MCE380: Measurements and Instrumentation Lab

Chapter 3: Statistical Data Analysis

Part II

Topics:
Normality Tests
Confidence Intervals for the Mean
The Student's T Distribution
Chauvenet's Criterion

Reference: Holman, CH 3.

Cleveland State University
Mechanical Engineering
Hanz Richter, PhD

Checking the Normality of Data

For perfect normal distributions, the chances that a reading will fall within a given number of standard deviations is known (Tables 3.3 and 3.4 in Holman).
We can check the normality of our data by evaluating how closely it follows these characteristics. We could simply plot a cumulative histogram (the "integral" of the histogram) and compare it with the integral of the normal PDF. But it would be difficult to compare the two curved traces.

Instead, we can plot the percentage of readings at or below x against x using a special scale that gives a straight line for perfectly normal data. We can use a "probability paper" to do this.

Confidence Intervals for the Mean

When we take a sample and calculate the average, this is an approximation of the true population mean.
We want to have an idea of how close our sample average and the true mean can be, in a statistical sense.
We wish to make statements like: "the length is 10 $\pm$ 0.01 mm with a confidence level of 99 percent".
The meaning of this statement is that the length will be between 9.99 and 10.01 mm 99% of the time.

Sometimes we express the interval in terms of standard deviations: 10$\pm z\sigma$ with a confidence level of p percent. For large data samples, the relationship between z and p is given in Table 3.4

Student's T Distribution

William S. Gosset, (a.k.a. Student, 1876-1937), found a reliable way of estimating confidence intervals.
Measurements are be expressed as $x_m \pm \frac{t\sigma}{\sqrt{n}}$ with a confidence level of p percent.
Find the values of t for the desired confidence level from Table 3.7 in Holman.
Note that the degrees of freedom is defined to be $\nu = n - 1$.

Also note that "level of significance" is just 100% minus confidence level.

Chauvenet's Criterion

This criterion is useful to determine if apparently "bad" data points should be removed from the sample.
Each reading has an actual deviation from the calculated mean. In the theoretical Gaussian distribution, these deviations have a well-defined probability. For example, a deviation of σ or more from the mean of has a probability of 31.7% (see Table 3.4). A deviation of 3.3σ or more from the mean has a probability of only 0.l%.
In Chauvenet's criterion, we find the probability of the actual deviations and compare them with $1/2n$. If the probability is lower than $1/2n$, we reject the corresponding data point.

The maximum allowable deviations are listed as a ratio to σ in Table 3.5.

Exercises

Use the data collected in class with the measuring microscope (wall thickness of sections of tubing). Use the supplied probability paper to check how close the distribution is to being Gaussian. Then use Student's T distribution to obtain confidence intervals for the wall thickness with levels of 50, 90 and 99.9%. Use Chauvenet's criterion to see if some points could be removed. Repeat the confidence interval calculation and see if there's improvement (tradeoff: more centralized data, but less points!)