Statistics MM4: Hypothesis testing 1

Lecturer: Israel Leyva-Mayorga

email: ilm@es.aau.dk

AALBORG UNIVERSITY
DENMARK

Connectivity

Schedule

1. Introduction to statistics
2. Parameter estimation
3. Confidence intervals
4. **Hypothesis testing 1**
5. Hypothesis testing 2
6. Regression
7. Workshop: wrap-up and exam problems

Outline

Recap on confidence intervals

Introduction to hypothesis testing

Tests with one normally distributed population and known variance

- Two-sided test for the mean
- One-sided test for the mean

Recap on confidence intervals

Confidence intervals

Our estimators are RVs, so they must have a distribution

The confidence interval (CI) is the best answer to the question:
What is the range of values $C_{1-\alpha} = (a, b)$ around the estimate $\hat{\theta}_n$ such that we are confident with probability $1 - \alpha$ that the true value θ is inside the range?

$$P(\theta \in C_{1-\alpha}) \geq 1 - \alpha$$

Usually $\alpha = 0.05$ so we look at the 95% confidence interval

$$C_{0.95} = (a, b)$$

Confidence intervals for normal RVs with known σ^2

If $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$, we know that the MLE of the μ is the sample mean

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Therefore, the variance of our estimator is $\mathrm{var}(\hat{\mu}_n) = \sigma^2/n$ and the CI is

$$C_{1-\alpha} = \left(\hat{\mu}_n - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\mu}_n + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

For normal RVs, 95% of outcomes are between $\hat{\mu}_n - \mathbf{1.96}\sigma$ and $\hat{\mu}_n + \mathbf{1.96}\sigma$

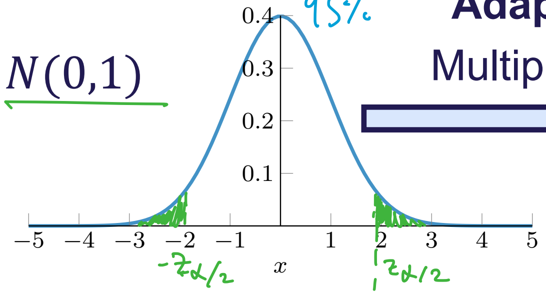
$$C_{0.95} = \left(\hat{\mu}_n - \mathbf{1.96}\frac{\sigma}{\sqrt{n}}, \hat{\mu}_n + \mathbf{1.96}\frac{\sigma}{\sqrt{n}}\right)$$

Why this formula?

$$C_{1-\alpha} = \left(\hat{\mu}_n - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\mu}_n + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

So we don't have to obtain the quantiles $1-\alpha$ for each distribution $N(\hat{\mu}_n, \hat{\sigma}_n^2/n)$

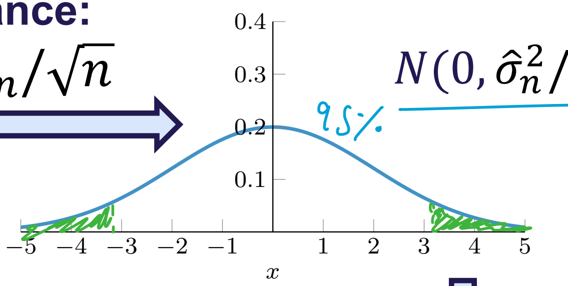
$N(0,1)$



Adapt variance:

