

CENG 222 - Chapter 9: Statistical Inference I

Burak Metehan Tuncel - June 2022

1 Parameter Estimation

We have learned a few elementary ways to determine the *family of distributions*. Now, we learn how to *estimate parameters of distributions*. As a result, a large family will be reduced to just one distribution that we can use for performance evaluation, forecasting, etc.

Statisticians developed a number of estimation techniques, each having certain optimal properties. Two rather popular methods are discussed in this section:

- method of moments, and
- method of maximum likelihood.

1.1 Method of Moments

1.1.1 Moments

First, let us define the moments.

Definition 1

The k -th **population moment** is defined as

$$\mu_k = \mathbf{E}(X^k)$$

The k -th **sample moment**

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

estimates μ_k from a sample (X_1, \dots, X_n) .

The first sample moment is the sample mean \bar{X} .

Central moments are computed similarly, after centralizing the data, that is, subtracting the mean.

Definition 2

For $k \geq 2$, the k -th **population central moment** is defined as

$$\mu'_k = \mathbf{E}((X - \mu_1)^k)$$

The k -th **sample central moment**

$$m'_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$$

estimates μ_k from a sample (X_1, \dots, X_n) .

Remark:

- The *second population central moment* is variance $\text{Var}(X)$.
- The *second sample central moment* is *sample variance*, although $(n-1)$ in its denominator is now replaced by n .

We mentioned that estimation methods are not unique. For unbiased estimation of $\sigma^2 = \text{Var}(X)$, we use

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

however, method of moments and method of maximum likelihood produce a different version,

$$S^2 = m'_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

1.1.2 Estimation

Method of moments is based on a simple idea. Since our sample comes from a family of distributions $\{F(\theta)\}$, we choose such a member of this family whose properties are close to properties of our data. Namely, we shall *match the moments*.

To estimate k parameters, equate the first k population and sample moments,

$$\begin{cases} \mu_1 = m_1 \\ \dots \dots \dots \\ \mu_k = m_k \end{cases}$$

The left-hand sides of these equations depend on the distribution parameters. The right-hand sides can be computed from data. The **method of moments estimator** is the solution of this system of equation.

Example 1 (Poisson)

To estimate parameter λ of Poisson(λ) distribution, we recall that

$$\mu_1 = \mathbf{E}(X) = \lambda$$

There is only one unknown parameter, hence we write one equation,

$$\mu_1 = \lambda = m_1 = \bar{X}$$

“Solving” it for λ , we obtain

$$\hat{\lambda} = \bar{X}$$

the method of moments estimator of λ .

Check the examples 9.4 and 9.5 from textbook.

On rare occasions, when k equations are not enough to estimate k parameters, we'll consider higher moments.

Example 2 (Normal)

Suppose we already know the mean μ of a Normal distribution and would like to estimate the variance σ^2 . Only one parameter σ^2 is unknown; however, the first method of moments equation

$$\mu_1 = m_1$$

does not contain σ^2 and therefore does not produce its estimate. We then consider the second equation, say,

$$\mu'_2 = \sigma^2 = m'_2 = S^2$$

which gives us the method of moments estimate immediately, $\hat{\sigma}^2 = S^2$.

Method of moments estimates are typically easy to compute. They can serve as a quick tool for estimating parameters of interest.

1.2 Method of Maximum Likelihood

Since the sample $\mathbf{X} = (X_1, \dots, X_n)$ has already been observed, we find such parameters that maximize the probability (likelihood) for this to happen. In other words, we make the event that has already happened to be as likely as possible. This is yet another way to make the chosen distribution consistent with the observed data.

Definition 3

Maximum likelihood estimator is the parameter value that maximizes the likelihood of the observed sample.

- For a discrete distribution, we maximize the joint pmf of data $P(X_1, \dots, X_n)$.
- For a continuous distribution, we maximize the joint density $f(X_1, \dots, X_n)$.

1.2.1 Discrete Case

For a discrete distribution, the probability of a given sample is the joint pmf of data,

$$\begin{aligned} P\{\mathbf{X} = (X_1, \dots, X_n)\} &= P(\mathbf{X}) = P(X_1, \dots, X_n) \\ &= \prod_{i=1}^n P(X_i) \end{aligned}$$

because in a simple random sample, all observed X_i are independent.

To maximize this likelihood, we consider the critical points by taking derivatives with respect to all unknown parameters and equating them to 0. The maximum can only be attained at such parameter values θ where the derivative $\frac{\partial}{\partial \theta} P(\mathbf{X})$ equals 0, where it does not exist, or at the boundary of the set of possible values of θ .

A nice computational shortcut is to take logarithms first.

Differentiating the sum

$$\ln \prod_{i=1}^n P(X_i) = \sum_{i=1}^n \ln P(X_i)$$

is easier than differentiating the product $\prod P(X_i)$. Besides, logarithm is an increasing function, so the likelihood $P(\mathbf{X})$ and the log-likelihood $\ln P(\mathbf{X})$ are maximized by exactly the same parameters.

Check the example 9.7 from textbook.

1.2.2 Continuous Case

In the continuous case, the probability to observe exactly the given number $X = x$ is 0. Instead, the method of maximum likelihood will maximize the probability of observing “almost” the same number.

For a very small h ,

$$P\{x-h < X < x+h\} = \int_{x-h}^{x+h} f(y)dy \approx (2h)f(x)$$

That is, the probability of observing a value close to x is proportional to the density $f(x)$ (see Figure 1). Then, for a sample $\mathbf{X} = (X_1, \dots, X_n)$, the maximum likelihood method will maximize the joint density $f(X_1, \dots, X_n)$.

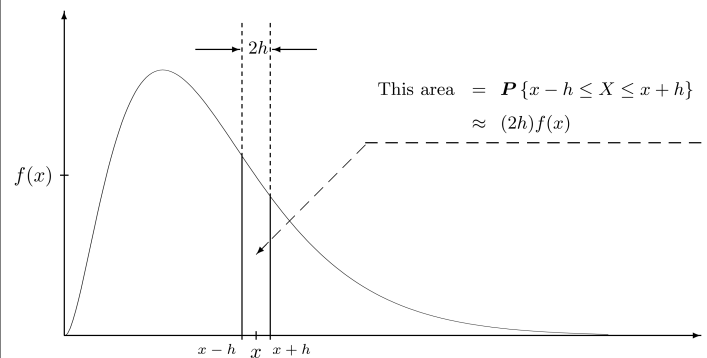


Figure 1

Check the examples 9.8 and 9.9 from textbook.

Sometimes the likelihood has no critical points inside its domain, then it is maximized at the boundary.

When we estimate more than 1 parameter, all the partial derivatives should be equal 0 at the critical point. If no critical points exist, the likelihood is again maximized on the boundary.

Maximum likelihood estimators are rather popular because of their nice properties. Under mild conditions, these estimators are consistent, and for large samples, they have an approximately Normal distribution. Often in complicated problems, finding a good estimation scheme may be challenging whereas the maximum likelihood method always gives a reasonable solution.

1.3 Estimation of Standard Errors

Standard errors can serve as measures of their accuracy. To estimate them, we derive an expression for the standard error and estimate all the unknown parameters in it.

Example 3 (Estimation of the Poisson Parameter)

By the method of moments and maximum likelihood estimators of the Poisson parameter λ is $\hat{\lambda} = \bar{X}$. Let us now estimate the standard error of $\hat{\lambda}$.

