ENV 710: Lecture 7

statistical power

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

statistical inference

decision errors and statistical power

learning goals

- articulate decisions errors - type I and type II errors that can result from hypothesis testing
- understand statistical power
- calculate power and type II error

stuff you should know

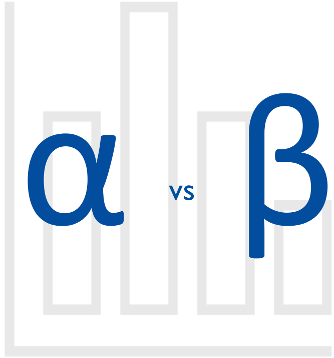
decision errors

two types of errors can result from a hypothesis test

type I error: probability that you will wrongly reject the null hypothesis

type II error: probability that you will wrongly retain the null hypothesis

error probabilities



α is the acceptable probability of committing a Type I error

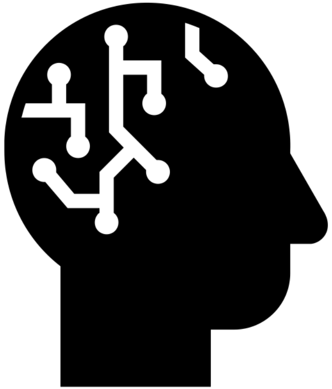
- probability that you will *wrongly reject* the null hypothesis
- α often set at 0.05 (1 out of 20 by chance alone)

β is the acceptable probability of committing a Type II error

- probability that you will *wrongly retain* the null hypothesis
- never really know β because we never know the "true" probability distribution
- α and β are inversely related

type I and type II errors

decision	true state: H_0	true state: H_A
do not reject H_0	correct decision $p = 1\text{-}\alpha$	**type II error** $p = \beta$
reject H_0	**type I error** $p = \alpha$	correct decision $p = 1\text{-}\beta$

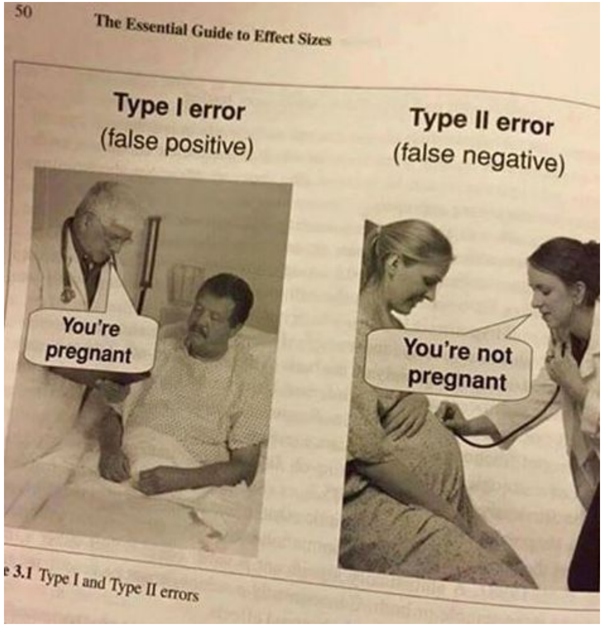


type I and type II errors

false negative

decision	true state: H_0	true state: H_A
do not reject H_0	correct decision $p = 1-\alpha$	**type II error** $p = \beta$
reject H_0	**type I error** $p = \alpha$	correct decision $p = 1-\beta$

false positive



https://www.pinterest.com/pin/281334307955541598/

example

A particular compound is not hazardous in drinking water if it is present at a rate of no more than 25ppm. A watchdog group believes that a certain water source exceeds this standard.

μ: mean amount of compound (ppm)

H_0: $\mu = 25$

H_a: $\mu > 25$

If the group gathers data and formally conducts this test, what are the Type I and II errors in this scenario and their consequences?

example

A particular compound is not hazardous in drinking water if it is present at a rate of no more than 25ppm. A watchdog group believes that a certain water source exceeds this standard.

μ: mean amount of compound (ppm)

H_0: $\mu = 25$

H_a: $\mu > 25$

Type I error

- stating that the water is contaminated when it is safe
- requires resources for unnecessary cleanup

