STATISTICS

MATH 414

6th March 2025

Vivek Tewary

Introduction to hypothesis testing

Introduction to hypothesis testing

We looked at point estimation using the maximum likelihood method.

Introduction to hypothesis testing

We looked at point estimation using the maximum likelihood method.

We also looked at confidence intervals for mean.

Introduction to hypothesis testing

We looked at point estimation using the maximum likelihood method.

We also looked at confidence intervals for mean.

Today we look at a third type of inference

– testing of hypotheses

Introduction to hypothesis testing

We are interested in a random variable X with density $f(x;\theta)$ where $\theta \in \Omega$.

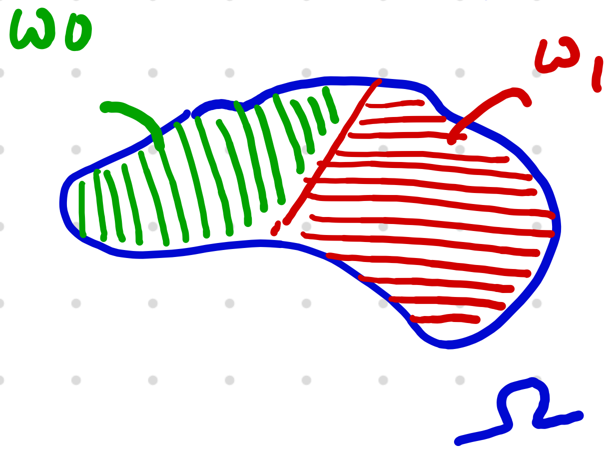
Introduction to hypothesis testing

We are interested in a random variable X with density $f(x;\theta)$ where $\theta \in \Omega$.

A previous experiment or theory has convinced us that the parameter

$$\theta \in \omega_0 \quad \text{or} \quad \theta \in \omega_1 \quad \text{where}$$

$$\Omega = \omega_0 \cup \omega_1$$



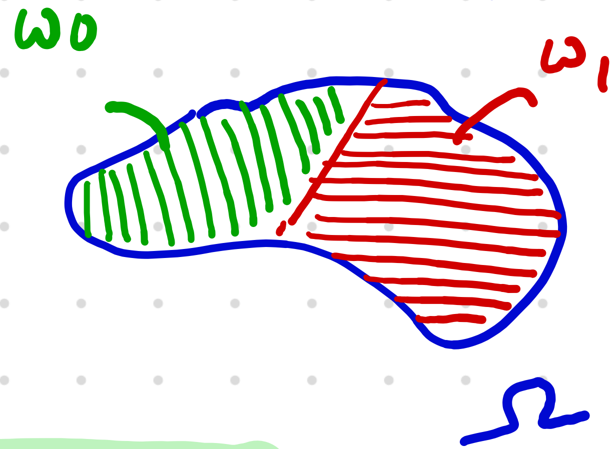
Introduction to hypothesis testing

We are interested in a random variable X with density $f(x;\theta)$ where $\theta \in \Omega$.

A previous experiment or theory has convinced us that the parameter

$\theta \in \omega_0$ or $\theta \in \omega_1$ where

$$\Omega = \omega_0 \cup \omega_1$$



We see this as two hypotheses

$H_0: \theta \in \omega_0$ versus $H_1: \theta \in \omega_1$

$H_0 : \theta \in \omega_0$ is referred to as the <u>null hypothesis</u>

and

$H_1 : \theta \in \omega_1$ is called the <u>alternative hypothesis</u>.

$H_0 : \theta \in \omega_0$ is referred to as the null hypothesis

and

$H_1 : \theta \in \omega_1$ is called the alternative hypothesis.

Based on a sample $X_1, X_2, \ldots, X_n$, we will rule for either H_0 or H_1. Clearly, our decision could be wrong.

$H_0 : \theta \in \omega_0$ is referred to as the <u>null hypothesis</u>

and

$H_1 : \theta \in \omega_1$ is called the <u>alternative hypothesis</u>.

Based on a sample $X_1, X_2, \ldots, X_n$, we will rule for either H_0 or H_1. Clearly, our decision could be wrong.

<u>Type I error</u> We decide $\theta \in \omega_1$ when actually $\theta \in \omega_0$

<u>Type II error</u> We decide $\theta \in \omega_0$ when actually $\theta \in \omega_1$.

$H_0 : \theta \in \omega_0$ is referred to as the null hypothesis

and

$H_1 : \theta \in \omega_1$ is called the alternative hypothesis.

Based on a sample $X_1, X_2, \ldots, X_n$, we will rule for either H_0 or H_1. Clearly, our decision could be wrong.

Type I error We decide $\theta \in \omega_1$ when actually $\theta \in \omega_0$

Type II error We decide $\theta \in \omega_0$ when actually $\theta \in \omega_1$.

We defer an analysis of these errors to the future.

Examples

1. Are temperatures on average higher now than they were a hundred years ago?

2. Are people with high glucose levels at age 30 more likely to develop diabetes at age 60?

3. Does smoking decrease life expectancy?

Binomial Test

We flip a coin n times.
Let X be the number of heads

$$\boxed{X \sim \text{Binom}(n, p)}$$

We want to know whether $p = 0.5$ or not.

