02409 Multivariate Statistics

Lecture B, September 8 2025

Anders Stockmarr
anst@dtu.dk

Course developers:

Anders Stockmarr
Anders Nymark Christensen

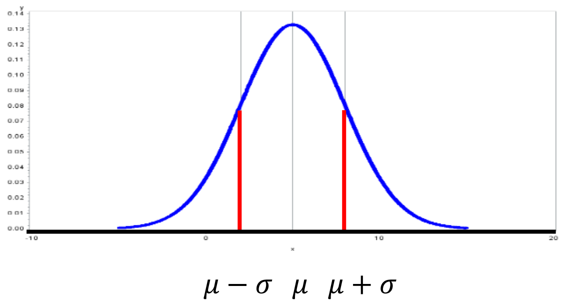
Today's lecture

1. Causation and correlation
2. Moments of univariate random variables revisited
3. Positive semidefinite matrices.
4. **Conditional distributions**
5. Hight-weight of children
6. **Multiple correlation coefficient**
7. **Partial correlation coefficients**
8. Case on cement

Recab: Uni- and bivariate normal distributions

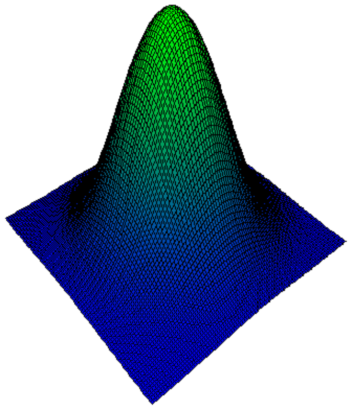
The univariate random variable X is normally distributed $X \in N\left(\mu, \sigma^2\right)$ if

$$f(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$



The bivariate random variable $\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is normally distributed $\boldsymbol{X} \in N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if the density is

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{2\pi}^2}\frac{1}{\sqrt{\det \boldsymbol{\Sigma}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

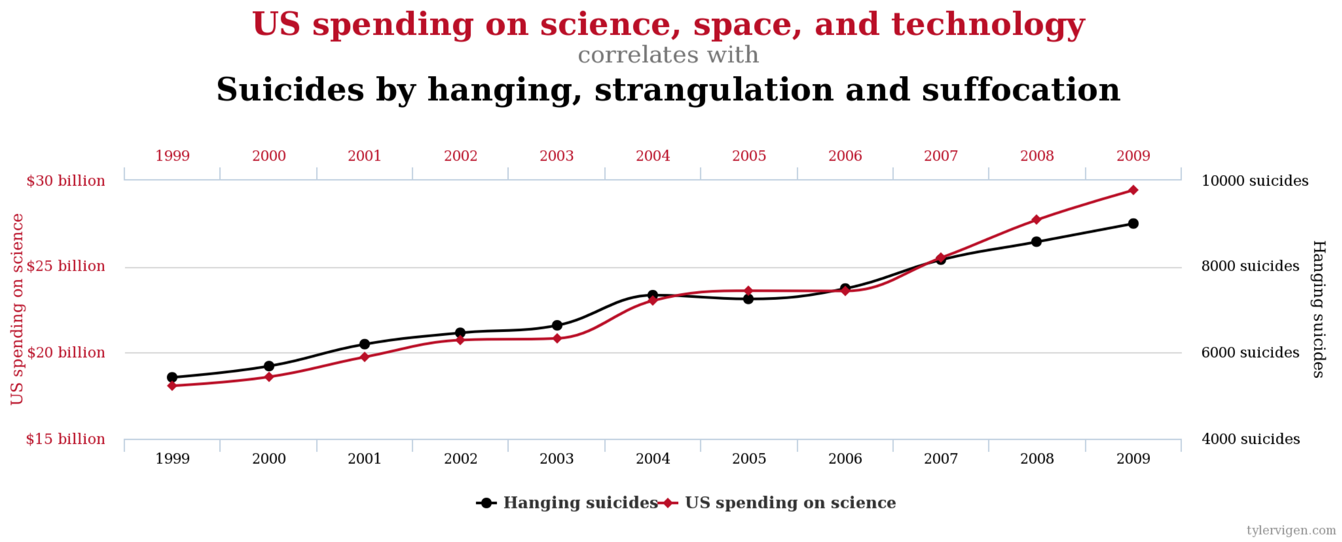


Example: $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$; independent, identically distributed standard normal:

$$f(x_1, x_2) = \frac{1}{2\pi} exp\left(-\frac{1}{2}\left(x_1^2 + x_2^2\right)\right)$$

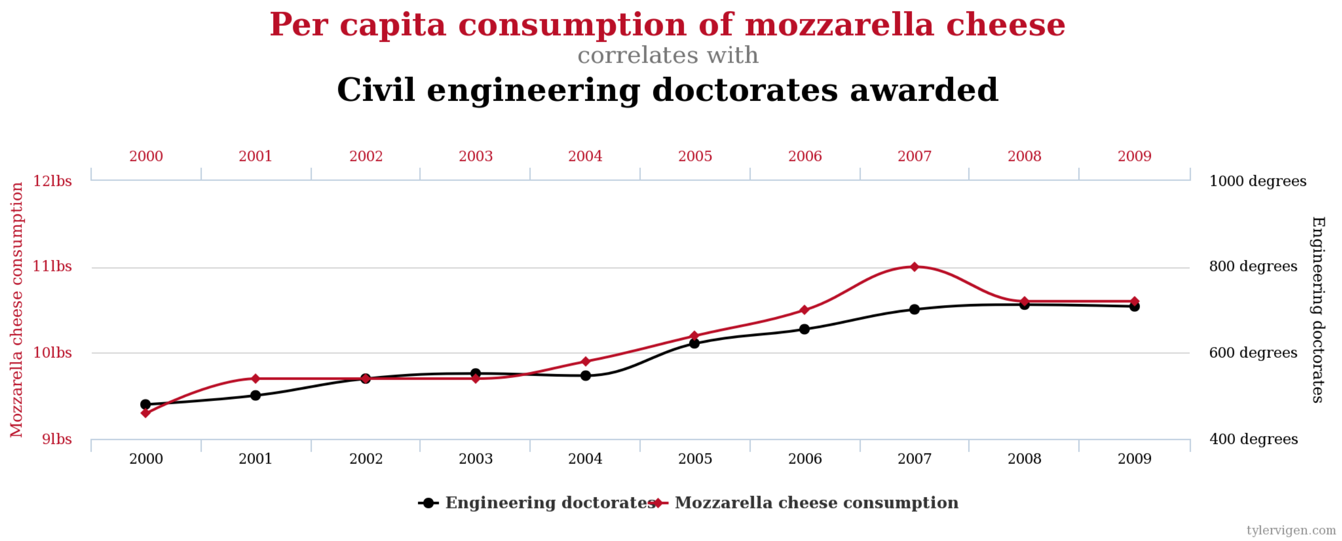
$$= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_1^2\right) \times \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_2^2\right)$$