Solution: There are at least two ways to do it.

On one hand, $\sigma = \sqrt{\lambda}$ for the $\text{Poisson}(\lambda)$ distribution, so $\sigma(\hat{\lambda}) = \sigma(\bar{X}) = \sigma/\sqrt{n} = \sqrt{\lambda/n}$, as we know before (*textbook (8.2) on p. 213*). Estimating λ by \bar{X} , we obtain

$$s_1(\hat{\lambda}) = \sqrt{\frac{\bar{X}}{n}} = \frac{\sqrt{\sum X_i}}{n}$$

On the other hand, we can use the sample standard deviation and estimate the standard error of the sample mean,

$$s_2(\hat{\lambda}) = \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n(n-1)}}$$

Apparently, we can estimate the standard error of $\hat{\lambda}$ by two good estimators, s_1 and s_2 .

2 Confidence Intervals

When we report an estimator $\hat{\theta}$ of a population parameter θ , we know that most likely

$$\hat{\theta} \neq \theta$$

due to a sampling error. We realize that we have estimated θ up to *some error*.

Then how much can we trust the reported estimator? How far can it be from the actual parameter of interest? What is the probability that it will be reasonably close? And if we observed an estimator $\hat{\theta}$, then what can the actual parameter θ be?

To answer these questions, statisticians use **confidence intervals**, which contain parameter values that deserve some confidence, given the observed data.

Definition 4

An interval $[a, b]$ is a $(1-\alpha)100\%$ **confidence interval** for the parameter θ if it contains the parameter with probability $(1-\alpha)$,

$$P\{a \leq \theta \leq b\} = 1 - \alpha$$

The **coverage probability** $(1-\alpha)$ is also called a **confidence level**.

Let us take a moment to think about this definition. The probability of a random event $\{a \leq \theta \leq b\}$ has to be $(1-\alpha)$.

What randomness is involved in this event?

The population parameter θ is not random. It is a *population feature, independent of any random sampling procedure*, and therefore, it remains constant. On the other hand, the interval is computed from random data, and therefore, it is random. The coverage probability refers to the chance that our interval covers a constant parameter θ .

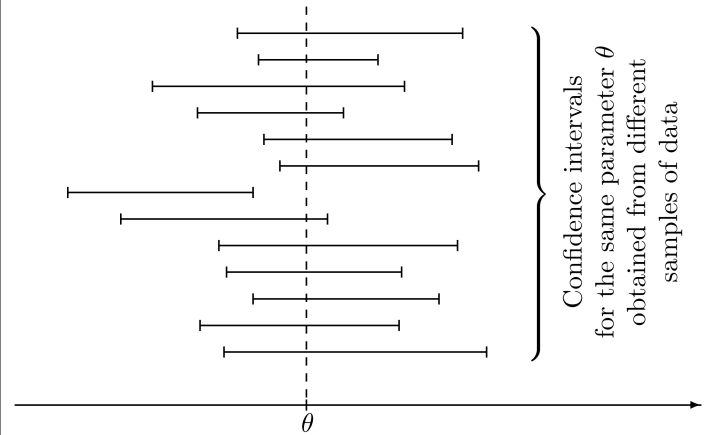


Figure 2

This is illustrated in Figure 2. Suppose that we collect many random samples and produce a confidence interval from each of them. If these are $(1-\alpha)100\%$ confidence intervals, then we expect $(1-\alpha)100\%$ of them to cover θ and $100\alpha\%$ of them to miss it. In Figure 2, we see one interval that does not cover θ . No mistake was made in data collection and construction of this interval. It missed the parameter only due to a *sampling error*.

It is therefore wrong to say, “I computed a 90% confidence interval, it is $[3, 6]$. Parameter belongs to this interval with probability 90%”. The parameter is constant; it either belongs to the interval $[3, 6]$ (with probability 1) or does not. In this case, 90% refers to the proportion of confidence intervals that contain the unknown parameter in a long run.

2.1 Construction of Confidence Intervals: A General Method

Given a sample of data and a desired confidence level $(1-\alpha)$, how can we construct a confidence interval $[a, b]$ that will satisfy the coverage condition

$$P\{a \leq \theta \leq b\} = 1 - \alpha$$

in definition 4.

We start by estimating parameter θ . Assume there is an unbiased estimator $\hat{\theta}$ that has a Normal distribution. When we standardize it, we get a Standard Normal variable

$$Z = \frac{\hat{\theta} - \mathbf{E}(\hat{\theta})}{\sigma(\hat{\theta})} = \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})}$$

where $\mathbf{E}(\hat{\theta}) = \theta$ because $\hat{\theta}$ is unbiased, and $\sigma(\hat{\theta}) = \sigma(\hat{\theta})$ is its standard error.

This variable falls between the Standard Normal quantiles $q_{\alpha/2}$ and $q_{1-\alpha/2}$, denoted by

$$\begin{aligned} -z_{\alpha/2} &= q_{\alpha/2} \\ z_{\alpha/2} &= q_{1-\alpha/2} \end{aligned}$$

with probability $(1 - \alpha)$, see Figure 3.

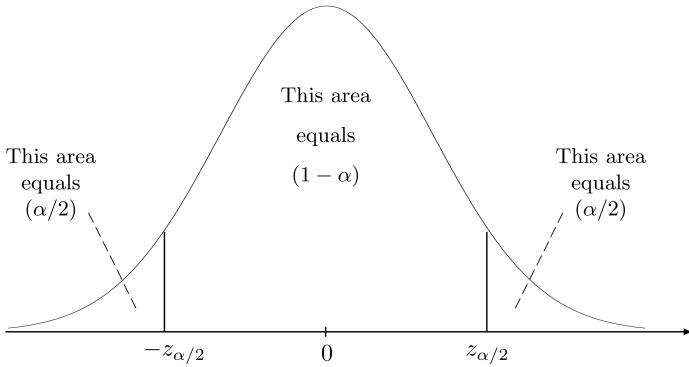


Figure 3

Then,

$$P \left\{ -z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})} \leq z_{\alpha/2} \right\} = 1 - \alpha$$

Solving the inequality inside $\{\dots\}$ for θ , we get

$$P \left\{ \hat{\theta} - z_{\alpha/2} \cdot \sigma(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\alpha/2} \cdot \sigma(\hat{\theta}) \right\} = 1 - \alpha$$

The problem is solved! We have obtained two numbers

$$\begin{aligned} a &= \hat{\theta} - z_{\alpha/2} \cdot \sigma(\hat{\theta}) \\ b &= \hat{\theta} + z_{\alpha/2} \cdot \sigma(\hat{\theta}) \end{aligned}$$

such that

$$P \{a \leq \theta \leq b\} = 1 - \alpha$$

Confidence interval, Normal distribution, Eq. 3

If parameter θ has an unbiased, Normally distributed estimator $\hat{\theta}$, then

$$\hat{\theta} \pm z_{\alpha/2} \cdot \sigma(\hat{\theta}) = \left[\hat{\theta} - z_{\alpha/2} \cdot \sigma(\hat{\theta}), \hat{\theta} + z_{\alpha/2} \cdot \sigma(\hat{\theta}) \right]$$

is a $(1 - \alpha)100\%$ confidence interval for θ .

