Chapter 7

Statistical Intervals Based on a Single Sample

Jing Zhang
Department of Mathematics and Statistics

DALHOUSIE UNIVERSITY

Point Estimation and Unbiased Estimators

- Statistical inference is almost always directed toward drawing some type of conclusion about one or more parameters (population characteristics).
- A **point estimate** of a parameter θ is a single number that can be regarded as a sensible value for θ. It is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ.
- A point estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ.

Unbiased Estimators:

- When X is a binomial rv with parameters n and p, the sample proportion $\hat{p} = X/n$ is an unbiased estimator of p.
- Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then the estimator

$$\hat{\sigma}^2 = s^2 = \frac{\sum(X_i - \mu)^2}{n-1}$$

is unbiased for estimating σ^2.

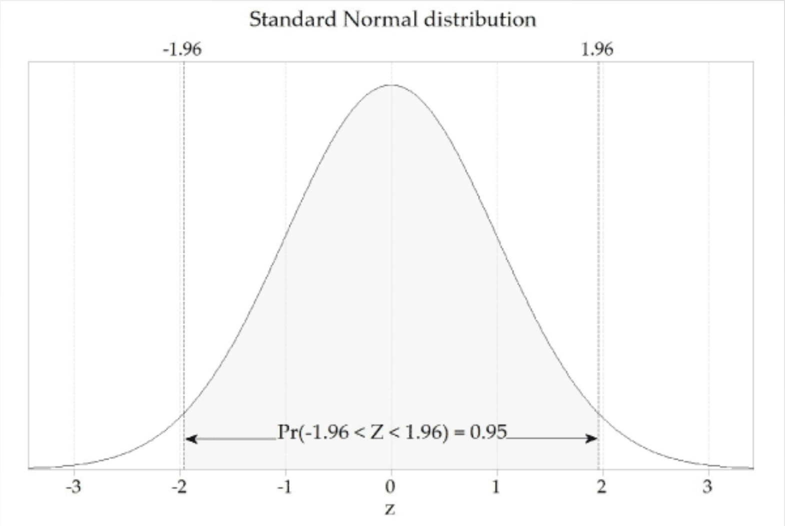
The estimator $\bar{X}$ is an unbiased estimator of μ.

Interval Estimate or Confidence Interval (CI)

- A point estimate, because it is a single number, by itself provides no information about the precision and reliability of estimation.
- An alternative to reporting a single sensible value for the parameter being estimated is to calculate and report an entire interval of plausible values – an interval estimate or confidence interval (CI).
- A confidence interval is always calculated by first selecting a confidence level, which is a measure of the degree of reliability of the interval.

Recall Chapter 5:

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a **normal** distribution with mean value μ and standard deviation σ. Then for any n, $\bar{X}$ is normally distributed (with mean μ and standard deviation $\sigma/\sqrt{n}$.
- $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$



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

Then

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

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

To interpret above equation, consider it as a random interval because the two endpoints involve a random variable $\bar{X}$.

$$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \qquad \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

The probability is 0.95 that a random interval includes or covers the true value of μ.

95% Confidence Interval for μ.

If, after observing $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, we compute the observed sample mean $\bar{x}$ and then substitute $\bar{x}$ into $\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ in place of $\bar{X}$, the resulting fixed interval is called a 95% confidence interval for μ. This CI can be expressed either as

$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ is a 95% CI for μ

Or as

$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$ with 95% confidence

A concise expression for the interval is $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$.

Interpreting a Confidence Level

- It is incorrect to write the statement

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

since by substituting $\bar{x}$ for $\bar{X}$, all randomness disappears.

- A correct interpretation of “95% confidence” relies on the long-run relative frequency interpretation of probability: To say that an event A has probability 0.95 is to say that if the experiment on which A is defined is performed over and over again, in the long run A will occur 95% of the time.
- If we collect many simple random samples in the say way, with the same sample size, and compute 95% CI for each sample, we expect 95% of the computed CIs cover the true μ.

Confidence Interval for μ with known σ

- A 100(1-α)% confidence interval for the mean μ of a normal population when the value of σ is known is give by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \qquad \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

or, equivalently, by $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.

