Chapter 9 Sample Estimation Problems

Classical Methods of Estimation

- **Point Estimation**

  A point estimate of some population parameter \( \theta \) is a single value \( \hat{\theta} \) of a statistic \( \Theta \).

  For example, the value \( \bar{x} \) of the statistic \( \bar{X} \) computed from a sample of size \( n \) is a point estimate of the population parameter \( \mu \).

**Definition.** A statistic \( \hat{\Theta} \) is said to be an unbiased estimator of the parameter \( \theta \) if

\[
\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta.
\]

**Example 1.** Show that \( \bar{X} \) is an unbiased estimator of the parameter \( \mu \).

  We need to show that \( E(\bar{X}) = \mu \). That is,

\[
E\left( \frac{\sum_{i=1}^{n} X_i}{n} \right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu.
\]

**Example 2.** \( S^2 \) is an unbiased estimator of the parameter \( \sigma^2 \).

**Variance of a Point Estimator**

**Definition.** If we consider all possible unbiased estimators of some parameter \( \theta \), the one with the smallest variance is called the **most efficient estimator** of \( \theta \).

**Example 1.** For normal populations one can show that both \( \bar{X} \) and \( \tilde{X} \) are unbiased estimators of the population mean \( \mu \), but the variance of \( \bar{X} \) is smaller than the variance of \( \tilde{X} \). Hence the estimator \( \bar{X} \) is more efficient than \( \tilde{X} \).

**Example 2.** If \( X \) is a binomial random variable, show that

(a) \( \hat{P} = \frac{X}{n} \) is an unbiased estimator of \( P \).

**Solution.**

\[
E(\hat{P}) = E\left( \frac{X}{n} \right) = \frac{1}{n} E(X) = \frac{1}{n} np = P.
\]

(b) \( P' = \frac{X + \sqrt{n}/2}{n + \sqrt{n}} \) is a biased estimator of \( P \).
Solution.

\[
E(P') = E\left( \frac{X + \sqrt{n} / 2}{n + \sqrt{n}} \right) = \frac{1}{n + \sqrt{n}}E\left( X + \sqrt{n} / 2 \right)
\]

\[
= \frac{1}{n + \sqrt{n}}\left\{ E(X) + \sqrt{n} / 2 \right\} = \frac{nP + \sqrt{n} / 2}{n + \sqrt{n}} \neq P
\]

Therefore \( P' \) is a biased estimator of \( P \).

Example 3. Let \( X_1, X_2, \) and \( X_3 \) be a random sample of size 3. We consider three estimators of \( \mu \) given by
\[
\hat{\theta}_1 = \frac{2X_1 + X_2 + X_3}{4}, \quad \hat{\theta}_2 = \frac{X_1 + 2X_3}{3} \quad \text{and} \quad \hat{\theta}_3 = \frac{X_1 + X_2 + X_3}{6}.
\]

(a) Find out if \( \hat{\theta}_1, \hat{\theta}_2 \) and \( \hat{\theta}_3 \) are unbiased or not.

(b) If unbiased, which one is more efficient? Give reasons.

Solution. (a)

\[
E\left( \hat{\theta}_1 \right) = E\left( \frac{2X_1 + X_2 + X_3}{4} \right) = \frac{1}{4}\left\{ E\left( 2X_1 \right) + E\left( X_2 \right) + E\left( X_3 \right) \right\}
\]

\[
= \frac{1}{4}\left\{ 2E\left( X \right) + E\left( X_2 \right) + E\left( X_3 \right) \right\} = \frac{2\mu + \mu + \mu}{4} = \mu
\]

Thus, \( \hat{\theta}_1 \) is an unbiased estimator of \( \mu \).

\[
E\left( \hat{\theta}_2 \right) = E\left( \frac{X_1 + 2X_3}{3} \right) = \frac{1}{3}\left\{ E\left( X_1 \right) + 2E\left( X_3 \right) \right\}
\]

\[
= \frac{\mu + 2\mu}{3} = \mu
\]

Thus, \( \hat{\theta}_2 \) is also an unbiased estimator of \( \mu \).

\[
E\left( \hat{\theta}_3 \right) = E\left( \frac{X_1 + X_2 + X_3}{6} \right) = \frac{1}{6}\left\{ E\left( X_1 \right) + E\left( X_2 \right) + E\left( X_3 \right) \right\}
\]

\[
= \frac{\mu + \mu + \mu}{6} = \frac{\mu}{2} \neq \mu
\]

Thus, \( \hat{\theta}_3 \) is not an unbiased estimator of \( \mu \).

