Chapter 2

Estimation and Hypothesis Testing

2.1 Point Estimation

Example 2.1. Cholesterol levels continued. Suppose we want to make inference on the mean cholesterol level of a population of people in a north eastern American state on the second day after a heart attack. We have data of 28 patients, which are a realization of a random sample of size \( n = 28 \). The first thing we can do is to calculate the estimate of the population mean (\( \mu \)) and of the population variance (\( \sigma^2 \)). To do this we can use some functions of the random sample, such as the sample mean (\( \bar{X} \)) and the sample variance (\( S^2 \)), respectively, where

\[
\bar{X} = \frac{1}{28} \sum_{i=1}^{28} X_i, \quad S^2 = \frac{1}{27} \sum_{i=1}^{28} (X_i - \bar{X})^2.
\]

Say, we have

\[ \bar{x} = 257 \text{ and } s^2 = 32. \]

Note that \( \bar{X} \) and \( S^2 \) are random variables, as they are functions of random variables, while \( \bar{x} \) and \( s^2 \) are their values obtained for the particular values of the rvs; in this case, for the 28 patients who took part in the study. A different group of 28 patients who suffered a
heart attack, would give different values of $\overline{X}$ and $S^2$.

Let $X_1, \ldots, X_n$ be a random sample from a population (distribution) with a parameter $\vartheta$. A random variable which is a function of the random sample, $T(X_1, \ldots, X_n)$, is called an estimator of the population parameter $\vartheta$, while its value is called an estimate of the population parameter $\vartheta$.

**Notation**

*Special symbols* such as $\overline{X}$ or $S^2$, are used to denote estimators of some common parameters, in these cases, of the population mean and variance. Then their “small letter” counterparts denote the respective estimates.

*Hat over a symbol* of the parameter of interest, for example $\hat{\vartheta}$, indicates an estimator of $\vartheta$. Here we have to be careful because $\hat{\vartheta}$ is often used to denote an estimate of $\vartheta$ as well. It is better to write $\hat{\vartheta}_{\text{obs}}$ to indicate that this is a value of the estimator $\hat{\vartheta}$ obtained for the observed values of random sample.

A random sample will often be denoted by a **bold** capital letter, while its realization by a **bold** small case letter. For example, $X$ denotes a random sample $(X_1, \ldots, X_n)$ and $x$ denotes its realization $(x_1, \ldots, x_n)$.

### 2.1.1 Properties of Point Estimators

An estimator $\hat{\vartheta}$ of a parameter $\vartheta$ is a random variable (a function of rvs $X_1, \ldots, X_n$) and the estimate $\hat{\vartheta}_{\text{obs}}$ is a single value taken from the distribution of $\hat{\vartheta}$. Since we want our estimate to be close to $\vartheta$, the random variable $\hat{\vartheta}$ should be centred close to $\vartheta$ and have a small variance. Also, we would want our estimator to be such that, as
2.1. POINT ESTIMATION

$n \to \infty$, $\hat{\vartheta} \to \vartheta$ with probability one. Below are definitions of some measures of goodness of the estimators.

**Definition 2.1.** Let $X_1, X_2, \ldots, X_n$ be a random sample from a population (distribution) with a parameter $\vartheta$ and let $\hat{\vartheta}$ be its point estimator. Then,

1. **The error** in estimating $\vartheta$ by estimator $\hat{\vartheta}$ is $\hat{\vartheta} - \vartheta$;
2. **The bias** of estimator $\hat{\vartheta}$ for $\vartheta$ is

   $$\text{bias}(\hat{\vartheta}) = E(\hat{\vartheta}) - \vartheta,$$

   i.e, it is the expected error;
3. Estimator $\hat{\vartheta}$ is **unbiased** for $\vartheta$ if

   $$E(\hat{\vartheta}) = \vartheta,$$

   that is, it has zero bias.

A weaker property if unbiasedness is defined below.

**Definition 2.2.** The estimator $\hat{\vartheta}$ is asymptotically unbiased for $\vartheta$ if

$$E(\hat{\vartheta}) \to \vartheta$$

as $n \to \infty$, that is, the bias tends to zero as $n \to \infty$.

Another popular measure of goodness of an estimator is given in the following definition.