- The formula for the CI can also be expressed in words as

Point estimate of μ $\pm$ (z critical value)(standard error of the mean)

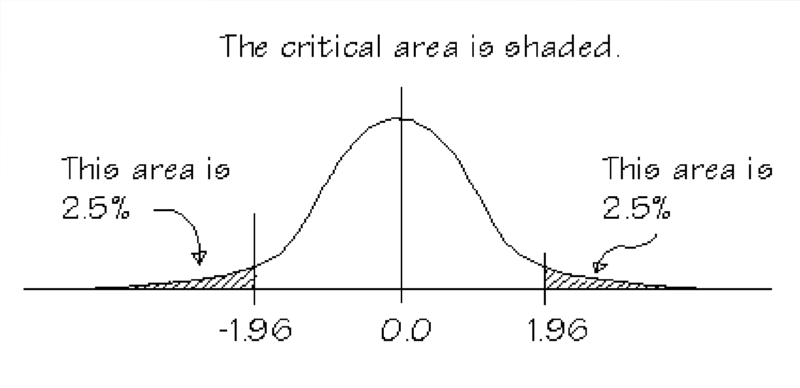
Z critical values

- $z_{\alpha/2}$ can be found at Z table.
- How to find z critical value for 95% CI?

$100(1-\alpha)\% = 95\% \leftrightarrow \alpha = 0.05$

$z_{\alpha/2} = z_{0.05/2} = z_{0.025}$

$P(Z > z_{0.025}) = 0.025$



From the Z table,

If the upper tail probability is 0.025, the z value must be equal to 1.96. Therefore, the z critical value for 95%
$z_{\alpha/2} = z_{0.05/2} = z_{0.025}$ = **1.96.**

Z critical values

- Commonly used confidence levels are 90%, 95% and 99%.

Some Common Critical Values

Confidence level	z critical value
80%	1.28
90%	1.645
95%	1.96
98%	2.33
99%	2.58
99.8%	3.09
99.9%	3.29

Example 7.1

To estimate the mean lifetime of a tire produced by a tire company, 10 tires were tested (on a test wheel simulating normal road conditions) with the following lifetime (thousands of miles):

42, 36, 46, 43, 41, 35, 43, 45, 40, 39

Prior experience has shown that the lifetime follow a normal distribution with variance 12.96 $thousands\ of\ miles^2$.

1. What is the point estimate of the mean lifetime of the tire?
2. Compute a 95% CI for the mean lifetime of the tire.

Solutions

1. Sample mean $\bar{X}$ is the point estimator of population mean μ. $\bar{x} = \frac{\sum x}{n}$ = 41.

2. 95% CI for μ

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} = 41, z_{\alpha/2} = 1.96, \sigma = \sqrt{12.96} = 3.6, n = 10$$

$$41 \pm 1.96 \cdot \frac{3.6}{\sqrt{10}} = 41 \pm 2.23 = (38.77, 43.23)$$

We are 95% confident that the mean lifetime of the tire is between 38.77 and 43.23 thousands of miles.

Confidence Level, Precision, and Sample Size

- The higher the desired degree of confidence, the wider the resulting interval will be.
- If we think of the width of the interval as specifying its precision or accuracy, then the confidence level (or reliability) of the interval is inversely related to its precision.
- An appealing strategy is to specify both the desired confidence level and interval width and then determine the necessary sample size.

The general formula for the sample size n necessary to ensure an interval with width w (or margin of error: ME = half of width = w/2), for a normal population with known population standard deviation σ, is

$$n = (2z_{\alpha/2} \cdot \frac{\sigma}{w})^2$$

Note: **Round up** the result to an **integer** number!

Example 7.2: What is the sample size needed to calculate a 95% confidence interval within 1.65 (with width 3.3 thousands miles) for the mean lifetime of a randomly selected tire from the company in previous example?

Large-sample Confidence Interval for a Population Mean

Case I: Large Sample and Known σ

Let $X_1, X_2, \dots, X_n$ be a random sample from a population having a mean μ and standard deviation σ. If n is sufficiently large (n>40), then

100(1-α)% Confidence Interval for population mean μ

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Large-sample Confidence Interval for a Population Mean

Case II: Large Sample and Unknown σ

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population having a mean μ. If the population standard σ is unknown and sample size n is sufficiently large (n>40), then replace σ with the sample standard deviation s.

100(1-α)% Confidence Interval for population mean μ

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Point estimator $\pm$ (z critical value)(estimated standard error of the mean)

Example 7.3

A random sample of 49 students, from Ontario secondary school students in 2007-2008, showed that the mean travel time to school and standard deviation are 17 and 9.66 minutes, respectively. Find a 90% confidence interval for the true mean travel time to school.