Correlation does not imply causation I



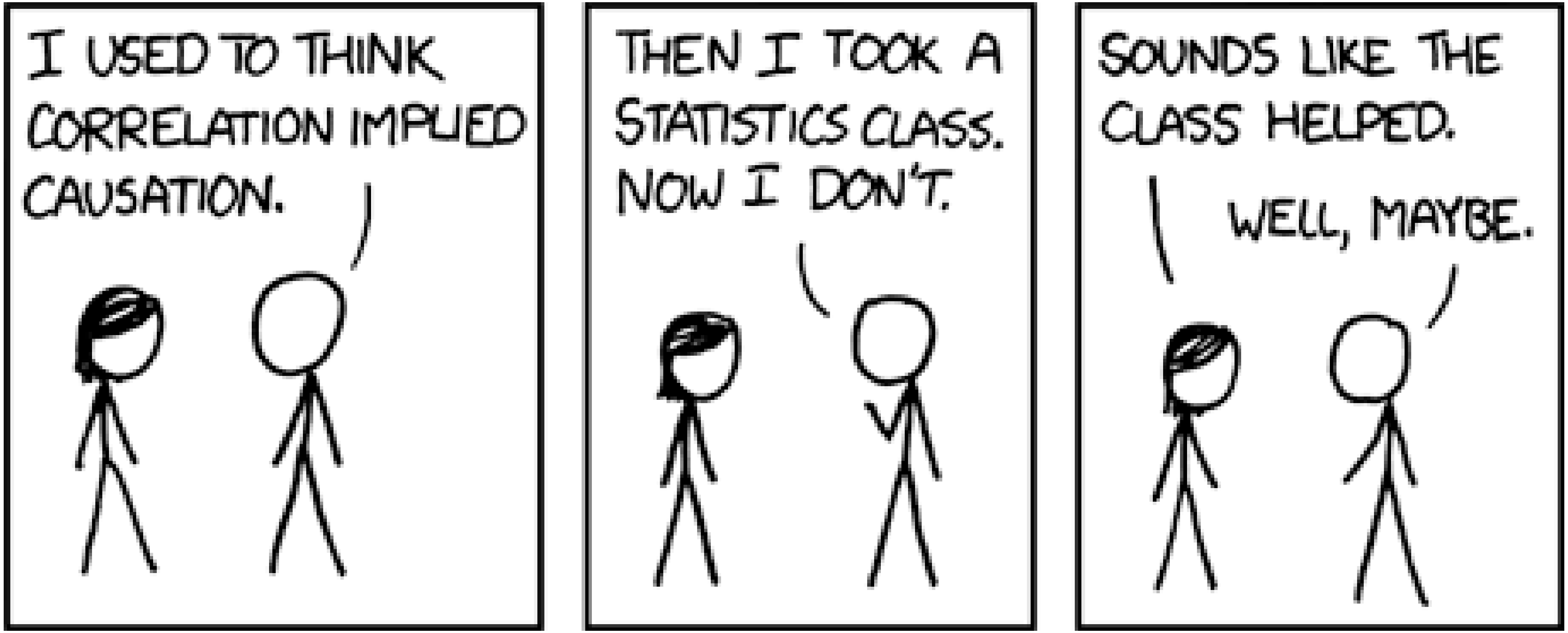
ρ=0.998

[An earlier posting at http://www.tylervigen.com/spurious-correlations](http://www.tylervigen.com/spurious-correlations)



ρ=0.959

Correlation does not imply causation II



- Correlation doesn't imply causation, but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'. Sometimes however, correlation can be terribly wrong due to confounding.

Moments of a univariate random variable revisited

The **k'th moment** E_k is defined as:

$$E_k = E\left(X^k\right) = \int_{-\infty}^{\infty} t^k f(t)dt$$

The **k'th central moment** M_k is defined as:

$$M_k = E\left((X-\mu)^k\right) = \int_{-\infty}^{\infty} (t-\mu)^k f(t)dt$$

where μ is the first moment. Specific forms:

E_1 is the **Expectation** (also called the **mean** (typically for empirical distrubutions);

M_2 is the **Variance**; $\sqrt{M_2}$ is the **Standard Deviation**;

Standardiazation

$M_3/M_2^{3/2}$ is the **Skewness**; a measure of how much the mode is to the left or the right of the mean;

M_4/M_2^2 is the **Kurtosis**; a measure for how fat the tails are in the distribution.

Moments of a univariate random variable revisited

- Example: The standard Normal distribution $N(0,1)$.

$$E_1 = 0, E_2 = 1, E_3 = 0, E_4 = 3.$$

Thus:

$$M_2 = 1, M_3 = 0, M_4 = 3$$

- Mean: 0. For $N\left(\mu, \sigma^2\right)$: μ.
- Variance: 1. For $N\left(\mu, \sigma^2\right)$: σ^2.
- Skewness: 0. For $N\left(\mu, \sigma^2\right)$: 0.
- Kurtosis: 3. For $N\left(\mu, \sigma^2\right)$: 3.

- These values can be used for reference, when evaluating other distributions.

Variance matrix – positive semidefinite

- What does it mean?

