MET CS 555 - Data Analysis and Visualization

Module-2: Confidence Interval, Hypotheses Tests and z-test
Lecture - 3

Kia Teymourian

Boston University

Table of contents

Statistical Inference

Statistical inference provides methods for drawing conclusions about a population from sample data.

The goals of statistical inference are:

- Draw conclusion about a population based on sample data
- Provide a statement, expressed in the language of probability, of how much confidence to be placed in the conclusion

Two of the most common types of statistical inference include:

- Confidence Intervals for estimating the value of a population parameter
- Tests of Significance or Hypothesis Tests which assess the evidence for a scientific claim.

Confidence Intervals

Example - Call Center Calculating Confidence Interval

`Suppose a company would like to estimate the average amount of time callers are on hold before they reach a customer service representative.`

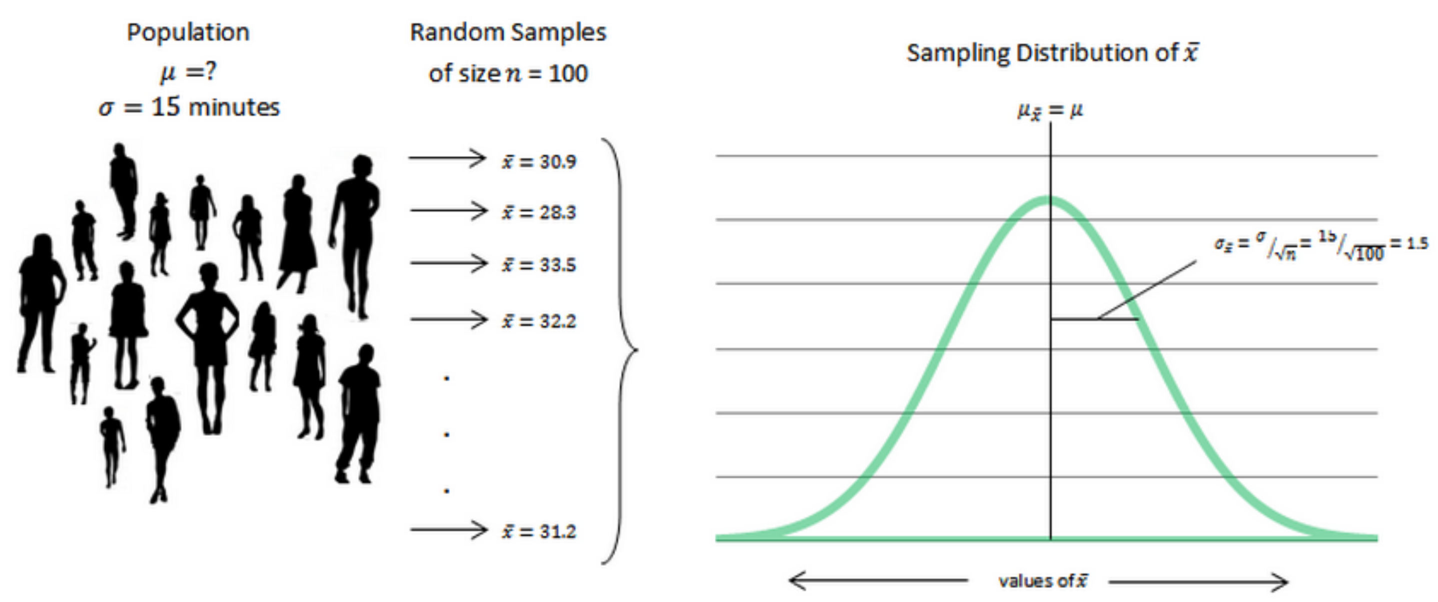
Executives from the company plan to randomly sample 100 calls and use the sample mean $\bar{x}$, to estimate population mean μ (the average of all callers wait times).

Assume that $\sigma = 15$ minutes and
sample mean of the wait times is $\bar{x} = 30.9$ minutes.
The parameter of interest is the population mean, μ.

Executives also want to compute a confidence interval to assess the accuracy of the point estimate (the sample mean).