Multiply by $\hat{\sigma}_n/\sqrt{n}$

$N(0, \hat{\sigma}_n^2/n)$



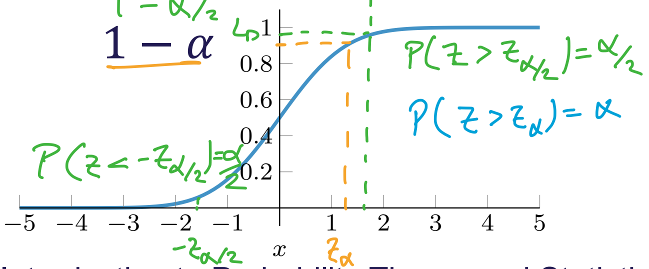
$\alpha = 0.05$

$\alpha/2 = 0.025$

$1-\alpha/2$

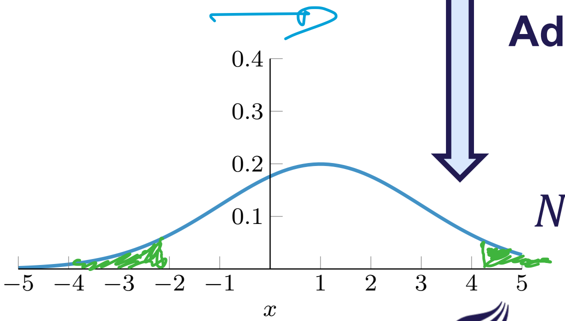
$1-\alpha$

$Z_{\alpha/2}$



Adapt mean:

Add $\hat{\mu}_n$



$N(\hat{\mu}_n, \hat{\sigma}_n^2/n)$

Introduction to hypothesis testing

Types of tests based on the populations

Parameter estimation with 1 population

There is some idea about the value of a parameter

Is that idea correct?

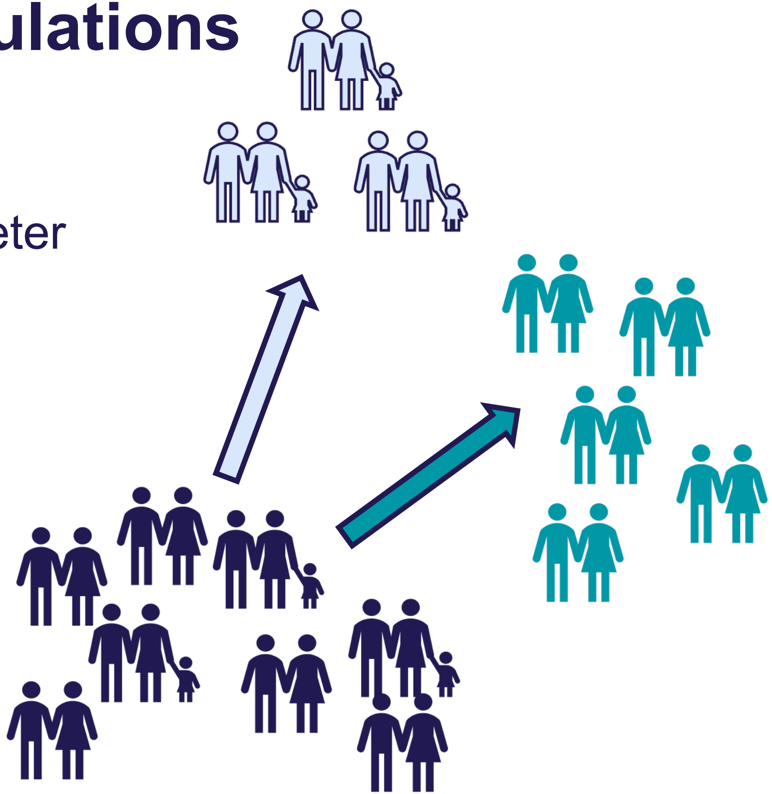
Example: Is it true that the average age is 20?

Compare 2 populations with each other

No parameter known a priori

- Begin with different populations
- One initial population divided into 2

Can we find differences between populations?



How do we conduct the tests?

We begin with two opposing hypotheses:

H_0: The null hypothesis, the one assumed to be true

H_1: The alternative hypothesis, which contradicts H_0

We try to find evidence to support H_0

If we cannot, then we say we can **reject** H_0

Accepting a hypothesis does not mean it is true, but that the data support it

Example in a pharmaceutical setup:

H_0: The new drug is **not** effective → We know something about the baseline case

H_1: The new drug is effective → We may not know how effective it is

How do we formulate the hypotheses?

