Chapter 8 - Statistical intervals for a single sample

8-1 Introduction

In statistics, no quantity estimated from data is known for certain. All estimated quantities have probability distributions of their own, for example, the mean of normally distibuted data.

Definition

Bounds that represent an range of plausible values for a parameter are known as **interval estimates**.

Interval estimates are useful for quantifying uncertainty in our parameter estimates.

8-2 Confidence Interval on the Mean of a Normal Distribution, Variance Known

8-2.1 Development of confidence intervals

Suppose $X_1, X_2, \ldots, X_n$ are samples from a normal distribution: $\mathrm{N}(\mu, \sigma^2)$. We know that $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$. Recall, the sample mean can be standardized

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

Where now $Z \sim \mathrm{N}(0, 1)$, a standard normal distribution.

A confidence interval (CI) estimate for μ is an interval of the form $[l, u]$ where l and u are computed from the data. The endpoints, l and u are samples of random variables L and U, respectively, because different samples will produce different values of l and u. Suppose we knew L and U such that

$$P(L \leq \mu \leq U) = 1 - \alpha$$

where $0 \leq \alpha \leq 1$. There is a probability $1 - \alpha$ of selecting a sample for which the CI will contain the true value of μ. Given a sample $(x_1, x_2, \ldots, x_n)$, the confidence interval for μ is

$$l \leq \mu \leq u$$

l and u are known as the lower and upper confidence limits, respectively.

Since $Z \sim \mathrm{N}(0,1)$,

$$P\left(z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Definition

If $\bar{x}$ is the sample mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval on μ is

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}$$

where $100(1-\alpha)\%$ is the **confidence level**.

Example

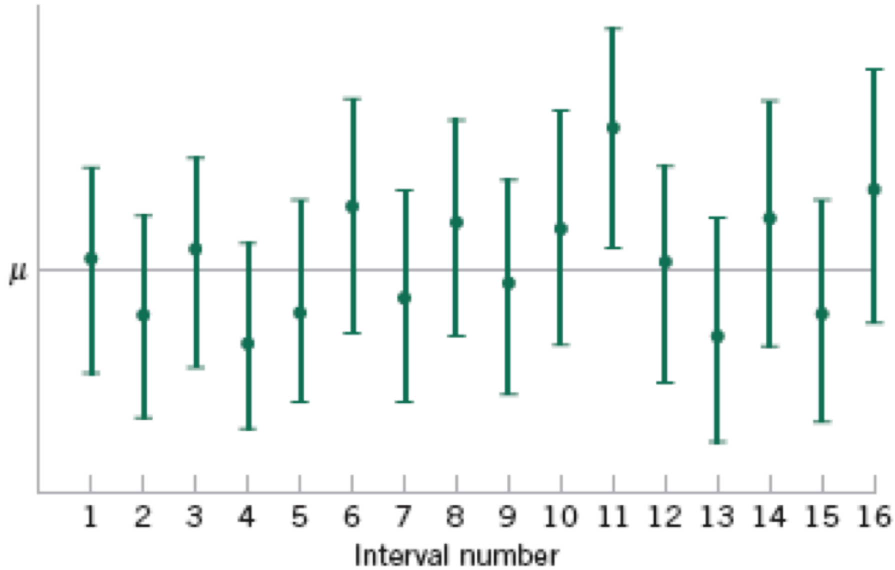
We want a 95% CI for μ, so $\alpha = 1 - 0.95 = 0.05$. $\bar{x} = 64.46$, $n = 10$, $\sigma = 1$, $z_{alpha/2} = z_{0.025} = 1.96$

$$\begin{gathered}\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha/2}\sigma/\sqrt{n} \\ 64.46 - 1.96/\sqrt{10} \le \mu \le 64.46 + 1.96/\sqrt{10} \\ 63.84 \le \mu \le 65.08\end{gathered}$$

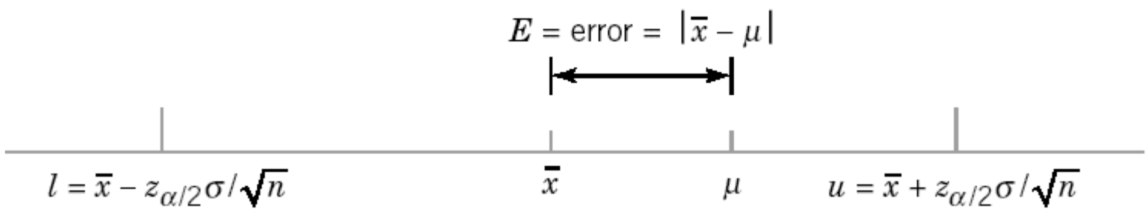
There is a 95% chance that the true population mean lies in the interval $[63.84, 65.08]$.

Interpreting a confidence interval:

- The CI is a random interval
- the observed interval $[l, u]$ brackets the true value of μ, with confidence level $100(1-\alpha)\%$



The length of a confidence interval is a measure of the precision of estimation.



