Round Bearing Experiment

Measurement Equation (reconsidered)

$$X_i = x_t + \cancelto{0}{b} + E_i \qquad i = 1, 2, \cdots N$$

This equation is appropriate when all the round bearings have the same diameter, x_t, and measurement errors are the only source of variablity in the data.

However, the real problem is that the variablity in bearing diameters is caused by the manufacturing process and by measurement errors. A new measurement model is required.

$$X_i = D + \cancelto{0}{b} + E_i \qquad i = 1, 2, \cdots N$$

where

1. D is a normally distributed RV with mean μ_D a variance σ_D^2. Shorthand: $D = N(d, \mu_D, \sigma_D^2)$
2. b is the bias; assume $b = 0$.

3. E_i are independent and identically distributed RV's with $N(e, \mu_E, \sigma_E^2)$.

4. $X_i = N(x, \mu_x, \sigma_x^2)$

The next step is to determine values of μ_x and σ_x^2 from the measured data.

The function of Statistics is to provide the equations for computing the best estimates of μ_x and σ_x^2 from measured data. These equations are called Estimators.

Since our data set contains a finite number of measurements, $N = 20$, the statistical is called Finite Statistics.

Estimators

Sample Mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{1}$$

Sample Variance

$$s_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad (2)$$

Sample Standard Deviation

$$s_x = \sqrt{s_x^2}$$

Equations (1) and (2) have a dual meaning. If we are given a data set drawn from a single random sample, each equation gives a single number. However, if we consider a large number of random samples, each data set generated by the random samples will give a different value for $\bar{x}$ and s_x^2. Therefore, $\bar{x}$ and s_x^2 are random variables, and each data set gives a different realization of the RV. Our notation requires

$$\bar{x} \rightarrow \bar{X} \text{ and } s_x^2 \rightarrow S_x^2$$

Paper House

EXAMPLE 4.4

Consider the data of Table 4.1. (a) Compute the sample statistics for this data set. (b) Estimate the interval of values over which 95% of the measurements of the measurand should be expected to lie. (c) Estimate the true mean value of the measurand at 95% probability based on this finite data set.

KNOWN

Table 4.1
$N = 20$

ASSUMPTIONS

Data set follows a normal distribution

FIND

$\bar{x}, \bar{x} \pm t S_x$, and $\bar{x} \pm t S_{\bar{x}}$

SOLUTION

The sample mean value is computed for the $N = 20$ values by the relation

$$\bar{x} = \frac{1}{20} \sum_{i=1}^{20} x_i = 1.02$$

This, in turn, is used to compute the sample standard deviation

$$S_x = \sqrt{\frac{1}{19} \sum_{i=1}^{20} (x_i - 1.02)^2} = 0.16$$

The degrees of freedom in the standard deviation are $\nu = N - 1 = 19$. From Table 4.4 at 95% probability, $t_{19,95}$ is 2.093. Then, the interval of values in which 95% of the measurements of x should lie is given by equation (4.15):

$$x_1 = \bar{x} \pm (2.093 \times 0.16) = 1.02 \pm 0.33 \quad (95\%)$$

Accordingly, if a 21st data point were to be taken, there is a 95% probability that its value would lie between 0.69 and 1.35.

The true mean value is estimated by the sample mean value. However, the precision interval for this estimate is $\pm t_{19,95} S_{\bar{x}}$, where

$$S_{\bar{x}} = \frac{S_{\bar{x}}}{N^{1/2}} = \frac{0.16}{(20)^{1/2}} = 0.04$$

Then, from equation (4.17),

$$x' = \bar{x} \pm t_{19,95} S_{\bar{x}} = 1.02 \pm 0.08 \quad (95\%) \tag{4.18}$$