Q3

$$\chi^2 = \sum \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

$$= \sum \frac{(x_i - \mu)^2}{\sigma^2}$$

(Since all data points have the same μ and σ)

$$\frac{d\chi^2}{d\mu} = -2 \sum \frac{(x_i - \mu)}{\sigma^2} = 0$$

$$\Rightarrow \mu = \frac{\sum x_i}{N}$$ Maximum likelihood estimate of mean

(where N: n is the number of data points).

The error in the maximum likelihood estimate of mean $\Rightarrow$

$$Var(\mu) = \langle \mu \rangle^2 - \langle \mu^2 \rangle$$

$$= Var\left(\frac{\sum x_i}{N} \right)$$

$$= \frac{1}{N^2} \left(\sum Var(x_i) \right)$$

$$= \frac{1}{N^2} (N\sigma^2) = \sigma^2 / N$$

(Using properties of variance)

Error on maximum likelihood estimate of the mean =

$$\sqrt{\frac{\sigma^2}{N}} = \frac{\sigma}{\sqrt{N}}$$

For a general case of each data point having average μ and std σ_i,

$$\chi^2 = \sum \frac{(x_i - \mu)^2}{\sigma_i^2}$$

$$\frac{d\chi^2}{d\mu} = 0 \Rightarrow -2\sum \frac{(x_i - \mu)}{\sigma_i^2} = 0$$

$$\Rightarrow \sum \frac{x_i}{\sigma_i^2} - \sum \frac{\mu}{\sigma_i^2} = 0$$

$$\Rightarrow \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum 1/\sigma_i^2} = \mu \Leftarrow$$ Maximum likelihood estimate of mean.

$$= \frac{\sum w_i x_i}{\sum w_i} \quad \text{where } w_i = 1/\sigma_i^2$$

If half the data points have an error of $2\sigma^2 \Rightarrow$

$$\mu = \frac{\sum_{i=1}^{N/2} \frac{x_i}{2\sigma^2} + \sum_{j=N/2+1}^{N} \frac{x_j}{\sigma^2}}{\sum_{i=1}^{N/2} 1/2\sigma^2 + \sum_{j=N/2+1}^{N} 1/\sigma^2}$$

$$\mu = \frac{\sum_{i=1}^{N/2} \frac{x_i}{2} + \sum_{j=N/2+1}^{N} x_j}{\frac{N}{4} + \frac{N}{2}}$$

$$\mu = \frac{4}{3N}\left(\sum_{i=1}^{N/2} \frac{x_i}{2} + \sum_{j=N/2+1}^{N} x_j\right)$$

$$\mathrm{Var}(\mu) =$$

$$\frac{16}{9N^2}\left(\sum_{i=1}^{N/2} \frac{2\sigma^2}{4} + \sum_{j=N/2+1}^{N} \sigma^2\right)$$

$$= \frac{16}{9N^2}\left(\frac{N}{2}\frac{\sigma^2}{2} + \frac{N}{2}\sigma^2\right)$$

$$= \frac{16}{9N^2}\left(\frac{3N\sigma^2}{4}\right) = \frac{4N\sigma^2}{3N^2} = \frac{4}{3}\frac{\sigma^2}{N}$$

True error in this case $= \sigma' =$

$$\sqrt{\frac{4}{3}\frac{\sigma^2}{N}} = \frac{2\sigma}{\sqrt{3N}} \simeq 1.15\frac{\sigma}{\sqrt{N}}$$

Thus, this error is 1.15 times of the original error of $\frac{\sigma}{\sqrt{N}}$

Underweighting 1% of data by a factor of 100 $\Rightarrow$

$$\mu = \frac{\sum w_i x_i}{\sum w_i}$$

$$= \frac{\sum_{i=1}^{N/100} \frac{x_i}{100\sigma^2} + \sum_{j=N/100+1}^{N} \frac{x_j}{\sigma^2}}{\sum_{i=1}^{N/100} \frac{1}{100\sigma^2} + \sum_{j=N/100+1}^{N} \frac{1}{\sigma^2}}$$

$$= \frac{\sum_{i=1}^{N/100} \frac{x_i}{100} + \sum_{j=N/100+1}^{N} x_j}{\left(\frac{N}{100}\right)\left(\frac{1}{100}\right) + \left(\frac{99N}{100}\right)}$$