8-2.2 Choice of sample size

Definition

If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error $|\bar{x}-\mu|$ will not exceed a specified amount E when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

Example

$E = 0.05$, $\sigma = 1$, $z_{\alpha/2} = 1.96$

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{1.96}{0.5}\right)^2 = 15.37.$$

Round up, so $n = 16$.

8-2.3 One sided confidence bounds

Definition

A $100(1-\alpha)$ **upper confidence bound** for μ is

$$\mu \le u = \bar{x} + z_\alpha \sigma/\sqrt{n}$$

A $100(1-\alpha)$ **lower confidence bound** for μ is

$$\bar{x} - z_\alpha \sigma/\sqrt{n} = u \le \mu$$

8-2.4 General method to derive a CI

Let $X_1, X_2, \ldots, X_n$ be a random sample of n observations. We can construct a CI for an unknown parameter θ if we can find a random variable, $g(X_1, X_2, \ldots, X_n; \theta)$ with the following properties:

1. $g(X_1, X_2, \ldots, X_n; \theta)$ depends on both the sample and θ
2. The probability distribution of $g(X_1, X_2, \ldots, X_n; \theta)$ does not depend on θ or any other unknown parameter.

For the population mean, $\theta = \mu$, $g(X_1, X_2, \ldots, X_n; \theta) = (\bar{x} - \mu)/(\sigma/\sqrt{n})$, satisfies both conditions; It depends on μ and since σ is known, $g \sim \mathrm{N}(0, 1)$ which does not depend on μ.

To construct a CI:

1. Find constants C_L and C_U such that

   $$P(C_L \le g(X_1, X_2, \ldots, X_n; \theta) \le C_U) = 1 - \alpha$$

   Because of property 2, C_L and C_U do not depend on θ. For population mean $C_L = -z_{\alpha/2}$ and $C_U = z_{\alpha/2}$.

2. Manipulate the inequalities in the probability statement so that

$$P(L(X_1, X_2, \ldots, X_n) \leq \theta \leq U(X_1, X_2, \ldots, X_n)) = 1 - \alpha$$

which gives $L(X_1, X_2, \ldots, X_n)$ and $U(X_1, X_2, \ldots, X_n)$ as the lower and upper confidence limits defining a $100(1-\alpha)$ CI for θ.

8-2.5 Large sample CI for μ

Definition

When n is large, the quantity

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

has a $N(0, 1)$ distribution. Consequently,

$$\bar{x} - z_{\alpha/2} s/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha/2} s/\sqrt{n}$$

is a **large sample confidence interval** for μ with confidence level $100(1-\alpha)\%$.

Rule of thumb: Valid for $n > 40$.

Important note: CI for the population mean does not require the underlying data to be normally distributed.

General large sample confidence interval

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$$

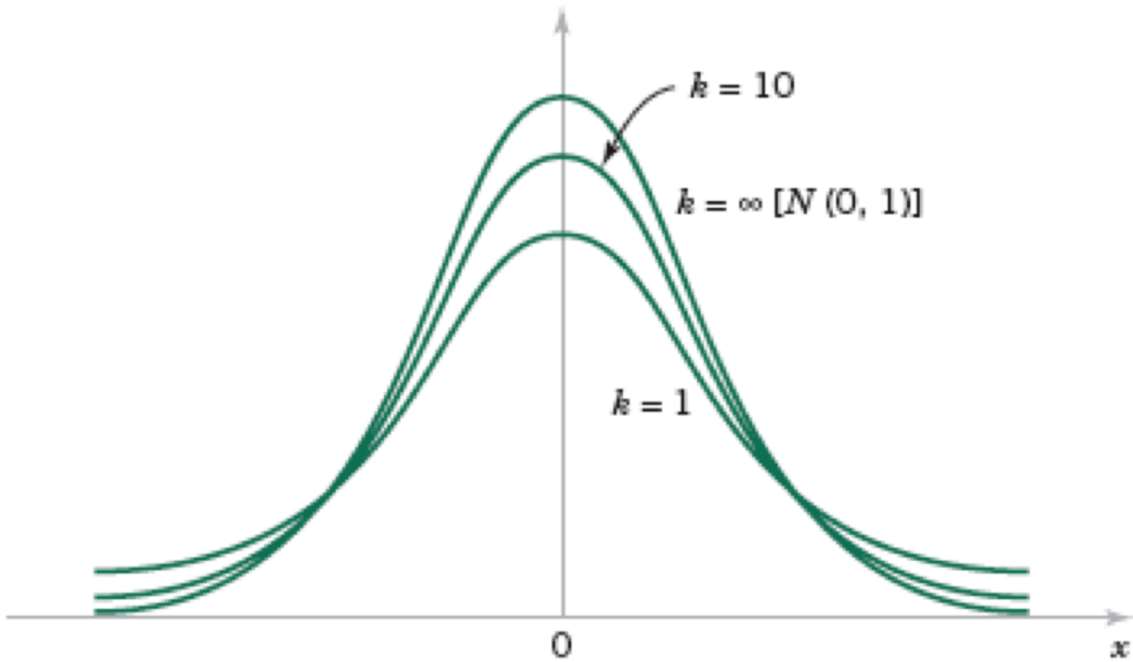
8-3 Confidence Interval on the Mean of a Normal Distribution, Variance Unknown