Suppose we get 40 heads out of 100.
What can we conclude?

We look at more ways to do hypothesis testing later. For now, we focus on a simple template for testing.

We look at more ways to do hypothesis testing later. For now, we focus on a simple template for testing.

Step 1 Write down the null and alternative hypothesis.

We look at more ways to do hypothesis testing later. For now, we focus on a simple template for testing.

Step 1 — Write down the null and alternative hypothesis.

In this case

$$\boxed{H_0 : p \text{ is } 0.5}$$ ~ null hypothesis.

$$\boxed{H_1 : p \text{ is not } 0.5}$$ ~ alternative hypothesis.

Step 2 Calculate a test statistic.

A statistic is a numerical output of an experiment. In this case, it would be the number of heads, 40.

Since we are using this statistic in a statistical test, we call it a

test statistic.

The further this number is from 50, the stronger the evidence against the null hypothesis.

Step 3 Compute the p-value.

As we have said before, the statistic is a random variable.

Now we ask: How unusual would the test statistic be if the null hypothesis were true?

We answer this question from our knowledge of the distribution of X if the null hypothesis were true. In this case, $X \sim \text{Binom}(100, 0.5)$

In this case $P(X \leq 40) = \text{pbinom}(40, 100, 0.5)$

$= 0.0284$.

We also need to look at the other tail,

Viz. $P(X \geq 60) = 0.0284$.

Therefore, the probability of getting the test statistic if null hypothesis were true is 0.0568. This is called a p-value.

In this case, $P(X \leq 40) = \text{pbinom}(40, 100, 0.5)$
$= 0.0284.$

We also need to look at the other tail,

viz. $P(X \geq 60) = 0.0284.$ (by symmetry)

Therefore, the probability of getting the test statistic if null hypothesis were true is 0.0568. This is called a p-value.

a definition will have to wait.

Step 4 Draw a conclusion.

The conclusion is that if we flipped a coin 100 times in multiple experiments, 5.7% of the time we will obtain fewer than 41 or more than 59 heads.

We will decide whether to keep H_0 or reject H_0 depending on a pre-determined threshold — which is conventionally taken as 5%.

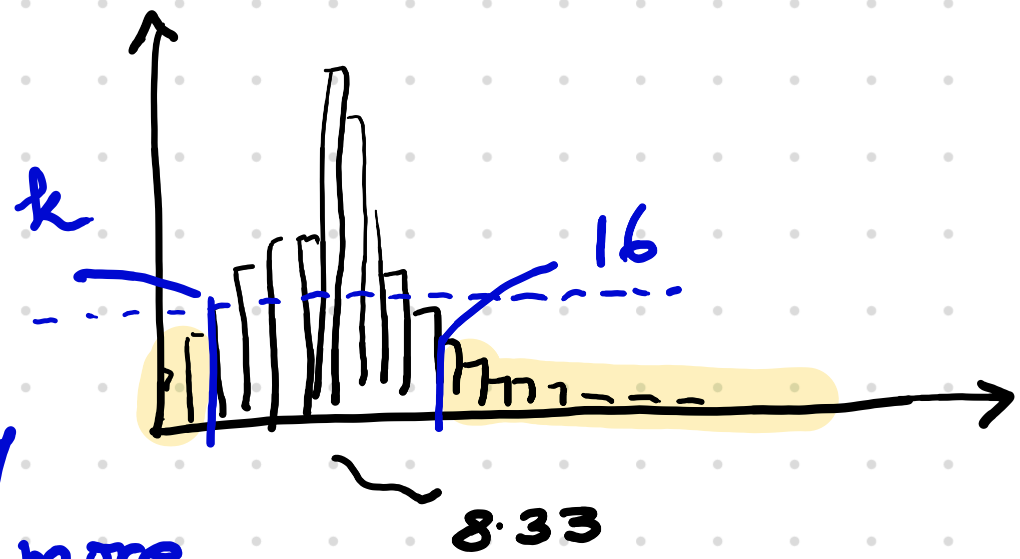
Example 2

Suppose our coin has a probability $1/6$ of turning up heads. Then 50 tosses follow $\text{Binom}(50, 1/6)$.

Suppose in an experiment, we get 16 heads.

① $H_0: p = 1/6$

$H_1: p \neq 1/6$



② Test Statistic is 16

③ p-value is the probability of getting 16 heads or more but also getting k heads or fewer where k is unknown.

we find all values x such that

$$P(X=x) \leq P(X=16)$$

we find all values x such that

$$P(X=x) \le P(X=16)$$

we can do this on R

```
probs <- dbinom(0:50, 50, 1/6)
sum(probs[probs <= dbinom(16, 50, 1/6)])
```

0.0069

we find all values x such that

$$P(X=x) \leq P(X=16)$$

We can do this on R

```
probs <- dbinom(0:50, 50, 1/6)
sum(probs[probs <= dbinom(16, 50, 1/6)])
```

0.0069

In this situation, we might want to reject the null hypothesis.

