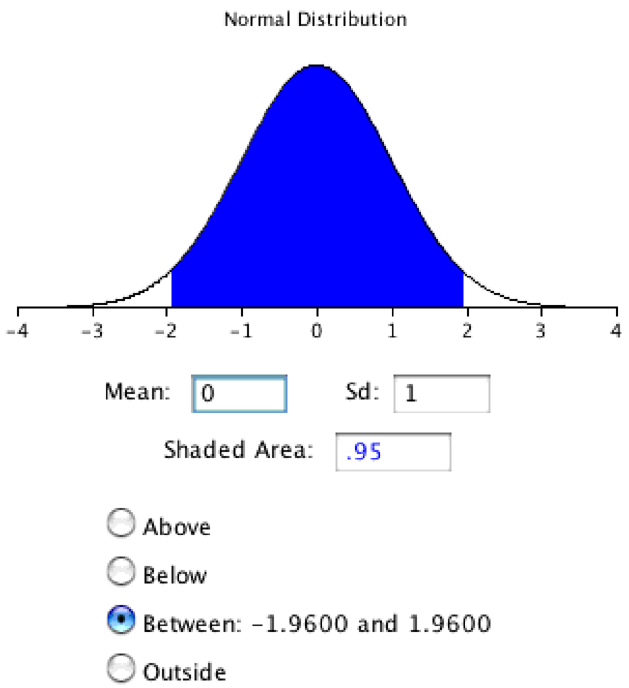
In the sections that follow the determination of L and U when estimating the parameters μ, p and σ^2 will be discussed.

7.4 Confidence interval for the population mean (population variance known)

The determination of the confidence limits is based on the central limit theorem. This theorem states that for sufficiently large samples

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \qquad \text{or} \qquad Z = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0;1)$$

Formulae for the lower and upper confidence limits can be constructed in the following way (using a confidence coefficient of 0.95 as an example).



Since Z ~ N(0;1), it follows from the graph that

$$P(-1.96 \le Z \le 1.96) = 0.95$$

and so

$$P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

By a few steps of mathematical manipulation (not shown here), the above part in brackets can be changed to have only the parameter μ between the inequality signs. This will give

$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Let $L = \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}$ and $U = \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$.

Then the above formula can be written as $P(L \le \mu \le U) = 0.95$.

This formula is interpreted in the following way:

> Since both L and U are determined by the sample values (which determine $\bar{X}$), they (and the confidence interval) will change for different samples. Since the parameter μ that is being estimated remains constant, these intervals will either include or exclude μ. The central limit theorem states that such intervals will include the parameter μ with probability 0.95 (95 out of 100 times).
>
> In a practical situation the confidence interval will not be determined by many samples, but by only one sample. Therefore the confidence interval that is calculated in a practical situation will involve replacing the random variable $\bar{X}$ by the sample value $\bar{x}$. Then the above formulae for a 95% confidence interval for the population mean μ becomes

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \; ; \; \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) \qquad \text{or} \qquad \bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

The percentage of confidence associated with the interval is determined by the value (called the z – multiplier) obtained from the standard normal distribution. In the above formula a z-multiplier of 1.96 determines a 95% confidence interval.

If a different percentage of confidence is required, the z – multiplier needs to be changed. The table below gives examples of z-multipliers and their corresponding confidence percentages.

confidence percentage	99	95	90
z-multiplier	2.576	1.96	1.645
α	0.01	0.05	0.10

Calculation of confidence interval for μ (σ^2 known)
Step 1 : Calculate $\bar{x}$. Values of n, σ^2 and confidence percentage are given
Step 2 : Look up z-multiplier for given a confidence percentage.
Step 3 : Confidence interval is $\bar{x} \pm$ z-multiplier $\frac{\sigma}{\sqrt{n}}$

<u>Example</u>

The actual content of cool drink in a 500 milliliter bottle is known to vary. The standard deviation is known to be 5 milliliters. Thirty (30) of these 500 milliliter bottles were selected at random and their mean content found to 498.5. Calculate 95% and 99% confidence intervals for the population mean content of these bottles.

Solution:

95% confidence interval
Substituting $\bar{x}$ = 498.5, n = 30, σ = 5, z = 1.96 into the above formula gives
$498.5 \pm 1.96 \frac{5}{\sqrt{30}}$ = (496.71, 500.29).

99% confidence interval
Substituting $\bar{x}$ = 498.5, n = 30, σ = 5, z = 2.576 into the above formula gives
$498.5 \pm 2.576 \frac{5}{\sqrt{30}}$ = (496.15, 500.85).

7.5 Confidence interval for the population mean (population variance not known)

When the population variance (σ^2) is not known, it is replaced by the sample variance (S^2) in the formula for Z mentioned in the previous section. In such a case, when n is small, the quantity

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$ follows a t-distribution with
degrees of freedom = df = n – 1.

The confidence interval formula used in the previous section is modified by replacing the z-multiplier by the t-multiplier that is looked up from the t-distribution.

Calculation of confidence interval for μ (σ^2 not known), small n
Step 1 : Calculate $\bar{x}$ and S. Values of n and confidence percentage are given
Step 2 : Look up t-multiplier for a given confidence percentage and degrees of freedom = n–1
Step 3 : Confidence interval is $\bar{x} \pm$ t-multiplier $\frac{S}{\sqrt{n}}$

Example

The time (in seconds) taken to complete a simple task was recorded for each of 15 randomly selected employees at a certain company. The values are given below.

38.2	43.9	38.4	26.2	41.3	42.3	37.5	37.2	41.2	42.3	31	50.1	37.3	36.7	31.8

Calculate 95% and 99% confidence intervals for the population mean time it takes to complete this task.

Solution:

n = 15 (given) , $\bar{x} = 38.36$, S = 5.78 (Calculated from the data)

95% confidence interval:

α = 0.05 therefore α/2 = 0.025 and 1 – α/2 = 0.975
degrees of freedom = df = ν = 15 – 1 = 14
t-multiplier = $t_{14;\,0.975}$ = 2.145 (from t-table)

Substituting $\bar{X}$ = 38.36, n = 15, S = 5.78, t = 2.145 into the above formula gives

$$38.36 \pm 2.145\frac{5.78}{\sqrt{15}} = (35.16\,,41.56)$$

99% confidence interval:

α = 0.01 therefore α/2 = 0.005 and 1 – α/2 = 0.995
degrees of freedom = df = ν = 15 – 1 = 14
t-multiplier = $t_{14;\,0.995}$ = 2.977 (from t-table)

Substituting $\bar{X}$ = 38.36, n = 15, S = 5.78, t = 2.977 into the above formula gives

$$38.36 \pm 2.977\frac{5.78}{\sqrt{15}} = (33.92\,,42.80)$$