Simple statistics

Allan Aasbjerg Nielsen
Technical University of Denmark
DTU Space – National Space Institute
http://www.imm.dtu.dk/~aa
aa@space.dtu.dk

Mean and variance

- Mean value (Danish: middelværdi)

$$\hat{\mu} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Variance (Danish: varians)

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

- $\hat{\sigma}$ is standard deviation (Danish: spredning)

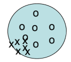
Precision and accuracy

- Precision (Danish: præcision), relates to repeatability

$$\hat{\sigma}_p = s_p = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2}$$

- Accuracy (Danish: nøjagtighed), relates to reality

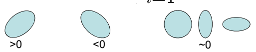
$$\hat{\sigma}_a = s_a = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$



Covariance and correlation

Covariance (Danish: kovarians)

$$\hat{\sigma}_{xy} = s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$



Correlation (Danish: korrelation)

$$\hat{\rho}_{xy} = r_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$$

Dispersion matrix

Variance-covariance matrix

$$\hat{\Sigma} = S = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_2^2 & \cdots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & \cdots & \hat{\sigma}_p^2 \end{bmatrix}$$

a.k.a. dispersion matrix ($\hat{\Sigma}^{-1}$ is precision)

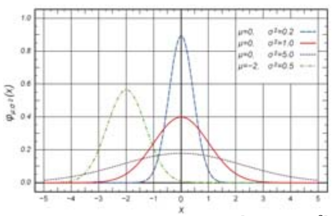
measurements → ∞

$$n \to \infty : \hat{\mu} \to \mu,\ \hat{\sigma}^2 \to \sigma^2,\ \hat{\sigma}_{xy} \to \sigma_{xy},\ \hat{\Sigma} \to \Sigma$$

Stochastic variable. Notation not particularly computational.

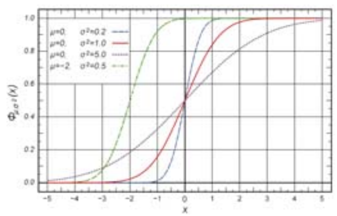
$$\begin{aligned} \mu &= \mathsf{E}\{X\} \\ \sigma^2 &= \mathsf{V}\{X\} \\ &= \mathsf{E}\{(X - \mathsf{E}\{X\})^2\} \\ &= \mathsf{E}\{X^2 + \mathsf{E}\{X\}^2 - 2X\mathsf{E}\{X\}\} \\ \mathsf{V}\{X\} &= \mathsf{E}\{X^2\} - \mathsf{E}\{X\}^2 \\ \sigma_{xy} &= \mathsf{Cov}\{X, Y\} \\ &= \mathsf{E}\{(X - \mathsf{E}\{X\})(Y - \mathsf{E}\{Y\})\} \\ &= \mathsf{E}\{XY\} - \mathsf{E}\{X\}\mathsf{E}\{Y\} \end{aligned}$$

Normal distribution



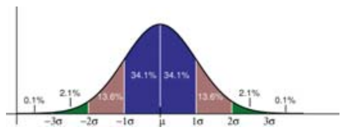
$$\varphi(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

Normal distribution



$$\Phi(X) = \int_{-\infty}^{X} \varphi(x)\,dx$$

Normal distribution



$$\varphi(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

$$P(a < X < b) = \int_a^b \varphi(x)\,dx$$

$\mu \mp 1\sigma : 68.3\%$
$\mu \mp 2\sigma : 95.4\%$
$\mu \mp 3\sigma : 99.7\%$

Mean of *ax+b*

$$\begin{aligned}\widehat{\mu}_{ax+b} &= \frac{1}{n}\sum_{i=1}^{n}(ax_i+b)\\ &= \frac{a}{n}\sum_{i=1}^{n}x_i + \frac{b}{n}\sum_{i=1}^{n}1\\ &= a\widehat{\mu}_x + b\end{aligned}$$

$$\mathsf{E}\{aX+b\} = a\mathsf{E}\{X\}+b$$

Variance of *ax+b*

$$\begin{aligned}\widehat{\sigma}^2_{ax+b} &= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i+b-a\widehat{\mu}_x-b)^2\\ &= \frac{a^2}{n-1}\sum_{i=1}^{n}(x_i-\widehat{\mu}_x)^2\\ &= a^2\widehat{\sigma}^2_x\end{aligned}$$

$$\mathsf{V}\{aX+b\} = a^2\mathsf{V}\{X\}$$

Variance of a sum

$$\begin{aligned}\widehat{\sigma}^2_{ax+by+c} &= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i+by_i+c-a\widehat{\mu}_x-b\widehat{\mu}_y-c)^2\\ &= \frac{1}{n-1}\sum_{i=1}^{n}[a(x_i-\widehat{\mu}_x)+b(y_i-\widehat{\mu}_y)]^2\\ &= \frac{1}{n-1}[a^2\sum_{i=1}^{n}(x_i-\widehat{\mu}_x)^2+b^2\sum_{i=1}^{n}(y_i-\widehat{\mu}_y)^2\\ &\quad + 2ab\sum_{i=1}^{n}(x_i-\widehat{\mu}_x)(y_i-\widehat{\mu}_y)]\\ &= a^2\widehat{\sigma}^2_x+b^2\widehat{\sigma}^2_y+2ab\widehat{\sigma}_{xy}\\ &= [a\ b]\begin{bmatrix}\widehat{\sigma}^2_x & \widehat{\sigma}_{xy}\\ \widehat{\sigma}_{yx} & \widehat{\sigma}^2_y\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix} \geq 0\end{aligned}$$

$$\mathsf{V}\{aX+bY+c\} = a^2\mathsf{V}\{X\}+b^2\mathsf{V}\{Y\}+2ab\mathsf{Cov}\{X,Y\}$$

Variance of a sum

$$\widehat{\sigma}^2_{x+y} = \widehat{\sigma}^2_x+\widehat{\sigma}^2_y(+2\widehat{\sigma}_{xy})$$

$$\widehat{\sigma}^2_{x-y} = \widehat{\sigma}^2_x+\widehat{\sigma}^2_y(-2\widehat{\sigma}_{xy})$$

Difference between two large, nearly equal, independent observations?

Variance of the mean

Variance of a weighted sum of n variables

$$\begin{aligned}\widehat{\sigma}^2_{a_1x_1+\cdots+a_nx_n+b} &= a_1^2\widehat{\sigma}_1^2+\cdots+a_n^2\widehat{\sigma}_n^2+\cdots+2a_ia_j\widehat{\sigma}_{ij}+\cdots\\ &= [a_1\ a_2\ \cdots\ a_n]\begin{bmatrix}\widehat{\sigma}_1^2 & \widehat{\sigma}_{12} & \cdots & \widehat{\sigma}_{1n}\\ \widehat{\sigma}_{21} & \widehat{\sigma}_2^2 & \cdots & \widehat{\sigma}_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \widehat{\sigma}_{n1} & \widehat{\sigma}_{n2} & \cdots & \widehat{\sigma}_n^2\end{bmatrix}\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}\end{aligned}$$

Standard deviation of the mean of n independent samples with same $\widehat{\sigma}$: $a_i=1/n,\ b=0$

$$\widehat{\sigma}^2_{(x_1+\cdots+x_n)/n} = \frac{1}{n^2}(\widehat{\sigma}^2+\cdots+\widehat{\sigma}^2) = \frac{\widehat{\sigma}^2}{n}$$

$$\widehat{\sigma}_{\widehat{\mu}} = \frac{\widehat{\sigma}}{\sqrt{n}}$$