$$= \frac{10^4}{9901N}\left(\sum_{i=1}^{N/100} \frac{x_i}{100} + \sum_{j=N/100+1}^{N} x_j\right)$$

$$\approx \frac{100}{99N}\left(\sum_{i=1}^{N/100} \frac{x_i}{100} + \sum_{j=N/100+1}^{N} x_j\right)$$

$$\mathrm{Var}(\mu) =$$

$$\mathrm{Var}\left(\frac{100}{99N} \left(\sum_{i=1}^{N/100} \left(\frac{x_i}{100} \right) + \sum_{j=N/100+1}^{N} x_j \right) \right)$$

$$= \frac{100^2}{99^2 N^2} \left(\frac{1}{100^2} \sum_{i=1}^{N/100} \mathrm{Var}(x_i) + \sum_{j=N/100+1}^{N} \mathrm{Var}(x_j) \right)$$

$$= \frac{100^2}{99^2 N^2} \left(\frac{1}{100^2} \times \frac{N}{100} \times \sigma^2 + \frac{99 N \sigma^2}{100} \right)$$

$$= \frac{100^2}{99^2 N^2} \frac{1}{100^2} \left(\frac{N}{100} \sigma^2 + 9900 N \sigma^2 \right)$$

$$= \cancel{100^2} \frac{1}{99^2 N^2} \left(9900.01 N \sigma^2 \right)$$

$$\simeq \frac{9900 \sigma^2}{99^2 N} = \frac{100}{99} \frac{\sigma^2}{N}$$

$$\text{Error} = \sqrt{\frac{100}{99} \frac{\sigma^2}{N}} = \cancel{\frac{10\sigma}{\sqrt{99N}}} \quad \frac{10\sigma}{\sqrt{99 N}}$$

$$= 1.005 \frac{\sigma}{\sqrt{N}}$$

Overweighting 1% of data points by a factor of 100 ⟹

$$\mu = \frac{\sum w_i x_i}{\sum w_i} = \frac{\sum_{i=1}^{N/100} \frac{100 x_i}{\sigma^2} + \sum_{j=N/100+1}^{N} \frac{x_j}{\sigma^2}}{\sum_{i=1}^{N/100} \frac{100}{\sigma^2} + \sum_{j=N/100+1}^{N} \frac{1}{\sigma^2}}$$

$$= \frac{\sum_{i=1}^{N/100} 100 x_i + \sum_{j=N/100+1}^{N} x_j}{\sum_{i=1}^{N/100} 100 + \sum_{j=N/100+1}^{N} 1}$$

$$= \frac{\sum_{i=1}^{N/100} 100 x_i + \sum_{j=N/100+1}^{N} x_j}{(100)\left(\frac{N}{100}\right) + \frac{99N}{100}}$$

$$= \frac{100}{199N} \left(\sum_{i=1}^{N/100} 100 x_i + \sum_{j=N/100+1}^{N} x_j \right)$$

$$\mathrm{Var}(\mu) =$$

$$\mathrm{Var}\left(\frac{100}{199N}\left(\sum_{i=1}^{N/100} 100x_i + \sum_{j=N/100+1}^{N} x_j\right)\right)$$

$$= \frac{100^2}{199^2 N^2}\left(100^2 \sum_{i=1}^{N/100} \sigma^2 + \sum_{j=N/100+1}^{N} \sigma^2\right)$$

$$= \frac{100^2}{199^2 N^2}\left(100^2 \times \frac{N}{100}\sigma^2 + \frac{99N}{100}\sigma^2\right)$$

$$= \frac{100}{199^2 N^2}\left(100^2 N\sigma^2 + 99N\sigma^2\right)$$

$$= \frac{100}{199^2 N}\left(10099\sigma^2\right)$$

$$= \frac{100 \times 10099}{199^2 N}\sigma^2$$

~~Dividing by 100^2~~

$$= \frac{100.99}{1.99 \times 1.99 N}\frac{\sigma^2}{}$$

$$\approx \frac{101}{2 \times 2}\frac{\sigma^2}{N} = \frac{101}{4}\frac{\sigma^2}{N}$$

$$\text{Error} = \sqrt{\frac{101}{4}}\frac{\sigma}{\sqrt{N}}$$

$$= 5.02\frac{\sigma}{\sqrt{N}}$$

As we can see, the error when overweighting the data is much larger than the original error of $\sigma/\sqrt{N}$

Thus, overweighting of data should be avoided.