Example calculating confidence interval

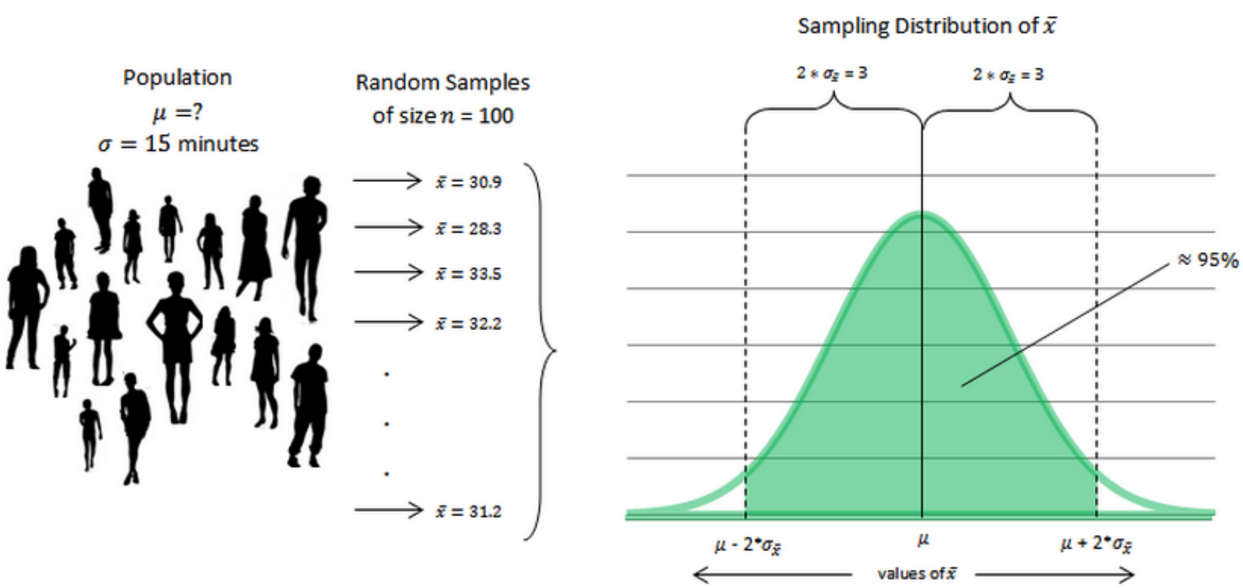
Since $n \geq 30$ (here, n=100), we know that the sample mean is approximately normally distributed with a mean $\mu_{\bar{x}} = \mu$ and a standard deviation of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{100}} = 1.5$ minutes.



An example calculating confidence interval

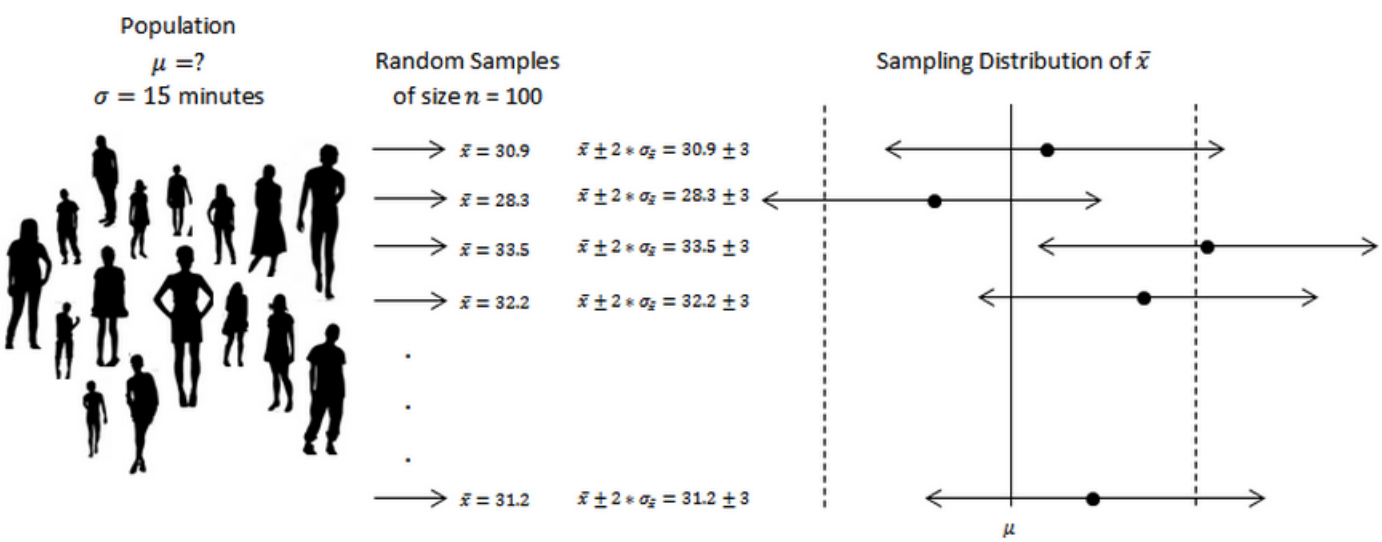
In a normal distribution, 95% of the sample means are within 2 SDs of the population mean ($z = 1.96$ corresponds to an area of 95% under the curve). The sample mean, $\bar{x}$, and the population mean, μ, are within three minutes (2 SDs) of each other 95% of the time.

If our estimate of the population mean is between $\bar{x} \pm 3$
$\bar{x} - 3 = 30.9 - 3 = 27.9$ and
$\bar{x} + 3 = 30.9 + 3 = 33.9$ then we will be right 95% of the time.

