t-Test

When the sample size is small, we use the t-test and the t distribution instead of the z-test and the normal distribution to test hypotheses about the unknown population (box model). Here are the steps.

1. When the number of measurements n is small, use SD^+ instead of SD.

$$\mathrm{SD}^+ = \mathrm{SD}\sqrt{\frac{n}{n-1}}$$

2. Calculate the SE for the sum, average, or percent using SD^+ and the number of measurements.

$$\mathrm{SE}_{\mathrm{sum}} = \sqrt{n}\,\mathrm{SD}^+ \qquad \mathrm{SE}_{\mathrm{av}} = \frac{\mathrm{SD}^+}{\sqrt{n}} \qquad \mathrm{SE}_{\%} = \frac{\mathrm{SD}^+}{\sqrt{n}}$$

3. Calculate the test statistic.

$$\frac{\text{observed} - \text{expected}}{\mathrm{SE}}.$$

4. Calculate the P-value using the t-distribution with $n-1$ degrees of freedom.

Example: The five calibration readings of a spectrophotometer are: 78, 83, 68, 72 and 88. If the machine reads close to 70, then it is ready for use. We can use the t-test to test the null hypothesis "the average of the machine is 70" as follows.

$$\begin{aligned}
\text{average of the readings} &= \frac{78+83+68+72+88}{5} \\
&= 77.8 \\
\text{SD of the readings} &= \sqrt{\frac{(78-77.8)^2+\cdots+(88-77.8)^2}{5}} \\
&= 7.22 \\
\mathrm{SD}^+ \text{ of the readings} &= \sqrt{(5/4)}\ \ 7.22 \\
&= 8.07 \\
\mathrm{SE}_{\mathrm{av}} &= \frac{8.07}{\sqrt{5}} \\
&= 3.6 \\
\text{test statistic} &= \frac{77.8-70}{3.6} \\
&= 2.2 \\
P\text{-value} &= 1 - \mathrm{pt}(2.2,\ \ \mathrm{df}=4) \\
&= 4.6\%
\end{aligned}$$

If we had used the z-test, we would have calculated $\mathrm{SE}_{\mathrm{av}} = 3.22$, $z = 2.4$ and $P = 1 - \mathrm{pnorm}(2.4) = 0.9\%$. Use of the t-test instead of the z-test corrects for the small sample size and makes a big difference in the computed P-value.

Worksheet: t-Test

1. Compare the t-distribution to the standard normal distribution.

   a) Calculate the area under the normal curve between ± 1 using $\textbf{pnorm}(1) - \textbf{pnorm}(-1)$

   b) Calculate the corresponding area under the t-distribution with 5 degrees of freedom using $\textbf{pt}(1,\ \textbf{df}{=}5) - \textbf{pt}(-1,\ \textbf{df}{=}5)$

   c) How many degrees of freedom do you need before the areas under the normal and t-distributions agree to two decimal places?

   d) Repeat b)–c) for the area between ± 2.

2. Find the specified area under the t distribution with 3 degrees of freedom.

   a) $-1 < t < 1$

   b) $-2 < t < 2$

   c) left of 2.02

   d) between -2.02 and 2.02

   e) right of 0.5

   f) between -3 and 1

3. Repeat #2 for 10 degrees of freedom.

   a) $-1 < t < 1$

   b) $-2 < t < 2$

   c) left of 2.02

   d) between -2.02 and 2.02

   e) right of 0.5

   f) between -3 and 1

4. Each (hypothetical) data set below represents spectrophotometer readings. In each case, do a t-test to see whether the instrument is properly calibrated or not.

   a) 71, 68, 79

   b) 71, 68, 79

   c) 71