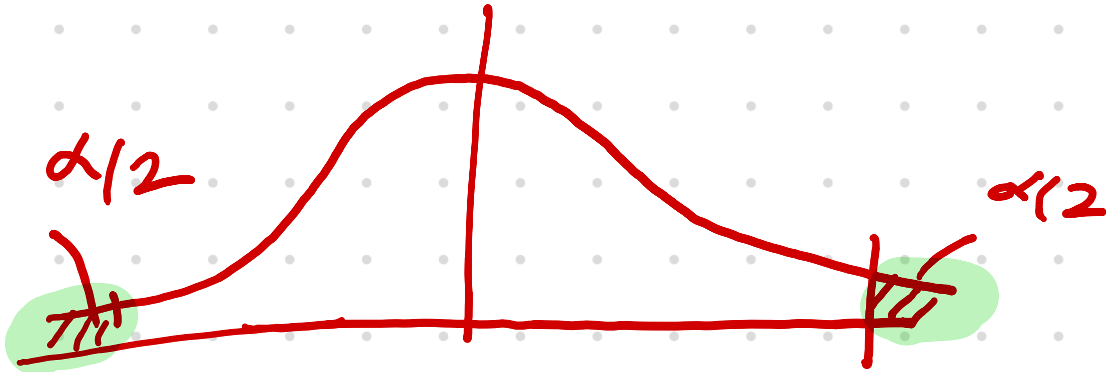
Types of Error and Statistical Power

Type I error

This is when we reject the null hypothesis when it is true.

The threshold α that we choose is called a significance level of the test.

The probability of making a type I error is α when the null hypothesis is true. (sort of!) ~ for cts things!



Type II error It can happen that null hypothesis is false but we do not reject it.

The probability of a type II error is not as straightforward as it depends on

- the pre-determined significance level α
- how wrong is the null hypothesis.

We expect that if the alternate hypothesis is very far from the null then we are not very likely to make this error.

Let us see this in an example.

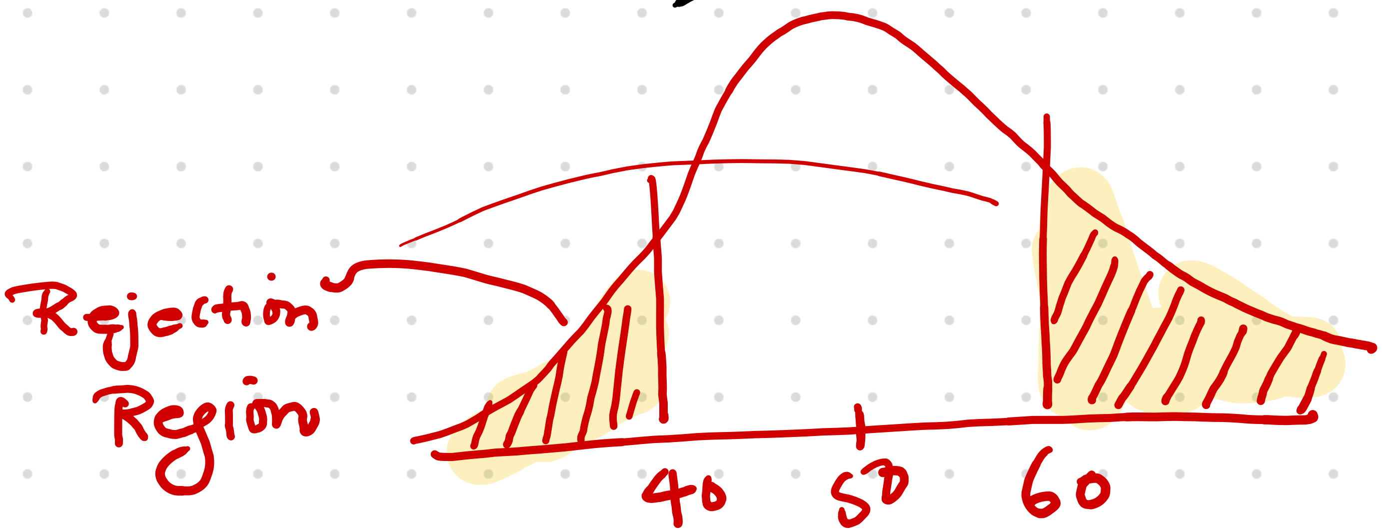
let $\alpha = 0.05$

We make 100 Coin tosses.

The region in which we reject the null-hypothesis is $\{\underbrace{0, 1, \ldots, 39}, \underbrace{61, \ldots, 100}\}$

Since qbinom(0.025, 100, 0.5)