**Definition 2.3.** The **mean square error** of $\hat{\vartheta}$ as an estimator of $\vartheta$ is

$$\text{MSE}(\hat{\vartheta}) = E\left\{ (\hat{\vartheta} - \vartheta)^2 \right\}.$$
Note: The mean square error can be written as a sum of variance and squared bias of the estimator.

\[
\text{MSE}(\hat{\vartheta}) = E \left( \left[ \{\hat{\vartheta} - E(\hat{\vartheta})\} + \{E(\hat{\vartheta}) - \vartheta\} \right]^2 \right)
\]

\[
= E \left[ \{\hat{\vartheta} - E(\hat{\vartheta})\}^2 + 2\{\hat{\vartheta} - E(\hat{\vartheta})\}\{E(\hat{\vartheta}) - \vartheta\} + \{E(\hat{\vartheta}) - \vartheta\}^2 \right]
\]

\[
= E \left[ \{\hat{\vartheta} - E(\hat{\vartheta})\}^2 \right] + 2\{E(\hat{\vartheta}) - E(\hat{\vartheta})\}\{E(\hat{\vartheta}) - \vartheta\} + \{E(\hat{\vartheta}) - \vartheta\}^2
\]

\[
= \text{var}(\hat{\vartheta}) + \{\text{bias}(\hat{\vartheta})\}^2.
\]

It seems natural that if we increase the sample size \(n\) we should get a more precise estimate of \(\vartheta\). The estimators which have this property are called consistent. In the following definition we use the notion of convergence in probability:

A sequence of rvs \(X_1, X_2, \ldots\) converges in probability to a random variable \(X\) if for every \(\varepsilon > 0\)

\[
\lim_{n \to \infty} P(|X_n - X| < \varepsilon) = 1.
\]

**Definition 2.4.** A sequence \(T_n = T_n(X_1, X_2, \ldots, X_n), n = 1, 2, \ldots\), is called a consistent sequence of estimators for \(\vartheta\) if \(T_n\) converges in probability to \(\vartheta\), that is for all \(\varepsilon > 0\) and for all \(\vartheta \in \Theta\)

\[
\lim_{n \to \infty} P(|T_n - \vartheta| < \varepsilon) = 1.
\]

Note: Consistency means that the probability of our estimator being within some small \(\varepsilon\) of \(\vartheta\) can be made as close to one as we like by making the sample size \(n\) sufficiently large.

The following Lemma gives a useful tool for establishing consistency of estimators.
Lemma 2.1. A sufficient condition for consistency is that $\text{MSE}(T_n) \to 0$ as $n \to \infty$, or, equivalently, $\text{var}(T_n) \to 0$ and $\text{bias}(T_n) \to 0$ as $n \to \infty$. □

Example 2.2. Suppose that $Y_1, Y_2, \ldots$ are independent Poisson($\lambda$) random variables and consider using the sample mean to estimate $\lambda$. Denote $\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$. Then we have

\[
E(\overline{Y}_n) = \frac{1}{n} E(Y_1 + \ldots + Y_n) = \frac{1}{n} \{E(Y_1) + \ldots + E(Y_n)\} = \frac{1}{n} n \lambda = \lambda,
\]

and so $\overline{Y}_n$ is an unbiased estimator of $\lambda$. Next, by independence, we have

\[
\text{var}(\overline{Y}_n) = \frac{1}{n^2} \{\text{var}(Y_1) + \ldots + \text{var}(Y_n)\} = \frac{1}{n^2} n \lambda = \frac{\lambda}{n},
\]

so that $\text{MSE}(\overline{Y}_n) = \lambda/n$. Thus, $\text{MSE}(\overline{Y}) \to 0$ as $n \to \infty$, and so $\overline{Y}$ is a consistent estimator of $\lambda$. □

A more general result about the mean is given in the following theorem, which states that the sample mean approaches the population mean with probability 1 as $n \to \infty$, and it holds for any distribution.

Theorem 2.1. Weak Law of Large Numbers
Let $X_1, X_2, \ldots$ be IID random variables with $E(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2 < \infty$. Define $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then, for every $\varepsilon > 0$,

\[
\lim_{n \to \infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1,
\]

that is, $\overline{X}_n$ converges in probability to $\mu$. 

