Hypothesis Testing

Statistical Hypothesis Testing

Inference

Samples → Populations

Starting Point:

Ask, is our population mean equal to zero? Or not?

$\mu = 0$ or $\mu \neq 0$

We have two hypotheses:

$\mu = 0$ this one is called the Null Hypothesis
"the mean is not different from 0"

$\mu \neq 0$ this one is called the Alternative Hypothesis
"the mean is not equal to 0"

You could also ask, is the population mean equal to 80kg?

$\mu = 80$ the Null Hypothesis
"the mean is not different from 80"

$\mu \neq 80$ the Alternative Hypothesis
"the mean is not equal to 80"

We usually shorten our statements to:

- H_0: $\mu = 0$
- H_A: $\mu \neq 0$

- H_0: $\mu = 80$
- H_A: $\mu \neq 80$

Important points …

- Hypotheses
  - before data are collected, set up to answer
  - specific research questions.

- Hypotheses
  - statements about the POPULATION
  - NEVER about the SAMPLE

In other words…

- Does the sample we've collected come from a population with a mean = μ ?
  - Two possible answers
    - Yes
    - No

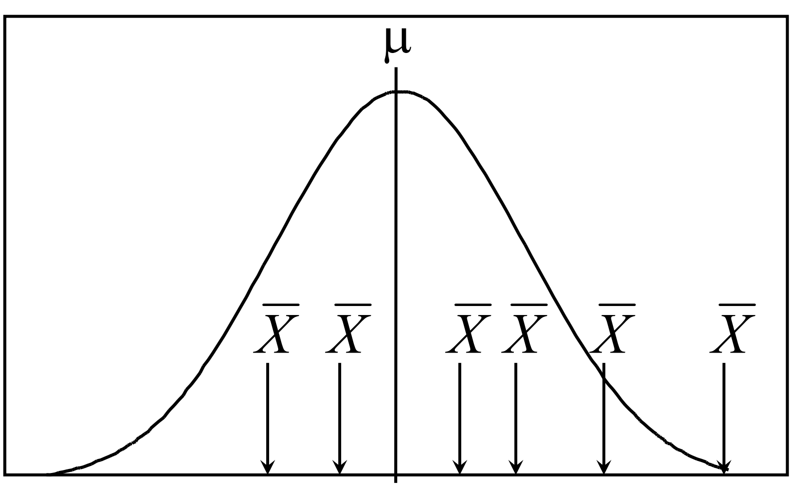
Well, not quite so cut and dried. Really what we end up saying is likely yes, or likely no.

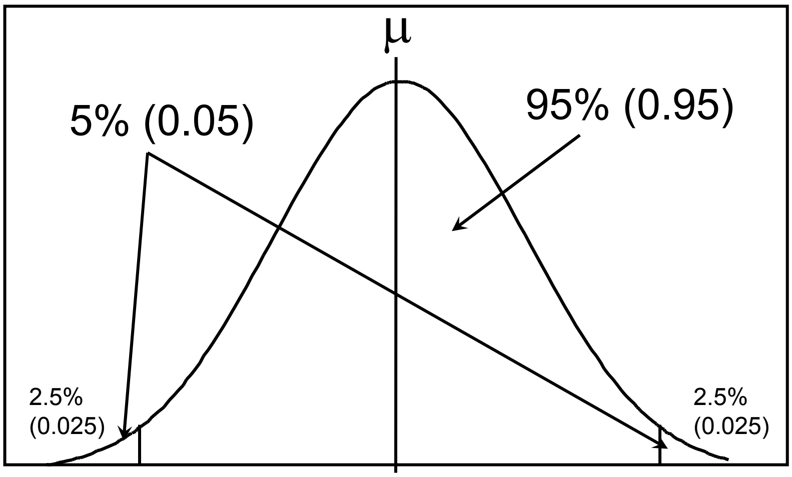
In practice …

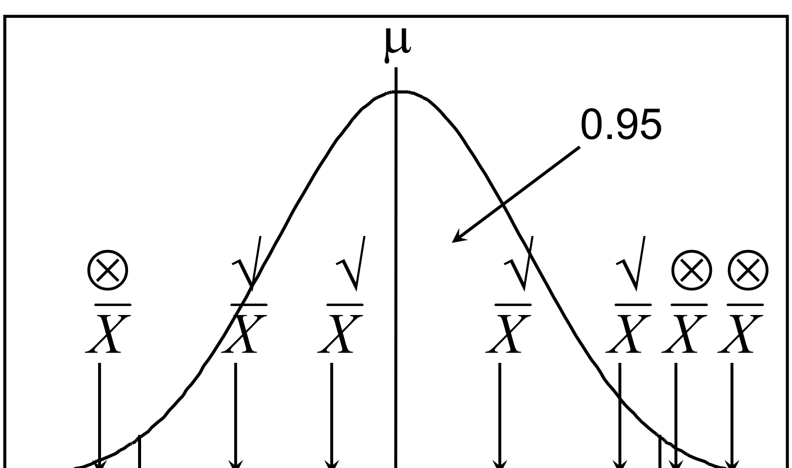
- We never measure/count/… every individual from a population, conclusions about the population