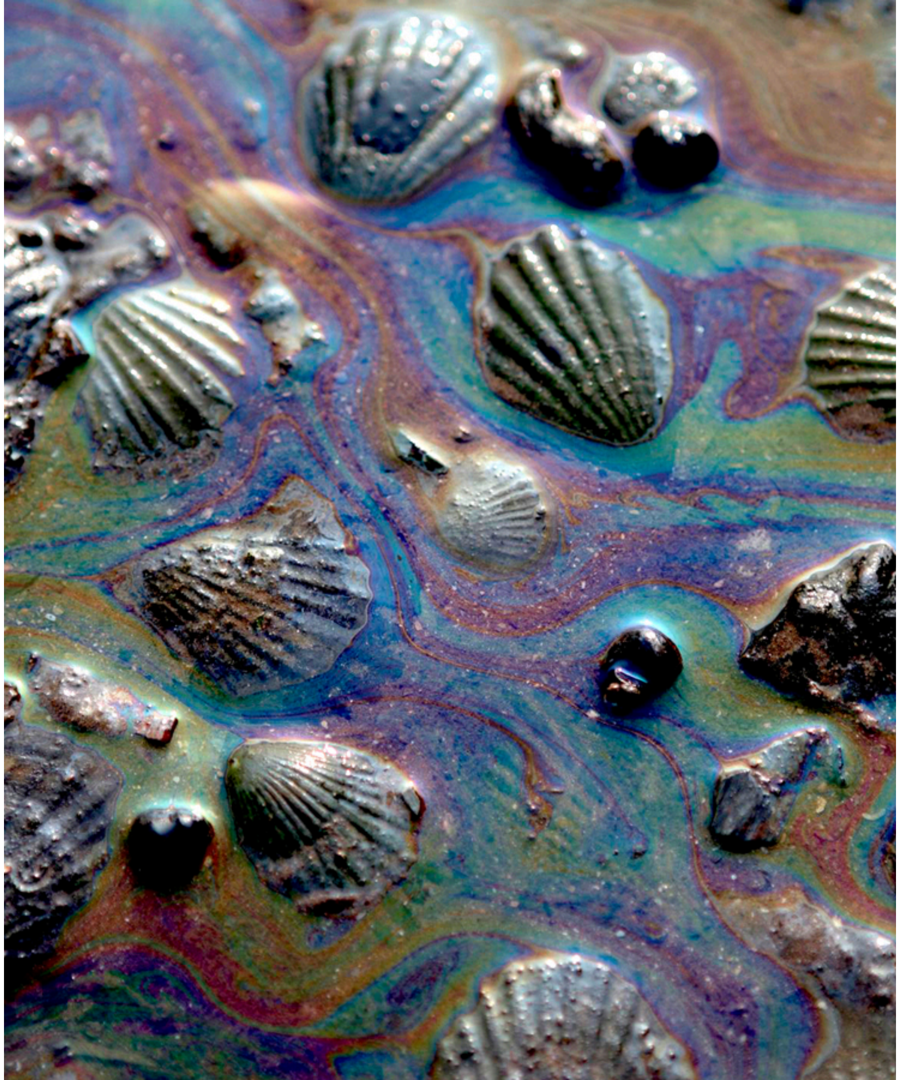
Type II error

- stating that the water is above standard when the water is contaminated (unsafe).
- potential for public health risk

precautionary principle

- what type of error do companies want to minimize?

- what type of error do watchdog groups want to minimize?



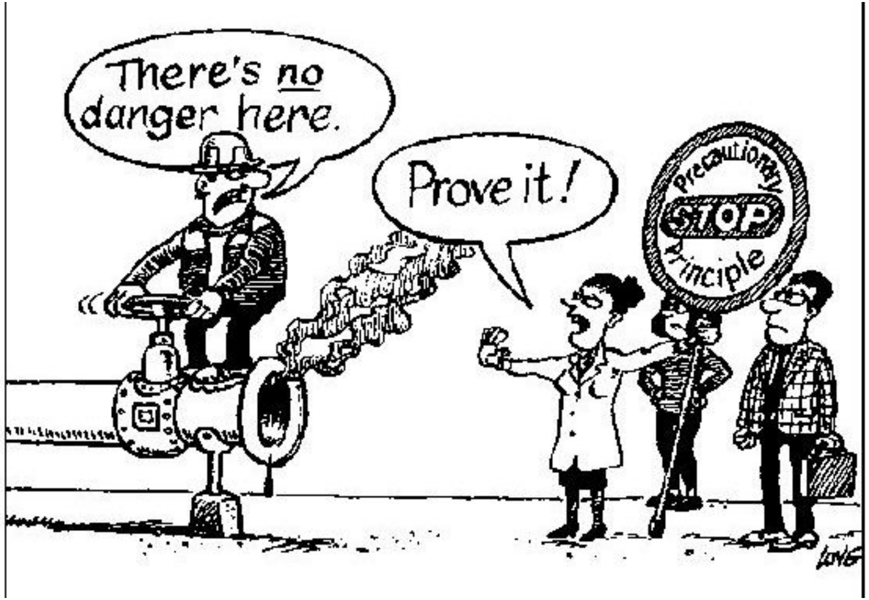
precautionary principle

- what type of error do companies want to minimize?

type I – producer error

- what type of error do watchdog groups want to minimize?

type II – consumer error



https://cagj.org/2020/03/letter-to-editor-regarding-thalidomides-legacy-and-the-precautionary-principle/