If the distribution of $\hat{\theta}$ is *approximately* Normal, we get an approximately $(1 - \alpha)100\%$ confidence interval.

In this formula, $\hat{\theta}$ is the **center of the interval**, and $z_{\alpha/2} \cdot \sigma(\hat{\theta})$ is the **margin**. The margin of error is often reported along with poll and survey results. In newspapers and press releases, it is usually computed for a 95% confidence interval.

NOTATION

$$z_{\alpha} = q_{1-\alpha} = \Phi^{-1}(1 - \alpha)$$

is the value of a Standard Normal variable Z that is exceeded with probability α

Several important applications of this general method are discussed below. In each problem, we

1. Find an unbiased estimator of θ ,
2. Check if it has a Normal distribution,
3. Find its standard error $\sigma(\hat{\theta}) = \text{Std}(\hat{\theta})$,
4. Obtain quantiles $\pm z_{\alpha/2}$ from the table of Normal distribution (Table A4 in the Appendix in textbook), and finally,
5. Apply the rule (3).

2.2 Confidence Interval for the Population Mean

Let us construct a confidence interval for the population mean

$$\theta = \mu = \mathbf{E}(X)$$

Start with an estimator

$$\hat{\theta} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

The rule (3) is applicable in two cases.

1. If a sample $\mathbf{X} = (X_1, \dots, X_n)$ comes from Normal distribution, then \hat{X} is also Normal, and rule (3) can be applied.
2. If a sample comes from any distribution, but the sample size n is large, then \hat{X} has an approximately Normal distribution according to the Central Limit Theorem. Then rule (3) gives an approximately $(1 - \alpha)100\%$ confidence interval.

The followings can be derived

$$\begin{aligned} \mathbf{E}(X) &= \mu && \text{(unbiased estimator)} \\ \sigma(\bar{X}) &= \sigma/\sqrt{n} \end{aligned}$$

Then, (3) reduces to the following $(1 - \alpha)100\%$ confidence interval for μ

Confidence interval for the mean; σ is known

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (5)$$

Example 4

Construct a 95% confidence interval for the population mean based on a sample of measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

if measurement errors have Normal distribution, and the measurement device guarantees a standard deviation of $\sigma = 2.2$.

Solution: This sample has size $n = 6$ and sample mean $\bar{X} = 6.50$. To attain a confidence level of

$$1 - \alpha = 0.95$$

we need $\alpha = 0.05$ and $\alpha/2 = 0.025$. Hence, we are looking for quantiles

$$q_{0.025} = -z_{0.025} \quad \text{and} \quad q_{0.975} = z_{0.025}$$

From Table 4 in textbook, we find that $q_{0.975} = 1.960$. Substituting these values into (9.5), we obtain a 95% confidence interval for μ ,

$$\begin{aligned} \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} &= 6.50 \pm 1.960 \frac{2.2}{\sqrt{6}} \\ &= 6.50 \pm 1.76 \text{ or } [4.74, 8.26] \end{aligned}$$

2.3 Confidence Interval for the Difference Between Two Means

Under the same conditions as in the previous section,

- Normal distribution of data or
- Sufficiently large sample size,

we can construct a confidence interval for the difference between two means.

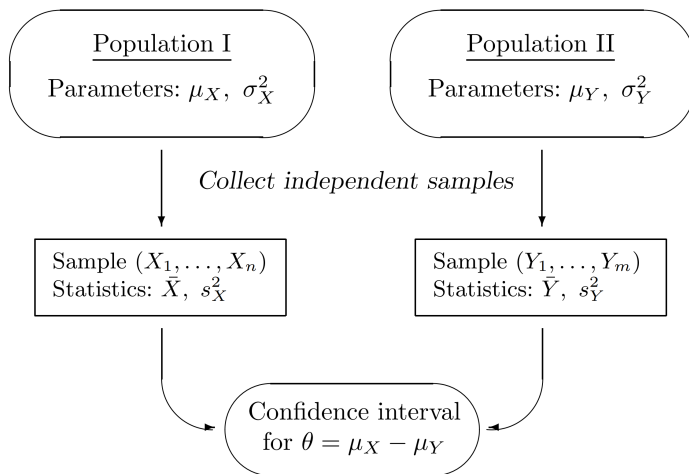


Figure 4: Comparison of two populations

Suppose that the two samples are collected *independently* of each other. To construct a confidence interval for the difference between population means

$$\theta = \mu_X - \mu_Y$$

we complete the usual steps (a)-(e) below:

- (a) Propose an estimator of θ ,

$$\hat{\theta} = \bar{X} - \bar{Y}$$

It is natural to come up with this estimator because \bar{X} estimates μ_X and \bar{Y} estimates μ_Y .

- (b) Check that $\hat{\theta}$ is unbiased. Indeed,

$$\mathbf{E}(\hat{\theta}) = \mathbf{E}(\bar{X} - \bar{Y}) = \mathbf{E}(\bar{X}) - \mathbf{E}(\bar{Y}) = \mu_X - \mu_Y = \theta$$

- (c) Check that $\hat{\theta}$ has a Normal or approximately Normal distribution. This is true if the observations are Normal or both sample sizes m and n are large.

- (d) Find the standard error of $\hat{\theta}$ (using independence of \mathbf{X} and \mathbf{Y}),

$$\begin{aligned} \sigma(\hat{\theta}) &= \sqrt{\text{Var}(\bar{X}) - \text{Var}(\bar{Y})} = \sqrt{\text{Var}(\bar{X}) + \text{Var}(\bar{Y})} \\ &= \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \end{aligned}$$

- (e) Find quantiles $\pm z_{\alpha/2}$ and compute the confidence interval according to (3). This results in the following formula.

Confidence interval for the difference of means; known standard deviations

$$\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \quad (6)$$

Example 5 (Effect of an upgrade)

A manager evaluates effectiveness of a major hardware upgrade by running a certain process 50 times before the upgrade and 50 times after it. Based on these data, the average running time is 8.5 minutes before the upgrade, 7.2 minutes after it. Historically, the standard deviation has been 1.8 minutes, and presumably it has not changed. Construct a 90% confidence interval showing how much the mean running time reduced due to the hardware upgrade.

Solution: We have $n = m = 50$, $\sigma_X = \sigma_Y = 1.8$, $\bar{X} = 8.5$, and $\bar{Y} = 7.2$. Also, the confidence level $(1 - \alpha)$ equals 0.9, hence $\alpha/2 = 0.05$, and $z_{\alpha/2} = 1.645$.

The distribution of times may not be Normal; however, due to large sample sizes, the estimator

$$\hat{\theta} = \bar{X} - \bar{Y}$$

is approximately Normal by the Central Limit Theorem. Thus, formula (6) is applicable, and a 90% confidence interval for the difference of means $(\mu_X - \mu_Y)$ is

$$\begin{aligned} 8.5 - 7.2 \pm 1.645 \sqrt{(1.8)^2 \left(\frac{1}{50} + \frac{1}{50} \right)} \\ = 1.3 \pm 0.6 \text{ or } [0.7, 1.9] \end{aligned}$$

We can say that the hardware upgrade resulted in a 1.3-minute reduction of the mean running time, with a 90% confidence margin of 0.6 minutes.