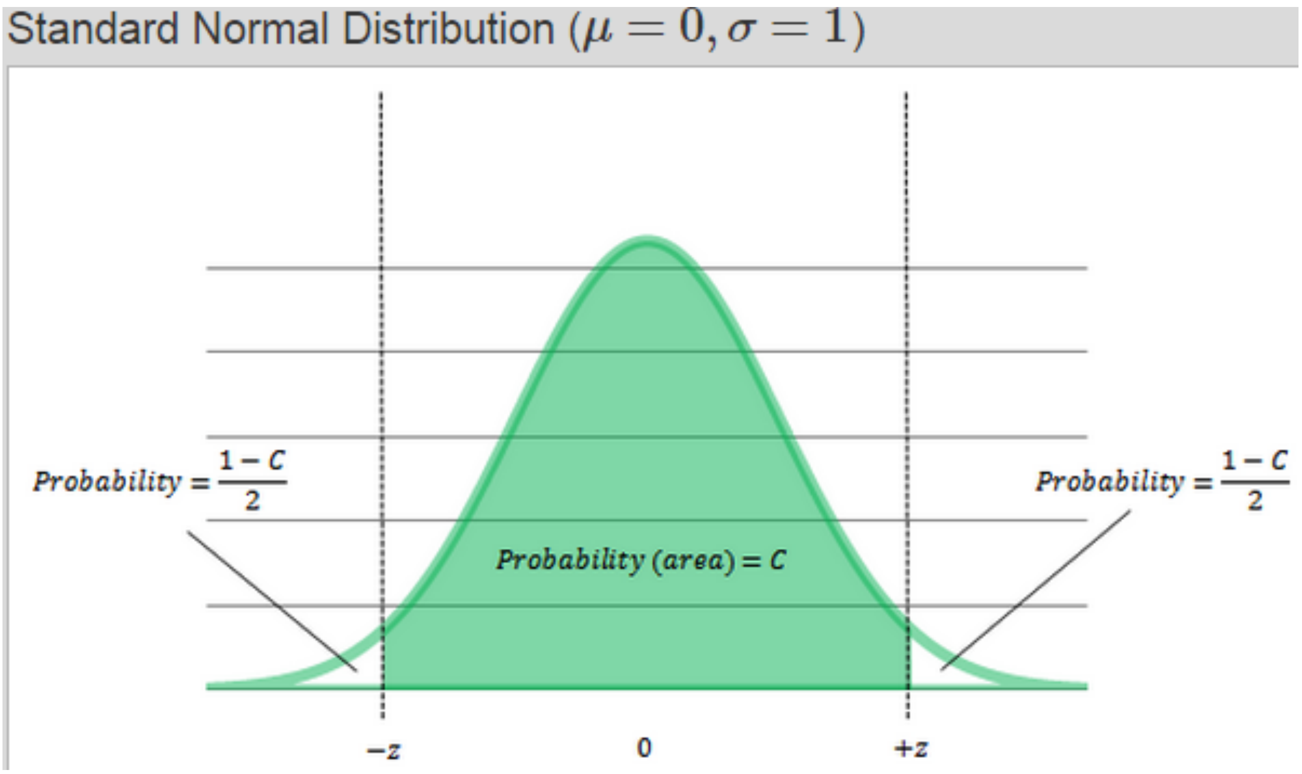


An example calculating confidence interval

- We know that **95% of the sample means will be within 2 SDs of the population mean.**
- The sample mean 2 times the SD ($\bar{x} \pm 2 \cdot \sigma_{\bar{x}}$) is called the 95% confidence interval for the population mean.
- The two-way arrows in the figure represent the confidence intervals from each random sample. The center point is the sample mean and the end of the arrows show the edges of each interval.



Confidence interval



Confidence interval

- Confidence intervals are of the form: **estimate ± margin of error**
- To calculate a confidence interval with a confidence level of C for the population mean, μ, we use the following formula:

$$\bar{x} \pm z.\frac{\sigma}{\sqrt{n}}$$

  z is the appropriate critical value corresponding to the confidence level.
- Typical values of C with their associated values of z are:
- Generally, 95% confidence intervals are standard. Other common intervals are 90% and 99%.

Confidence Level, C	90%	95%	99%
Critical Value, z	1.645	1.960	2.576

Exercise 1 - Call Center Data Waiting Time

`Suppose the executives would like to estimate the average amount of time callers are on hold before they reach a customer service representative by computing a 99% confidence interval. The population standard deviation of wait times,` σ`, is 15 minutes.`

The sample mean of the wait times from the 100 random calls in the sample is 30.9 minutes.

Exercise 1 - Call Center Data Waiting Time

`Suppose the executives would like to estimate the average amount of time callers are on hold before they reach a customer service representative by computing a 99% confidence interval. The population standard deviation of wait times,` σ`, is 15 minutes.`

The sample mean of the wait times from the 100 random calls in the sample is 30.9 minutes.