null hypothesis

H_0: Defendant is innocent

alternative hypothesis

H_A: Defendant is guilty

burden of proof

present the evidence

collect data

judge the evidence

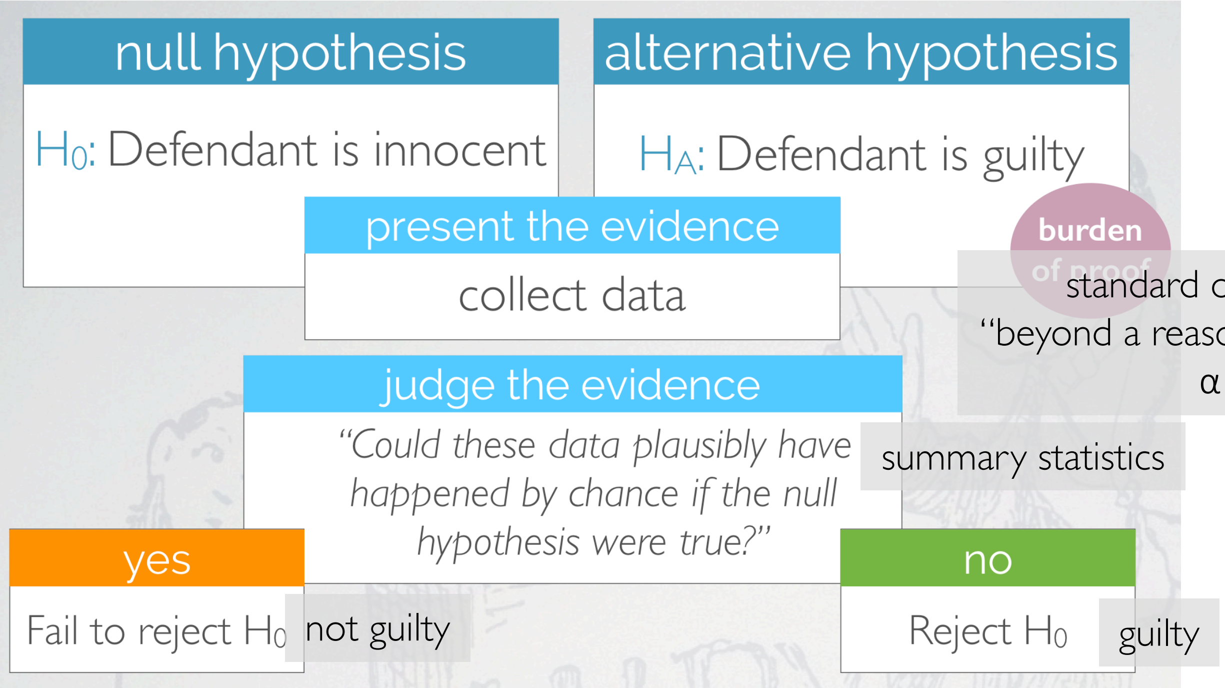
"Could these data plausibly have happened by chance if the null hypothesis were true?"

yes

Fail to reject H_0

no

Reject H_0



What is the Type I error? Convict an innocent person

What is the Type II error? Fail to convict a guilty person

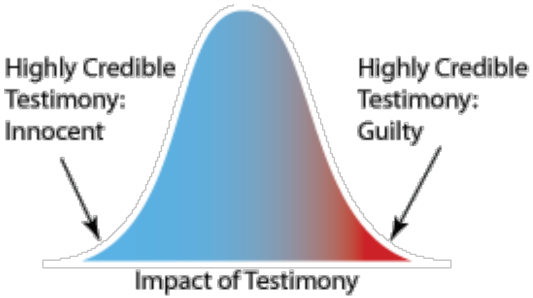
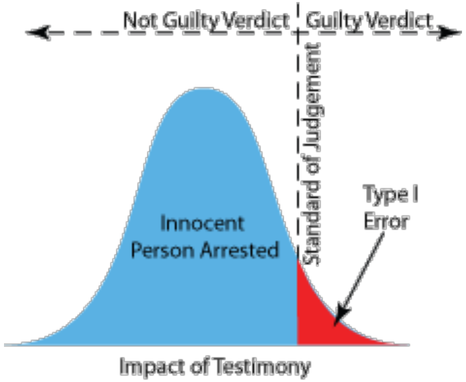


figure 1. Distribution of possible witnesses in a trial when the accused is innocent



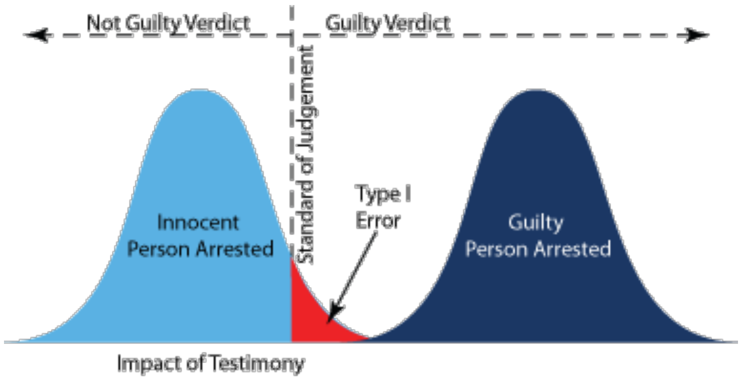


figure 3. Distribution of possible witnesses in a trial showing the probable outcomes with a single witness if the accused is innocent or obviously guilty..

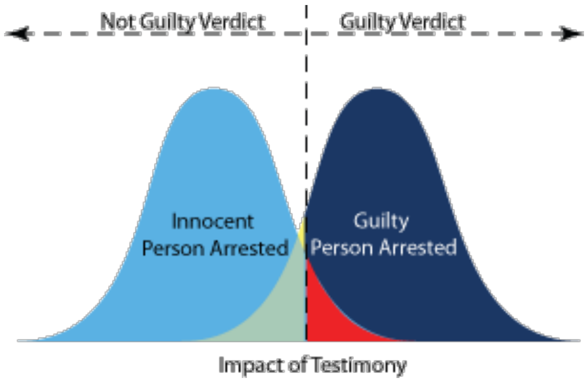


figure 4. Distribution of possible witnesses in a trial showing the probable outcomes with a single witness if the accused is innocent or not clearly guilty..

