Statistical Relations

We start by positing a set of N observations (e.g. operator databases), with values x_i and observational uncertainties σ_i, $i = 1, \ldots, N$. In consequence each observation also has an associated Z score $Z_i = x_i/\sigma_i$.

The overall mean is $\bar{x} = \sum(x_i/\sigma_i^2)/\sum(1/\sigma_i^2)$. This can be derived by minimizing the total squared error $E = \sum\left([x_i - \bar{x}]/\sigma_i^2\right)^2$.

The uncertainty of $\bar{x}$ is $\sigma = 1/\sqrt{\sum(1/\sigma_i^2)}$. The derivation is as follows, using $V(x)$ to denote the variance of x:

$$\sigma^2 \equiv V(\bar{x}) = V\left(\frac{\sum(x_i/\sigma_i^2)}{\sum(1/\sigma_i^2)}\right) = \frac{V(\sum(x_i/\sigma_i^2))}{\left(\sum(1/\sigma_i^2)\right)^2} = \frac{1}{\left(\sum(1/\sigma_i^2)\right)^2}\sum V(x_i/\sigma_i^2)$$
$$= \frac{1}{\left(\sum(1/\sigma_i^2)\right)^2}\sum\left(\frac{V(x_i)}{\sigma_i^4}\right) = \frac{1}{\left(\sum(1/\sigma_i^2)\right)^2}\sum\frac{\sigma_i^2}{\sigma_i^4} = \frac{\sum(1/\sigma_i^2)}{\left(\sum(1/\sigma_i^2)\right)^2} = \frac{1}{\sum(1/\sigma_i^2)}. \quad (1)$$

By definition the composite Z is the overall mean $\bar{x}$ divided by its uncertainty: $Z_c \equiv \bar{x}/\sigma$. This means that

$$Z_c \equiv \frac{\bar{x}}{\sigma} = \frac{\sum(x_i/\sigma_i^2)/\sum(1/\sigma_i^2)}{1/\sqrt{\sum(1/\sigma_i^2)}} = \frac{\sum(x_i/\sigma_i^2)}{\sqrt{\sum(1/\sigma_i^2)}}. \quad (2)$$

The raw χ^2, with N degrees of freedom, is just the sum of the squares of the individual Z scores.

$$\chi^2 = \sum Z_i^2 = \sum(x_i^2/\sigma_i^2). \quad (3)$$

The variability χ^2, with $N-1$ degrees of freedom, measures the degree of departure of the individual observations relative to their expected degree of variation. A Z score for each observation's departure from the mean is constructed as $(x_i - \bar{x})/\sigma_i$; these Z scores are then squared and summed:

$$\chi_v^2 = \sum\left(\frac{x_i - \bar{x}}{\sigma_i}\right)^2 \quad (4)$$

Although there are N such Z scores, one degree of freedom is lost due to the fact that $\bar{x}$ is constructed as a minimizing fit to the values of the x_i. If all of the σ_i have the same value, χ_v^2 is just $N-1$ times the ratio of sample variance to theoretical variance; thus $\chi_v^2/(N-1)$ in the full form can be seen to be a generalization of the variance concept to observations with different intrinsic precisions.

The raw χ^2, χ_v^2, and Z_c are closely related as shown by derivation 5.

$$\chi_v^2 = \sum\left(\frac{x_i-\bar{x}}{\sigma_i}\right)^2 = \sum\frac{x_i^2 - 2\bar{x}x_i + \bar{x}^2}{\sigma_i^2} = \sum\frac{x_i^2}{\sigma_i^2} - 2\bar{x}\sum\frac{x_i}{\sigma_i^2} + \bar{x}^2\sum\frac{1}{\sigma_i^2}$$
$$= \chi^2 - 2\bar{x}\left(\bar{x}\sum\frac{1}{\sigma_i^2}\right) + \bar{x}^2\sum\frac{1}{\sigma_i^2} = \chi^2 - \bar{x}^2\sum\frac{1}{\sigma_i^2} = \chi^2 - \frac{\bar{x}^2}{\sigma^2} = \chi^2 - Z_c^2. \quad (5)$$

Thus, to obtain the variability chi-squared χ_v^2 from the raw χ^2, simply subtract Z_c^2 and remove one degree of freedom. The raw χ^2 can, in essence, be partitioned into the contribution from mean shift (Z_c^2) and the contribution by inter-observation variability (χ_v^2).