H_0: The null hypothesis

The baseline case: Oftentimes, it has a simple formulation

Example: A parameter θ may take values in the set Θ and we test the value of θ

H_0: $\theta = \theta_0 \in \Theta$

H_1: $\theta \neq \theta_0$

If we reject H_0, then it means that **H_1:** $\theta \neq \theta_0$ is true but still $\theta \in \Theta$

We could also define the hypotheses as **H_0:** $\theta \in \Theta_0$ versus **H_1:** $\theta \in \Theta_1$

As long as $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$

Types of tests with a single population

One-sided tests

The null hypothesis is that the true value lies in an interval with one finite limit

$\mathbf{H_0}$: $\theta \in \Theta$ where $\Theta = (-\infty, b]$ or $\Theta = [a, \infty)$

The same as $\mathbf{H_0}$: $\theta \leq b$ or $\theta \geq a$

Two-sided tests

The null hypothesis is that the true value lies in an interval with finite limits

$\mathbf{H_0}$: $\theta \in \Theta = [a, b]$

The same as $\mathbf{H_0}$: $a \leq \theta \leq b$

There is a special case: If Θ has a single element $\theta = a = b$

Procedure for testing

1. Choose a parameter for testing θ
2. Formulate the null hypothesis H_0 and the alternative hypothesis H_1 about θ
3. Design the test and define the rejection region $R = \{x: \underbrace{T(x)}_{\text{test statistic}} > \overset{\text{threshold}}{c}\}$
4. Collect a sample $X_1, X_2, X_3, \dots, X_n$ of i.i.d. RVs
5. If the observation regarding θ is close enough to the value(s) in H_0

    H_0 cannot be rejected

6. Else:

    Reject H_0 and accept H_1

Test outcomes and errors

Errors can occur even when we follow a proven methodology

If a test statistic is in the rejection region R, $\mathbf{H_0}$ is rejected

	Accept	Reject
$\mathbf{H_0}$ is True	**All good!**	**Type I error**
$\mathbf{H_0}$ is False	**Type II error**	**All good!**

Significance level α

The probability of Type I error should not exceed α

The probability of Type I error is equal to the probability of being outside of the CI

We want α to be small: We don't want to reject $\mathbf{H_0}$ when it's true

Trade-off between Type I and Type II errors

If we reduce α, we have fewer Type I errors, but more Type II errors

The choice of α depends on the application:

Computer security: H_0 is that the user fingerprint is correct (unlock the phone)

Medicine: H_0 is that the patient has cancer

We have to evaluate the consequences of Type I and Type II errors

Finally, the coin toss example from the first lecture

A fair dice should have the exact same probability of rolling any of the numbers

Can you say if a dice is fair by rolling it once?

Twice?

After rolling it 100 times, how confident are you that it is fair?

→ Coin example

Procedure

1. Choose a parameter for testing θ

Each coin toss is a RV $X_i \sim \text{Bernoulli}(p)$, so $\theta = p$

2. Formulate the null hypothesis $\mathbf{H_0}$ and the alternative hypothesis $\mathbf{H_1}$

A fair coin should have $p = 1/2$, so $\mathbf{H_0}$: $p = 1/2$ and $\mathbf{H_1}$: $p \neq 1/2$

Finally, the coin toss example from the first lecture

3. Design the test and define the rejection region

Toss the coin n times and count the number of heads

Generate an estimate for p as

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

We define our **test statistic** as $\boxed{T = |\hat{p}_n - p_0|}$ where $p_0 = 1/2$

If $T = |\hat{p}_n - p_0| \leq c$, we cannot reject H_0

Otherwise, if $T = |\hat{p}_n - p_0| > c$, we reject H_0

Rejection region: $R = \{\hat{p}_n : T > c\}$

Finally, the coin toss example from the first lecture

4. Collect a sample $X_1, X_2, X_3, \dots, X_n$ of i.i.d. RVs

5. If the observation regarding θ is close enough to the value(s) in H_0

H_0 cannot be rejected

6. Else:

Reject H_0 and accept H_1

How do we choose c and α?

We're not done yet

We still need to define the probability of Type I error α

$$P(\text{Type I error}) = P(T > c \mid \mathrm{H}_0 \text{ is True}) = \alpha$$

This means that the probability that we fall in the rejection region given that H_0 is True should be, at most, equal to α

How do we choose c and α?