= 40

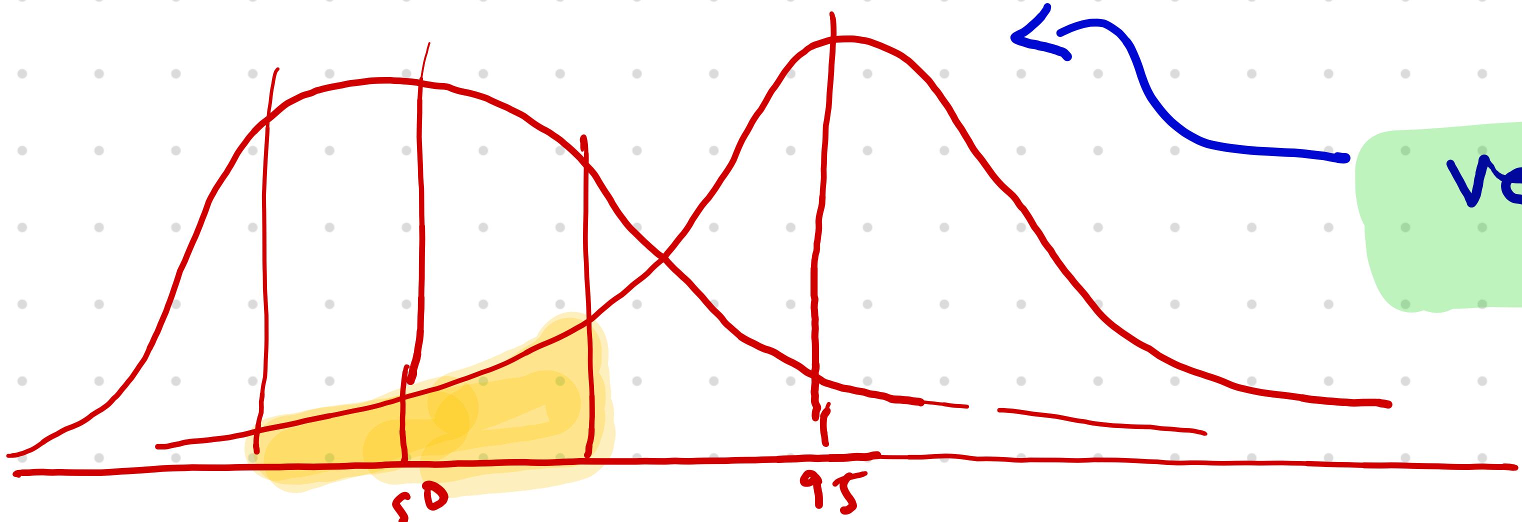


① If the true $p = 0.95$.

then the probability that we will fail to reject the null hypothesis is

$$pbinom(60, 100, 0.95) - pbinom(39, 100, 0.95)$$

$$= 6.24 \times 10^{-26}.$$

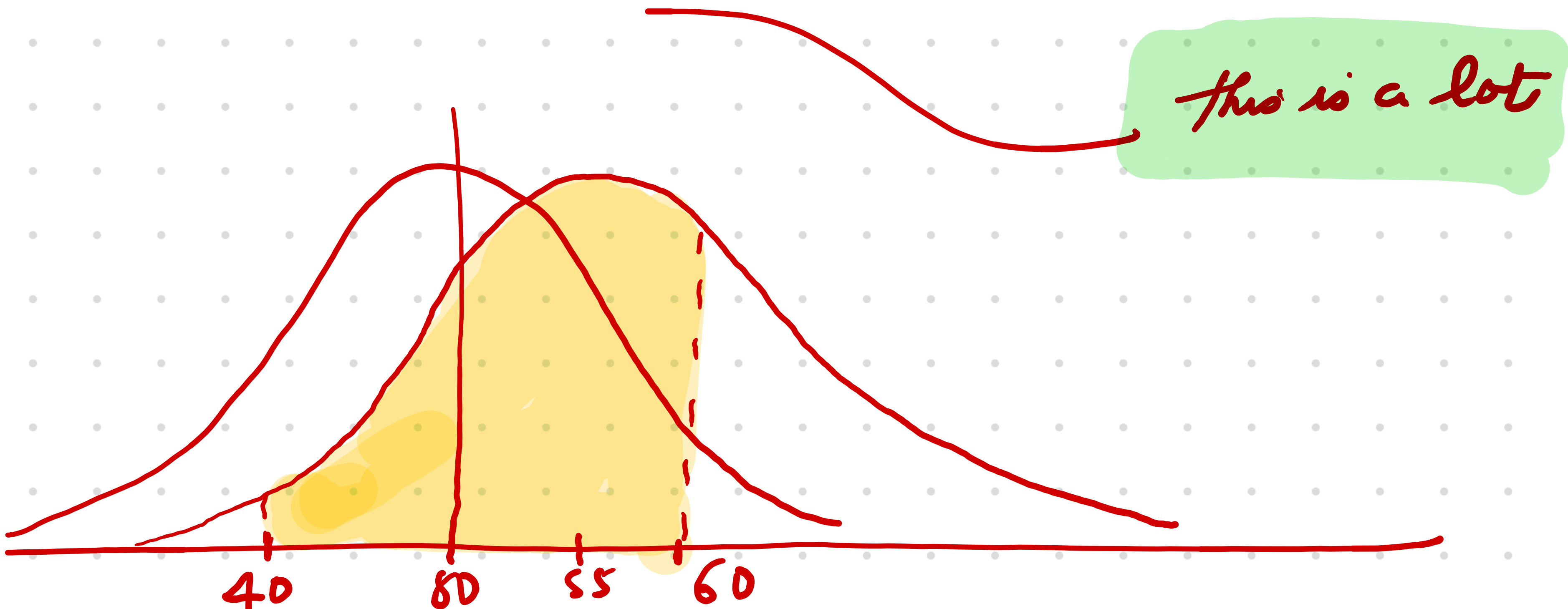


very small !

