A motivating example

- Scenario [see Sec 26.1] A senator is sponsoring a bill that will change the tax laws, maintaining that the changes proposed will be revenue neutral.

- To examine this claim, the Treasury Department randomly selects 100 files from a database of 100,000 returns. For each they record the revenue difference by computing (tax paid under proposed law) - (tax paid under current law). From this sample they observe an average revenue difference of -$219 and an SD of $725.

Questions

- How significant is the Treasury Department’s sample against a claim of revenue neutrality?

- One way you might argue the significance of the sample is to compute the 95% CI for the sample: $-219 +/- 2*72.5$. Since this interval, which runs from -364 to -74, misses “0,” the expected difference [ie EV] for a claim of neutrality, it would seem that the neutrality claim is questionable. There is some chance that the sample was poorly chosen so as to miss the actual difference, but this should happen only 5 times in 100 samples.
Questions

• Rephrasing the last question suggests a more structured and refined way of measuring ‘significance.’ Assuming the claim of revenue neutrality, what are the chances of seeing a sample average of -219 or something more extreme in a randomly selected sample?
  
  – This is more structured because it makes clear a fixed benchmark against which we will make a measure of significance. A skeptic could argue that any confidence interval is bad based on a belief that the sample, which produced the confidence interval, was strongly influenced by chance error.

A test of significance

• The form of the last question gives the basis for many tests of significance. Each requires a null hypothesis [NH]
  
  – The null hypothesis expresses the idea that any observed event is due to chance variability [in the sampling procedure].
  
  – Every NH stands in opposition to an alternate hypothesis: The observed difference is due to something real.

• To complete the test, the NH is used to describe a box model. With a box model, methods of normal approximation can be used to estimate the chance, or P-value, of the observed event, called a test or z-statistic, or something more extreme.

Notes on P-values

• The P-value of a test is the chance of getting a particular test statistic or one more extreme assuming the null hypothesis is correct.
  
  – It is not the chance of the null hypothesis being correct.

• By common agreement a P-value of less than 5% allows us to reject the hypothesis. In this case the observed event or result is called significant.

• If the P-value is less than 1%, then the result is called highly significant.