- $\Sigma \in Mat_k(\mathbb{R})$ positive semidefinite $\underset{def}{\Longleftrightarrow} \forall x \in \mathbb{R}^k\colon x^T\Sigma x \geq 0.$

Note that if $\mathbb{Y}$ is a random variable with variance Σ, then for any $x \in \mathbb{R}^k$, the variance of $x^T\mathbb{Y}$ is

$$V\left(x^T\mathbb{Y}\right) = x^T V(\mathbb{Y})x = x^T\Sigma x,$$

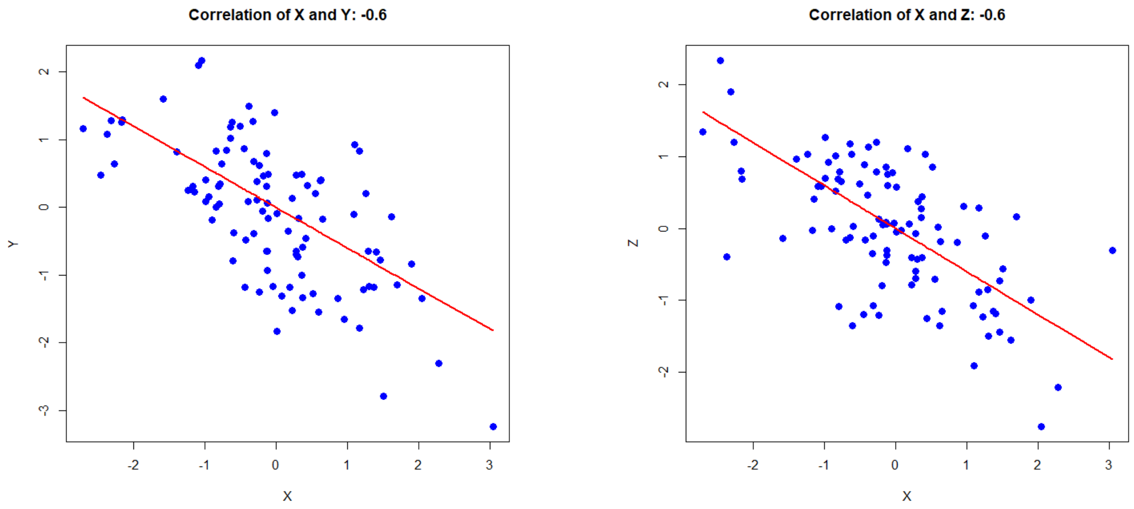
which necessarily has to be ≥ 0. So, **any variance matrix is positive semidefinite**.

Conversely, it is possible to show that if Σ is positively semidefinite, then there is as distribution that has Σ as its variance matrix (in fact infinitely many; one is $N_k(0, \Sigma)$).

Variance matrix – positive semidefinite

- What does it mean?

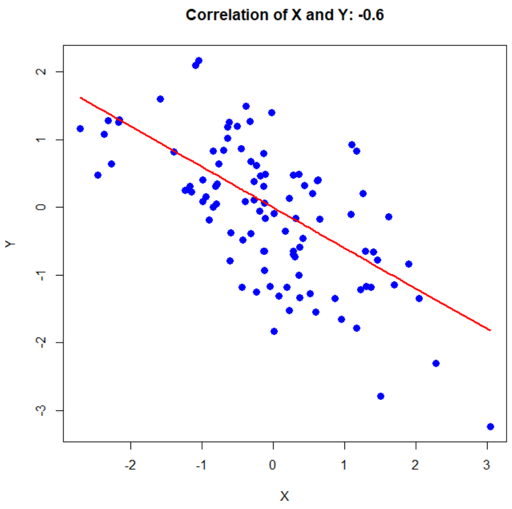
- $\mathbf{V}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$. Does $\mathbf{\Sigma}$ as a variance matrix make sense, if $\rho < -0.5$? NO (Exercise 1.1). $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ has the eigenvalue $1 + 2\rho < 0$.

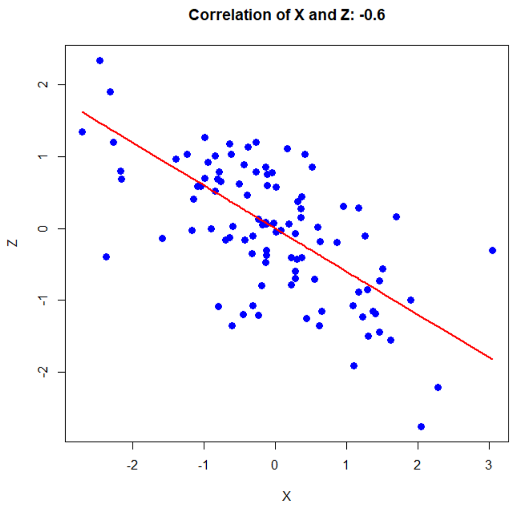


Variance matrix – positive semidefinite

- What does it mean?

- $\mathbf{V}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$. Does $\mathbf{\Sigma}$ as a variance matrix make sense, if $\rho < -0.5$? NO (Exercise 1.1). $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ has eigenvalue $1 - 2\rho < 0$.
- Assume that $\rho < -0.5$, and that $Cov(X, Y) = \rho$, $Cov(X, Z) = \rho$. What values can $\gamma = Cov(Y, Z)$ attain?





Variance matrix – positive semidefinite

- What does it mean?

