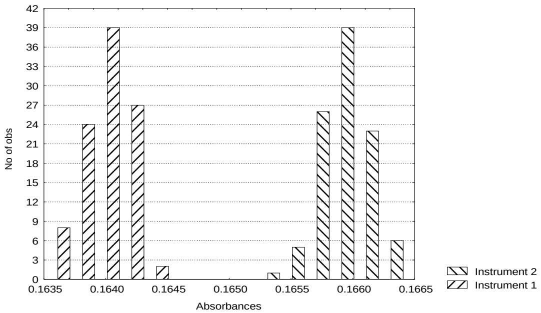
5.4. Statistical significance, Equivalence and Importance

The result of any statistical test can only be that a difference or effect is **statistically** significant.

Statistical significance has **no relation with practical significance or importance**. Judging this requires background knowledge.

<u>Example</u>: Comparing two spectrometers



Equivalence testing

- Define an "**acceptable difference**" or **acceptance interval**
- If the confidence interval is **completely included** in the acceptance interval, the two groups are declared **statistically equivalent**

E.g. Spectrometer comparison

95% CI for the difference in absorbance: [0.00193, 0.00207]

*0 is not included ⇒ **statistically significant***

But if any difference smaller than 0.005 is not relevant

⇒ acceptance interval is [-0.005, 0.005]

⇒ the confidence interval is completely included in the acceptance interval

*⇒ instruments are **equivalent***

5.4. Statistical significance, Equivalence and Importance

If you'd had an infinite number of data, any difference will be statistically significant, since nearly nothing will be exactly equal to something else. If your sample size is one or two, only huge differences will be declared statistically significant. So "statistically significant" means no more or no less than "I had **enough data to denote the observed difference as too unlikely to be pure coincidence**". A better term would be statistically **discernible.**
Unfortunately, very often the word "statistical" is dropped and significant is often interpreted as relevant or important. Whether something is relevant or important is decided by the investigator, not by statistics.

In the comparison on the left, there is an obvious difference between the two instruments and it is definitely statistically significant (see next page) but at the same time it is very small in size and very likely not important (although we would need background information to be sure).

<u>Note</u>:

When a "no difference" hypothesis is rejected at the 5% level, it is statistically incorrect to say that you are **95% sure that there <u>is</u>** a difference. This can be readily seen if you write down the full statistical statement: "Under assumption of H_0 and the distribution associated with H_0, we are 95% sure that H_0 and its associated distribution can be rejected". This cannot be correct: first you define an area covering 95% under a curve and in the same phrase you say that the curve is not appropriate.
However, if you want your conclusions to be read and understood: just report that you are 95% "sure".

Inferential Statistics

A much more relevant concept than statistical significance is that of **Equivalence**. Given that nothing is exactly the same as something else, the true question then becomes: "**is the difference acceptable**". This is precisely what is being investigated in a test of equivalence.
Although the associated classical testing procedures can be quite a bit more complex to understand (e.g. because H_0 now becomes "there is NO equivalence") again the use of confidence intervals is almost trivial. Since the confidence interval consists of all acceptable values for the true value of a difference (at a specified confidence level), if the entire confidence interval is included in the **acceptance interval** (the range of acceptable differences), we can conclude that the items under comparison are equivalent.

<u>Note</u>: it can be shown that for equivalence testing a 90% confidence interval yields an equivalence test on the 95% confidence level.
(cfr. "Testing statistical hypotheses of equivalence", S. Wellek, Chapmann & Hall, 2003)

Although equivalence testing in the classical hypothesis testing framework is quite a bit more complex than the "simple" classical tests, it is again trivial when based on confidence intervals.