2.4 Selection of a Sample Size

Formula (3) describes a confidence interval as

$$\text{center} \pm \text{margin}$$

where

$$\begin{aligned}\text{center} &= \hat{\theta} \\ \text{margin} &= z_{\alpha/2} \cdot \sigma(\hat{\theta})\end{aligned}$$

We can revert the problem and ask a very practical question: *How large a sample should be collected to provide a certain desired precision of our estimator?*

To answer this question, we only need to solve the inequality

$$\text{margin} \leq \Delta \quad (7)$$

in terms of n . Typically, parameters are estimated more accurately based on larger samples, so that the standard error $\sigma(\hat{\theta})$ and the margin are decreasing functions of sample size n . Then, (7) must be satisfied for sufficiently large n .

2.5 Estimating Means with a Given Precision

When we estimate a population mean, the margin of error is

$$\text{margin} = z_{\alpha/2} \cdot \sigma / \sqrt{n}$$

Solving inequality (7) for n results in the following rule.

Sample size for a given precision

In order to attain a margin of error Δ for estimating a population mean with a confidence level $(1 - \alpha)$, a sample of size

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{\Delta} \right)^2 \quad (8)$$

is required.

When we compute the expression in (8), it will most likely be a fraction. Notice that we can only round it up to the nearest integer sample size. If we round it down, our margin will exceed Δ .

Looking at (8), we see that a large sample will be necessary

- to attain a narrow margin (small Δ);
- to attain a high confidence level (small α); and
- to control the margin under high variability of data (large σ).

In particular, we need to quadruple the sample size in order to half the margin of the interval.

Example 6

In Example 4, we constructed a 95% confidence with the center 6.50 and margin 1.76 based on a sample of size 6. Now, that was too wide, right? How large a sample do we need to estimate the population mean with a margin of at most 0.4 units with 95% confidence?

Solution: We have $\Delta = 0.4$, $\alpha = 0.05$, and $\sigma = 2.2$ (from Example 4). By (8), we need a sample of

$$n \geq \left(\frac{z_{0.05/2} \sigma}{\Delta} \right)^2 = \left(\frac{(1.960)(2.2)}{0.4} \right)^2 = 116.2$$

Keeping in mind that this is the minimum sample size that satisfies Δ , and we are only allowed to round it up, we need a sample of at least 117 observations.

3 Unknown Standard Deviation

A rather heavy condition was assumed when we constructed all the confidence intervals. We assumed a *known standard deviation* σ and used it in all the derived formulas.

Sometimes this assumption is perfectly valid and we may know the variance. However, generally, the population variance is *unknown*. We'll then estimate it from data and see if we can still apply methods of the previous section.

Two broad situations will be considered:

- large samples from any distribution,
- samples of any size from a Normal distribution.

3.1 Large Samples

A large sample should produce a rather accurate estimator of a variance. We can then replace the true standard error $\sigma(\hat{\theta})$ in (3) by its estimator $s(\hat{\theta})$, and obtain an approximate confidence interval

$$\hat{\theta} \pm z_{\alpha/2} \cdot s(\hat{\theta})$$

Example 7 (Delays at nodes)

Internet connections are often slowed by delays at nodes. Let us determine if the delay time increases during heavy-volume times.

Five hundred packets are sent through the same network between 5 pm and 6 pm (sample **X**), and three hundred packets are sent between 10 pm and 11 pm (sample **Y**). The early sample has a mean delay time of 0.8 sec with a standard deviation of 0.1 sec whereas the second sample has a mean delay time of 0.5 sec with a standard deviation of 0.08 sec. Construct a 99.5% confidence interval for the difference between the mean delay times.

Solution: We have $n = 500$, \bar{X} , $s_X = 0.1$; $m = 500$, \bar{Y} , $s_Y = 0.1$. Large sample sizes allow us to replace unknown population standard deviations by their estimates and use an approximately Normal distribution of sample means.

For a confidence level of $1 - \alpha = 0.995$, we need

$$z_{\alpha/2} = z_{0.0025} = q_{0.9975}$$

Look for the *probability* 0.9975 in the body of Table A4 (from textbook) and find the corresponding value of z ,

$$z_{0.0025} = 2.81$$

Then, a 99.5% confidence interval for the difference of mean execution times is

$$\begin{aligned}\bar{X} - \bar{Y} \pm z_{0.0025} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} \\ = (0.8 - 0.5) \pm (2.81) \sqrt{\frac{(0.1)^2}{500} + \frac{(0.08)^2}{300}} \\ = 0.3 \pm \text{ or } [0.282, 0.318]\end{aligned}$$

3.2 Confidence Intervals for Proportions

In particular, we surely don't know the variance when we estimate a population proportion.

Definition 5

We assume a subpopulation A of items that have a certain *attribute*. By the **population proportion** we mean the probability

$$p = P\{i \in A\}$$

for a randomly selected item i to have this attribute. A **sample proportion**

$$\hat{p} = \frac{\text{number of sampled items from } A}{n}$$

is used to estimate p .

Let us use the *indicator* variables

$$X_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

Each X_i has Bernoulli distribution with parameter p . In particular

$$\mathbf{E}(X_i) = p \quad \text{and} \quad \text{Var}(X_i) = p(1 - p)$$

Also,

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$$

is nothing but a sample of X_i .

Therefore,

$$\mathbf{E}(\hat{p}) = p \quad \text{and} \quad \text{Var}(\hat{p}) = \frac{p(1 - p)}{n}$$

We conclude that

1. a sample proportion \hat{p} is unbiased for the population proportion p ;
2. it has approximately Normal distribution for large samples, because it has a form of a sample mean;
3. when we construct a confidence interval for p , we do not know the standard deviation $\text{Std}(\hat{p})$.

Indeed, knowing the standard deviation is equivalent to knowing p , and if we know p , why would we need a confidence interval for it?

Thus, we estimate the unknown standard error

$$\sigma(\hat{p}) = \sqrt{\frac{p(1 - p)}{n}}$$

by

$$s(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

and use it in the general formula

$$\hat{p} \pm z_{\alpha/2} \cdot s(\hat{p})$$

to construct an approximate $(1 - \alpha)100\%$ confidence interval.

Confidence interval for a population proportion

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

Similarly, we can construct a confidence interval for the difference between two proportions. In two populations, we have proportions p_1 and p_2 of items with an attribute. Independent samples of size n_1 and n_2 are collected, and both parameters are estimated by sample proportions \hat{p}_1 and \hat{p}_2 .

Summarizing, we have

Parameter of interest: $\theta = p_1 - p_2$

Estimated by: $\hat{\theta} = \hat{p}_1 - \hat{p}_2$

Its standard error: $\sigma(\hat{\theta}) = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$

Estimated by: $s(\hat{\theta}) = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$

Confidence interval for the difference of proportions

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

Example 8 (Pre-election poll)

A candidate prepares for the local elections. During his campaign, 42 out of 70 randomly selected people in town A and 59 out of 100 randomly selected people in town B showed they would vote for this candidate. Estimate the difference in support that this candidate is getting in towns A and B with 95% confidence. Can we state affirmatively that the candidate gets a stronger support in town A?