- $\mathbf{V}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \gamma \\ \rho & \gamma & 1 \end{bmatrix}$

Assume that $\rho < -0.5$ and

$$Y = \rho \cdot X + \sqrt{1-\rho^2} W_1, \qquad W_1 \perp\!\!\!\perp X, W_1 \sim N(0,1);$$
$$Z = \rho \cdot X + \sqrt{1-\rho^2} W_2, \qquad W_2 \perp\!\!\!\perp X, W_2 \sim N(0,1).$$

Then

$$\gamma = Cov(Y,Z) = Cov\left(\rho \cdot X + \sqrt{1-\rho^2} W_1, \rho \cdot X + \sqrt{1-\rho^2} W_2\right)$$
$$= \rho^2 + \left(1-\rho^2\right) \cdot Cov(W_1, W_2)$$

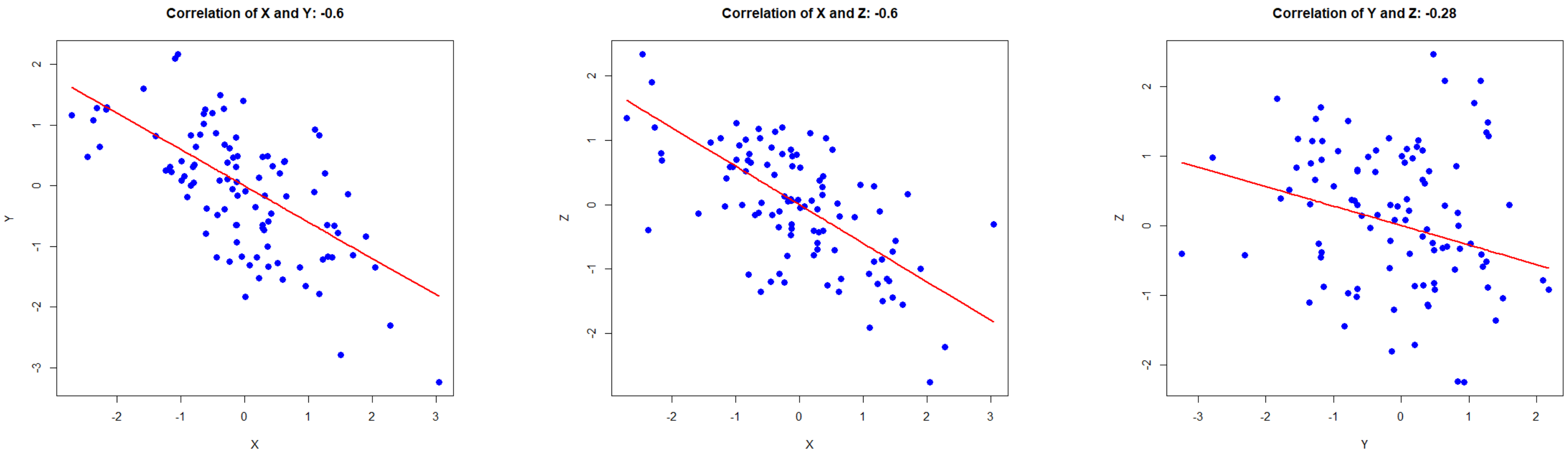
Range: $W_1 = W_2$: $\gamma = 1$. $W_1 = -W_2$: $\gamma = 2\rho^2 - 1 > -0.5$. Range is $[2\rho^2 - 1; 1] \subset (-0.5; 1]$!

Variance Matrix – positive semidefinite

- What does it mean?

- $\mathbf{V}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}\right) = \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \gamma \\ \rho & \gamma & 1 \end{bmatrix}$

Example: $W_2 = -W_1, \rho = -0.6$: $\gamma = 2\rho^2 - 1 = -0.28$.



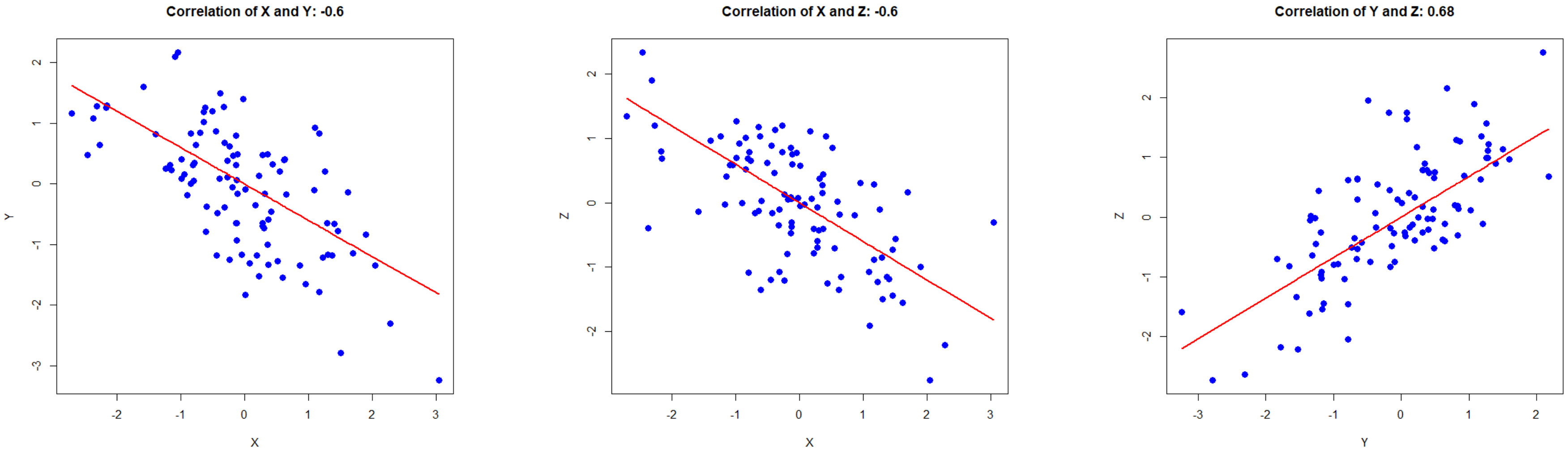
Variance Matrix – positive semidefinite

- What does it mean?

Even though both Y and Z correlate strongly negatively to the same variable, there is (nearly) no telling what the internal correlation is.

- $\boldsymbol{D}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \gamma \\ \rho & \gamma & 1 \end{bmatrix}$

Example: $W_2 = \frac{1}{2}W_1 + \frac{\sqrt{3}}{2}W_3$, with $W_3 \perp\!\!\!\perp X, W\_1$, $\rho = -0.6$: $\gamma = \frac{1}{2}(1+\rho^2) = 0.68$.