② On the other hand, if the bias is smaller, say $p = 0.55$ then

pbinom(60, 100, 0.55) − pbinom(39, 100, 0.55)

= 0.865



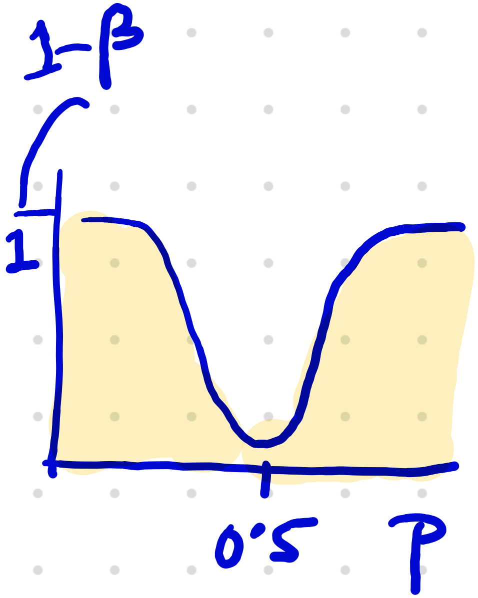
The point is that for any possible value of probability p, we get a different type II error.

The **power** of the test is the probability that we will make the correct decision (reject null hypothesis) when the alternative is true. This is $1-\beta$ where β is the probability of type II error.

Now, we can use R to plot the power as a function of the alternate probability.

```
p <- seq(0,1, by = 0.01)
power <- 1 - binom(60,100, p)
          + binom(39,100, p)
gf_line(power~p, size = 1) %>%
  gf_labs(x = p)
```

Now, we can use R to plot the power as a function of the alternate probability.




```
p <- seq(0,1, by = 0.01)
power <- 1 - binom(60,100, p)
            + binom(39,100, p)
gf_line(power~p, size = 1) %>%
    gf_labs(x= p)
```

Given a fixed alternate probability, we can plot the power against the size of the sample.

A test is said to be under-powered if we have too little data to detect an effect of some desired magnitude.

Later on, we will look at power of tests in more detail.

- We can now repeat this for normal distributions under the assumption of

knowledge of variance — z-test

ignorance of variance - t-test

An example with normal distribution

(1) Data. In an experiment, the lifetime of car tyres was recorded. It may be assumed that the mean is θ and standard deviation is 5000.

An example with normal distribution

① **Data.** In an experiment, the lifetime of car tyres was recorded. It may be assumed that the mean is θ and standard deviation is 5000. From past experience, we think that $\theta = 30000$ miles. The manufacturer claims that tyres made by a new process have $\theta > 30000$.

(2) The null hypothesis $H_0: \theta = 30000$

The alternative hypothesis $H_1: \theta > 30000$

(3)

(2) The null hypothesis $H_0: \theta = 30000$

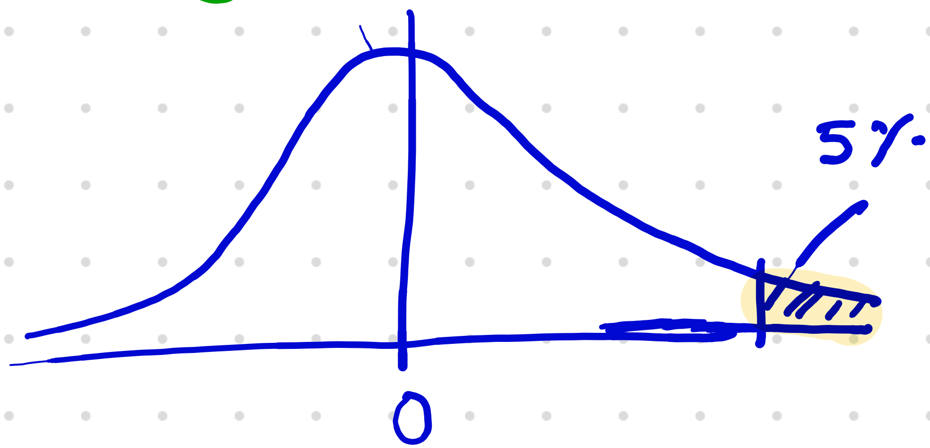
The alternative hypothesis $H_1: \theta > 30000$

(3) Suppose that 10 independent values of X, say $x_1, x_2, \ldots, x_{10}$ were observed and the sample mean was found to be 35000. Do we reject the null hypothesis at 95% significance level?

② The null hypothesis $H_0: \theta = 30000$

The alternative hypothesis $H_1: \theta > 30000$

③ Suppose that 10 independent values of X, say $x_1, x_2, \ldots, x_{10}$ were observed and the sample mean was found to be 35000. Do we reject the null hypothesis at 95% significance level?