Solution: We have $n_1 = 70$, $n_2 = 100$, $\hat{p}_1 = 42/70 = 0.6$, and $\hat{p}_2 = 59/100 = 0.59$. For the confidence interval, we have

$$\text{center} = \hat{p}_1 - \hat{p}_2 = 0.01$$

and

$$\begin{aligned} \text{margin} &= z_{0.05/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= (1.960) \sqrt{\frac{(0.6)(0.4)}{70} + \frac{(0.59)(0.41)}{100}} = 0.15 \end{aligned}$$

Then

$$0.01 \pm 0.15 = [-0.14, 0.16]$$

is a 95% confidence interval for the difference in support ($p_1 - p_2$) in the two towns.

So, is the support stronger in town A? On one hand, the estimator $\hat{p}_1 - \hat{p}_2 = 0.01$ suggests that the support is 1% higher in town A than in town B. On the other hand, the difference could appear positive just because of a sampling error. As we see, the 95% confidence interval includes a large range of negative values too. Therefore, the obtained data *does not* indicate affirmatively that the support in town A is stronger.

Example 9

A sample of size

$$n \geq 0.25 \left(\frac{1.960}{0.1} \right)^2 = 96.04$$

that is, at least 97 observations always guarantees that a population proportion is estimated with an error of at most 0.1 with a 95% confidence.

3.4 Small Samples: Student's t distribution

Having a small sample, we can no longer pretend that a sample standard deviation s is an accurate estimator of the population standard deviation σ . Then, how should we adjust the confidence interval when we replace σ by s , or more generally, when we replace the standard error $\sigma(\hat{\theta})$ by its estimator $s(\hat{\theta})$?

A famous solution was proposed by *William Gosset* (1876–1937), he derived the **T-distribution**.

He replaced the true but unknown standard error of $\hat{\theta}$ by its estimator $s(\hat{\theta})$ and concluded that the **T-ratio**

$$t = \frac{\hat{\theta} - \theta}{s(\hat{\theta})}$$

the *ratio* of two random variables, no longer has a Normal distribution!

For the problem of estimating the mean based on n Normal observations X_1, \dots, X_n , this was **T-distribution** with $(n-1)$ *degrees of freedom*. Table A5 (from textbook) gives critical values t_α of the T-distribution that we'll use for confidence intervals.

So, using *T-distribution* instead of Standard Normal and estimated standard error instead of the unknown true one, we obtain the confidence interval for the population mean.

Confidence interval for the mean σ is unknown

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad (9)$$

where $t_{\alpha/2}$ is a critical value from T-distribution with $n-1$ degrees of freedom.

The density of *Student's T-distribution* is a bell-shaped symmetric curve that can be easily confused with Normal. Comparing with the Normal density, *its peak is lower and its tails are thicker*. Therefore, a larger number t_α is generally needed to cut area α from the right tail. That is

$$t_\alpha > z_\alpha$$

for small α . As a consequence, the confidence interval (9) is wider than the interval (5) for the case of known σ . This wider margin is the price paid for not knowing the standard deviation σ . When we lack a certain piece of information, we cannot get a more accurate estimator.

3.3 Estimating Proportions with a Given Precision

Our confidence interval for a population proportion has a margin

$$\text{margin} = z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

A standard way of finding the sample size that provides the desired margin Δ is to solve the inequality

$$\text{margin} \leq \Delta \quad \text{or} \quad n \geq \hat{p}(1-\hat{p}) \left(\frac{z_{\alpha/2}}{\Delta} \right)^2$$

However, this inequality includes \hat{p} . To know \hat{p} , we first need to collect a sample, but to know the sample size, we first need to know \hat{p} !

Since $\hat{p}(1-\hat{p})$ does not exceed 0.25, we can replace this unknown by 0.25 and find a sample size n , perhaps larger than we actually need, that will ensure that we estimate \hat{p} with a margin not exceeding Δ . That is, choose a sample size

$$n \geq 0.25 \left(\frac{z_{\alpha/2}}{\Delta} \right)^2$$

However, we see in Table A5 that

$$t_\alpha \rightarrow z_\alpha$$

as the number of degrees of freedom ν tends to infinity. Indeed, having a large sample (hence, large $\nu = n - 1$), we can count on a very accurate estimator of σ , and thus, the confidence interval is almost as narrow as if we knew σ in this case.

Degrees of freedom ν is the parameter of T-distribution controlling the shape of the T-density curve. Its meaning is the dimension of a vector used to estimate the variance. Here we estimate σ^2 by a sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

and thus, we use a vector

$$\mathbf{X}' = (X_1 - \bar{X}, \dots, X_n - \bar{X})$$

The initial vector $\mathbf{X} = (X_1, \dots, X_n)$ has dimension n ; therefore, it has n degrees of freedom. However, when the sample mean \bar{X} is subtracted from each observation, there appears a linear relation among the elements,

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

We lose 1 degree of freedom due to this constraint; the vector \mathbf{X}' belongs to an $(n - 1)$.

In many similar problems, degrees of freedom can be computed as

$$\begin{array}{lcl} \text{number of} & = & \text{sample size} - \text{number of} \\ \text{degrees of} & & \text{estimated} \\ \text{freedom} & & \text{location} \\ & & \text{parameters} \end{array} \quad (10)$$

3.5 Comparison of Two Populations with Unknown Variances

We now construct a confidence interval for the difference of two means $\mu_X - \mu_Y$, comparing the population of X 's with the population of Y 's.

Again, independent random samples are collected,

$$\mathbf{X} = (X_1, \dots, X_n) \quad \text{and} \quad \mathbf{Y} = (Y_1, \dots, Y_m)$$

one from each population, as in Figure 4. This time, however, population variances σ_X^2 and σ_Y^2 are unknown to us, and we use their estimates.

Two important cases need to be considered here. In one case, there exists an exact and simple solution based on T-distribution. The other case suddenly appears to be a famous Behrens-Fisher problem, where no exact solution exists, and only approximations are available.

3.5.1 Case 1. Equal Variances

Suppose there are reasons to assume that the two populations have equal variances,

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2$$

For example, two sets of data are collected with the same measurement device, thus, measurements have different means but the same precision. In this case, there is only one variance σ^2 to estimate instead of two. We should use both samples \mathbf{X} and \mathbf{Y} to estimate their common variance. This estimator of σ^2 is called a **pooled sample variance**, and it is computed as

$$\begin{aligned} s_p^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n + m - 2} \\ &= \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n + m - 2} \end{aligned} \quad (11)$$

Substituting this variance estimator in (6) for σ_X^2 and σ_Y^2 , we get the following confidence interval.

Confidence interval for the difference of means; equal, unknown standard deviations

$$\bar{X} - \bar{Y} \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

where s_p is the pooled standard deviation, a root of the pooled variance in (11) and $t_{\alpha/2}$ is a critical value from T-distribution with $(n + m - 2)$ degrees of freedom.

3.5.2 Case 2. Unequal Variances

The most difficult case is when both variances are unknown and unequal. Confidence estimation of $\mu_X - \mu_Y$ in this case is known as the *Behrens-Fisher* problem. Certainly, we can replace unknown variances σ_X^2 , σ_Y^2 by their estimates s_X^2 , s_Y^2 and form a T-ratio

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

However, it won't have a T-distribution.