http://www.intuitor.com/statistics/T1T2Errors.html

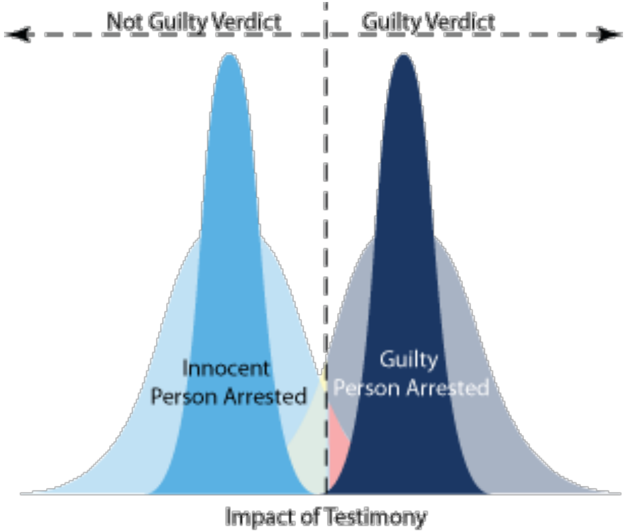


figure 5. The effects of increasing sample size or in other words, number of independent witnesses.

β (type II error) depends on:

- sample size
- population σ^2
- effect size: difference between actual and hypothesized means

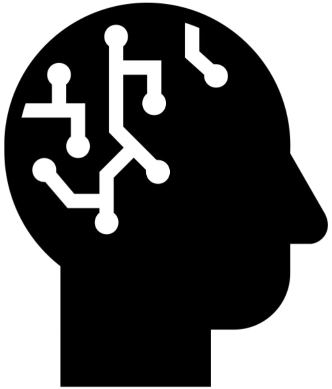
power

statistical power is the probability of correctly rejecting a false null hypothesis

true state

decision

	H_0	H_A
do not reject H_0	correct decision $p = 1\text{-}\alpha$	**type II error** $p = \beta$
reject H_0	**type I error** $p = \alpha$	correct decision $p = 1\text{-}\beta$



statistical power

statistical power is the probability of correctly rejecting a false null hypothesis

- we want statistical tests to have good power to capture the "effect" we are testing (~0.80)
- power of a test increases by…
  - making α larger
  - using a 1-tailed test, rather than 2-tailed test
  - decreasing variance (increase sample size, reduce error)
  - increasing the effect size

without increasing the chances of a Type II error, the only way to increase power is to increase sample size

