Answer on Question #84693 – Math – Statistics and Probability

Question

Calculate CM, MST, SST in any data.

Solution

Suppose we draw one sample of each of the k populations. Then n_i denotes the sample size of the sample from population i $(i = 1, 2, \dots, k)$.
Let x_{ij} be the jth measurement $(j = 1, 2, \dots, n_i)$ in the ith sample.

The different sum of squares:
The total variance in the experiment is the variance of all k samples together. It is based on the total sum of squares

$$Total\ SS = \sum \left(x_{ij} - \overline{x}\right)^2 = \sum {x_{ij}}^2 - \frac{\left(\sum x_{ij}\right)^2}{n}$$

where $\overline{x}$ is the overall mean, from all k samples and $n = n_1 + n_2 + \cdots + n_k$.

Let the grand total G be the sum of all observations from all samples

$$G = \sum x_{ij}$$

Then the correction for the mean CM is given by

$$CM = \frac{\left(\sum x_{ij}\right)^2}{n} = \frac{G^2}{n}$$

Now the $Total\ SS$ is partitioned in two components. The first component, the sum of squares for treatments (SST), measures the variation among the k sample means (from one sample to the others):

$$SST = \sum n_i \left(x_{ij} - \overline{x}\right)^2 = \sum \frac{{T_i}^2}{n_i} - CM$$

where T_i is the total of the observations for treatment i:

$$T_i = \sum_j x_{ij}$$

The second component, called the sum of squares for error (SSE) is used to measure the pooled variation within the k samples:

$$SSE = (n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2 + \cdots + (n_k - 1){s_k}^2$$

Assuming the variation in all k populations is the same, than this is an estimate for the variation inside the populations.
Now we can proof algebraically, that

$$Total\ SS = SST + SSE$$

Therefore, you only need to calculate two of them and can find the third one with this equation. Each of the sum of squares, when divided by its appropriate degrees of freedom, provides an estimate of the variatio in the experiment.
Since $Total\ SS$ involves n squares its degrees of freedom are $df = n - 1$.

Since SST involves k squares its degrees of freedom are $df = k - 1$.
Since SSE involves $(n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = n - k$ squares its degrees of freedom are $df = n - k$.

Find that

$$df(total) = df(SST) + df(SSE)$$

Now the mean squares (MS) are calculated by dividing the sum of squares by the degrees of freedom $MS = SS/df$.

ANOVA Table for k independent Random Samples

Source				
Treatments	$k-1$	SST	$MST = SST/(k-1)$	MST/MSE
Error	$n-1$	SSE	$MSE = SSE/(n-k)$	
$Total$	$n-1$	$Total\ SS$		

$$Total\ SS = \sum x_{ij}{}^2 - CM = (\text{Sum of squares of all x} - \text{values}) - CM$$

$$CM = \frac{\left(\sum x_{ij}\right)^2}{n} = \frac{G^2}{n}$$

$$SST = \sum \frac{T_i{}^2}{n_i} - CM \qquad MST = \frac{SST}{k-1}$$

$$SSE = Total\ SS - SST \qquad MSE = \frac{SSE}{n-k}$$

G = Grand total of all n observations
T_i = Total of all observations in sample i
n_i = Number of observations in sample i
$n = n_1 + n_2 + \cdots + n_k$.