An approximate solution was proposed in the 1940s by Franklin E. Satterthwaite. Satterthwaite used the method of moments to estimate degrees of freedom ν of a T-distribution that is "closest" to this T-ratio. This number depends on unknown variances. Estimating them by sample variances, he obtained the formula that is now known as *Satterthwaite approximation*

$$\nu = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m} \right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}} \quad (12)$$

This number of degrees of freedom often appears non-integer. There are T-distributions with non-integer ν . To use Table A5, just take the closest ν that is given in that table.

Formula (12) is widely used for t-intervals and t-tests.

Confidence interval for the difference of means; unequal, unknown standard deviation

$$\bar{X} - \bar{Y} \pm t_{\alpha/2} \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$$

where $t_{\alpha/2}$ is a critical value from T-distribution with ν degrees of freedom given by formula (12).

Check the examples of 9.3.4 from book.

Definition 7

A **type I error** occurs when we reject the true null hypothesis.

A **type II error** occurs when we accept the false null hypothesis.

A type I error is often considered *more dangerous and undesired* than a type II error. For this reason, we shall design tests that bound the probability of type I error by a preassigned small number α . Under this condition, we may want to minimize the probability of type II error.

Definition 8

Probability of a type I error is the **significance level** of a test,

$$\alpha = P \{ \text{reject } H_0 \mid H_0 \text{ is true} \}$$

Probability of rejecting a false hypothesis is the **power of the test**,

$$p(\theta) = P \{ \text{reject } H_0 \mid H_A \text{ is true} \}$$

It is usually a function of the parameter θ because the alternative hypothesis includes a set of parameter values. Also, the power is the probability to avoid a Type II error.

4 Hypothesis Testing

4.1 Hypothesis and Alternative

To begin, we need to state exactly what we are testing. These are hypothesis and alternative.

NOTATION

H_0 = hypothesis (the null hypothesis)

H_A = alternative (the alternative hypothesis)

H_0 and H_A are simply two mutually exclusive statements. Each test results either in acceptance of H_0 or its rejection in favor of H_A .

Definition 6

Alternative of the type $H_A: \mu \neq \mu_0$ covering regions on both sides of the hypothesis ($H_0: \mu = \mu_0$) is a **two-sided alternative**.

Alternative $H_A: \mu < \mu_0$ covering the region to the left of H_0 is **one-sided, left-tail**.

Alternative $H_A: \mu > \mu_0$ covering the region to the right of H_0 is **one-sided, right-tail**.

Check the examples 9.22, 9.23, 9.24.

4.2 Type I and Type II errors: Level of Significance

When testing hypotheses, we realize that all we see is a random sample. Therefore, with all the best statistics skills, our decision to accept or to reject H_0 may still be wrong

	Result of the test	
	Reject H_0	Accept H_0
H_0 is true	Type I error	correct
H_0 is false	correct	Type II error

4.3 Level α Tests: General Approach

A standard algorithm for a level α test of a hypothesis H_0 against an alternative H_A consists of 3 steps.

Step 1. Test Statistic

Testing hypothesis is based on a **test statistic** T , a quantity computed from the data that has some known, tabulated distribution F_0 if the hypothesis H_0 is true.

Step 2. Acceptance Region and Rejection Region

Next, we consider the **null distribution** F_0 . This is the distribution of test statistic T when the hypothesis H_0 is true. If it has a density f_0 , then the whole area under the density curve is 1, and we can always find a portion of it whose area is α , as shown in Figure 5. It is called **rejection region** (\mathcal{R}).

The remaining part, the complement of the rejection region, is called **acceptance region** ($\mathcal{A} = \bar{\mathcal{R}}$). By the complement rule, its area is $(1 - \alpha)$.

Step 3. Result and Its Interpretation

Accept the hypothesis H_0 if the test statistic T belongs to the acceptance region. Reject H_0 in favor of the alternative H_A if T belongs to the rejection region.

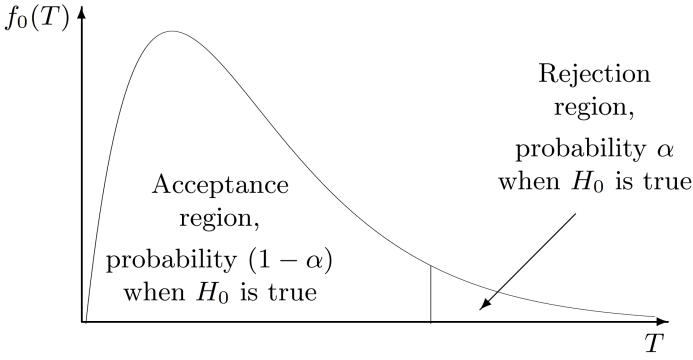


Figure 5

4.4 Rejection Regions and Power

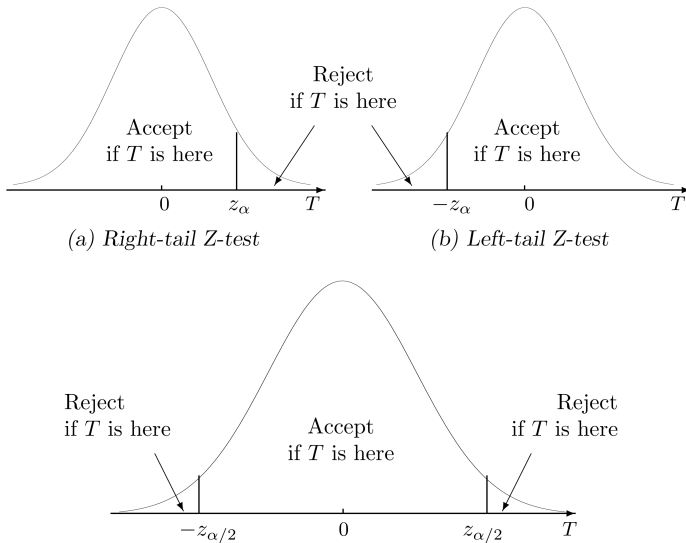


Figure 6: Acceptance and rejection regions for a Z-test with (a) a one-sided right-tail alternative; (b) a one-sided left-tail alternative; (c) a two-sided alternative.

- For a **right-tail alternative**, the rejection region \mathcal{R} should consist of large values of T . Choose \mathcal{R} on the right, \mathcal{A} on the left (Figure 6a).
- For a **left-tail alternative**, the rejection region \mathcal{R} should consist of small values of T . Choose \mathcal{R} on the left, \mathcal{A} on the right (Figure 6b).
- For a **two-sided alternative**, the rejection region \mathcal{R} should consist of very small and very large values of T . Let \mathcal{R} consist of two extreme regions, while \mathcal{A} covers the middle (Figure 6c).

4.5 Z-Test: Standard Normal Null Distribution

An important case, in terms of a large number of applications, is when the null distribution of the test statistic is *Standard Normal*.

The test in this case is called a **Z-test**, and the test statistic is usually denoted by Z .

- (a) A level α test with a **right-tail alternative** should
- $$\begin{cases} \text{reject } H_0 & \text{if } Z \geq z_\alpha \\ \text{accepts } H_0 & \text{if } Z < z_\alpha \end{cases} \quad (14)$$

The rejection region in this case consists of large values of Z only,

$$\mathcal{R} = [z_\alpha, +\infty), \quad \mathcal{A} = (-\infty, z_\alpha)$$

Under the null hypothesis, Z belongs to \mathcal{A} and we reject the null hypothesis with probability

$$P\{T \geq z_\alpha \mid H_0\} = 1 - \Phi(z_\alpha) = \alpha$$

making the probability of false rejection (type I error) equal α .