Estimation of parameters

$$\bar{X} = \frac{1}{n} X^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{n} X^T \mathbf{1}$$

$$(n-1)\mathbf{S} = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T = \mathbf{X}^T\mathbf{X} - n\bar{X}\bar{X}^T = \mathbf{X}^T\mathbf{X} - \frac{1}{n} X^T \mathbf{1}\mathbf{1}^T X.$$

Theorem 1.32

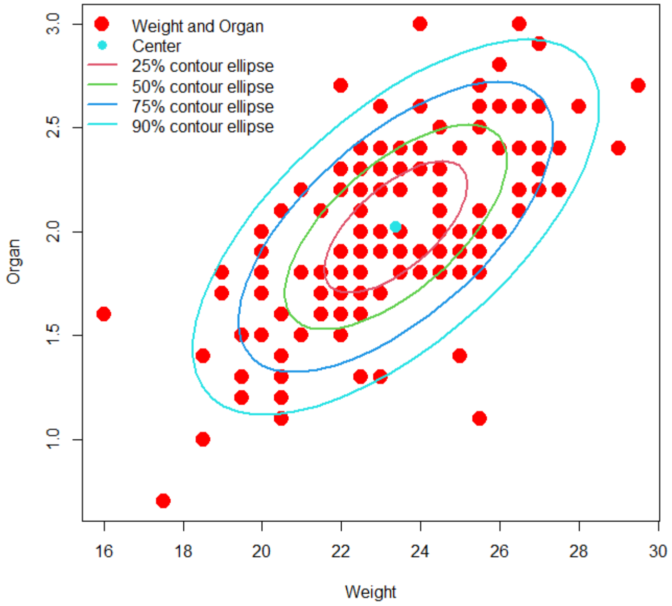
Let the situation be as stated above. Then the $100(1-\alpha)\%$ confidence ellipsoid for the unknown mean $\boldsymbol{\mu}$ is

$$\{\boldsymbol{\mu} | (\mu - \bar{\mathbf{x}})^T \mathbf{s}^{-1} (\mu - \bar{\mathbf{x}}) \le \frac{p(n-1)}{(n-p)n} F(p, n-p)_{1-\alpha}\}$$

and the $100(1-\alpha)\%$ prediction ellipsoid for a coming observation $\mathbf{x}$ is

$$\{\mathbf{x} | (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{s}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \le \frac{p(n-1)(n+1)}{(n-p)n} F(p, n-p)_{1-\alpha}\}$$

Organ data (last lecture): Contour ellipses for actual data (not the mean and not predictions):



Example – heiwei data

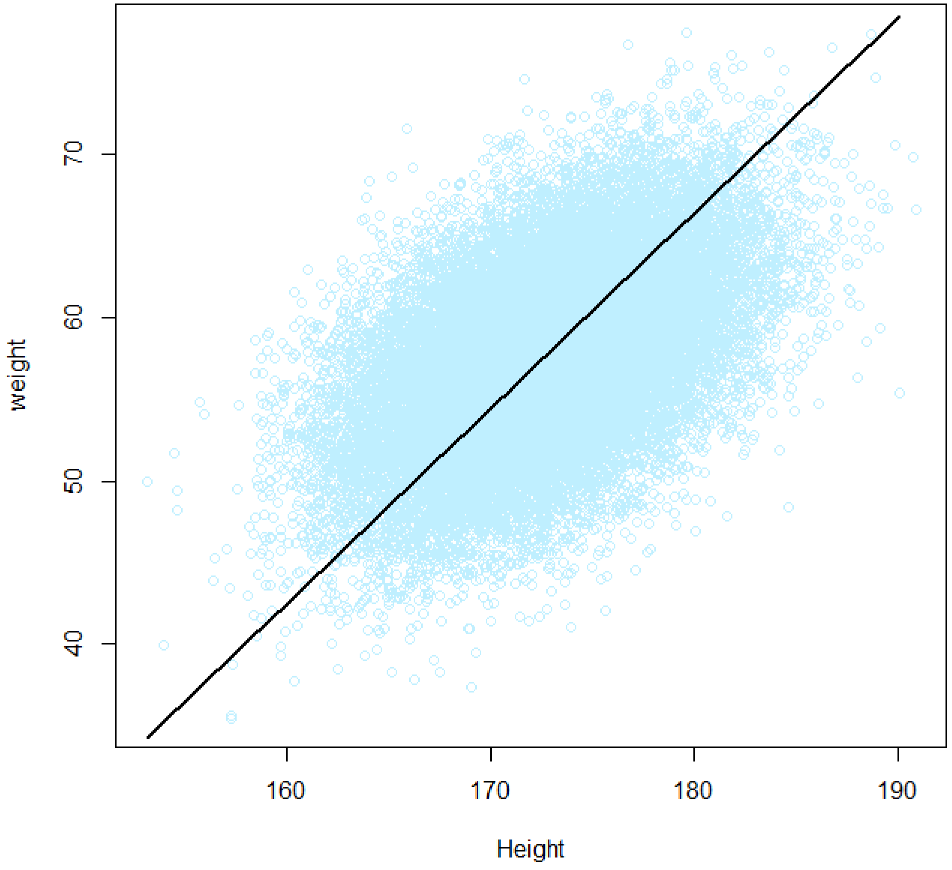
$$\bar{X} = \frac{1}{n} X^T \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{1}{n} X^T \mathbf{1}$$

$$(n-1)\mathbf{S} = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T = \mathbf{X}^T\mathbf{X} - n\bar{X}\bar{X}^T = \mathbf{X}^T\mathbf{X} - \frac{1}{n} X^T \mathbf{1}\mathbf{1}^T X.$$

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_{height} \\ \hat{\mu}_{weight} \end{pmatrix} = \bar{X} = \begin{pmatrix} 172.70251 \\ 57.64221 \end{pmatrix},$$

$$S = \hat{\Sigma} = \begin{pmatrix} 23.33145 & 12.84736 \\ 12.84736 & 27.97659 \end{pmatrix}$$

Example – heiwei data



Independence in the multivariate normal distribution

Theorem 1.22

Let

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right).$$

Then

$$X_i \sim \mathrm{N}(\mu_i, \Sigma_{ii}),$$

and

$$X_1, X_2 \text{ are stochastically independent} \quad \Leftrightarrow \quad \Sigma_{12} = \Sigma_{21}^T = \mathbf{0},$$

where $\mathbf{0}$ is a null matrix.

