Example of the fallibility of hypothesis testing

A researcher suspects that people who are overweight will exaggerate how many minutes of jogging it takes to burn off 100 calories.

To find out, she samples 50 overweight and 50 average weight students from around campus. She asks them "how long would a typical 25 year old male have to jog for to burn off 100 calories". She obtains the following:

overweight: mean = 62 minutes
average weight: mean = 28 minutes

Obviously, the sample means are different. The real question is: why are they different?

Ho: $\mu_{overweight} = \mu_{average}$ (i.e. difference is due to sampling error alone)
Ha: $\mu_{overweight} \neq \mu_{average}$ (i.e. difference is due to something in addition to SE)

She uses *Hypothesis testing* to find out whether the data support Ho or Ha. Theoretically, she knows that 1 of 4 outcomes are possible:

stats say ↓	**Ho is true in reality** overweight & average weight people estimate the same	**Ho is false in reality** overweight & average weight people estimate differently
retain Ho (sampling error only)	she correctly concludes that weight makes no difference	she mistakenly concludes that weight makes no difference (i.e. weight *does* make a difference) **type II**
reject Ho (SE + some-thing else)	she mistakenly concludes that weight makes a difference (i.e. weight does *not* make a difference) **type I**	she correctly concludes that overweight & average weight people do *not* estimate the same

Whether she rejects or retains Ho, there is a chance she will make a mistake. If she says there's a weight effect when in fact there is none, she's made a **type I error**. If she says there's no weight effect and there is, she's made a **type II error**.