④ The test statistic is

$$3.162 = \frac{\sqrt{10}\,(35000 - 30000)}{5000}$$

5 What is the probability?

$P(Z > 3.162) = 1 - \text{pnorm}(3.162, 0, 1)$

$= 0.00078\ldots$

(5) What is the Probability?

$P(Z > 3.162) = 1 - pnorm(3.162, 0, 1)$

$= 0.00078...$

Therefore, we reject the null hypothesis.

⑤ What is the probability?

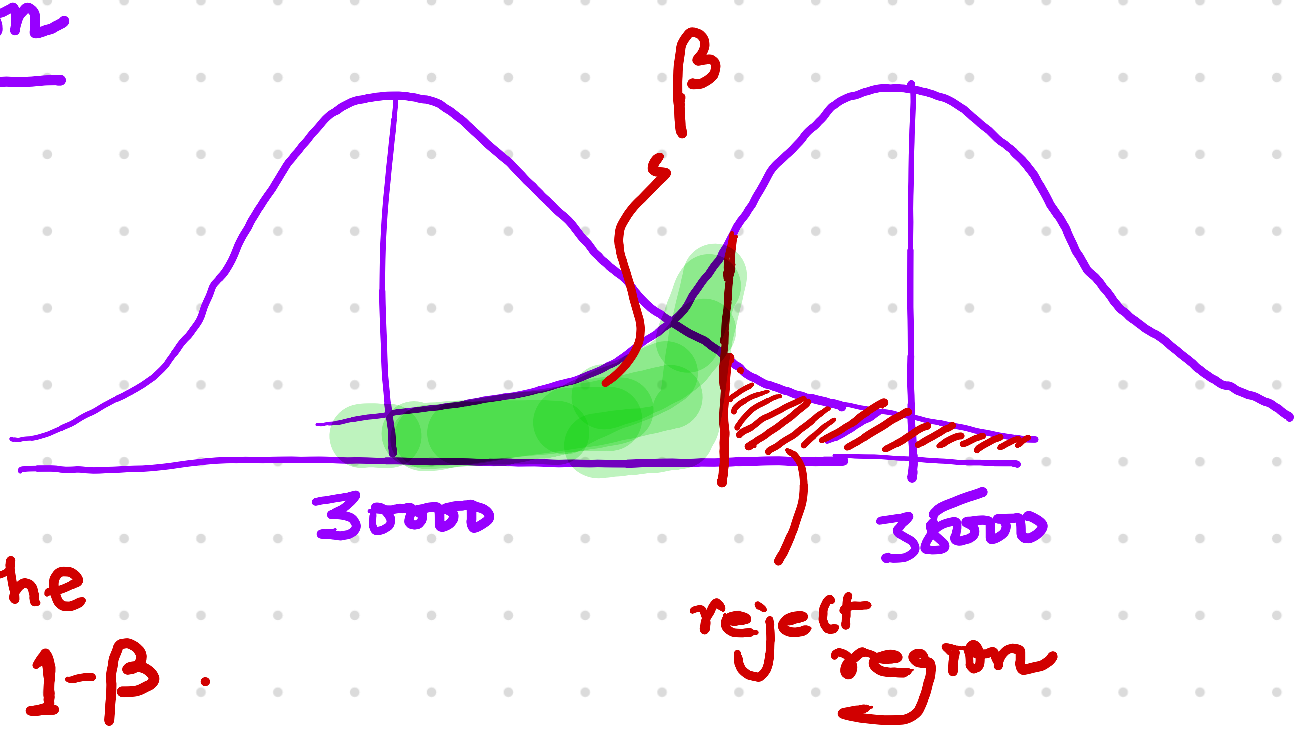
$P(Z > 3.162) = 1 - \text{pnorm}(3.162, 0, 1)$

$= 0.00078\ldots$

Therefore, we reject the null hypothesis.