To calculate a confidence interval with a confidence level of C for the population mean, μ, we use the formula: $\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}}$

The 99% confidence interval is calculated as follows:

$$\bar{x} \pm z \times \frac{\sigma}{\sqrt{n}} =$$

$$30.9 \pm 2.576 \times \frac{15}{\sqrt{100}} =$$

$$30.9 \pm 2.576 \times 1.5 = \{27.036, 34.764\}$$

The Sample size and the margin of error

In the case of the confidence interval for the population mean, the margin of error is equal to

$$m = z \times \frac{\sigma}{\sqrt{n}}$$

We like our confidence intervals to be narrow (small in width) with a high amount of confidence. The precision (width) of the confidence interval decreases as:

- z **gets smaller.** Smaller values of z are associated decreased confidence levels. There is a trade off between the confidence level and the margin of error.
- σ **gets smaller.** When the variability of the individual observations is reduced, it becomes easier to estimate the population mean with higher precision.
- n **gets larger.** Increasing the sample size n reduces the margin of error for a fixed confidence level C.

The Sample size and the margin of error

In the case of the confidence interval for the population mean, the margin of error is equal to

$$m = z \times \frac{\sigma}{\sqrt{n}}$$

To obtain a specific **margin of error, m**, for a desired confidence level, C, the number of observations needed is:

$$n = \left(\frac{z \times \sigma}{m}\right)^2$$

Exercise 2 - How many callers must they sample?

`Suppose the executives would like to estimate the average amount of time callers are on hold before they reach a customer service representative by computing a 95% confidence interval.`

As before, the population standard deviation of wait times, σ, is 15 minutes. The executives would like the half width of the confidence interval to be 1.0 minutes (that is, they'd like the margin of error to be 1.0).

How many callers must they sample to get their 95% confidence interval to have the desired width?

Exercise 2 - How many callers must they sample?

How many callers must they sample to get their 95% confidence interval to have the desired width?

To obtain a specific margin of error, m, for a desired confidence level, C, we use the formula to determine the number of observations needed:

$$n = \left(\frac{z \times \sigma}{m}\right)^2 = \left(\frac{1.96 \times 15}{1}\right)^2 = 29.4^2 = 864.36$$

Note: In calculations involving sample size we always round up (instead of down) to achieve the requested margin of error. In our example we need to have samples of 865 callers

Confidence Level, C	90%	95%	99%
Critical Value, z	1.645	1.960	2.576

Hypotheses Tests

Set up the hypotheses

- We formally call the claim that we are **hoping to disprove the null hypothesis**. The null hypothesis is generally denoted as H_0.
- The test of significance is designed to assess the strength of the evidence against the null hypothesis.
- Usually the null hypothesis is a statement of no effect or no difference.
- The conclusion we'd like to make is captured in the alternative hypothesis. The alternative hypothesis is generally denoted H_a or H_1. This is written as the opposite of the null hypothesis and generally suggests that there is an effect or difference.

Set up the hypotheses

- Hypotheses always refer to some population or distribution, **not to a particular sample outcome.** Thus we state H_0 and H_1 in terms of population parameters and not in terms of sample statistics.

- The alternative hypothesis states that a parameter differs from its null value in a specific direction (an one-sided alternative) or in either direction (a two-sided alternative).

Set up the hypotheses - one-sided alternative

Example: `A gym is interested in whether or not a 6-week weight loss training program they launched has been successful in helping their clients lose weight. To assess this, they took a sample of 30 participants. State the null and alternative hypotheses.`

Let's denote the mean weight change, μ, as the population parameter of interest. The gym is specifically interested in whether program participants on average lost weight. Thus they would be interested in an **one-sided alternative** where they seek to claim that program participants lose weight on average.

A shorthand notation to capture these in an easy to read form:
$H_0 : \mu = 0$ (there is **no effect on weight change** of participants)
$H_1 : \mu < 0$ (**participants lose weight** on average)

Set up the hypotheses - two-sided alternative

Example: *County officials are interested in measuring a particular chemical in water sources in the county. High levels of this chemical are harmful as are low levels. Either indicate a need for intervention.*

Normal levels of this chemical are **15 parts per million (ppm)**. County officials will start an intervention program if the chemical mean level in the water sources is different than 15 ppm.

State the null and alternative hypotheses.

Let's denote the mean level of the chemical, , as the population parameter of interest. The alternative hypothesis is that the levels are different than 15 ppm. It is a two-sided alternative.

$H_0 : \mu = 15$ (the mean level of the chemical is normal)
$H_1 : \mu \neq 15$ (the mean level of the chemical is abnormal)

Test Statistics

Significance tests use data from a sample to evaluate the evidence against the null hypothesis.

Significance tests generally compare the value of the population parameter in the null hypothesis to the value of the estimate of the population parameter from the sample.

The **test statistic** is generally a measurement of how far the point estimate is from the expected value under the null.

The test statistic for hypotheses about the mean μ is the z statistic:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

z-test

n is Large ($n \geq 30$) and σ is unknown

A z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution.

Because of the central limit theorem, many test statistics are approximately normally distributed for large samples.