→ samples

– collect sample,
– calculate mean of sample and
– compare to hypothesized population mean




μ
5% (0.05)
95% (0.95)
2.5%
(0.025)
2.5%
(0.025)


μ
0.95

With regard to hypotheses…

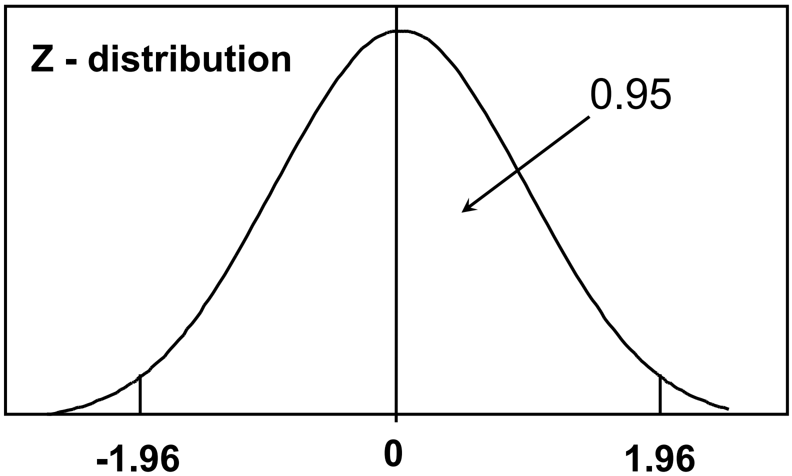
- √ - **Means that would support H_0**
  - **Do Not reject H_0**

- ⊗ - **Means that would not support H_0**
  - **Reject H_0**

A formal test

- We can use our Z to do a formal test

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$



An example

- CO detectors must sound an alarm at 10mg/m^3
- A CO detector manufacturer has sampled 18 of his CO detectors from the assembly line.
- Record the CO concentration at which the alarm sounded

Do the detectors 'alarm' at 10 mg /m^3?

10.25	10.54
10.37	10.33
10.66	10.48
10.47	10.68
10.56	10.40
10.22	10.39
10.44	10.26
10.38	10.32
10.63	10.54

H_0: μ = 10.0 mg/m^3

H_A: $\mu \neq$ 10.0 mg/m^3

10.25	10.54
10.37	10.33
10.66	10.48
10.47	10.68
10.56	10.40
10.22	10.39
10.44	10.26
10.38	10.32
10.63	10.54

H_0: μ = 10.0 mg/m^3

H_A: $\mu \neq$ 10.0 mg/m^3

$\sigma^2 = 1.0434\ (\text{mg/m}^3)^2$

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

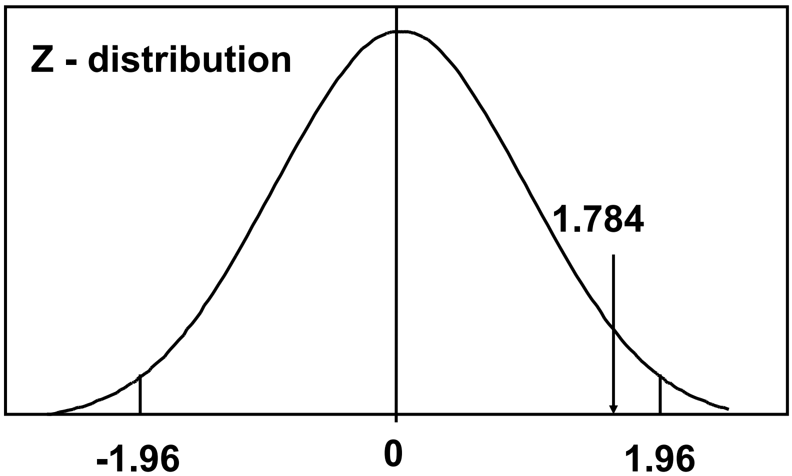
Preliminary calculations

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{1.0434}{18}} = 0.241 mg/m^3$$