Solutions:

$$n = 49, \bar{x} = 17, s = 9.66 \; and \; z_{\alpha/2} = z_{0.1/2} = z_{0.05} = 1.645$$

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$17 \pm 1.645 \cdot \frac{9.66}{\sqrt{49}} = 17 \pm 2.27 = (14.73, 19.27)$$

We are 90% confident that the true mean travel time is between 14.73 and 19.27 minutes.

Confidence Interval for a Population Proportion

- Let p denote the proportion of “successes” in a population, where success identifies an individual or object that has a specified property.
- A random sample of n individuals or objects is to be selected, and X is the number of successes in the sample.
- If the sample size is large enough, $np \geq 10$ and $nq \geq 10$ where $q=1-p$.
- $100(1-\alpha)\%$ confidence interval for population proportion p is

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Sample Size

- The sample size needed to ensure an confidence interval with width w is

$$n = \frac{4(z_{\alpha/2})^2 \hat{p}(1-\hat{p})}{w^2}$$

- Notes:

1. Round up to an integer number
2. When $\hat{p}$ is unknown, use 0.5 instead.

Example 7.4

Suppose that in 48 trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette. Find the 95% confidence interval of the long-run proportion of all such trails that would result in ignition?

Solutions:

$$n = 48, \hat{p} = \frac{16}{48} = 0.33 \; and \; z_{\alpha/2} = 1.96$$

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.33 \pm 1.96 \times \sqrt{\frac{0.33 \times (1-0.33)}{48}}$$

$$= 0.33 \pm 0.13 = (0.2, 0.46)$$

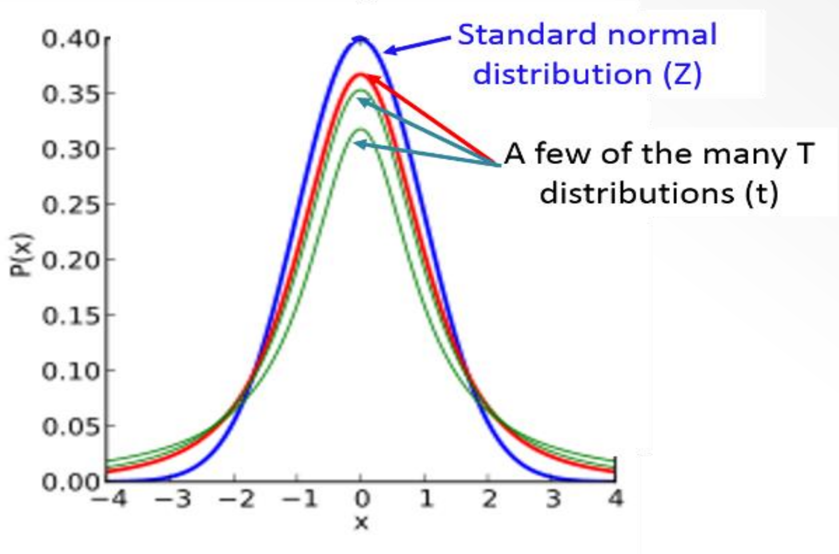
We are 95% confident that the true proportion is between 20% and 46%.

T distribution

When $\bar{X}$ is the mean of a random sample of size n from a Normal distribution with mean μ, the rv

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

Has a probability distribution called a t distribution with n-1 degrees of freedom (df).



Properties of t distributions

Let t_v denote the t distribution with v df.

1. Each t_v curve is bell-shaped and centered at 0.
2. Each t_v curve is more spread out than then standard normal (z) curve.
3. As v increases, the spread of the corresponding t_v curve decreases.
4. As $v \to \infty$, the sequence of t_v curves approaches the standard normal curve (so the z curve is often called the t curve with $\mathrm{df} = \infty$.

The One-Sample t Confidence Interval

Let $\bar{x}$ and s be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ. Then a 100(1-α)% confidence interval for μ is

$$\bar{x} \pm t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}$$

$t_{\frac{\alpha}{2},n-1}$ is called t critical value and it can be found from t distribution table.

T distribution table (Partial)

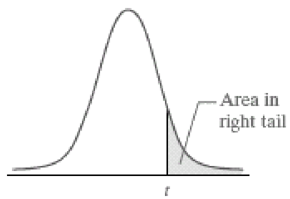


TABLE III

t-Distribution

Area in Right Tail

df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787

Example 7.5

The fat content (in percentage) on 10 randomly selected hot dogs are:

25.2, 21.3, 22.8, 17.0, 29.8, 21.0, 25.5, 16.0, 20.9, 19.5

Find the 95% CI of the population mean fat content of the hot dogs, assuming the fat content follows the normal distribution.