Many statistical tests can be conveniently performed as approximate Z-tests if

- the sample size is large $n \geq 30$
- the population variance is known or unknown

Confidence interval is $\bar{x} \pm z_{CL} \times \frac{\sigma}{\sqrt{n}}$ or $\bar{x} \pm z_{CL} \times \frac{S}{\sqrt{n}}$

Note: If the population variance is unknown (and therefore has to be estimated from the sample itself) and the sample size is not large ($n < 30$), the Student's t-test may be more appropriate.

p-value of the test

One of the ways to measure how far the point estimate is from the expected value of the population parameter under the null hypothesis is to **calculate the test statistic's associated p-value.**

The **p-value of the test is the probability**, computed assuming H_0 is true, that the test statistic would take a value as extreme or more extreme than that actually observed.

The p-value is a measure of the strength of the evidence against H_0 .

- **A small p-value** suggests that the observed result was **unlikely to occur if the null hypothesis is in fact true.**
- **Large p-values**, on the other hand, **do not give evidence against the null hypothesis**.

Meaning of Large p-value

- If the resulting p-value is large (meaning that the test statistic is small), then the sample did not give evidence against the null hypothesis.
- A large p-value only means that the data are inconsistent with the null hypothesis, not that we have clear evidence that the null hypothesis is untrue. Statistical tests are set up to look for evidence against the null, not to prove that the null hypothesis is untrue.

Calculating p-value of the test

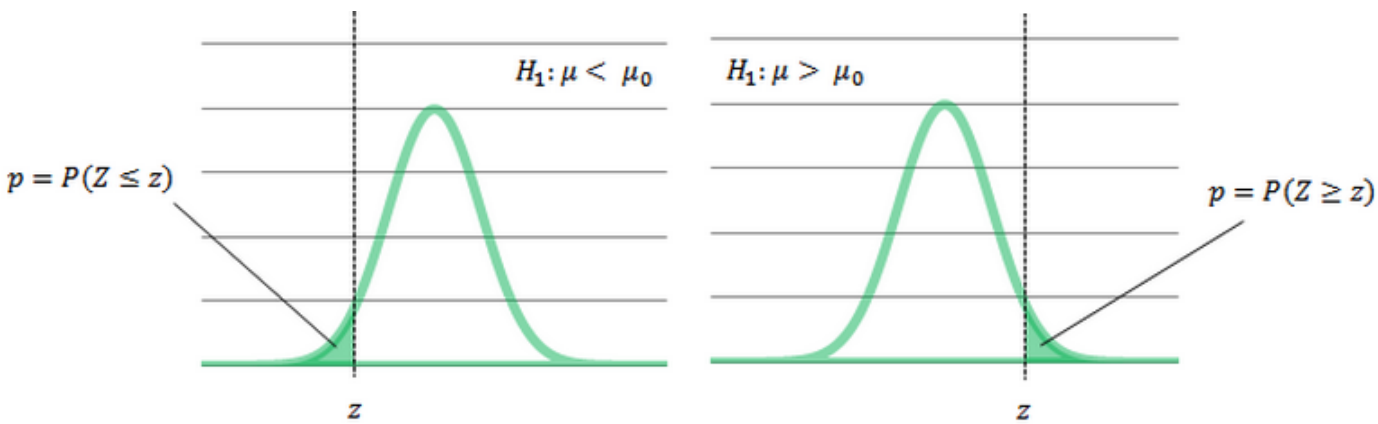
The p-value for the z statistic is calculated using the standard normal distribution.

Depending on the alternative hypothesis, the p-value is calculated as follows:

An one-sided alternative hypothesis.