$$\overline{X} = \frac{\Sigma X_i}{n} = \frac{187.92}{18} = 10.44 mg/m^3$$

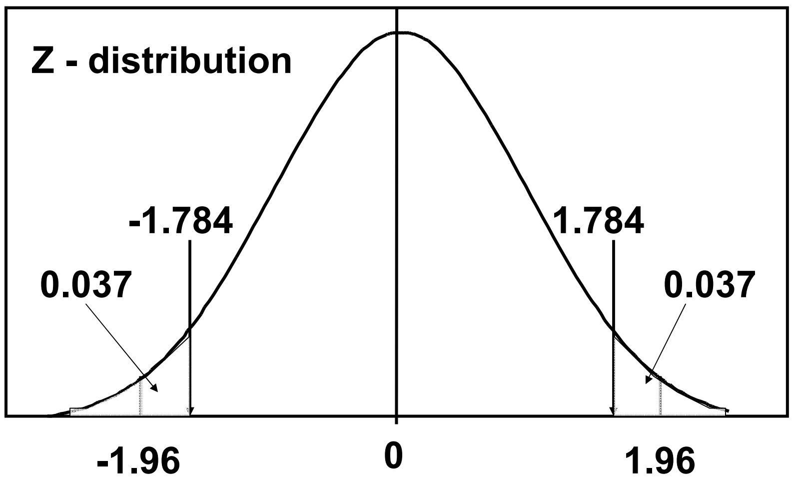
Calculate Z …

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{10.44 - 10}{0.241} = 1.784$$



Alternatively …

- Could ask what is the
  - probability of Z ≥ 1.784 and
  - probability of Z ≤ -1.784 ?



Therefore the probability …

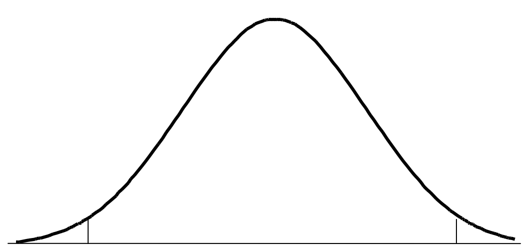
- Of getting a |Z| of that value or more extreme is 0.037 + 0.037 = 0.074
- 0.074 > 0.05
- And we DO NOT reject Ho.

--> Why did we use both ‘tails’ in our test?

We were concerned with knowing whether the alarm values were **different** from 10mg/m^3.

We didn't really care if they were larger or smaller, just **different**

--> were the readings in the range that we would call the 0.05 extremes - those in the upper 0.025 and lower 0.025 of the distribution - or were they in the 0.95 'bulk' of the distribution?

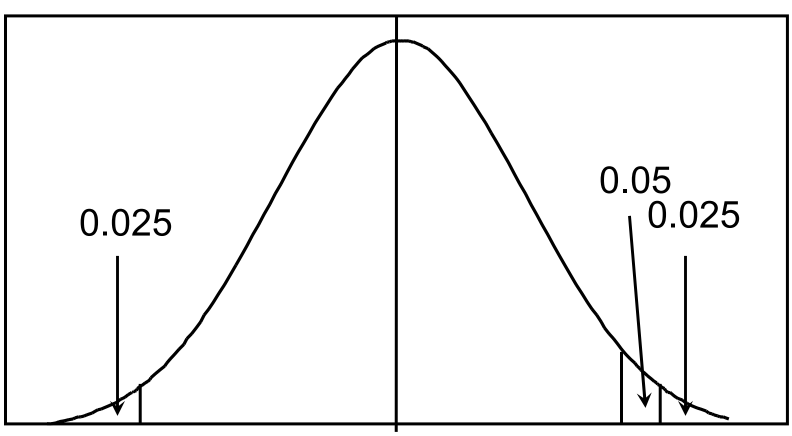


We could have phrased our question differently.

Really, as consumers, we want to know if the alarm goes off at a concentration that is too high for safety.

So, instead of asking ‘are the values **different** from 10 mg/m^3?’ you would as ‘are the values **greater** than 10mg /m^3?’

Translates to - are the values in the upper extreme?



One tailed hypotheses …

look slightly different.

- H_0: $\mu \leq 10$ mg/m^3
- H_A: $\mu > 10$ mg/m^3

Previously, we saw that

- the probability of $Z \geq 1.784$ was 0.037

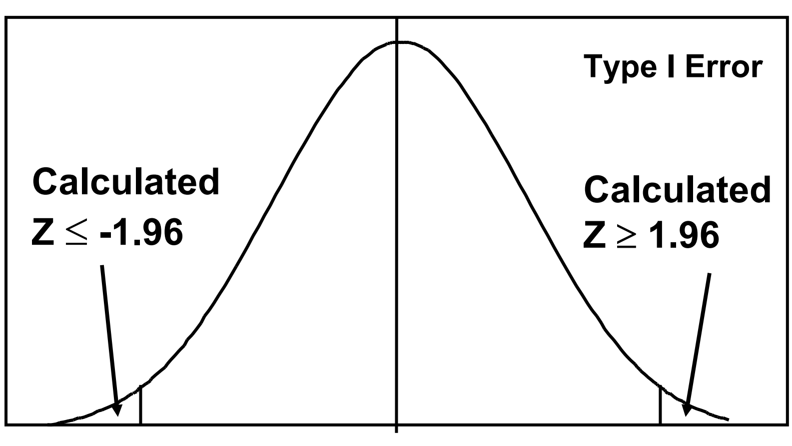
0.037 < 0.05, so this value is in the 0.05 (5%) in the upper 'tail',

therefore, **REJECT** your H_o, the value is greater than 10mg/m^3.