**Proof.** The proof is a straightforward application of Markov’s Inequality, which says the following: Let $X$ be a random variable and let $g(\cdot)$ be a non-negative function. Then, for any $k > 0$,

$$P(g(X) \geq k) \leq \frac{E[g(X)]}{k}. \tag{2.1}$$

[Markov’s Inequality]

Proof of Markov’s inequality (continuous case)

For any rv $X$ and a nonnegative function $g(\cdot)$ we can write

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \geq \int_{x:g(x) \geq k} g(x)f_X(x)dx$$

$$\geq \int_{x:g(x) \geq k} kf_X(x)dx = kP(g(X) \geq k).$$

Hence (2.1) holds. Then,

$$P(|X_n - \mu| < \varepsilon) = 1 - P(|X_n - \mu| \geq \varepsilon)$$

$$= 1 - P((X_n - \mu)^2 \geq \varepsilon^2)$$

$$\geq 1 - \frac{E((X_n - \mu)^2)}{\varepsilon^2}$$

$$= 1 - \frac{\text{var}(X_n)}{\varepsilon^2} = 1 - \frac{\sigma^2}{n\varepsilon^2} \to 1 \text{ as } n \to \infty.$$

\[\square\]

Markov’s inequality is a generalization of Chebyshev’s inequality. If we take $g(X) = (X - \mu)^2$ and $k = \sigma^2t^2$ for some positive $t$, then the Markov’s inequality states that

$$P((X - \mu)^2 \geq \sigma^2t^2) \leq \frac{E[(X - \mu)^2]}{\sigma^2t^2} = \frac{\sigma^2}{\sigma^2t^2} = \frac{1}{t^2}$$

This is equivalent to

$$P(|X - \mu| \geq \sigma t) \leq \frac{1}{t^2}$$

which is known as *Chebyshev’s inequality.*
Example 2.3. The rule of three-sigma.
This rule basically says that a chance that an rv will have values outside the interval \((\mu - 3\sigma, \mu + 3\sigma)\) is close to zero. Chebyshev’s inequality, however, says that for any random variable the probability that it will deviate from its mean by more than three standard deviations is at most \(1/9\), i.e.,

\[
P(|X - \mu| \geq 3\sigma) \leq \frac{1}{9}.
\]

This suggests that the rule should be considered with caution. The rule of three-sigma works well for normal rvs or for the variables whose distributions are close to normal. Let \(X \sim \mathcal{N}(\mu, \sigma^2)\). Then

\[
P(|X - \mu| \geq 3\sigma) = 1 - P(|X - \mu| \leq 3\sigma) = 1 - P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)
\]

\[
= 1 - P(-3\sigma \leq X - \mu \leq 3\sigma) = 1 - P(-3 \leq \frac{X - \mu}{\sigma} \leq 3)
\]

\[
= 1 - [\Phi(3) - \Phi(-3)] = 1 - [\Phi(3) - (1 - \Phi(3))]
\]

\[
= 2(1 - \Phi(3)) = 2(1 - 0.99865) \approx 0.0027.
\]

\[\square\]

Another important result, which is widely used in statistic is the Central Limit Theorem (CLT) which uses the notion of convergence in distribution. We have already used this notion, but here we have its formal definition.

**Definition 2.5.** A sequence of random variables \(X_1, X_2, \ldots\) converges in distribution to a random variable \(X\) if

\[
\lim_{n \to \infty} F_{X_n}(x) = F_X(x),
\]

at all points \(x\) where the cdf \(F_X(x)\) is continuous. \[\square\]

The following theorem represents the large sample behavior of the sample mean (compare the Weak Law of Large Numbers).
Theorem 2.2. The Central Limit Theorem
Let \( X_1, X_2, \ldots \) be a sequence of independent random variables such that \( E(X_i) = \mu \) and \( \text{var}(X_i) = \sigma^2 < \infty \). Define
\[
U_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}.
\]
Then the sequence of random variables \( U_1, U_2, \ldots \) converges in distribution to a random variable \( U \sim N(0, 1) \). □

The CLT means that the sample mean is asymptotically normally distributed whatever the distribution of the original rvs is. However, we have no way of knowing how good the approximation is. This depends on the original distribution of \( X_i \).