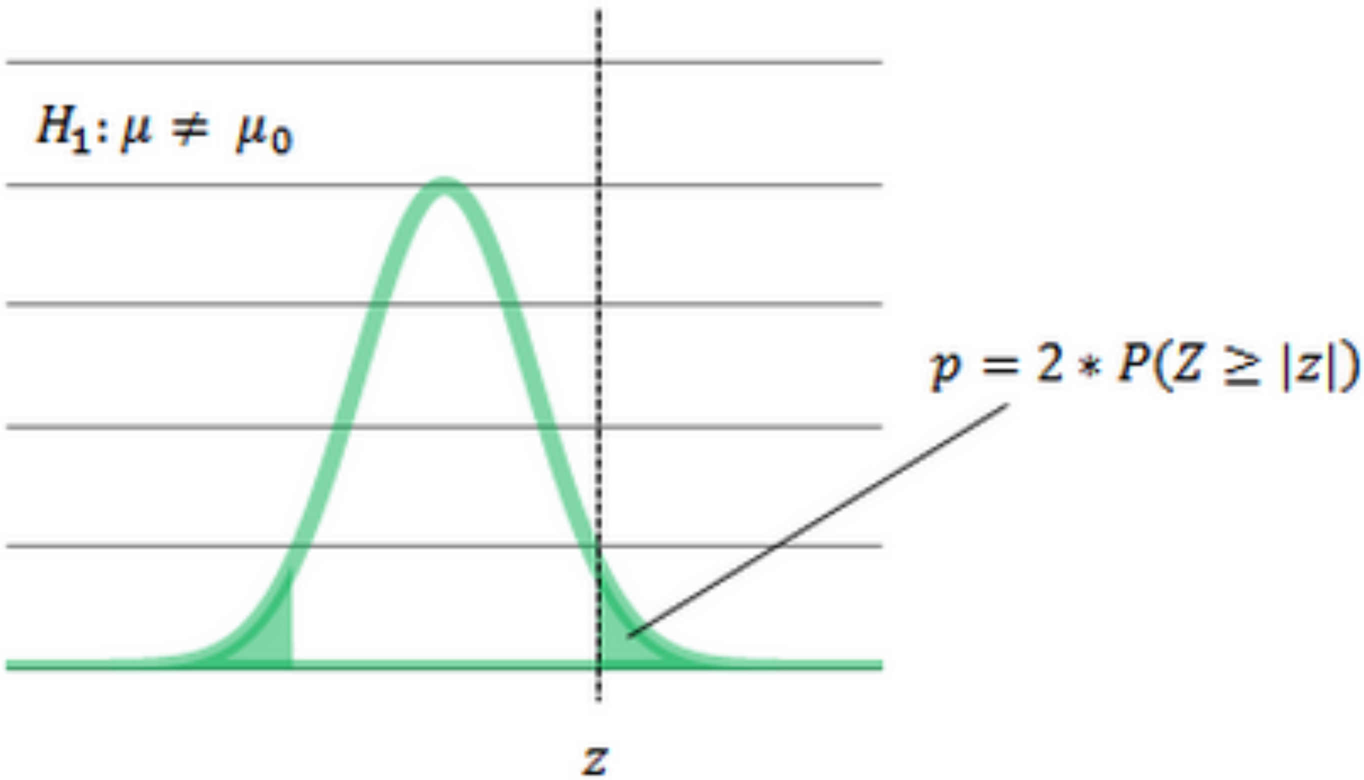
Mean is less than a specified value.

Mean is greater than a specified value.



Calculating p-value of the test - Two Sided

A two-sided alternative hypothesis stating that the population mean is different from a specified value.



Example - Calculating P-value for an one-sided test

`A gym is interested in whether a 6-week weight loss training program they launched has been successful in helping their clients lose weight. To assess this, they took a sample of 30 participants.`

They are interested in testing the following hypotheses:

- **$H_0 : \mu = 0$ (there is no effect on weight change of program participants)**
- **$H_1 : \mu < 0$ (program participants lose weight on average)**

Suppose we know that for the general population, the standard deviation of changes in weights over a six-week interval is 6 pounds.

The sample mean of the change in weight for the 30 participants in the sample was -2.98 pounds.

Calculate the value of the test statistic and the associated p-value.

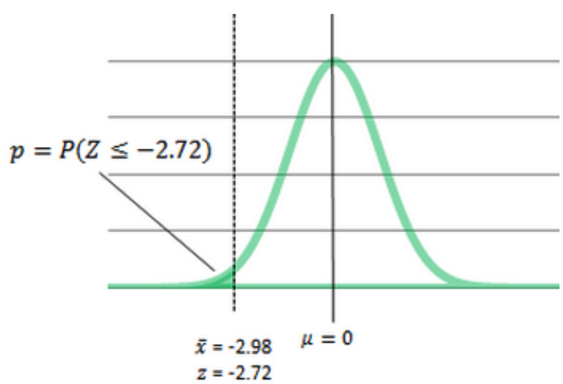
Calculating the value of the test statistic and p-value

$\sigma = 6$, $n = 30$, $\bar{x} = -2.98$, $\mu_{H_0} = 0$

The p-value is the probability that the test statistic is -2.72 or more extreme, which is the probability that $Z \leq -2.72$. Using the standard normal table, we can calculate: $P = P(Z \leq -2.72) = 0.0033$

$$z = \frac{\bar{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}} = \frac{-2.98 - 0}{\frac{6}{\sqrt{30}}} \approx \frac{-2.98}{1.0954} \approx -2.72$$

This is a small p-value. It appears that the sample mean (= -2.98) is highly unlikely to have occurred if the true population mean $\mu = 0$. **Thus we have strong evidence against the null hypothesis.**



Example: Chemical in Water - Two-Sided

Normal levels of this chemical are 15 parts per million (ppm).

Samples from 50 water sources throughout the county are taken and the levels of this chemical are measured.

They are interested in testing the following hypotheses:

$H_0 : \mu = 15$ (the mean level of the chemical is normal)

$H_1 : \mu \neq 15$ (the mean level of the chemical is abnormal)

Suppose we know that the population standard deviation is 6.2.

The sample mean from the 50 samples was 16.4 ppm.

Calculate the value of the test statistic and the associated p-value.