Solutions:

$$\bar{x} = \frac{\sum x_i}{n} = 21.9$$

$$s = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}} = 4.134$$

$$df = n - 1 = 10 - 1 = 9, t_{\frac{\alpha}{2},df} = t_{0.025,9} = 2.262$$

$$\bar{x} \pm t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}} = 21.9 \pm 2.262 \cdot \frac{4.134}{\sqrt{10}} = 21.9 \pm 2.96$$
$$= (18.94, 24.86)$$

We are 95% confident that the percentage of the fat content in the hot dogs is between 18.94% and 24.86%.

Confidence Interval for the Variance and Standard Deviation of a Normal Population

Chi-squared (χ^2) distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with parameters μ and σ. Then the rv

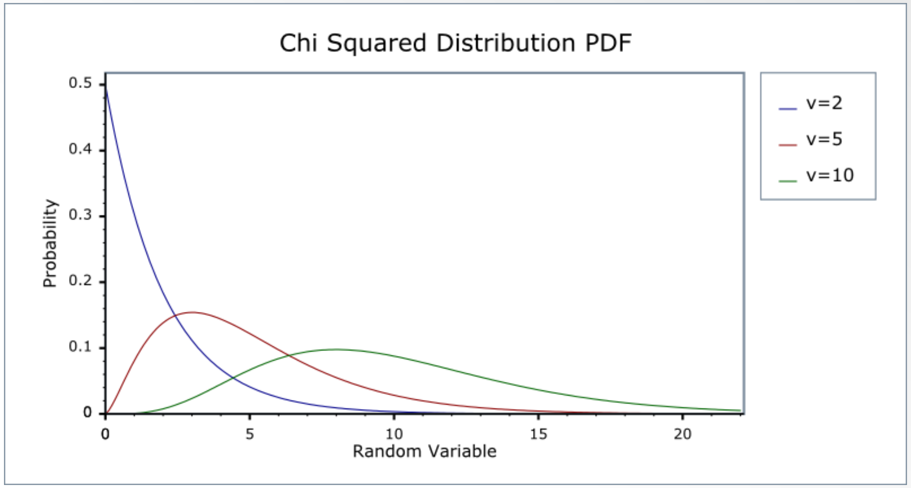
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X-\bar{X})^2}{\sigma^2}$$

has a chi-squared probability distribution with n-1 df.

Notes:

1. χ_v^2 denotes chi-squared distributed rv with df = v.
2. χ_v^2 is a special case from gamma, $G(\alpha, \beta)$, when $\alpha = \frac{v}{2}, \beta = v$.
3. Mean = v and variance = 2v.

χ_v^2 Distribution Curve



1. χ_v^2 is a right skewed distribution
2. The smaller v value, the larger skewness.

A 100(1-α)% confidence interval for the variance σ^2 of a normal population is

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$$

where s^2 is the sample variance.

A 100(1-α)% confidence interval for the standard deviation σ of a normal population is

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}}$$

Example 7.6

The article "Concrete Pressure of Formwork" gave the following observations on maximum concrete pressure (kN/m2): 33.2, 41.8, 37.3, 40.2, 36.7, 39.1, 36.2, 41.8, 36.0, 35.2, 36.7, 38.9, 35.8, 35.2, 40.1. Prior experience showed that the maximum pressures are normally distributed.

1. Calculate the point estimator of the population variance of maximum pressure.
2. Find the 95% CI for the population variance.
3. Find the 95% CI for the population standard deviation of maximum pressure.

Solutions:

1. Sample variance is the point estimator of the population variance.

$$\bar{x} = \frac{\sum x_i}{n} = 37.6$$

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = 7.69$$

2. $\chi^2_{\frac{\alpha}{2},n-1} = \chi^2_{0.025,14} = 26.119,\ \chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.975,14} = 5.629$

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$$

$$4.12 < \sigma^2 < 19.13$$

3. 95% CI for the population standard deviation is

$$\sqrt{4.12} < \sigma < \sqrt{19.13}$$
$$2.03 < \sigma < 4.37$$

We are 95% confident that the population standard deviation of maximum pressure is between 2.03 kN/m2 and 4.37 kN/m2.