2.2 Interval Estimation and Hypothesis Testing

Let \( Y = (Y_1, \ldots, Y_n) \) be a random sample from a distribution which belongs to a family of distributions with a parameter \( \vartheta \in \Theta \).

Definition 2.6. A random interval \( [T_1, T_2] = [T_1(Y), T_2(Y)] \), where \( P(T_1 < T_2) = 1 \), such that
\[
P(T_1 \leq \vartheta \leq T_2) = 1 - \alpha \quad \text{for } \vartheta \in \Theta, \ \alpha \in (0, 1),
\]
is called a 100(1 − \( \alpha \))% Confidence Interval (CI) for \( \vartheta \). Random variables \( T_1 \) and \( T_2 \) are called the lower and upper limit, respectively; \( 1 - \alpha \) is called the confidence coefficient. □

Interpretation
\( T_1(Y) \) and \( T_2(Y) \) are rvs, hence their value may be different for different realizations of the random sample \( Y \). For some observed samples \( y \) the intervals \( [T_1(y), T_2(y)] \) may not be covering the true
unknown parameter \( \vartheta \), however, most of the observed samples will give the confidence limits covering the parameter.

For a large number of realizations of \( Y \) the frequency of the event \( \{ \vartheta \in [T_1(y), T_1(y)] \} \) should be approximately equal to \( 1 - \alpha \).

Note that here the lower and upper limits of the interval are realizations of the rvs \( T_1(Y) \) and \( T_2(Y) \) while \( \vartheta \) is an unknown constant.

\([T_1(Y), T_2(Y)]\) is an Interval Estimator of the unknown parameter \( \vartheta \). For a given realization, \( y \), of the random sample we obtain the Interval Estimate of \( \vartheta \). \( 1 - \alpha \) is usually taken to be 0.9, 0.95 or 0.99.

### 2.2.1 Pivotal method of constructing CIs

**Definition 2.7.** Let \( Y = (Y_1, \ldots, Y_n) \) be a random sample from a distribution with a parameter \( \vartheta \in \Theta \). A function \( Q(Y; \vartheta) \) is called a **pivot** for \( \vartheta \) if its distribution does not depend on \( \vartheta \). □

**Example 2.4.** Let \( Y_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma_0^2) \), \( \sigma_0^2 \) known. Then

\[
Q(Y; \mu) = \frac{\bar{Y} - \mu}{\sigma_0 / \sqrt{n}} \sim \mathcal{N}(0, 1)
\]

is a pivot for \( \mu \). It is a function of statistic \( \bar{Y} \) and parameter \( \mu \), but its distribution does not depend on \( \mu \). □

**Example 2.5.** Let \( Y_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2) \), where both parameters are unknown. Then

\[
\frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)
\]

is not a pivot for \( \mu \), since it depends on an unknown \( \sigma \) as well as on \( \mu \).
However, if we replace $\sigma^2$ by its estimator $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$, then

$$Q_1(Y; \mu) = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

depends only on $Y$ and $\mu$ and its distribution does not depend on $\mu$. So, it is a pivot for $\mu$. Also,

$$Q_2(Y; \sigma^2) = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

is a pivot for $\sigma^2$. Note, that these pivots are functions of the estimators of $\mu$ and $\sigma^2$, respectively.

\[\square\]

**Construction of CIs**

To construct a CI we choose values $a$ and $b$ such that for a given $\alpha \in (0, 1)$ we have

$$P(a \leq Q(Y; \vartheta) \leq b) = 1 - \alpha.$$ 

If $Q$ is a strictly monotonic and continuous function of $\vartheta$ then this can be written as

$$P(T_1(Y; a, b) \leq \vartheta \leq T_2(Y; a, b)) = 1 - \alpha.$$ 

Then, the 100$(1 - \alpha)$\% CI for $\vartheta$ is

$$[T_1(Y; a, b), T_2(Y; a, b)].$$

\[\square\]

**Example 2.6. CI for $\mu$ when $\sigma^2$ is unknown, normal population.**

Suppose that $Y_1, \ldots, Y_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables. Then,

$$Q(Y; \vartheta) = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t_{\nu}$$

is a pivot for $\mu$. 
2.2. INTERVAL ESTIMATION AND HYPOTHESIS TESTING

Take the value $b$ as the upper $100\alpha\%$ point of $t_\nu$ distribution and denote it by $t_{\nu, \alpha/2}$, then let $a = -b$. That is $a$ and $b$ are such that

$$P(a < Q < b) = 1 - \alpha,$$

where $a = -t_{\nu, \alpha/2}$ and $b = t_{\nu, \alpha/2}$. Here $\nu = n - 1$. Then we have

$$P\left(-t_{n-1, \alpha/2} < Q(Y, \vartheta) < t_{n-1, \alpha/2}\right) = 1 - \alpha.$$