(b) To check the efficiency we need to compute their variances.

\[
\text{Var}\left( \hat{\theta}_1 \right) = \text{Var}\left( \frac{2X_1 + X_2 + X_3}{4} \right) = \frac{1}{16}\left\{ 4\text{Var}\left( X_1 \right) + \text{Var}\left( X_2 \right) + \text{Var}\left( X_3 \right) \right\}
\]

\[
= \frac{1}{16}\left\{ 4\sigma^2 + \sigma^2 + \sigma^2 \right\} = \frac{6}{16}\sigma^2
\]
\[
Var(\hat{\theta}_2) = Var\left(\frac{X_1 + 2X_3}{3}\right) = \frac{1}{9}\{Var(X_1) + 4Var(X_3)\} \\
= \frac{1}{9}\{\sigma^2 + 4\sigma^2\} = \frac{5}{9}\sigma^2
\]

Since \(Var(\hat{\theta}_1) < Var(\hat{\theta}_2)\), \(\hat{\theta}_1\) is more efficient than \(\hat{\theta}_2\).

## Interval Estimation

Suppose that the response variable \(X\) is normally distributed with unknown mean \(\mu\) and unknown variance \(\sigma^2\). The goal of interval estimation is to obtain a probability interval for the true mean \((\mu)\) based on a random sample of \(n\) observations.

Recall that sampling from the population may be considered as an experiment. When \(n\) is large, any particular sample of size \(n\) is one of many possible samples of size \(n\) that we could have obtained. It follows that the sample mean \((\bar{X})\) obtained from any particular sample of size \(n\) is one of many possible sample means we could have obtained from the same size samples.

According to the central limit theorem:

- Sample means are distributed normally if \(X\) is normal or if \(n\) is greater than about 30.
- The sample means are centered around the true mean. That is, the expected value of \(\bar{X}\) \((\mu_\bar{X})\) is equal to \(\mu\), meaning that \(\bar{X}\) is an unbiased estimate of \(\mu\).
- The variance in the distribution of means is \(\frac{\sigma^2}{n}\).

When \(\sigma^2\) is unknown, we must substitute the unbiased estimate of this quantity. An unbiased estimate of \(\sigma^2\) is given by

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

When \(\sigma^2\) is unknown, the sample means follow a t-distribution rather than a \(\mathcal{N}\)-distribution. As the sample size increase, however, the t-distribution converges to a \(\mathcal{N}\)-distribution.
Estimating the Mean

**Confidence Interval for \( \mu (\sigma \text{ known}) \).** If \( \bar{X} \) is the mean of a random sample of size \( n \) from a population (need not be normal) with known variance \( \sigma^2 \), then a \((1 - \alpha)100\%\) confidence interval for \( \mu \) is given by

\[
\bar{X} - e \leq \mu \leq \bar{X} + e
\]

where

\[
e = z_{a/2} \frac{\sigma}{\sqrt{n}}, \quad z_{a/2} \text{ is the } z\text{-value leaving an area of } \frac{\alpha}{2} \text{ to the right and } \frac{\sigma}{\sqrt{n}} \text{ is the standard error of } \bar{X}.
\]

**Example 1.** An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours, find a 96% confidence interval for the population mean \( \mu \) of all bulbs produced by this firm.

**Solution.** Given \( \sigma = 40 \text{ hours}, n = 30, \bar{X} = 780 \text{ hours}, 1 - \alpha = 0.96, \alpha/2 = 0.02 \), From table A.4 \( z_{0.02} = 2.05 \) and the confidence interval for \( \mu \) is

\[
\bar{X} - e < \mu < \bar{X} + e \Rightarrow 780 - 2.05 \frac{40}{\sqrt{30}} < \mu < 780 + 2.05 \frac{40}{\sqrt{30}}
\]

Thus, the 96% confidence interval is \( 765.1 < \mu < 794.9 \).

**Theorem.** If \( \bar{X} \) is used as an estimate of \( \mu \), we can be \((1 - \alpha)100\%\) confident that the error will not exceed a specified amount \( e \) when the sample size is

\[
n = \left( \frac{z_{a/2} \sigma}{e} \right)^2.
\]

**Example 2.** The average zinc concentration recovered from a sample of zinc measurements is found to be 2.6 grams per milliliter, where the population standard deviation is 0.3. How large a sample is required if we want to be 95% confident that our estimate of \( \mu \) is off by less than 0.05?

**Solution.** Given \( \sigma = 0.3, 1 - \alpha = 0.95, \alpha/2 = 0.025, e = 0.05 \), From table A.4 \( z_{0.025} = 1.96 \) and

\[
n = \left( \frac{z_{a/2} \sigma}{e} \right)^2 = \left( \frac{1.96(0.3)}{0.05} \right)^2 = 138.3 \Rightarrow n = 138.
\]

**Example 3.** How large a sample is needed in Example 1, if we wish to be 96% confident that our sample mean will be within 10 hours of the true mean?
We are attempting to estimate the mean of a population when the variance $\sigma^2$ is unknown. If we have a random sample chosen from a normal population, then the random variable

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

has a student’s t-distribution with degrees of freedom $\nu = n - 1$.