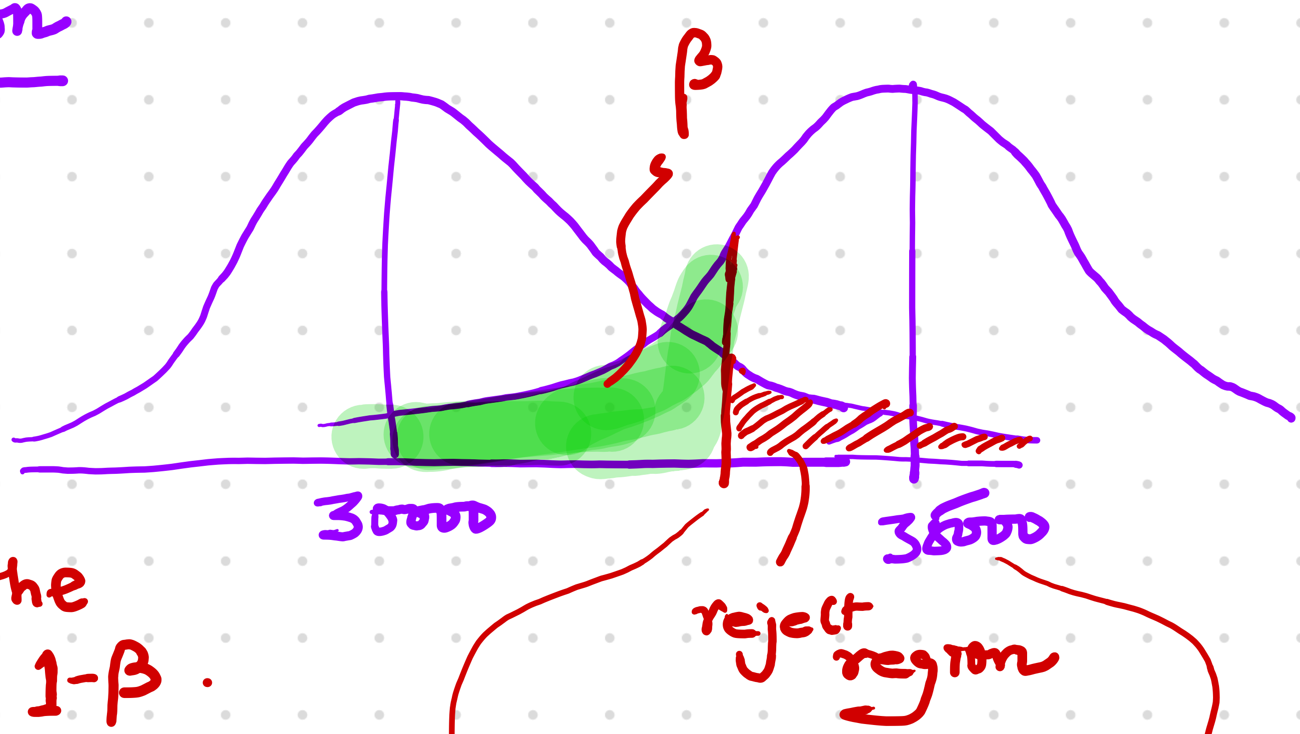
⑥ Power Function

probability that we reject null hypothesis when alternative hypothesis is true. In the diagram it is $1-\beta$.



⑥ Power Function

probability that we reject null hypothesis when alternative hypothesis is true. In the diagram it is $1-\beta$.



In the example,

power is $1 - \text{pnorm}(1.64485, 3.162, 1)$

$= \underline{\underline{0.684}}$

We have made the assumption that the standard deviation is the same for the new distribution!

One more example - In this case, we do not assume any knowledge of the variance.

One more example – In this case, we do not assume any knowledge of the variance.

Charles Darwin once recorded data for the growth of a plant and wanted to see if there was any effect of cross-fertilization vs self-fertilization. He compared 15 pairs of plants.

Data

Pot	1	2	3	4	5	6	7	
CROSS	23.5	12.0	21.000	22	19.125	21.500	22.125	
SELF	17.375	20.375	20.000	20.000	18.375	18.625	18.625	

Pot	8	9	10	11	12	13	14	15
CROSS	20.375	18.25	21.625	23.25	21.00	22.125	23.00	12.00
SELF	15.250	16.500	18.00	16.25	18.00	12.75	15.5	18

① In this case, we could do two things. Consider the differences. Assume that the differences are normally distributed. Compute $\hat{\mu}$ and s^2, the sample mean and sample variance of the differences.

1. In this case, we could do two things. Consider the differences. Assume that the differences are normally distributed. Compute $\hat{\mu}$ and s^2, the sample mean and sample variance of the differences.

2. Then $$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

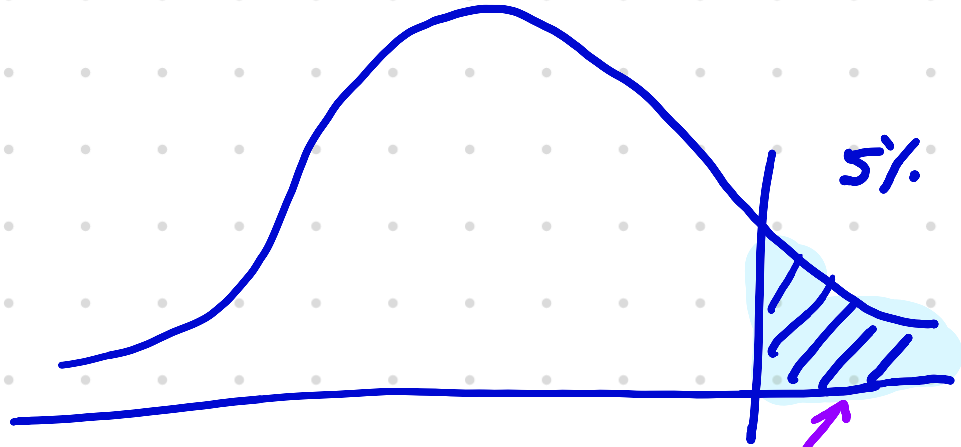
3. <u>Null hypothesis</u>: $\mu = 0$

   <u>Alternative hypothesis</u>: $\mu > 0$

4

In this case

$$\boxed{\begin{aligned} \bar{x} &= 2.62 \\ S_x &= 4.72 \end{aligned}}$$



Reject Region for the null hypothesis.