Equivalently,

$$P\left\{-t_{n-1, \alpha/2} < \overline{Y} - \mu \frac{S}{\sqrt{n}} < t_{n-1, \alpha/2}\right\} = 1 - \alpha.$$

Thus, we may write

$$P\left\{\overline{Y} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}} < \mu < \overline{Y} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right\} = 1 - \alpha.$$

Hence, a $100(1 - \alpha)\%$ confidence interval for $\mu$ is

$$[T_1, T_2] = \left[\overline{Y} - t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}, \overline{Y} + t_{n-1, \alpha/2} \frac{S}{\sqrt{n}}\right],$$

where $S^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2/(n - 1)$.

Note, that the length of the CI, in this case, is

$$2t_{n-1, \alpha/2} \sqrt{\frac{S^2}{n}}.$$

For a fixed value of $\sqrt{\frac{S^2}{n}}$, the length is a non-decreasing function of $\alpha$; on the other hand, for fixed $\alpha$, the length will usually decrease with increasing sample size $n$.

**Example 2.7.** Twenty pilots were tested in a flight simulator and the time for each of them to complete a certain corrective action was measured in seconds. The results are following:

5.2 5.6 7.6 6.8 4.8 5.7 9.0 6.0 4.9 7.4 6.5 7.9 6.8 4.3 8.5 3.6 6.1 5.8 6.4 4.0
Assuming that the sample is a realization of $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, 20$, calculate 95% CI for the mean time to complete the corrective action. Is there evidence that the mean time is different than 6 seconds?

Here we have $\bar{y} = 6.145$, $s = 1.467$, $n = 20$ and $t_{19;0.025} = 2.093$. These give the 95% CI for $\mu$ equal to

$$
\left[ \bar{y} - t_{19;0.025} \frac{s}{\sqrt{n}}, \bar{y} + t_{19;0.025} \frac{s}{\sqrt{n}} \right] = \left[ 6.145 - 2.093 \frac{1.467}{\sqrt{20}}, 6.145 + 2.093 \frac{1.467}{\sqrt{20}} \right] = [5.458, 6.832].
$$

We have 95% confidence that this interval covers the true, unknown mean. The interval includes the hypothesized value of $\mu_0 = 6$, hence, there is no evidence in the data that the mean time to complete the corrective action is not equal to 6 seconds. □

**Example 2.8. CI for $\sigma^2$ of a normal population.**

As we have seen in Example 2.5 the function

$$
Q_2 = \frac{(n - 1)S^2}{\sigma^2} \sim \chi^2_{n-1}
$$

is a pivot for $\sigma^2$ when $Y_i \sim iid N(\mu, \sigma^2)$.

Let $\chi^2_{\nu, \frac{\alpha}{2}}$ and $\chi^2_{\nu, 1 - \frac{\alpha}{2}}$ denote the respective upper and lower $100\frac{\alpha}{2}$% points of the $\chi^2_{\nu}$ distribution. Here $\nu = n - 1$. Then, we have

$$
P \left( \frac{\chi^2_{\nu, 1 - \frac{\alpha}{2}}}{(n - 1)S^2} < \frac{\chi^2_{\nu, \frac{\alpha}{2}}}{\sigma^2} < \chi^2_{\nu, 1 - \frac{\alpha}{2}} \right) = 1 - \alpha.
$$

Thus, we may write

$$
P \left( \frac{1}{\chi^2_{\nu, 1 - \frac{\alpha}{2}}} < \frac{\sigma^2}{(n - 1)S^2} < \frac{1}{\chi^2_{\nu, \frac{\alpha}{2}}} \right) = 1 - \alpha,
$$
which may be rearranged to yield
\[
P \left( \frac{(n-1)S^2}{\chi^2_{n-1, \frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}} \right) = 1 - \alpha.
\]
Hence,
\[
\left[ \frac{n-1S^2}{\chi^2_{n-1, \frac{\alpha}{2}}}, \frac{n-1S^2}{\chi^2_{n-1, 1-\frac{\alpha}{2}}} \right]
\]
is an exact 100(1 - \alpha)% confidence interval for \( \sigma^2 \).