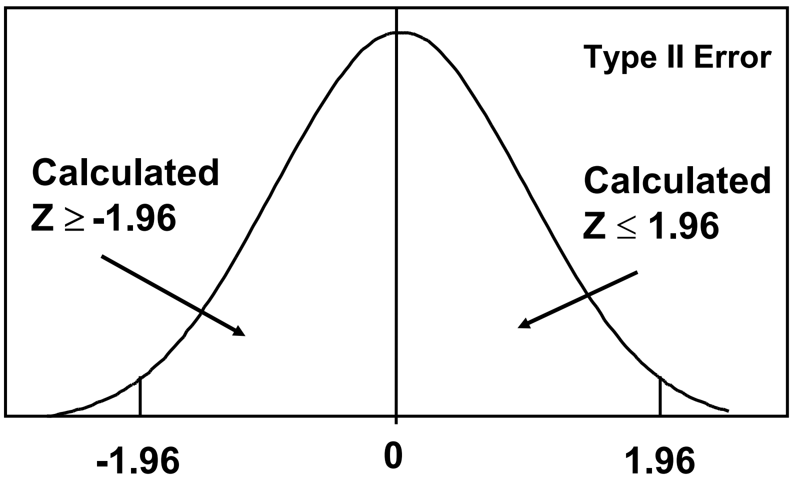
Mistake happen!!!

- We use 0.05 or 5% as our criterion for saying too extreme (reject null) or not extreme enough (do not reject null).
- The corresponding Z-value is 1.96 - called the critical Z.

Should NOT reject H_o, but you do.



Should reject H_o, but you Do Not.

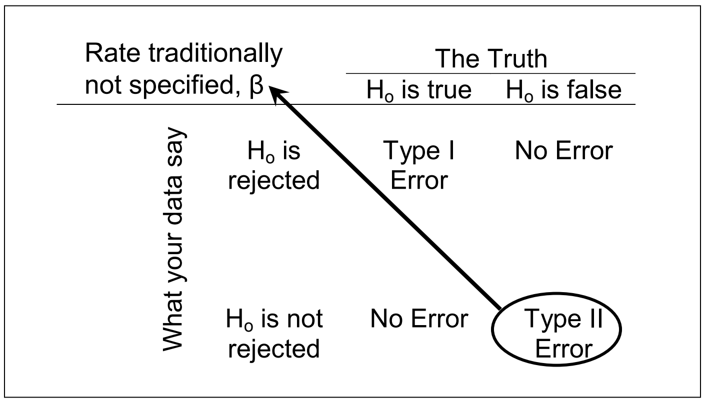


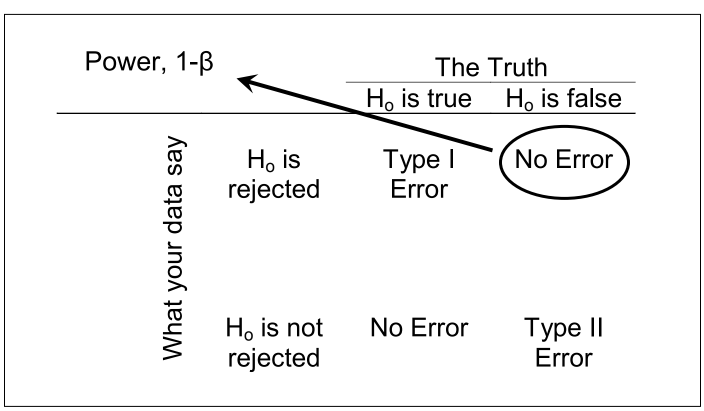
This will happen occasionally, just due to chance.

5% of the time, α

		The Truth	
		H_o is true	H_o is false
What your data say	H_o is rejected	Type I Error	No Error
	H_o is not rejected	No Error	Type II Error

This will happen occasionally, just due to chance.





Significance level
$1-\alpha$

		The Truth	
		H_o is true	H_o is false
What your data say	H_o is rejected	Type I Error	No Error
	H_o is not rejected	No Error	Type II Error