We know that

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \approx N\left(p, \frac{\sigma^2}{n}\right)$$

And we have a powerful definition of **Confidence interval (CI):**

Range of values $C_{1-\alpha} = (a, b)$ around the estimate $\hat{\theta}_n$ such that:

we are confident with probability $1 - \alpha$ that the **true value** θ is inside the range

$$P(\theta \in C_{1-\alpha}) \geq 1 - \alpha$$

So, we can use what we know about CIs for testing, at least in principle

Two-sided test for the mean:
Normal population with known variance

Defining hypothesis for μ

H$_0$: $\mu = \mu_0$

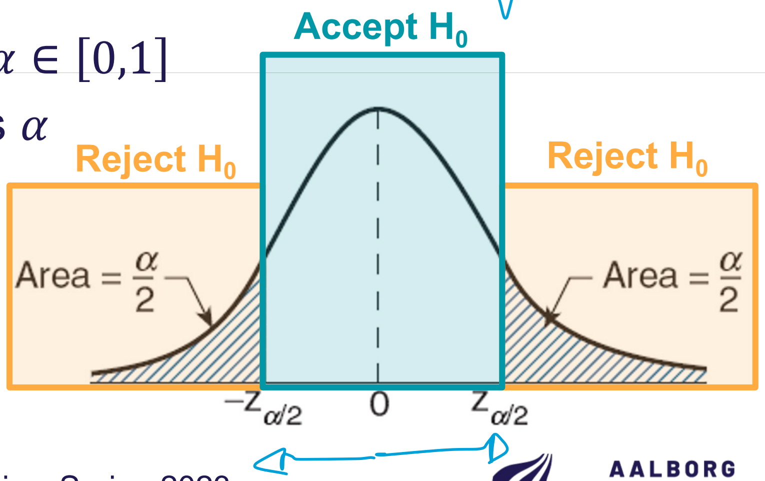
Our test statistic is derived from $|\hat{\mu}_n - \mu_0|$ and we aim for $P(\text{Type I error}) = \alpha$

Rejection region is $R = \{X_1, X_2, \dots, X_n : |\hat{\mu}_n - \mu_0| > c\}$ $-z_{\alpha/2}$ $z_{\alpha/2}$

How to choose c and α?

Recall that $Z_\alpha = \inf\{z \in R: \Phi(z) \geq 1 - \alpha\}, \quad \alpha \in [0,1]$

The total area of the region outside the CI is α



$$C_{1-\alpha} = \left(\hat{\mu}_n - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\mu}_n + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Reject H$_0$ if $|\hat{\mu}_n - \mu_0| > Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ c

Defining hypothesis for μ

H$_0$: $\mu = \mu_0$

Our test statistic is derived from $|\hat{\mu}_n - \mu_0|$ and we aim for $P(\text{Type I error}) = \alpha$

Rejection region is $R = \{X_1, X_2, \ldots, X_n : |\hat{\mu}_n - \mu_0| > c\}$

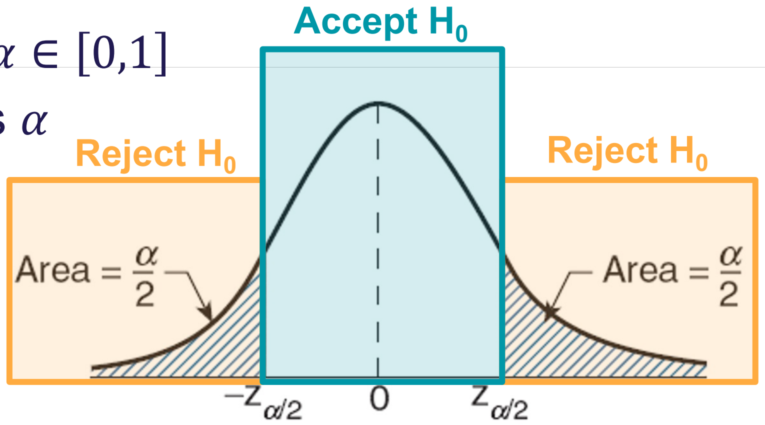
How to choose c and α?