\[\square\]

**Example 2.9.** Tested pilots - CI for the variance.

Here, for \( \nu = n - 1 = 19 \) and \( \alpha = 0.05 \) we have \( \chi^2_{19, 0.025} = 32.85 \), \( \chi^2_{19, 0.975} = 8.907 \). Also, \( s = 1.467 \). That is, the 95% CI for the unknown variance of the time pilots need to complete the corrective action is \([1.24, 4.59]\).

\[\square\]

### 2.2.2 Connection between confidence intervals and tests

**Example 2.10.** Suppose that \( Y_i \sim \mathcal{N}(\mu, \sigma^2) \), where \( \sigma^2 \) is known. In the course Introduction to Statistics, the null hypothesis \( H_0 : \mu = \mu_0 \) was tested against the two-sided alternative \( H_1 : \mu \neq \mu_0 \). The test statistic used was
\[
Z = \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}}.
\]
If \( H_0 \) is true, then \( Z \sim \mathcal{N}(0, 1) \).

We reject \( H_0 \) at the 100\( \alpha \)% level of significance if the observed value of \( Z \) is in the rejection region, i.e., \( z \in (-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty) \).

There is no evidence to reject \( H_0 \) if \( z \in (-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}) \) or in general, if
\[
-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}.
\]
It means that
\[-z_\alpha^2 < \frac{\bar{Y} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha^2.\]

Thus, rearranging, there is no evidence to reject \(H_0\) if
\[\bar{Y} - z_\alpha^2 \frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{Y} + z_\alpha^2 \frac{\sigma}{\sqrt{n}},\]
that is, if the 100(1 - \(\alpha\))% confidence interval for \(\mu\) includes \(\mu_0\). We reject \(H_0\) if the 100(1 - \(\alpha\))% confidence interval for \(\mu\) excludes \(\mu_0\).

\[\Box\]

This is an example of a link between confidence intervals and two-sided hypothesis tests. It provides an appealing interpretation of such tests. The confidence interval gives a range of plausible values of the parameter \(\vartheta\). If it excludes \(\vartheta_0\), then \(\vartheta_0\) is not a plausible value of \(\vartheta\), and so we reject the hypothesis that \(\vartheta = \vartheta_0\).

**Example 2.11. Test for \(\sigma^2\) of a normal population.**
Suppose that \(Y_i \sim \mathcal{N}(\mu, \sigma^2), i = 1, \ldots, n\), where \(\sigma^2\) is unknown.
Then we can also use this method for the test of
\[
\begin{align*}
H_0 &: \sigma^2 = \sigma_0^2 \\
H_1 &: \sigma^2 \neq \sigma_0^2
\end{align*}
\]

The test statistic is
\[
U = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2, \quad H_0
\]
that is, when \(H_0\) is true then \(U \sim \chi_{n-1}^2\). We reject \(H_0\) if the observed value of \(U\), say \(u\), is in the rejection region, i.e.,
\[
u \in \left(0, \chi_{n-1,1-\alpha/2}^2\right) \cup \left(\chi_{n-1,\alpha/2}^2, \infty\right)
\]
There is no evidence to reject $H_0$ if
$$u \in \left( \chi^2_{n-1, \frac{1}{2}}, \chi^2_{n-1, \frac{3}{2}} \right)$$
or, in general, if
$$\chi^2_{n-1, \frac{1}{2}} < \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{n-1, \frac{3}{2}}.$$ 
Thus, rearranging, there is no evidence to reject $H_0$ if
$$\frac{(n-1)S^2}{\chi^2_{n-1, \frac{3}{2}}} < \sigma_0^2 < \frac{(n-1)S^2}{\chi^2_{n-1, \frac{1}{2}}}.$$ 
that is, if the $100(1 - \alpha)\%$ confidence interval for $\sigma^2$ includes $\sigma_0^2$. We reject $H_0$ if the $100(1 - \alpha)\%$ confidence interval for $\sigma^2$ excludes $\sigma_0^2$. □

If $\sigma^2$ is unknown as in the example above, then the test statistic for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ is
$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}.$$ 
There is no evidence to reject $H_0$ if $-t_{n-1, \frac{1}{2}} < T < t_{n-1, \frac{3}{2}}$. That is, there is no evidence to reject $H_0$ at the $100\alpha\%$ level of significance if the $100(1 - \alpha)\%$ confidence interval for $\mu$ includes $\mu_0$.