8-3.1 The t distribution

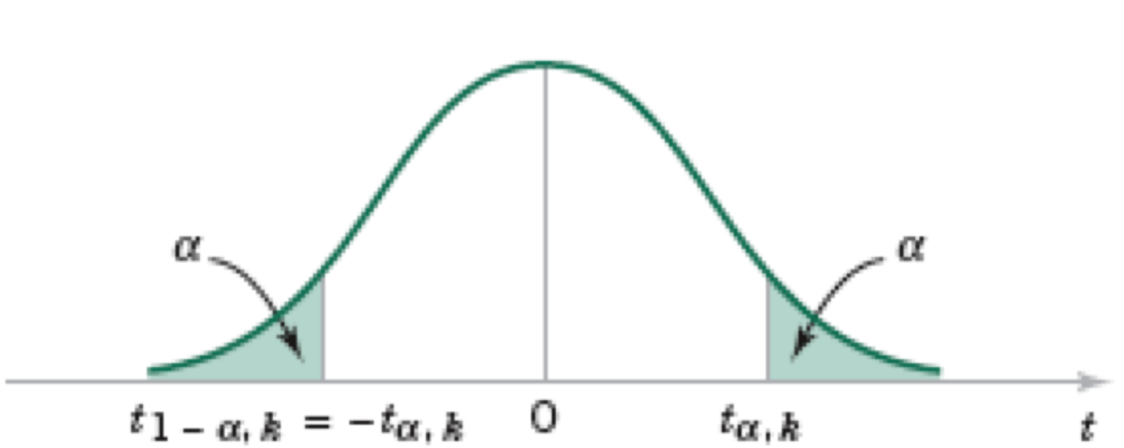
Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and variance σ^2. The random variable

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

has a standard t distribution with $\kappa = n - 1$ degrees of freedom.



t distribution approaches the standard normal for large κ



percentage points of the standard t distribution.

8-3.1 The t distribution

If $\bar{x}$ and s are the mean and standard deviation of a random sample $\mathrm{N}(\mu, \sigma^2)$ with unknown variance σ^2, a $100(1-\alpha)$ **percent confidence interval on** μ is given by

$$\hat{\theta} - t_{\alpha/2,n-1} s/\sqrt{n} \leq \mu \leq \hat{\theta} + t_{\alpha/2,n-1} s/\sqrt{n}$$

where $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ percentage point of the standard t distribution with $n-1$ degrees of freedom.

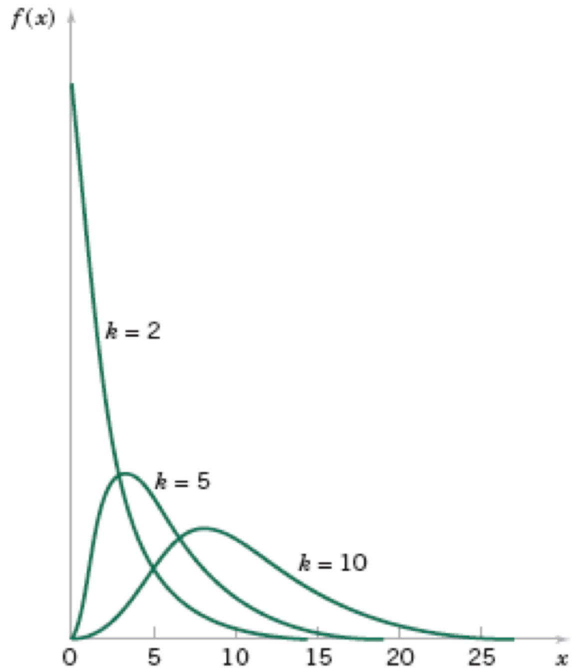
One sided confidence bounds are found by replacing $t_{\alpha/2}$ with t_α in the above equations.

8-4 Confidence Interval on the Variance and Standard Deviation of a Normal Distribution

Definition Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$ with σ^2 unknown, let S^2 be the sample variance. The random variable

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

has a chi-square distribution with $n-1$ degrees of freedom.



Definition

If s^2 is the sample variance from a random sample of n observations from a normal distribution with unknown variance, σ^2, then a $100(1-\alpha)\%$ is a **confidence interval on** σ^2

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

where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are upper and lower $100\alpha/2$ percentage points of the chi-square distribution with $n-1$ degrees of freedom. CI for σ is just the square root of both limits.