For example, we use this acceptance region to test the population mean,

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu > \mu_0$$

- (b) With a **left-tail alternative** should

$$\begin{cases} \text{reject } H_0 & \text{if } Z \leq -z_\alpha \\ \text{accepts } H_0 & \text{if } Z > -z_\alpha \end{cases} \quad (15)$$

The rejection region consists of small values of Z only,

$$\mathcal{R} = (-\infty, -z_\alpha], \quad \mathcal{A} = (-z_\alpha, +\infty)$$

Similarly, $P\{Z \in \mathcal{R}\} = \alpha$ under H_0 ; thus, the probability of type I error equals α . For example, this is how we should test

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu < \mu_0$$

- (c) With a **two-sided alternative**, we

$$\begin{cases} \text{reject } H_0 & \text{if } |Z| \geq z_{\alpha/2} \\ \text{accepts } H_0 & \text{if } |Z| < z_{\alpha/2} \end{cases} \quad (16)$$

The rejection region consists of very small and very large values of Z ,

$$\mathcal{R} = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, +\infty), \quad \mathcal{A} = (-z_{\alpha/2}, z_{\alpha/2})$$

Again, the probability of type I error equals α . For example, we use this test for

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_A : \mu \neq \mu_0$$

This is easy to remember:

- for a two-sided test, divide α by two and use $z_{\alpha/2}$;
- for a one-sided test, use z_α keeping in mind that the rejection region consists of just one piece.

Now consider testing a hypothesis about a population parameter θ . Suppose that its estimator $\hat{\theta}$ has Normal distribution, at least approximately, and we know $\mathbf{E}(\hat{\theta})$ and $\text{Var}(\hat{\theta})$ if the hypothesis is true.

Then the test statistic

$$Z = \frac{\hat{\theta} - \mathbf{E}(\hat{\theta})}{\sqrt{\text{Var}(\hat{\theta})}} \quad (17)$$

has Standard Normal distribution, and we can use (14), (15), and (16) to construct acceptance and rejection regions for a level α test. We call Z a **Z-statistic**.

4.6 Z-Test for Means and Proportions

As we already know,

- sample means have Normal distribution when the distribution of data is Normal;
- sample means have approximately Normal distribution when they are computed from large samples (the distribution of data can be arbitrary);

- sample proportions have approximately Normal distribution when they are computed from large samples;
- this extends to differences between means and between proportions

For all these cases, we can use a Z-statistic (17) and rejection regions (14)-(16) to design powerful level α tests.

Z-tests are summarized in Table 1.

Null hypothesis	Parameter, estimator	If H_0 is true:		Test statistic
		$\mathbf{E}(\hat{\theta})$	$\text{Var}(\hat{\theta})$	
H_0	$\theta, \hat{\theta}$			$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{\text{Var}(\hat{\theta})}}$
One-sample Z-tests for means and proportions, based on a sample of size n				
$\mu = \mu_0$	μ, \bar{X}	μ_0	$\frac{\sigma^2}{n}$	$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
$p = p_0$	p, \hat{p}	p_0	$\frac{p_0(1 - p_0)}{n}$	$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Two-sample Z-tests comparing means and proportions of two populations, based on independent samples of size n and m				
$\mu_X - \mu_Y = D$	$\mu_X - \mu_Y, \bar{X} - \bar{Y}$	D	$\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$	$\frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$
$p_1 - p_2 = D$	$p_1 - p_2, \hat{p}_1 - \hat{p}_2$	D	$\frac{p_1(1 - p_1)}{n} + \frac{p_2(1 - p_2)}{m}$	$\frac{\hat{p}_1 - \hat{p}_2 - D}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}}$
$p_1 = p_2$	$p_1 - p_2, \hat{p}_1 - \hat{p}_2$	0	$p(1 - p) \left(\frac{1}{n} + \frac{1}{m} \right),$ where $p = p_1 = p_2$	$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n} + \frac{1}{m} \right)}}$ where $\hat{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n + m}$

Table 1: Summary of Z-tests

Example 10 (Z-test about a population mean)

The number of concurrent users for some internet service provider has always averaged 5000 with a standard deviation of 800. After an equipment upgrade, the average number of users at 100 randomly selected moments of time is 5200. Does it indicate, at a 5% level of significance, that the mean number of concurrent users has increased? Assume that the standard deviation of the number of concurrent users has not changed.

Solution: We test the null hypothesis $H_0 : \mu = 5000$ against a one-sided right-tail alternative $H_A : \mu > 5000$, because we are only interested to know if the mean number of users μ has increased.

Step 1: Test statistic. We are given $\sigma = 800$, $n = 100$, $\alpha = 0.05$, $\mu_0 = 5000$, and from sample, $\bar{X} = 5200$. The test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{5200 - 5000}{800/\sqrt{100}} = 2.5$$

Step 2: Acceptance and rejection regions. The critical value is

$$z_\alpha = z_{0.05} = 1.645$$

(don't divide α by 2 because it is a one-sided test). With the right-tail alternative, we

$$\begin{cases} \text{reject } H_0 & \text{if } Z \geq 1.645 \\ \text{accept } H_0 & \text{if } Z < 1.645 \end{cases}$$

Step 3: Result. Our test statistic $Z = 2.5$ belongs to the rejection region; therefore, we reject the null hypothesis. The data (5200 users, on the average, at 100 times) provided sufficient evidence in favor of the alternative hypothesis that the mean number of users has increased.

Example 11 (Two-sample Z-test of proportions)

A quality inspector finds 10 defective parts in a sample of 500 parts received from manufacturer A. Out of 400 parts from manufacturer B, she finds 12 defective ones. A computer-making company uses these parts in their computers and claims that the quality of parts produced by A and B is the same. At the 5% level of significance, do we have enough evidence to disprove this claim?

Solution: We test $H_0 : p_A = p_B$, or $H_0 : p_A - p_B = 0$, against $H_A : p_A \neq p_B$. This is a two-sided test because no direction of the alternative has been indicated. We only need to verify whether or not the proportions of defective parts are equal for manufacturers A and B.

Step 1: Test statistic. We are given

$\hat{p}_A = 10/500 = 0.02$ from sample size $n = 500$; $\hat{p}_B = 12/400 = 0.03$ from sample size $m = 400$. The tested value is $D = 0$.

As we know, for these Bernoulli data, the variance depends on the unknown parameters p_A and p_B which are estimated by the sample proportions \hat{p}_A and \hat{p}_B .

The test statistic then equals

$$\begin{aligned} Z &= \frac{\hat{p}_A - \hat{p}_B - D}{\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n} + \frac{\hat{p}_B(1-\hat{p}_B)}{m}}} \\ &= \frac{0.02 - 0.03}{\sqrt{\frac{(0.02)(0.98)}{500} + \frac{(0.03)(0.97)}{400}}} = -0.945 \end{aligned}$$

Step 2: Acceptance and rejection regions. This is a two-sided test; thus we divide α by 2, find $z_{0.05/2} = z_{0.025} = 1.96$, and

$$\begin{cases} \text{reject } H_0 & \text{if } |Z| \geq 1.96 \\ \text{accept } H_0 & \text{if } |Z| < 1.96 \end{cases}$$

Step 3: Result. The evidence against H_0 is insufficient because $|Z| < 1.96$. Although sample proportions of defective parts are unequal, the difference between them appears too small to claim that population proportions are different.