probability of type II error

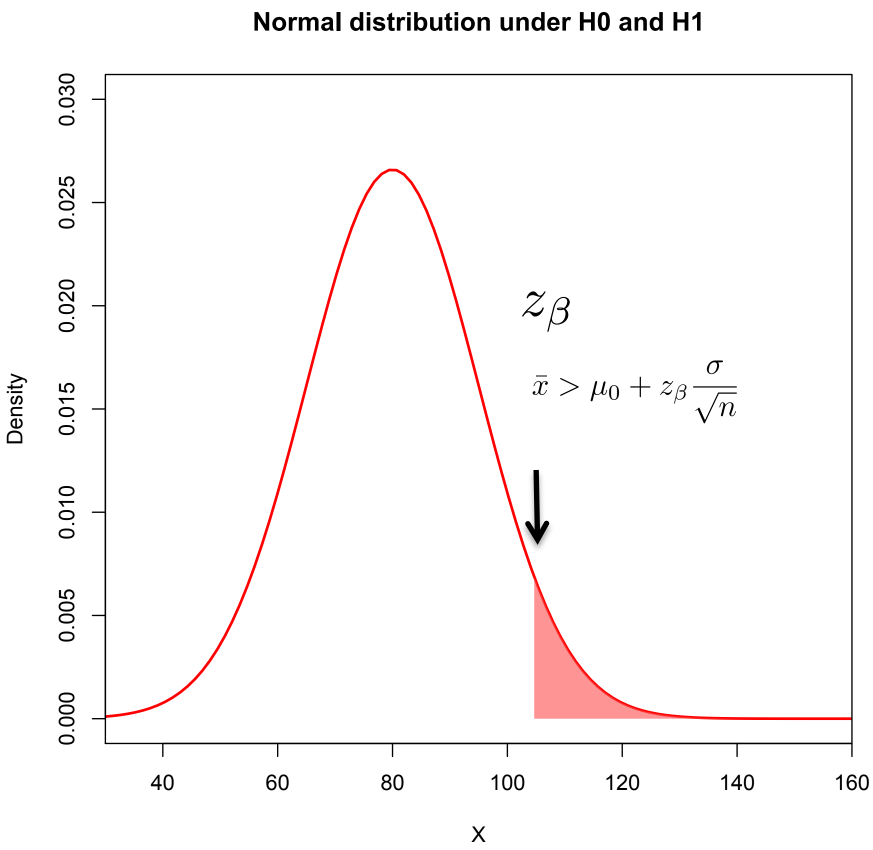
Assume $H_a : \mu_a > \mu_0$

Reject H_0 if

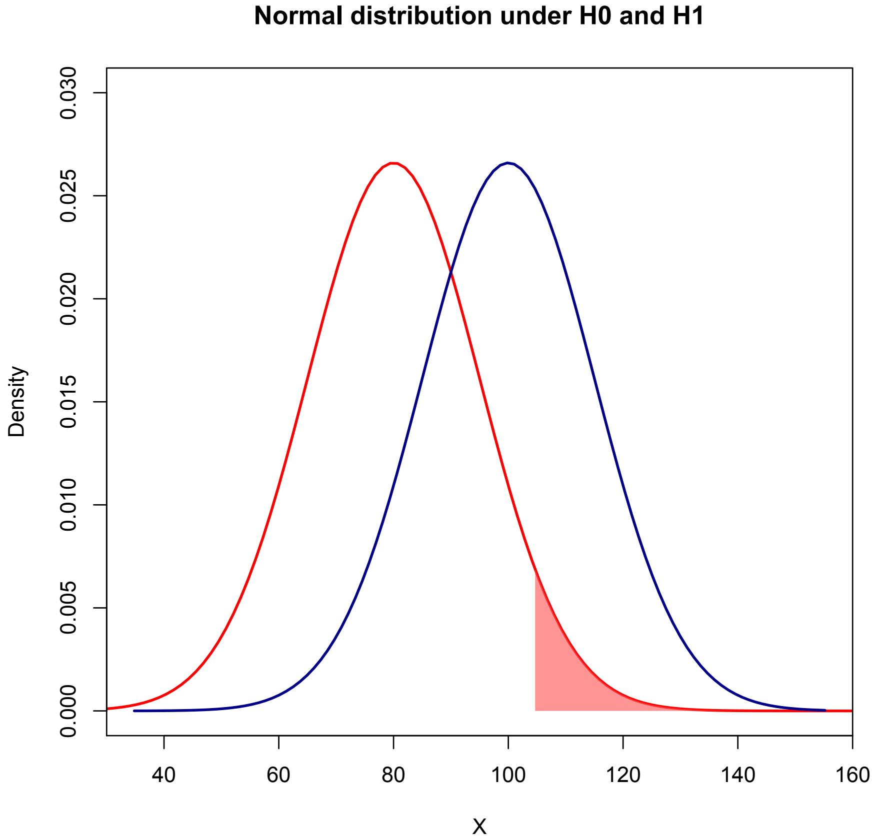
$$(\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}) > z_\beta$$

Equivalently, reject when

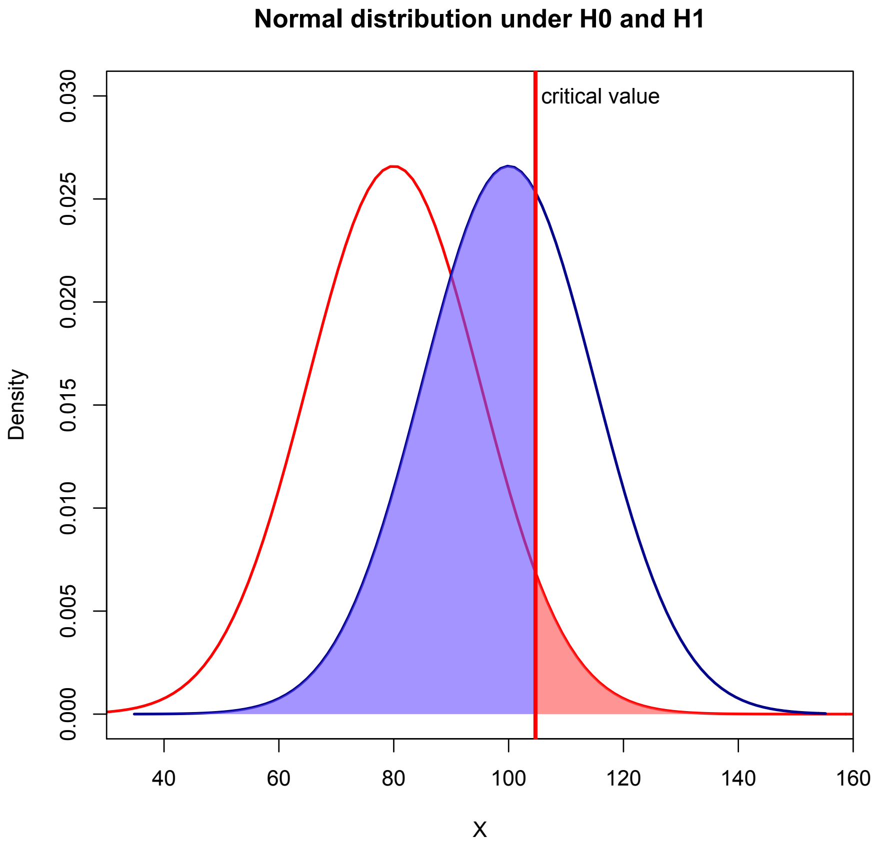
$$\bar{x} > \mu_0 + z_\beta \frac{\sigma}{\sqrt{n}}$$



probability of type II error



probability of type II error



To find β, we need to find the area to the left of:

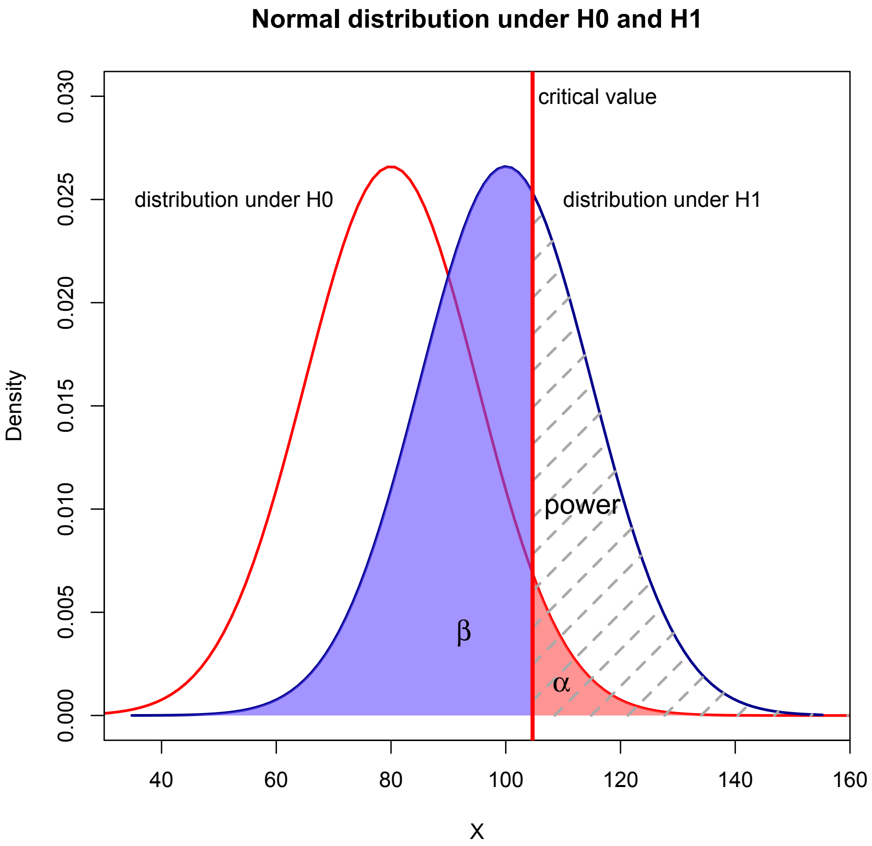
$$P(\bar{x} \leq \mu_0 + Z_\beta \frac{\sigma}{\sqrt{n}} | \mu_1)$$

standardize

$$\frac{\mu_0 + z_\beta \frac{\sigma}{\sqrt{n}} - \mu_1}{\sigma / \sqrt{n}}$$

simplify and we get:

$$\beta = P(z < \frac{(\mu_0 - \mu_1)}{\sigma / \sqrt{n}} + z_\beta)$$



example

EPA river sampling

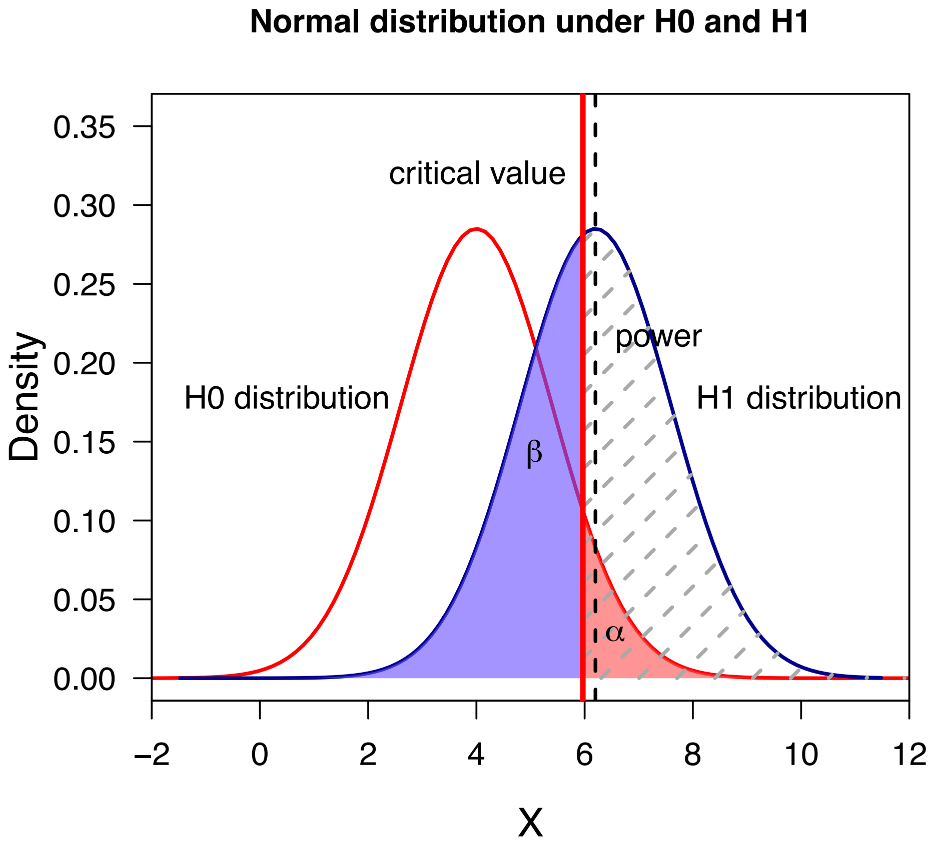
$$N(4, 1.4)$$

$$H_0 : \mu = 4$$

$$H_a : \mu_a = 6.2$$

what is the Type I error?
what is the Type II error?
what is the statistical power?