**Confidence Interval for $\mu$ ($\sigma$ unknown).** If $\bar{X}$ and $S$ are the mean and standard deviation of a random sample of size $n$ from a normal population with unknown variance $\sigma^2$, then a $(1 - \alpha)100\%$ confidence interval for $\mu$ is given by

$$\bar{X} - e \leq \mu \leq \bar{X} + e$$

where $e = t_{\alpha/2} \frac{S}{\sqrt{n}}$, $t_{\alpha/2}$ is the $t$-value with $\nu = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

**Example 1.** The heights of a random sample of 50 college students showed a mean of 174.5cm and a standard deviation of 6.9cm.

(a) What can we assert with 98% confidence about the possible size of our error if we estimate the mean height of all college students?

(b) Construct a 98% Confidence Interval for the mean height of all college students.

**Solution.** (a) Given $S = 6.9$cm, $n = 50$, $\bar{X} = 174.5$cm, $1 - \alpha = 0.98$, $\alpha/2 = 0.01$, From table A.3 $t_{0.01,49} = \frac{2.423 + 2.390}{2} = 2.406$, the size of the error is computed as

$$e = t_{\alpha/2} \frac{S}{\sqrt{n}} = 2.406 \frac{6.9}{\sqrt{50}} \approx 2.35.$$ 

(b) The confidence interval for $\mu$ is

$$\bar{X} - e < \mu < \bar{X} + e \Rightarrow 174.5 - 2.35 < \mu < 174.5 + 2.35.$$ 

Thus, the 98% confidence interval is $172.15 < \mu < 176.85$. 

---

**Sonuc Zorlu**

**Lecture Notes**
Estimating the Variance

**Confidence Interval for** $\sigma^2$. If $S^2$ is the variance of a random sample of size $n$ from normal population, then a $(1 - \alpha)100\%$ confidence interval for $\sigma^2$ is given by

$$\frac{(n-1)S^2}{\chi^2_{a/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-a/2}}$$

where $\chi^2_{a/2}$ and $\chi^2_{1-a/2}$ are $\chi^2$-values with $\nu = n - 1$.

**Example 1.** The following are the weights in decagrams of 10 packages of grass seed distributed by a certain company.

<table>
<thead>
<tr>
<th>Weight (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.4</td>
</tr>
<tr>
<td>46.1</td>
</tr>
<tr>
<td>45.8</td>
</tr>
<tr>
<td>47.0</td>
</tr>
<tr>
<td>45.9</td>
</tr>
<tr>
<td>46.1</td>
</tr>
<tr>
<td>45.8</td>
</tr>
<tr>
<td>46.9</td>
</tr>
<tr>
<td>45.2</td>
</tr>
<tr>
<td>46.0</td>
</tr>
</tbody>
</table>

Find a 95% confidence interval for the variance of all such packages of grass seeds, assuming normal population.

**Solution.** Given $n = 10, 1 - \alpha = 0.95, \alpha / 2 = 0.025$ and the sample variance is computed as

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)} = \frac{10(21273.12) - (461.2)^2}{10.9} = 0.286,$$

and from table A.5 we have $\chi^2_{0.025} = 19.023$ and $\chi^2_{0.975} = 2.7$. Thus, the 95% confidence interval for the variance is

$$\frac{(n-1)S^2}{\chi^2_{a/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-a/2}} \Rightarrow \frac{9(0.286)}{19.023} < \sigma^2 < \frac{9(0.286)}{2.7} \Rightarrow 0.135 < \sigma^2 < 0.953.$$  

**Example 2.** A random sample of 20 students obtained a mean of $\bar{X} = 72$ and a variance of $S^2 = 16$ on a college placement tests in maths. Assuming the scores are approximately normally distributed, construct a 98% confidence interval for the true variance $\sigma^2$.

**Solution.** Given $n = 20, \bar{X} = 72, S^2 = 16, \nu = 19$, $1 - \alpha = 0.98, \alpha / 2 = 0.01$. From table A.5, $\chi^2_{0.01,19} = 36.191$ and $\chi^2_{0.99,19} = 7.633$. Therefore the confidence interval for $\sigma^2$ is

$$\frac{(n-1)S^2}{\chi^2_{a/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-a/2}} \Rightarrow \frac{(19)16}{36.191} < \sigma^2 < \frac{(19)16}{7.633},$$

That is, $8.40 < \sigma^2 < 39.83$. 

Exercises

Exercise 1. An electrical firm manufactures light bulbs that have a length of life which is approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours find a 96% confidence interval for the population mean $\mu$ of all bulbs produced by this firm.

Exercise 2. How large a sample is needed in Q1 if we wish to be 96% confident that our sample mean will be within 10 hours of the true mean?