Conditional distributions in the multivariate normal distribution

Suppose that $X \sim N(\mu, \Sigma)$ decomposes into

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Theorem 1.27

If X_2 is regularly distributed, i.e. if $\boldsymbol{\Sigma}_{22}$ has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1 | X_2 = x_2) &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(x_2 - \boldsymbol{\mu}_2) \\ \mathrm{D}(X_1 | X_2 = x_2) &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned}$$

If $\boldsymbol{\Sigma}_{22}$ does not have full rank then the conditional distribution is still normal and $\boldsymbol{\Sigma}_{22}^{-1}$ in the above equations should be substituted by a generalised inverse $\boldsymbol{\Sigma}_{22}^{-}$.

Conditional distributions in the multivariate normal distribution

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

Proof for $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{22}$ regular: The general case easily follows:

Define $Z = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$. Then

$$Cov(Z, X_2) = Cov(X_1, X_2) - \Sigma_{12}\Sigma_{22}^{-1}Cov(X_2, X_2) = \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0$$

By Theorem 1.22, **$\boldsymbol{Z}$ and $\boldsymbol{X_2}$ are independent**.

Conditional distributions in the multivariate normal distribution

$$
\begin{aligned}
\mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\
\mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.
\end{aligned}
$$

Note that $X_1 = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 + \Sigma_{12}\Sigma_{22}^{-1}X_2 = Z + \Sigma_{12}\Sigma_{22}^{-1}X_2$. This establishes normality of $X_1|X_2 = x_2$.

Now,

$$
\begin{aligned}
E(X_1|X_2 = x_2) &= E\left(Z + \Sigma_{12}\Sigma_{22}^{-1}X_2 \middle| X_2 = x_2\right) \\
&= E\left(Z + \Sigma_{12}\Sigma_{22}^{-1}x_2 \middle| X_2 = x_2\right) \\
&= E\left(Z + \Sigma_{12}\Sigma_{22}^{-1}x_2\right) \\
&= E(Z) + \Sigma_{12}\Sigma_{22}^{-1}x_2 = \Sigma_{12}\Sigma_{22}^{-1}x_2.
\end{aligned}
$$

Conditional distributions in the multivariate normal distribution

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

Note that $X_1 = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 + \Sigma_{12}\Sigma_{22}^{-1}X_2 = Z + \Sigma_{12}\Sigma_{22}^{-1}X_2$.

Also,

$$\begin{aligned} V(X_1|X_2 = x_2) &= V\left(Z + \Sigma_{12}\Sigma_{22}^{-1}X_2 \middle| X_2 = x_2\right) \\ &= V\left(Z + \Sigma_{12}\Sigma_{22}^{-1}x_2 \middle| X_2 = x_2\right) \\ &= V\left(Z + \Sigma_{12}\Sigma_{22}^{-1}x_2\right) \\ &= V(Z) = V\left(\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right) \\ &= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \end{pmatrix}^T \\ &= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} \end{pmatrix} \begin{pmatrix} I \\ \Sigma_{22}^{-1}\Sigma_{21} \end{pmatrix} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \ \square \end{aligned}$$

Conditional distributions II

For the two-dimensional normal distribution

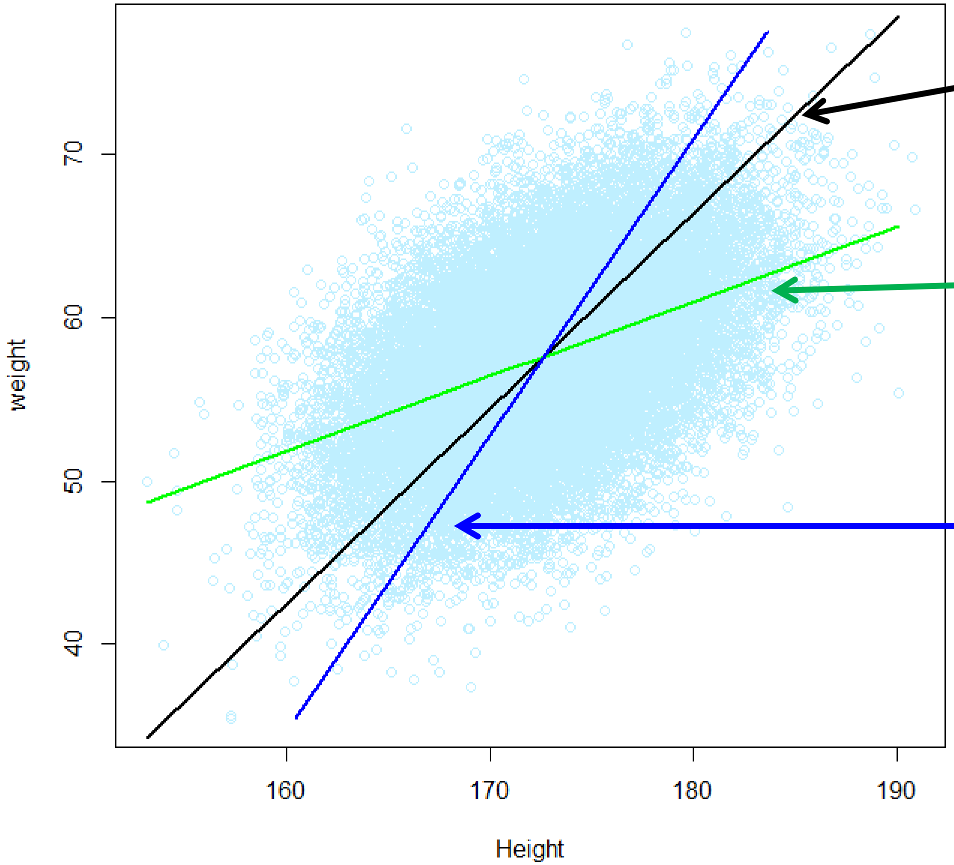
$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \in N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

with $\rho = Cor(X_1, X_2)$, the formula reads

$$E(X_1|X_2 = x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(X_2 - \mu_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)$$

$$V(X_1|X_2 = x_2) = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} = \sigma_1^2\left(1 - \rho^2\right)$$