Recall that $Z_\alpha = \inf\{z \in R \colon \Phi(z) \geq 1 - \alpha\}, \quad \alpha \in [0,1]$

The total area of the region outside the CI is α

$$C_{1-\alpha} = \left(\hat{\mu}_n - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \hat{\mu}_n + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Reject H$_0$ if $\boxed{\dfrac{|\hat{\mu}_n - \mu_0|\sqrt{n}}{\sigma}} > Z_{\alpha/2}$



Example

It is known that if a signal of value μ is sent from location A, then the received value at location B is normally distributed with mean μ and $\sigma = 2$. This means that Gaussian noise that is added to the signal is a RV with distribution $N(0,4)$. There is reason for people at location B to suspect that $\mu = 8$ will be sent today. Test this hypothesis when the signal value is sent 5 times and the sample average at location B is $\bar{X}_5 = 9.5$. Use a 5% level of significance.

1. Define the hypothesis $\alpha = 0.05$
2. Compute the test statistic
3. Compute the z-value
4. Decide on acceptance

Example

We know that the MLE for parameter μ is $\hat{\mu}_5 = \bar{X}_5 = 9.5$

1. H_0: $\mu = 8$
2. We accept H_0 if $\boxed{\frac{|\hat{\mu}_n - \mu_0|\sqrt{n}}{\sigma}} \leq Z_{\alpha/2}$. Therefore, the test statistic is $T = \frac{|\hat{\mu}_n - \mu_0|\sqrt{n}}{\sigma}$

   c

   $= \frac{|9.5-8|\sqrt{5}}{2}$
3. We recall that, for $\alpha = 0.05$, we have $Z_{\alpha/2} = 1.96$
4. We cannot reject H_0 because $T = \frac{|\hat{\mu}_n - \mu_0|\sqrt{n}}{\sigma} = 1.677 \leq 1.96$

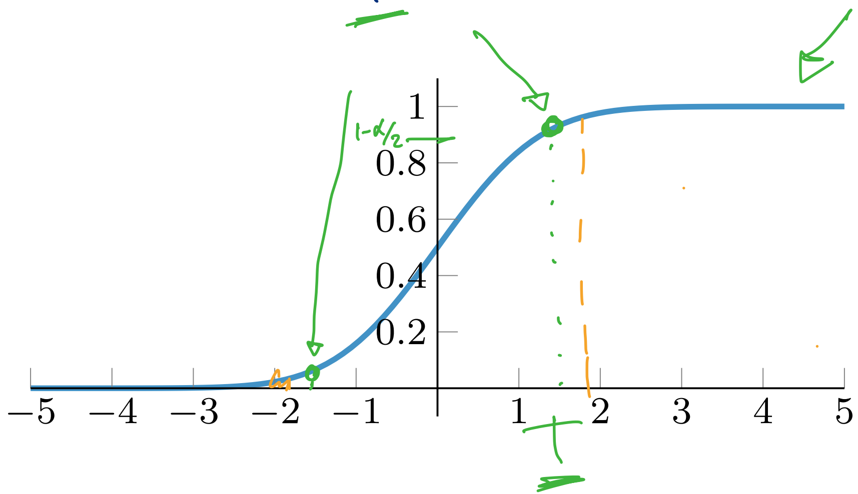
By doing this, our CI includes the real value of μ with 95% confidence

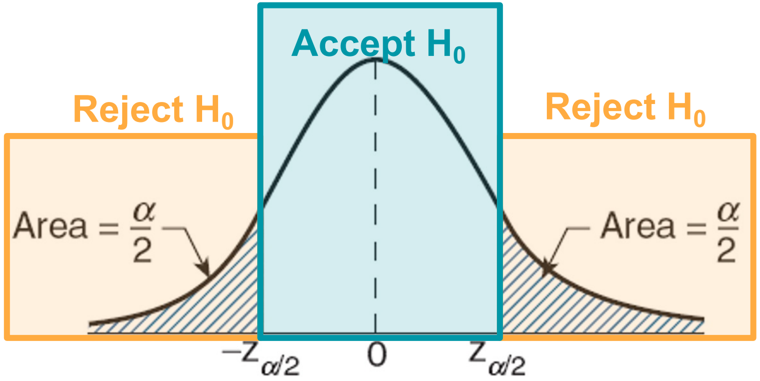
How close was the decision to reject or not?