Exercise 3. The weights of a random sample of 30 college students showed a mean of 74.5 cm and a standard deviation of 5.9 cm.
   (a) What can we assert with 98% confidence about the possible size of our error if we estimate the mean weight of all college students?
   (b) Construct a 98% confidence interval for the mean weight of all college students.

Exercise 4. Given a random sample of size 24 from a normal population, find $k$ such that $P(-2.069 < T < k) = 0.965$.

Exercise 5. A random sample of size $n=12$ from a normal population has the mean $\bar{x} = 27.8$ and $s^2 = 3.24$. Can we say that the given information supports the claim that the mean of the population if $\mu = 28.5$?

Exercise 6. Find $k$ such that $P(k < T < -1.761) = 0.045$ for a random sample of size 15 selected from a normal population.

Exercise 7. A maker of certain brand of low fat cereal bars claims that their average saturated fat content is 0.5 gr. In a random sample of 8 cereal bars of this brand the saturated fat content was 0.6,0.7,0.3,0.4,0.5,0.4 and 0.2. Would you agree with the claim? (If the computed $t$-value falls between $-t_{0.05}$ and $t_{0.05}$, teh maker is satisfied with his claim)

Exercise 8. The average zinc concentration recovered from a sample of zinc measurements is found to be 2.6 grams per milliliter, where the population standard deviation is 0.3. How large a sample is required if we want to be 95% confident that our estimate of $\mu$ is off by less than 0.05?

Exercise 9. A random sample of 12 graduates of certain secretarial school typed an average of 79.3 words per minute with a standard deviation of 7.8 words per minute. Assuming a normal distribution for the number of words typed per minute, find a 95% confidence interval for the average number of words typed by all graduates of this school.

Exercise 10. A paint manufacturer wants to determine the average drying time of a new wall paint. In 12 test areas of equal size painted by new product, he measured the drying times $X_1, X_2, \ldots, X_{12}$ in minutes. He obtained the following data: $\sum_{i=1}^{12} X_i = 798$ and $\sum_{i=1}^{12} X_i^2 = 53524$. Construct a 95% confidence interval for the true mean $\mu$. 
Exercise 11. The scores of a random sample of 20 students have a variance \( s^2 = 16 \) on a college placement test in mathematics. Assuming the scores are normally distributed, construct a 95% confidence interval for \( \sigma^2 \).

Exercise 12. (a) If \( X \) is a normal random variable with \( \mu = 20 \) and \( \sigma = 4 \), find \( P(\bar{X} < 13) \).

(b) Find \( P(T_{13} < 2.650) \). \( (T_{13} \) is a t random variable with \( \nu = 13 \) degrees of freedom)

(c) If in a random sampling problem the sample size is \( n = 11 \) and \( \sigma = 0.8 \), find 
\[
P\left(\sum_{i=1}^{n} (X_i - \bar{X})^2 > 3.1136\right).
\]

Exercise 13. The error made in measuring the heights of plants by a certain electronic instrument is accepted to be normally distributed with unknown mean \( \mu \) and unknown variance \( \sigma^2 \). Five measurements made by this instrument are checked later and found to have the following errors (in mm.s):

\[-3, \ 2, \ 4, \ 1, \ -2\]

Find a 98% confidence interval for the mean error \( \mu \).

Exercise 14. A random sample of 25 automobile owners shows that an automobile is driven on the average 20,000 km per year with a standard deviation of 1000 km. Construct a 99% confidence interval for the average number of kilometers an automobile is driven annually.

Exercise 15. The scores on a level determination test given to freshmen for the past five years are approximately normally distributed with a mean \( \mu = 62 \). Construct a 98% confidence interval for the population variance \( \sigma^2 \), if a random sample of 20 students who take this test this year obtain a value of \( S^2 = 12 \).

Exercise 16. A manufacturer of car batteries claim that his batteries will last on average 3 years with a variance of 1 year. If 5 of these batteries have lifetimes of 2.9, 2.4, 1.8, 3.5 and 4.2 years, 

(a) Construct a 96% confidence interval for \( \sigma^2 \).

(b) Decide whether the manufacturer’s claim that \( \sigma^2 = 1 \) is valid or not.

Exercise 17. Define \( \hat{P} = \frac{X + \sqrt{n}/2}{n + \sqrt{n}} \) where \( X \) is a Binomial random variable. Show that \( \hat{P} \) is a Biased Estimator of \( P \).

Exercise 18. (a) The weights (in grams) of certain soft drink bottles (filled) have a standard deviation of \( \sigma = 5 \) gr.s. A random sample of 16 bottles were tested and gave an average of \( \bar{X} = 290 \) gr.s. Find a 90% confidence interval for the population mean weight \( \mu \).

(b) If \( |\bar{X} - \mu| \leq 1 \), what should be our minimal sample size instead of 16 with 90% confidence level?