We compute the t value

$$qt(0.95, 14) = 1.76.$$

On the other hand, the test statistic is

$\bar{x}$

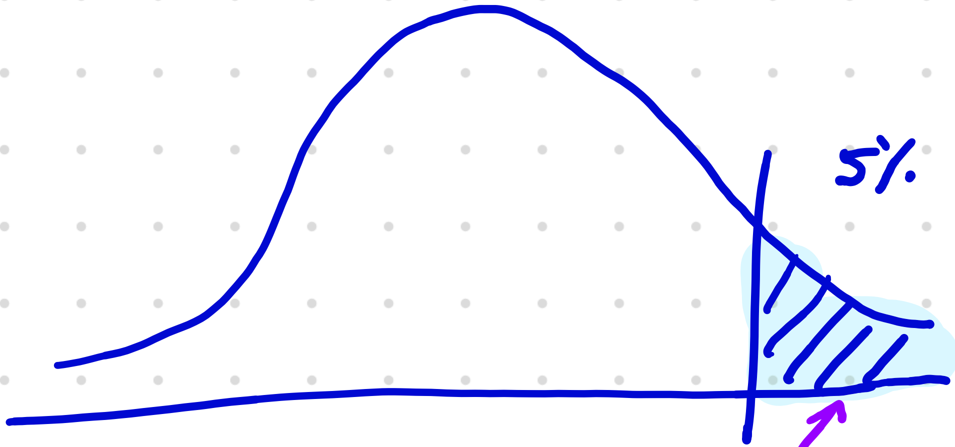
$$\boxed{\frac{2.62}{4.72} \times \sqrt{15}} = 2.15$$

S_x $\qquad$ $\sqrt{n}$

4

In this case

$$\boxed{\begin{aligned} \bar{x} &= 2.62 \\ S_x &= 4.72 \end{aligned}}$$



We compute the t value

$$qt(0.95, 14) = 1.76.$$

On the other hand, the test statistic is

$$\boxed{\frac{\overset{\bar{x}}{2.62}}{\underset{S_x}{4.72}} \times \underbrace{\sqrt{15}}_{\sqrt{n}}} = 2.15$$

Hence, we may reject the null hypothesis.

Is this correct ?

Possibly, a better analysis would be to assume that the two types of growth conditions individually represent two distinct normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.

In this situation, how to obtain the distribution of the differences ?

$$\bar{X} = \frac{\sum X_i}{n}, \quad \bar{Y} = \frac{\sum Y_i}{n}, \quad \hat{\Delta} = \bar{X} - \bar{Y}.$$

$\hat{\Delta}$ is an unbiased estimator.

The difference $\hat{\Delta} - \Delta$ will be the numerator of the pivot variable.

By independence, $\operatorname{Var}(\hat{\Delta}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Let $s_1^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2}{n_1 - 1}$ and $S_2 = \frac{\sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_2 - 1}$

Consider

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

which is a weighted average of S_1 and S_2.

Consider

$$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

which is a weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Consider

$$\boxed{S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$ which is a

weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Now, $\dfrac{(n_1-1)S_1^2}{\sigma^2}$ has a $\chi^2(n_1-1)$ distribution

$\dfrac{(n_2-1)S_2^2}{\sigma^2}$ has a $\chi^2(n_2-1)$ distribution

Consider

$$\boxed{S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

which is a weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Now,

$$\left\{ \begin{array}{l} \dfrac{(n_1-1)S_1^2}{\sigma^2} \text{ has a } \chi^2(n_1-1) \text{ distribution} \\[2ex] \dfrac{(n_2-1)S_2^2}{\sigma^2} \text{ has a } \chi^2(n_2-1) \text{ distribution} \end{array} \right.$$

these are independent

Consider

$$\boxed{S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

which is a weighted average of S_1 and S_2. (Check that this is an unbiased estimator for σ^2).

Now,

$$\frac{(n_1-1)S_1^2}{\sigma^2} \text{ has a } \chi^2(n_1-1) \text{ distribution}$$

$$\frac{(n_2-1)S_2^2}{\sigma^2} \text{ has a } \chi^2(n_2-1) \text{ distribution}$$

these are independent

$\therefore$ $\dfrac{(n-2)S_p^2}{\sigma^2}$ has a $\chi^2(n-2)$ distribution.

Due to the independence of S_1^2 and $\overline{X}$, S_2^2 and $\overline{Y}$, and the independence of the samples, S_p^2 is independent of $\overline{X}-\overline{Y}$.

Due to the independence of S_1^2 and $\overline{X}$, S_2^2 and $\overline{Y}$, and the independence of the samples, S_p^2 is independent of $\overline{X}-\overline{Y}$. Therefore, the statistic

$$T = \frac{\left[(\overline{X}-\overline{Y}) - (\mu_1-\mu_2)\right] / \sigma\sqrt{1/n_1 + 1/n_2}}{\sqrt{(n-2)\,S_p^2 / (n-2)\,\sigma^2}}$$

$$= \frac{(\overline{X}-\overline{Y}) - (\mu_1-\mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{n-2}$$

Similarly, we can obtain a distribution in the case when

$$\sigma_1 \neq \sigma_2 .$$

—x——x—