How close was the test?

$Z_{\alpha/2} = 1.96$

$T = 1.677$



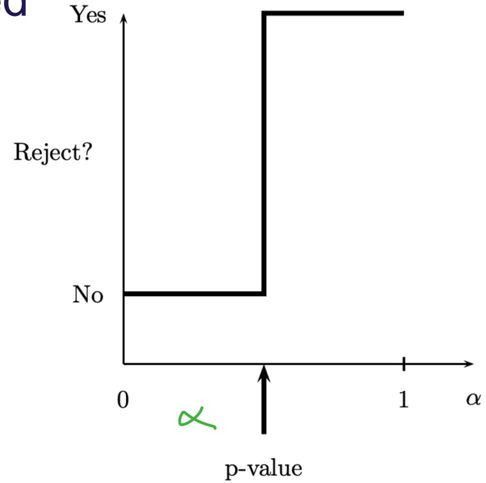


The p-value

If a test rejects the null hypothesis at significance level $\alpha \in (0,1)$,…

…then, it will also reject the test for any $\alpha' > \alpha$

We could calculate the smallest α at which a test is rejected



Definition:

The p-value is the smallest level at which we can reject H_0

The p-value: interpretation

The p-value is the **probability** (under H_0) of observing a value of the test statistic the same as or more extreme than what was actually observed

The p-value for a two-sided test with test statistic T and RV $Z \sim N(0,1)$ is

$$v = P(|Z| > T) = 2P(Z > T) = 2(1 - \underbrace{\Phi(T)})$$

For our example $v = 2\left(1 - \Phi(\underbrace{1.677}_{T})\right) = 0.0935 > \alpha$, so we cannot reject H_0

Type II error for two-sided test

The probability of Type II error is the probability of not rejecting H_0 given its False

$$P(\text{Type II error}) = P(T < c \mid \text{H}_0 \text{ is False}) = \beta$$

For our case, H_0 is false if $\mu \neq \mu_0$ and we can calculate

$$\beta(\mu) = P\left(|\hat{\mu}_n - \mu_0| < Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \,\middle|\, \text{H}_0 \text{ is False}\right)$$

$$= P\left(\mu_0 - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \hat{\mu}_n \leq \mu_0 + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \,\middle|\, \text{H}_0 \text{ is False}\right)$$

$$= \Phi\left(\frac{(\mu_0 - \mu)\sqrt{n}}{\sigma} + Z_{\alpha/2}\right) - \Phi\left(\frac{(\mu_0 - \mu)\sqrt{n}}{\sigma} - Z_{\alpha/2}\right)$$

One-sided test for the mean: Normal population with known variance

Defining hypothesis for μ

H$_0$: $\mu \leq \mu_0$ and so **H$_1$:** $\mu > \mu_0$

Our test statistic is derived from $\boxed{\hat{\mu}_n - \mu_0}$ and we aim for $P(\text{Type I error}) = \alpha$

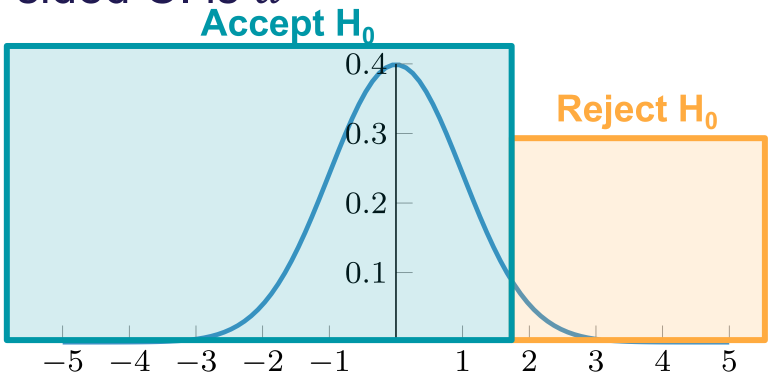
Rejection region is $R = \{X_1, X_2, \dots, X_n : \hat{\mu}_n - \mu_0 > c\}$

How to choose c and α?