Example: Chemical in Water - Two-Sided

Givens are:

$\bar{x} = 16.4$, $\mu_{H_0} = 15$, $\sigma = 6.2$, $n = 50$

Now, we can just plug these values in to calculate the value of z.

$$z = \frac{\bar{x} - \mu_{H_0}}{\frac{\sigma}{\sqrt{n}}} = \frac{16.4 - 15}{\frac{6.2}{\sqrt{50}}} \approx \frac{1.4}{0.8768} \approx 1.60$$

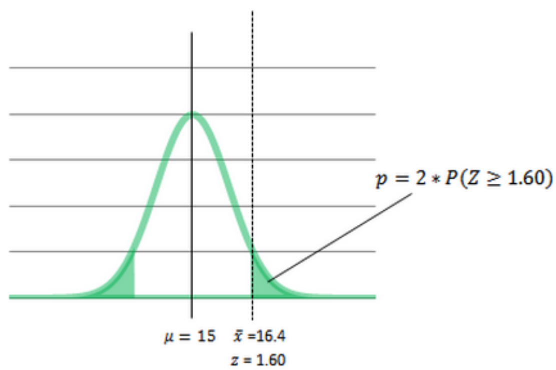
The p-value is the probability that the test statistic is 1.60 or more extreme.

$\boldsymbol{H_0 : \mu = 15}$

$\boldsymbol{H_1 : \mu \neq 15}$

Example: Chemical in Water - Two-Sided

That is, the p-value is the probability that $Z \geq 1.60$ or $Z \leq -1.60$.



$$P = P(Z \leq -1.60 \text{ or } Z \geq 1.60) = P(Z \geq 1.60) + P(Z \leq -1.60)$$
$$= 2 \times P(Z \geq 1.60) = 2 \times 0.0548 = 0.1096$$

It appears that the sample mean that we observed ($\bar{x} = 16.4$) is moderately likely to have occurred if the true population mean was 15 ppm (if $\mu = 15$).

We don't have strong evidence against the null hypothesis.

Evaluating Hypotheses using p-value - Significance Level

In order to decide whether to reject the null hypothesis, we can either compare: **the p-value to a pre-defined significance level** the test statistic to a critical value.

For either method, the **significance level** needs to be pre-specified. **It is denoted α (the Greek letter "alpha").**

The most common choice for **alpha is 0.05** (It is also called standard alpha level). The evidence from the data is so strong that the result obtained would only appear 5% of the time or less if the null hypothesis is in fact true.

- If the $\boldsymbol{p-value \leq \alpha}$, then we **reject the null hypothesis** in favor of the alternative hypothesis.
- If the $\boldsymbol{p-value > \alpha}$, then we **fail to reject the null hypothesis**. We do not have sufficient evidence that the null hypothesis is not true.

Evaluating Hypotheses using critical values

The critical value depends on the significance level and whether the test is one- or two-sided.

For a significance test with standard normal distribution ($\mu = 0, \sigma = 1$)

α	10%	5%	2%	1%
1-sided critical value	1.282	1.645	2.054	2.326
2-sided critical value	1.645	1.960	2.326	2.576

If the absolute value of the test statistic ≥ the critical value, then we **reject the null hypothesis** in favor of the alternative hypothesis.

If the absolute value of the test statistic < the critical value, then we **fail to reject** the null hypothesis.

Significance Test Result

If the $p - value \leq \alpha$ or (equivalently) the absolute value of the test statistic $\geq$ the critical value, then we say that the data are "statistically significant at the α level."

In statistics, "Significant" means "not likely to have happened by chance."

The significance level **quantifies** exactly how we are defining "not likely."

Generally, we should report **both the significance level as well as the p-value** as others may be interested in a more or less conservative level of significance.

By provided the exact p-value, the reader may compare the p-value with their own level of significance.

Test of Significance (hypothesis test)

There are 5 key steps in carrying out any significance test:

1. Set up the **hypotheses** and **select the alpha level**
2. Select the appropriate **test statistic**
3. State the **decision rule**
4. **Compute** the test statistic and the associated p-value
5. **State your conclusion**

Example: An One-Sided Test

`A gym is interested in whether a 6-week weight loss training program they launched has been successful in helping their clients lose weight. To assess this, they took a sample of 30 participants.`

We know that the standard deviation of changes in weights over a six-week interval is 6 pounds.

$s = 6$

The sample mean of the change in weight for the 30 participants in the sample was -2.98 pounds.

$\bar{x} = -2.98$, $n = 30$

Perform a significance test to determine whether the weight loss training program has been successful at the $\alpha = 0.05$ level of significance.

An one-sided test example

1. Set up the hypotheses and select the alpha level

$\boldsymbol{H_0 : \mu = 0}$ (there is no effect on weight change of program participants)

$\boldsymbol{H_1 : \mu < 0}$ (program participants lose weight on average)

$\boldsymbol{\alpha = 0.05}$

2. Select the appropriate test statistic

$$\boldsymbol{z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}}$$

3. State the decision rule

Determine the appropriate critical value from the standard normal distribution table associated with a right hand tail probability of $\boldsymbol{\alpha = 0.05}$.

