# Application of Chi-Square test in WEGO 2.0

Suppose we have 3 valid input files from samples s1, s2 and s3. To test if there is a sample difference in gene enrichments involved in biological process, WEGO constructed a contingency table of 3×2 (see Table 1) separating the genes by samples or annotated as GO terms of biological process.

If $O_{ij}$ is the observed number and $E_{ij}$ is the expected number of genes corresponding to the j$^{th}$ sample and the i$^{th}$ condition (whether annotated as GO terms), then chi-square is:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1),$$

where $E_{ij}$ is obtained as $E_{ij} = \frac{R_i \times C_j}{TotalSum}$ (2).

Table 1 Example of contingency table of 3 input samples

| | **s1** | **s2** | **d3** | **Total** |
|---|---|---|---|---|
| In biological process | g11 | g12 | g13 | R1=g11+g12+g13 |
| Not in biological process | g21 | g22 | g23 | R2=g21+g22+g23 |
| Total | C1=g11+g21 | C2=g12+g22 | C3=g13+g23 | Total Sum= g11+g12+g13+ g21+g22+g23 |

The P-value is obtained when projecting the value of chi-square to the chi-square distribution with 2 degrees of freedom.