We have seen that the usual two-sided tests correspond to the usual confidence intervals. We can also obtain confidence intervals corresponding to one-sided tests.

Example 2.12. **One-sided CI and test for $\mu$, normal population.** Suppose that $Y_1, \ldots, Y_n$ are independent $\mathcal{N}(\mu, \sigma^2)$ random variables, where $\sigma^2$ is unknown, and consider testing
$$H_0 : \mu = \mu_0 \quad H_1 : \mu > \mu_0$$
Then, the test statistic is
\[ T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}, \]
We reject \( H_0 \) if \( T > t_{n-1,\alpha} \), that is,
\[ \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} > t_{n-1,\alpha}. \]
Thus, rearranging the expression, we reject \( H_0 \) if
\[ \mu_0 < \bar{Y} - t_{n-1,\alpha} \frac{S}{\sqrt{n}}, \]
that is, if the interval \((\bar{Y} - t_{n-1,\alpha} S/\sqrt{n}, \infty)\) excludes \( \mu_0 \). This corresponds to a one-sided 100(1 - \( \alpha \))% confidence interval for \( \mu \). It gives a lower plausible bound for \( \mu \), so we reject the null hypothesis if the CI does not include the hypothesized value of the parameter.

Similarly, the interval \((-\infty, \bar{Y} + t_{n-1,\alpha} S/\sqrt{n})\) gives an upper plausible bound for \( \mu \) (when testing \( H_0 \) against \( H_1 : \mu < \mu_0 \)).

**Example 2.13.** Researchers have shown that cigarette smoking has a deleterious effect on lung function. In their study of the effect of smoking on the carbon monoxide diffusing capacity of the lung they found that current smokers had readings significantly lower than either ex-smokers or nonsmokers. The carbon monoxide diffusing capacity of a random sample of smokers was as follows:

| 103.7 | 88.6 | 73.0 | 123.1 | 91.1 |
| 92.3  | 61.7 | 90.7 | 84.0  | 76.0 |
| 100.6 | 88.0 | 71.2 | 82.1  | 89.2 |
| 102.8 | 108.6| 73.2 | 106.8 | 90.5 |

Do these data indicate that the mean diffusing capacity of current smokers is lower than 100, the population mean of nonsmokers?
Let $X$ denote the carbon monoxide diffusing capacity of the lung of current smokers and let $\mu$ denote the mean diffusing capacity of current smokers. The hypotheses are:

$$H_0 : \mu = 100$$
$$H_1 : \mu < 100$$

Assuming that the sample comes from a normal population with expectation $\mu = 100$ and unknown variance, we can use the one-sample t-test. Here we have a small sample size $n = 20$. We get

$$T_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{89.86 - 100}{14.91/\sqrt{20}} = -3.04.$$  

The rejection region is $(-\infty, -t_{19;0.05}) = (-\infty, -1.729)$. The observed value of the statistic is in the rejection region, hence there is evidence against the null hypothesis. At the significance level $\alpha = 0.05$ we reject the hypothesis that the mean diffusing capacity of current smokers is 100 versus the hypothesis that it is smaller than 100.

Alternatively we can use the one-sided confidence interval to make inference about the parameter. The 95% CI is

$$\left(-\infty, 89.86 + 1.729 \frac{14.91}{\sqrt{20}} \right] = (-\infty, 95.62].$$

That is the upper plausible bound for the true parameter is 95.62 (with 95% confidence), hence we reject the null hypothesis which says that it is 100. \hfill \square

Exactly the same methods give approximate confidence intervals and approximate hypothesis tests for the population mean when we do not know the distribution of the random sample but we have large samples. Then, the Central Limit Theorem gives the pivot for $\mu$ when the variance is known. We further approximate the CI and the test when we replace $\sigma^2$ by its estimator $S^2$. 