Conditional distributions II



Main axis of contour ellipse

$$E(X_2|X_1 = x_1) = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x_1 - \mu_1)$$

$$E(X_1|X_2 = x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)$$

Example

$$\hat{\mu} = \begin{pmatrix} 172.70251 \\ 57.64221 \end{pmatrix}, \hat{\Sigma} = \begin{pmatrix} 23.33145 & 12.84736 \\ 12.84736 & 27.97659 \end{pmatrix}$$

- $E(X_2|X_1 = x_1) = \mu_2 + \frac{\sigma_{21}}{\sigma_1^2}(x_1 - \mu_1)$ is estimated as $\boldsymbol{-37.46 + 0.55x_1}$
- $E(X_1|X_2 = x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2)$ is estimated as $\boldsymbol{146.23 + 0.46x_2}$

- Assume a height of 180 cm: Weight is expected to be $-37.46 + 0.55 * 180 = 61.66$ kilogram;
- Assume a weight of 65 kilogram: Height is expected to be $146.23 + 0.46 * 65$ = 176.08 cm.

Partial correlation coefficient

The partial correlation between some variables given others is simply

The correlation in the conditional distribution of the ’some’ variables given the ’other’

Variables:

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}$$

It follows from successive conditioning that

$$\rho_{ij|kl} = \frac{\rho_{ij|k} - \rho_{il|k} \cdot \rho_{jl|k}}{\sqrt{(1-\rho_{il|k}^2) \cdot (1-\rho_{jl|k}^2)}}$$

Example: Ice cream - I

We consider a two-dimensional random variable

$$\begin{bmatrix} D \\ I \end{bmatrix}, \mathbf{Cor}\left(\begin{bmatrix} D \\ I \end{bmatrix}\right) = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix},$$

where D is the number of drowning accidents, and I is the ice-cream sale.

Should ice-cream be prohibited? Is the ice-cream industry behind drownings to get more sales?

$$\begin{bmatrix} D \\ I \\ T \end{bmatrix}, Cor\left(\begin{bmatrix} D \\ I \\ T \end{bmatrix}\right) = \begin{bmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{bmatrix}$$

where T is the temperature

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}$$

Find the correlation between D and I conditioned on T, assuming normality:

Example: Ice cream - II

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}$$

$$\rho_{DI|T} = \frac{\rho_{DI} - \rho_{DT}\rho_{IT}}{\sqrt{(1-\rho_{DT}^2)(1-\rho_{IT}^2)}} = \frac{0.5 - 0.7 \cdot 0.7}{\sqrt{(1-0.7^2)(1-0.7^2)}} = 0.0196$$

$$\begin{bmatrix} D \\ I \\ T \end{bmatrix}, Cor\left(\begin{bmatrix} D \\ I \\ T \end{bmatrix}\right) = \begin{bmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1 & 0.7 \\ 0.7 & 0.7 & 1 \end{bmatrix}$$

- There is little correlation between drowings and ice cream sale, once we control for the temperatue.

Example: Ice cream - III

We consider a three-dimensional random variable

$$\begin{bmatrix} D \\ I \\ T \end{bmatrix}, \boldsymbol{\Sigma} = \left[\begin{array}{cc|c} \sigma_D^2 & 0.5\sigma_D\sigma_I & 0.7\sigma_D\sigma_T \\ 0.5\sigma_D\sigma_I & \sigma_I^2 & 0.7\sigma_I\sigma_T \\ \hline 0.7\sigma_D\sigma_T & 0.7\sigma_I\sigma_T & \sigma_T^2 \end{array}\right]$$

where D is the number of drowning accidents, and I is the ice-cream sale and T the temperature.

Find the variance, when conditioned upon temperature

Theorem 1.27

If X_2 is regularly distributed, i.e. if $\boldsymbol{\Sigma}_{22}$ has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1 | X_2 = x_2) &= \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \boldsymbol{\mu}_2) \\ \mathrm{D}(X_1 | X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If $\boldsymbol{\Sigma}_{22}$ does not have full rank then the conditional distribution is still normal and $\boldsymbol{\Sigma}_{22}^{-1}$ in the above equations should be substituted by a generalised inverse $\boldsymbol{\Sigma}_{22}^{-}$.

The variance of (D,I|T) is:

$\begin{pmatrix} \sigma_D^2 & 0.7\sigma_D\sigma_I \\ * & \sigma_I^2 \end{pmatrix}$ (A) 0%

$\begin{pmatrix} \sigma_D^2 & 0.7\sigma_T \\ * & \sigma_I^2 \end{pmatrix}$ (B) 0%

$\begin{pmatrix} 0.49\sigma_D^2 & 0.49\sigma_T \\ * & 0.49\sigma_I^2 \end{pmatrix}$ (C) 0%

$\begin{pmatrix} 0.51\sigma_D^2 & 0.51\sigma_D\sigma_I \\ * & 0.51\sigma_I^2 \end{pmatrix}$ (D) 0%

$\begin{pmatrix} 0.51\sigma_D^2 & 0.01\sigma_D\sigma_I \\ * & 0.51\sigma_I^2 \end{pmatrix}$ (E) 0%

Don't know (F) 0%

Example: Ice cream – IV

Find the variance, when conditioned upon temperature

We consider the three-dimensional normal random variable

$$\begin{bmatrix} D \\ I \\ T \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} \sigma_D^2 & 0.5\sigma_D\sigma_I & 0.7\sigma_D\sigma_T \\ 0.5\sigma_I\sigma_D & \sigma_I^2 & 0.7\sigma_I\sigma_T \\ 0.7\sigma_T\sigma_D & 0.7\sigma_I\sigma_T & \sigma_T^2 \end{bmatrix}$$

where D is the number of drowning accidents, and I is the ice-cream sale and T the temperature.