4.7 Pooled Sample Proportion

The test in Example 11 can be conducted differently and perhaps, more efficiently. Indeed, we standardize the estimator $\hat{\theta} = \hat{p}_A - \hat{p}_B$ using its expectation $\mathbf{E}(\hat{\theta})$ and variance $\text{Var}(\hat{\theta})$ under the null distribution, i.e., when H_0 is true. However, under the null hypothesis $p_A = p_B$. Then, when we standardize $(\hat{p}_A - \hat{p}_B)$, instead of estimating two proportions in the denominator, we only need to estimate one.

First, we estimate the common population proportion by the overall proportion of defective parts,

$$\begin{aligned} \hat{p}(\text{pooled}) &= \frac{\text{number of defective parts}}{\text{total number of parts}} \\ &= \frac{n\hat{p}_A + m\hat{p}_B}{n + m} \end{aligned}$$

Then we estimate the common variance as

$$\begin{aligned} \widehat{\text{Var}}(\hat{p}_A - \hat{p}_B) &= \frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{p}(1-\hat{p})}{m} \\ &= \hat{p}(1-\hat{p}) \left(\frac{1}{n} + \frac{1}{m} \right) \end{aligned}$$

and use it for the Z-statistic,

$$Z = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n} + \frac{1}{m} \right)}}$$

Example 12 (Example 11, continued)

Here the pooled proportion equals

$$\hat{p} = \frac{10 + 12}{500 + 400} = 0.0244$$

so that

$$Z = \frac{0.02 - 0.03}{\sqrt{(0.0244)(0.9756) \left(\frac{1}{500} + \frac{1}{400} \right)}} = -0.966$$

This does not affect our result. We obtained a different value of Z-statistic, but it also belongs to the acceptance region. We still don't have a significant evidence against the equality of two population proportions.

4.8 Unknown σ : T-tests

As we decided in Section 3, when we don't know the population standard deviation, we estimate it. The resulting T-statistic has the form

$$t = \frac{\hat{\theta} - \mathbf{E}(\hat{\theta})}{s(\hat{\theta})} = \frac{\hat{\theta} - \mathbf{E}(\hat{\theta})}{\sqrt{\widehat{\text{Var}}(\hat{\theta})}}$$

In the case when the distribution of $\hat{\theta}$ is Normal, the test is based on Student's T-distribution with acceptance and

rejection regions according to the direction of H_A :

(a) For a **right-tail alternative**,

$$\begin{cases} \text{reject } H_0 & \text{if } t \geq t_\alpha \\ \text{accepts } H_0 & \text{if } t < t_\alpha \end{cases} \quad (18)$$

(b) For a **left-tail alternative**,

$$\begin{cases} \text{reject } H_0 & \text{if } t \leq -t_\alpha \\ \text{accepts } H_0 & \text{if } t > -t_\alpha \end{cases} \quad (19)$$

(c) For a **two-sided alternative**,

$$\begin{cases} \text{reject } H_0 & \text{if } |t| \geq t_{\alpha/2} \\ \text{accepts } H_0 & \text{if } |t| < t_{\alpha/2} \end{cases} \quad (20)$$

Quantiles t_α and $t_{\alpha/2}$ are given in Table A5 (in textbook). As in Section 3.4, the number of degrees of freedom depends on the problem and the sample size, see Table 2 and formula (10)

Hypothesis H_0	Conditions	Test statistic t	Degrees of freedom
$\mu = \mu_0$	Sample size n ; unknown σ	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$n - 1$
$\mu_X - \mu_Y = D$	Sample sizes n, m ; unknown but equal standard deviations, $\sigma_X = \sigma_Y$	$t = \frac{\bar{X} - \bar{Y} - D}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$	$n + m - 2$
$\mu_X - \mu_Y = D$	Sample sizes n, m ; unknown, unequal standard deviations, $\sigma_X \neq \sigma_Y$	$t = \frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$	Satterthwaite approximation, formula (9.12)

Table 2: Summary of T-tests

As in Section 3.4, the **pooled sample variance**

$$\begin{aligned} s_p^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n + m - 2} \\ &= \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n + m - 2} \end{aligned}$$

is computed for the case of equal unknown variances. When variances are not equal, degrees of freedom are computed by

Satterthwaite approximation (12).

Check examples 9.28, 9.29, 9.30 in the textbook.

When the distribution of $\hat{\theta}$ is not Normal, the Student's T-distribution cannot be used. The distribution of a T-statistic and all its probabilities will be different from Student's T, and as a result, our test may not have the desired significance level.

4.9 Duality: Two-sided Tests and Two-sided Confidence Intervals

An interesting fact can be discovered if we look into our derivation of tests and confidence intervals. It turns out that we can conduct two-sided tests using nothing but the confidence intervals!

(Eq. 21)

A level α Z-test of $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$ accepts the null hypothesis

if and only if

a symmetric $(1 - \alpha)100\%$ confidence Z-interval for θ contains θ_0 .

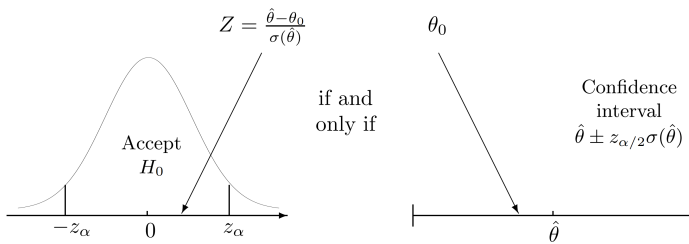


Figure 7

In fact, any two-sided test can be conducted this way. Accept $H_0 : \theta = \theta_0$ whenever a $(1 - \alpha)100\%$ confidence interval for θ covers θ_0 . Under θ_0 , this test will accept the null hypothesis as often as the interval will cover θ_0 , i.e., with probability $(1 - \alpha)$. Thus, we have a level α test.

Rule (21) applies only when

- we are testing against a two-sided alternative (notice that our confidence intervals are two-sided too);
- significance level α of the test matches confidence level $(1 - \alpha)$ of the confidence interval. For example, a two-sided 3% level test can be conducted using a 97% confidence interval.

Example 13

A sample of 6 measurements

2.5, 7.4, 8.0, 4.5, 7.4, 9.2

is collected from a Normal distribution with mean μ and standard deviation $\sigma = 2.2$. Test whether $\mu = 6$ against a two-sided alternative $H_A : \mu \neq 6$ at the 5% level of significance.

Solution: Solving Example 4, we have already constructed a 95% confidence interval for μ ,

[4.74, 8.26]

The value of $\mu_0 = 6$ belongs to it; therefore, at the 5% level, the null hypothesis is accepted.

Example 14

Use data in Example 13 to test whether $\mu = 7$.

Solution: The interval [4.74, 8.26] contains $\mu_0 = 7$ too; therefore, the hypothesis $H_0 : \mu = 7$ is accepted as well.

Check examples 9.33, 9.34, 9.35 in the textbook.

Similarly, for the case of unknown variance(s).

A level α T-test of $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$ accepts the null hypothesis

if and only if

a symmetric $(1 - \alpha)100\%$ confidence T-interval for θ contains θ_0 .