The total area of the region outside the one-sided CI is α

$$C_{1-\alpha} = \left(-\infty, \hat{\mu}_n + Z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

Reject H$_0$ if $\hat{\mu}_n - \mu_0 > Z_\alpha \frac{\sigma}{\sqrt{n}}$



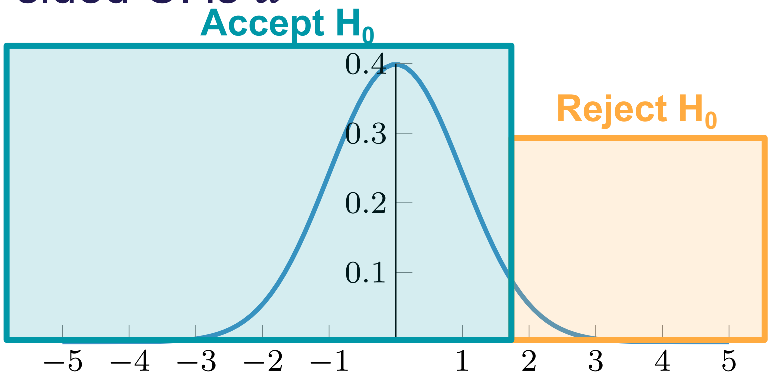
Defining hypothesis for μ

H$_0$: $\mu \leq \mu_0$ and so **H$_1$:** $\mu > \mu_0$
Our test statistic is derived from $\hat{\mu}_n - \mu_0$ and we aim for $P(\text{Type I error}) = \alpha$
Rejection region is $R = \{X_1, X_2, \dots, X_n : \hat{\mu}_n - \mu_0 > c\}$

How to choose c and α?

The total area of the region outside the one-sided CI is α

$$C_{1-\alpha} = \left(-\infty, \hat{\mu}_n + Z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$



Reject H$_0$ if $\underbrace{\frac{(\hat{\mu}_n - \mu_0)\sqrt{n}}{\sigma}}_{T} > Z_\alpha$

Example

Go back to the previous example of the transmitted signal.
Assume that we want to test whether the signal is 8 or less.
What is the conclusion?

1. Define H_0
2. Can we reject H_0?
3. What is the p-value?

The p-value for a one-sided test with test statistic T and RV $Z \sim N(0,1)$ is

$$v = P(Z > T) = 1 - \Phi(T)$$

Solution

We know that $\hat{\mu}_5 = \bar{X}_5 = 9.5$

1. Now H_0: $\mu \leq 8$ and H_1: $\mu > 8$
2. We get $T = \frac{(\hat{\mu}_n - \mu_0)\sqrt{n}}{\sigma} = 1.677 > Z_\alpha = 1.645$, so **we reject** H_0
3. The p-value is $v = 1 - \Phi(1.677) = 0.0467 < \alpha = 0.05$

Type II error for one-sided test

The probability of Type II error is the probability of not rejecting H_0 given its False

$$P(\text{Type II error}) = P(T < c \mid \mathrm{H}_0 \text{ is False}) = \beta$$

For our case, H_0 is false if $\mu \neq \mu_0$ and we can calculate

$$\beta(\mu) = P\left(\hat{\mu}_n - \mu_0 > Z_\alpha \frac{\sigma}{\sqrt{n}} \,\middle|\, \mathrm{H}_0 \text{ is False}\right)$$

$$= \Phi\left(\frac{(\mu_0 - \mu)\sqrt{n}}{\sigma} + Z_\alpha\right)$$

Summary

Summary

We have several types of tests based on:

- Number of populations involved, types of hypothesis, and assumptions

We need to choose the right one

If the distribution of the population is normal and the variance is known

- We can use the CIs with normal distribution for testing

Otherwise, we need more advanced mathematics… to be covered next lecture