Seven steps of hypothesis testing

Let us perform hypothesis testing through the following 7 steps of the procedure:

Step 1 : Specify the null hypothesis and the alternative hypothesis

Step 2 : What level of significance?

Step 3 : Which test and test statistic to be performed?

Step 4 : State the decision rule

Step 5 : Use the sample data to calculate the test statistic

Step 6 : Use the test statistic result to make a decision

Step 7 : Interpret the decision in the context of the original question

To guide us through the steps, let us use the following example.

Assume a food laboratory analyzed a certified reference freeze-dried food material with a stated sodium (Na) content of 250 mg/kg. It carried out 7 repeated analyses and obtained a mean value of 274 mg/kg of sodium with a sample standard deviation of 21 mg/kg. Now we want to know if the mean value of 274 mg/kg is significantly larger than the stated amount of 250 mg/kg. If so, we will conclude that the reported results of this batch of analysis were of bias and had consistently given higher values than expected.

Step 1 : Specify the null hypothesis and the alternative hypothesis

The **null hypothesis** H_o is the statement that we are interested in testing. In this case, the null condition is that the mean value is 250 mg/kg of sodium.

The **alternative hypothesis** H_1 is the statement that we accept if our sample outcome leads us to reject the null hypothesis. In our case, the alternative hypothesis is that the mean value is not equal to 250 mg/kg of sodium. In other words, it can be significantly larger or smaller than the value of 250 mg/kg.

So, our formal statement of the hypotheses for this example is as follows:

H_o : $\bar{x} = 250$ mg/kg (i.e., the certified value)

H_1 : $\bar{x} \neq 250$ mg/kg (i.e., indicating that the laboratory has a bias result)

Step 2 : What level of significance

The level of significance is the probability of rejecting the null hypothesis by chance alone. This could happen from sub-sampling error, methodology, analyst's technical competence, instrument drift, etc. So, we have to decide on the level of significance to reject the null hypothesis if the sample result was unlikely given the null hypothesis was true.

Traditionally, we define the *unlikely* (given by symbol α) as 0.05 (5%) or less. However, there is nothing to stop you from using $\alpha = 0.1$ (10%) or $\alpha = 0.01$ (1%) with your own justification or reasoning.

In fact, the significance level sometimes is referred to as the probability of a **Type I** error. A Type I error occurs when you falsely reject the null hypothesis on the basis of the above-mentioned errors. A **Type II** error occurs when you fail to reject the null hypothesis when it is false.

Step 3 : Which test and test statistic?

The test statistic is the value calculated from the sample to determine whether to reject the null hypothesis. In this case, we use Student's t-test statistic in the following manner:

$$\mu = \bar{x} \pm t_{(\alpha=0.05, v=n-1)} \frac{s}{\sqrt{n}}$$

or $$t_{(\alpha=0.05, v=n-1)} = \frac{|\bar{x}-\mu|\sqrt{n}}{s}$$

By calculation, we get a t-value of 3.024 at the significance level of $\alpha = 0.05$ and $v = (7-1)$ or 6 degrees of freedom for $n = 7$ replicates.

Step 4 : State the decision rule

The decision rule is always of the following form:

Reject H_o if …..

We reject the null hypothesis if the test statistic is larger than a critical value corresponding to the significance level in step 2.

There is now a question in H_1 on either one-tailed (> or <) or two-tailed (≠ not equal) tests to be addressed. If we are talking about either “greater than” or “smaller than”, we take the significance level at $\alpha = 0.05$ whilst for the unequal (that means the result can be either larger or smaller than the certified value), the significance level at $\alpha = 0.025$ on either side of the normal curve is to be studied.

As our H_1 is for the mean value to be larger or smaller than the certified value, we use the 2-tailed t-test for $\alpha = 0.05$ with 6 degrees of freedom. In this case, the t-critical value at $\alpha = 0.05$ and 6 degrees of freedom is 2.447 from the Student’s t-table or from using the Excel function “=T.INV.2T(0.05,6)” or “=TINV(0.05,6) in older Excel version.

That means the decision rule would be stated as below:

Reject H_o if t > 2.447

Step 5 : Use the sample data to calculate the test statistic

Upon calculation on the sample data, we have got a t-value of 3.024 at the significance level of $\alpha = 0.05$ and $v = (7-1)$ or 6 degrees of freedom for $n = 7$ replicates.

Step 6 : Use the test statistic to make a decision

When we compare the result of step 5 to the decision rule in step 4, it is obvious that 3.024 is greater than the t-critical value of 2.447, and so we reject the null hypothesis. In other words, the mean value of 274 mg/kg is significantly different from the certified value of 250 mg/kg. Is it really so? We must go to step 7.

Step 7 : Interpret the decision in the context of the original question

Since hypothesis testing involves some kind of probability under the disguise of significance level, we must interpret the final decision with caution. To say that a result is “statistically significant” sounds remarkable, but all it really means is that it is more than by chance alone.

To do justice, it would be useful to look at the actual data to see if there are one or more high outliers pulling up the mean value. Perhaps increasing the number of replicates might show up any undesirable data. Furthermore, we might have to take a closer look at the test procedure and the technical competence of the analyst to see if there were any lapses in the analytical process. A repeated series of experiment should be able to confirm these findings.