$$\alpha = 0.08$$



example

EPA river sampling

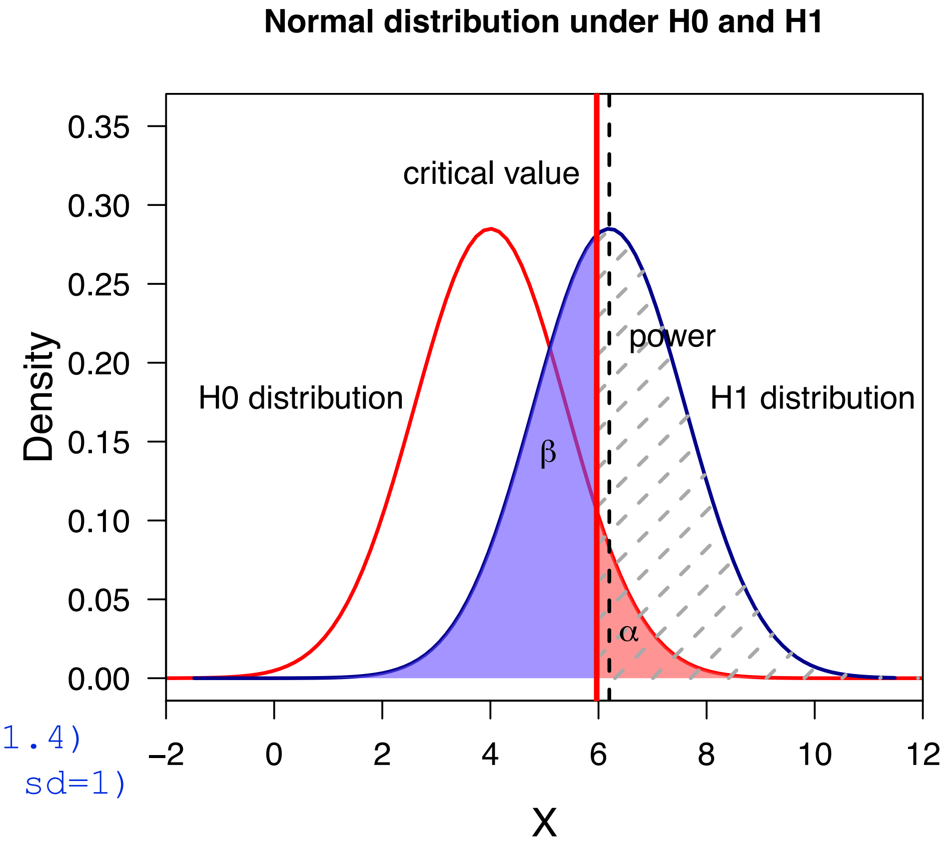
$$N(4, 1.4)$$

$$H_0 : \mu = 4$$

$$H_a : \mu_a = 6.2$$

what is the Type I error?
what is the Type II error?
what is the statistical power?

$$\alpha = 0.08$$




```
crit <- 1-pnorm(x=6, mean = 4, sd = 1.4)
crit <- 1-pnorm(x=(6-4)/1.4, mean=0, sd=1)
```

example

EPA river sampling

$$N(4, 1.4)$$

$$H_0 : \mu = 4$$

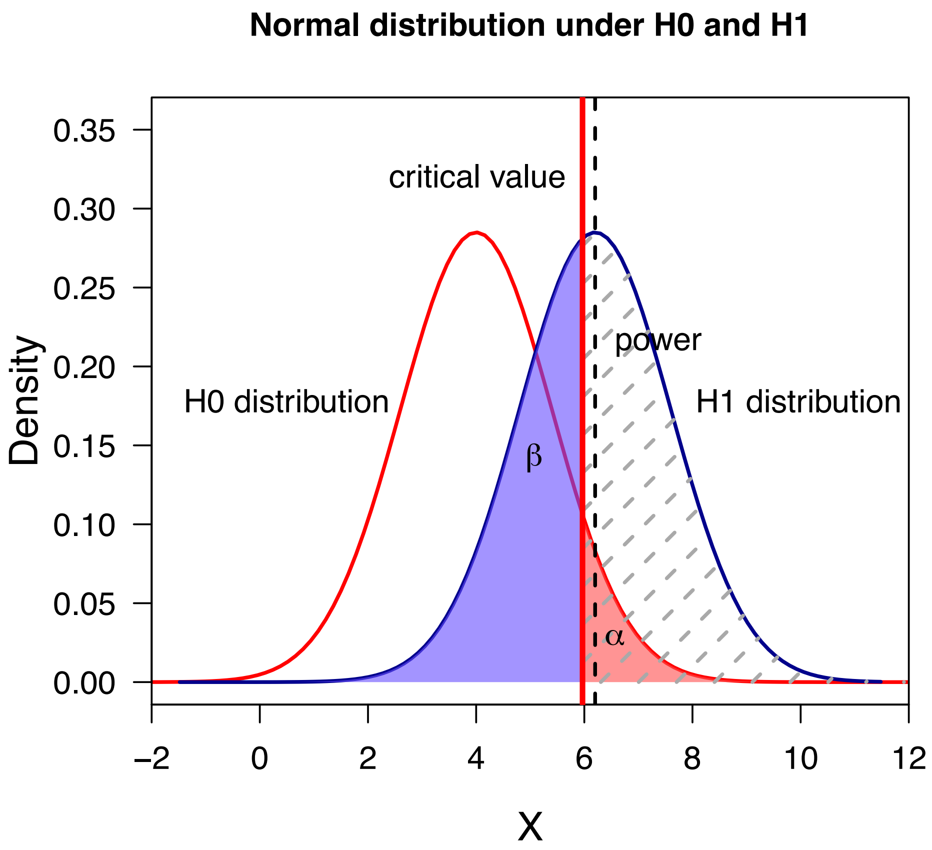
$$H_a : \mu_a = 6.2$$

what is the Type I error?
what is the Type II error?
what is the statistical power?