Using the table, the appropriate critical value is 1.645.

Decision Rule: Reject $\boldsymbol{H_0}$ if $\boldsymbol{|z| \geq 1.645}$

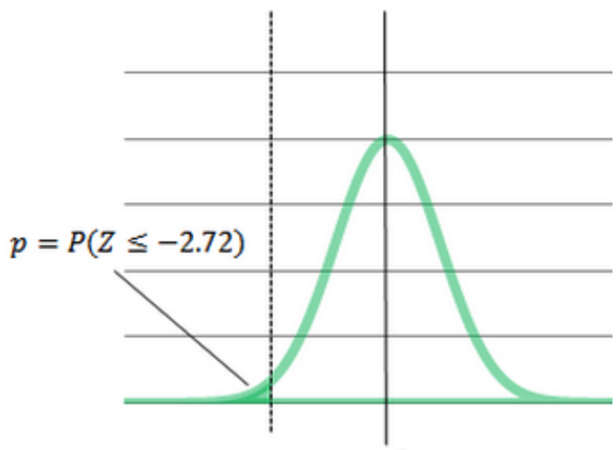
Otherwise, do not reject $\boldsymbol{H_0}$

An one-sided test example

4. Compute the test statistic and the associated p-value

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{-2.98-0}{\frac{6}{\sqrt{30}}} \approx \frac{-2.98}{1.0954} \approx -2.72$$

$$p = P(Z \leq -2.72) = 0.0033$$



5. Conclusion

Reject H_0 since $|2.72| \geq 1.645$

We have significant evidence at the $\alpha = 0.05$ level that $\mu < 0$.

We reject the null hypothesis that the weight loss program has no effect on weight change of program participants in favor of the alternative hypothesis that program participants lose weight on average (p = 0.0033).

A two-sided test example

`County officials were interested in measuring a particular chemical in water sources in the county. High levels of this chemical are as harmful as are low levels. Normal levels of this chemical are 15 parts per million (ppm).`

- ▷ Samples from 50 water sources throughout the county are taken and the levels of this chemical are measured.
- ▷ Suppose we know that the population standard deviation is 6.2. The sample mean from the 50 samples was 16.4 ppm.

A significance test was conducted to determine whether the mean levels are different than 15 ppm at the $\alpha = 0.05$ level of significance.

However, there was not enough evidence to suggest that the mean levels were not normal.

Given that the significance test did not reject the null hypothesis at the $\alpha = 0.05$ level, would you expect that the 95% confidence interval for the population mean to include 15 ppm or not?

A Two-Sided Test Example

Let's confirm by calculating the confidence interval. To calculate a confidence interval with a confidence level of C for the population mean, μ, we use formula:

$$\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

The 95% confidence interval is calculated:

$$\begin{aligned}\bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}} &= 16.4 \pm 1.960 \cdot \frac{6.2}{\sqrt{50}} \\ &= 16.4 \pm 1.960 \cdot 0.8768 = 16.4 \pm 1.7186 \\ &\approx (14.68, 18.12)\end{aligned}$$

We are 95% confident that the true mean level is between 14.69 ppm and 18.12 ppm.

The 95% confidence interval contained the null value of 15 ppm since the two-sided significance test at the $\alpha = 0.05$ level did not reject the null hypothesis that $\mu = 15$ (the mean levels are normal).

Z test R commands

Z-test (population SD is known = popsd)

```
z <-  (mean(data$variable) - mu )/(popsd/sqrt(nrow(data))

# Then pnorm function can be used to calculate the probability
pnorm(z) or 1- pnorm(z) or 2*(1- pnorm(abs(z)))

# Confidence Interval can be computed by calculating lower and upper
    bounds
lower <- mean(data$variable)-z*popsd/sqrt(nrow(data))
upper <- mean(data$variable)+z*popsd/sqrt(nrow(data))
```

Using ”asbio” package

```
install.packages("asbio")

library(asbio)

one.sample.z(null.mu=[mean under null], xbar=mean(data$variable),
sigma=popsd, n=nrow(data), alternative=[alternative], conf=[confidence
    level])

# [alternative] =  less  ,  greater  , or  two.sided
```

Significance Tests and Confidence Intervals

There is a relationship between two sided tests conducted with a significance level of α and $1 - \alpha$ confidence intervals.

- A level α significance test **rejects the null hypothesis** $H_0 : \mu = \mu_0$ when the value of μ_0 is not included in the $1 - \alpha$ confidence interval for μ.
- On the other hand, a level α significance test **fails to reject the null hypothesis** $H_0 : \mu = \mu_0$ when the value of μ_0 is included in the $1 - \alpha$ confidence interval for μ.

The conclusion of a significance test (whether or not the null hypothesis is rejected) at the α level of significance can be determined by checking if the null value of the mean as specified by the null hypothesis is contained within the $1 - \alpha$ confidence interval.