$$\alpha = 0.08$$

$$\beta = 0.443$$

$$P(\bar{x} < 6 | \mu_\alpha = 6.2) = P(\frac{\bar{x} - 6.2}{1.4} < \frac{6 - 6.2}{1.4} = \phi \cdot -0.143)$$



example

EPA river sampling

$$N(4, 1.4)$$

$$H_0 : \mu = 4$$

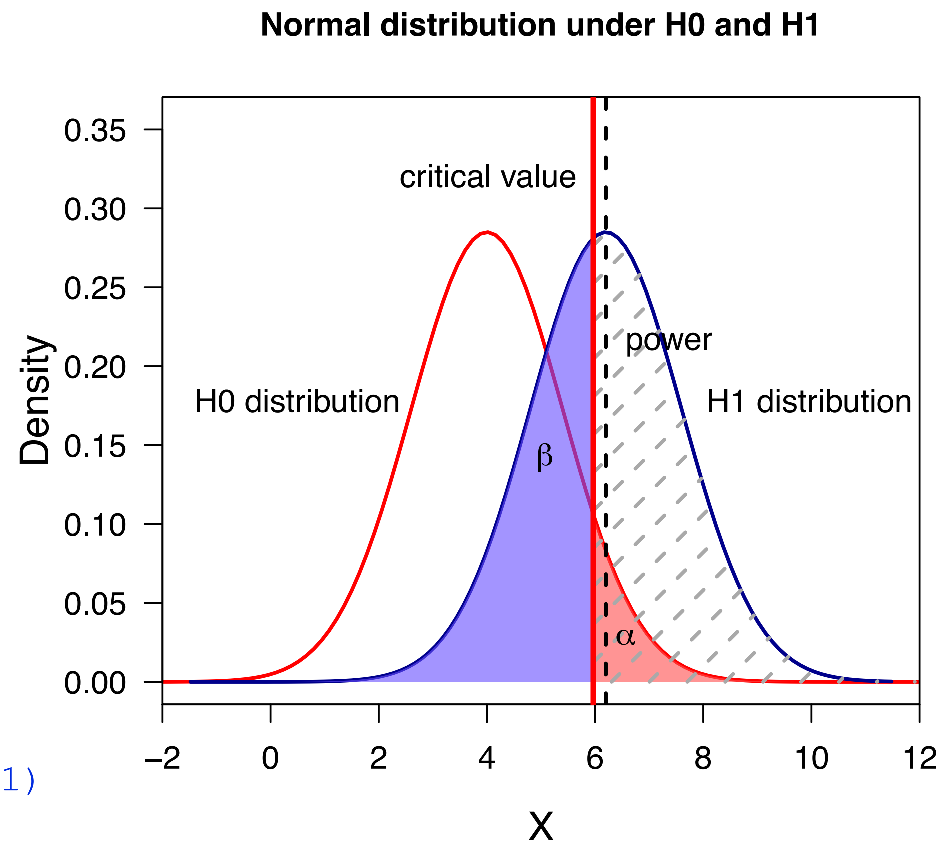
$$H_a : \mu_a = 6.2$$

what is the Type I error?
what is the Type II error?
what is the statistical power?

$$\alpha = 0.08$$

$$\beta = 0.443$$

```
pnorm(6, mean = 6.2, sd = 1.4)
pnorm((6-6.2)/1.4, mean = 0, sd = 1)
```



example

Drinking water

A watchdog group suspects that drinking water is contaminated. They take a sample of 35 water measurements to test:

$$H_0 : \mu = 25 \text{ ppm}$$

$$H_a : \mu > 25 \text{ ppm}$$

If the actual mean concentration is 27 ppm and the standard deviation is 4ppm, and $\alpha = 0.05$, what is the probability of a Type II error?

example

drinking water

A watchdog group suspects that drinking water is contaminated. They take a sample of 35 water measurements to test:

$$H_0 : \mu = 25 \text{ ppm}$$

$$H_a : \mu > 25 \text{ ppm}$$

If the actual mean concentration is 27 ppm and the standard deviation is 4ppm, and $\alpha = 0.05$, what is the probability of a Type II error?

```
qnorm(p=0.95, mean=0, sd=1)=1.645
```

$$\beta = P(Z < \frac{(\mu_0 - \mu_1)}{\sigma/\sqrt{n}} + Z_\beta)$$

$$\beta = P(Z < \frac{25 - 27}{4/\sqrt{35}} + 1.645) = \phi(-1.313) = 0.0946$$

example

drinking water

A watchdog group suspects that drinking water is contaminated. They take a sample of 35 water measurements to test:

$$H_0 : \mu = 25 \text{ ppm}$$

$$H_a : \mu > 25 \text{ ppm}$$

If the actual mean concentration is 27 ppm and the standard deviation is 4ppm, and $\alpha = 0.05$, what is the probability of a Type II error?

```
Xcrit <- qnorm(0.95, 25, 4/sqrt(35))

pnorm(Xcrit, mean = 27, sd = 4/sqrt(35))
```

Post your questions to be answered during lecture