Find the variance, when conditioned upon temperature:

$$V\left(\begin{bmatrix} D \\ I \end{bmatrix} \mid T\right) = \begin{bmatrix} \sigma_D^2 & 0.5\sigma_D\sigma_I \\ 0.5\sigma_I\sigma_D & \sigma_I^2 \end{bmatrix} - \begin{bmatrix} 0.7\sigma_D\sigma_T \\ 0.7\sigma_I\sigma_T \end{bmatrix} \left[\sigma_T^2\right]^{-1} \left[0.7\sigma_T\sigma_D \quad 0.7\sigma_I\sigma_T\right]$$

$$= \begin{bmatrix} \sigma_D^2 & 0.5\sigma_D\sigma_I \\ 0.5\sigma_I\sigma_D & \sigma_I^2 \end{bmatrix} - 0.49 \begin{bmatrix} \sigma_D^2 & \sigma_D\sigma_I \\ \sigma_D\sigma_I & \sigma_I^2 \end{bmatrix} = \begin{bmatrix} 0.51\sigma_D^2 & 0.01\sigma_D\sigma_I \\ 0.01\sigma_D\sigma_I & 0.51\sigma_I^2 \end{bmatrix}$$

$$corr(D, I|T) = \frac{0.01\sigma_D\sigma_I}{\sqrt{0.51\sigma_D^2 \cdot 0.51\sigma_I^2}} = 0.0196$$

Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}; \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}; \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Theorem 1.27

If X_2 is regularly distributed, i.e. if Σ_{22} has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If Σ_{22} does not have full rank then the conditional distribution is still normal and Σ_{22}^{-1} in the above equations should be substituted by a generalised inverse Σ_{22}^{-}.

Option E

Independent of σ_T^2!

Example: Ice cream – V

R code:

As seen from the previous slide, we can work directly with the correlations:

```
Sigma <- matrix(c(1, 0.5, 0.7,
                    0.5, 1, 0.7,
                    0.7, 0.7, 1),ncol=3)
Sigma.11<-Sigma[1:2,1:2]
Sigma.12<-Sigma[1:2,3]
Sigma.21<-Sigma[3,1:2]
Sigma.22<-Sigma[3,3]
(Sigma1.2<-Sigma.11-Sigma.12%*%solve(Sigma.22)%*%Sigma.21)
           [,1]       [,2]
[1,] 1.00000000 0.01960784
[2,] 0.01960784 1.00000000
```

Partial Correlation Matrix		
	D	I
D	1.0000	0.0196
I	0.0196	1.0000

Multiple correlation coefficient I

(Better known as R^2)

Let

$$Z = \begin{bmatrix} Y \\ X \end{bmatrix}; \quad \mu = \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}.$$

We now define the multiple correlation coefficient between Y_i , $i = 1, \ldots, m$ and X as the maximal correlation between Y_i and a linear combination of X's elements. It is denoted $\rho_{y_i|x}$.

Multiple correlation coefficient II

$$Z = \begin{bmatrix} Y \\ X \end{bmatrix}; \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}$$

Take β_i to be the *i*'th row of $\Sigma_{yx}\Sigma_{xx}^{-1}$, so that

$$Y_i = Y_i - \beta_i X + \beta_i X = Z_i + \beta_i X$$

where Z_i and X are **stochastically independent**. Thus, for any unit length vector α,

$$Cov(Y_i, \alpha X) = Cov(\beta_i X, \alpha X)$$

Obviously highest when α and β_i are proportional.

Have the highst correlation when looking in the direction beta_i

Multiple correlation coefficient III

$$Z = \begin{bmatrix} Y \\ X \end{bmatrix}; \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{bmatrix}; \quad \Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}$$

Thus, with β_i the i'th row of $\Sigma_{yx}\Sigma_{xx}^{-1}$:

$$\rho_{y_i|x} = \frac{Cov(Y_i, \beta_i X)}{\sqrt{V(Y_i)V(\beta_i X)}} = \frac{Cov(\beta_i X, \beta_i X)}{\sqrt{V(Y_i)V(\beta_i X)}} = \frac{V(\beta_i X)}{\sqrt{V(Y_i)V(\beta_i X)}} = \sqrt{\frac{V(\beta_i X)}{V(Y_i)}}$$

$$= \sqrt{\frac{\left(\Sigma_{yx}\Sigma_{xx}^{-1}\right)_{i,}\Sigma_{xx}\left(\Sigma_{yx}\Sigma_{xx}^{-1}\right)_{i,}^{T}}{\sigma_{Y_i}^2}} = \sqrt{\frac{\left(\Sigma_{yx}\right)_{i,}\Sigma_{xx}^{-1}\left(\Sigma_{yx}\right)_{i,}^{T}}{\sigma_{Y_i}^2}}$$

<— Standardize

Note that

$$V(Y_i|X) = V(Y_i) - \left(\Sigma_{yx}\right)_{i,}\Sigma_{xx}^{-1}\left(\Sigma_{yx}\right)_{i,}^{T} = V(Y_i) - V(\beta_i X)$$

Multiple correlation coefficient IV

Theorem 1.42

We consider the situation above. Let σ_i be the i'th column in $\mathbf{\Sigma}_{xy}$, i.e. σ_i^T is the i'th row in $\mathbf{\Sigma}_{yx}$. Further, let σ_{ii} denote the i'th diagonal element, i.e. the variance of Y_i

Then

$$\rho_{y_i|x} = \frac{\sqrt{\sigma_i^T \mathbf{\Sigma}_{xx}^{-1} \sigma_i}}{\sqrt{\sigma_{ii}}}.$$

If we let

$$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{ii} & \sigma_i^T \\ \sigma_i & \mathbf{\Sigma}_{xx} \end{bmatrix},$$

then

$$1 - \rho_{y_i|x}^2 = \frac{\det \mathbf{\Sigma}_i}{\sigma_{ii} \det \mathbf{\Sigma}_{xx}} = \frac{\mathrm{V}(Y_i|\mathbf{X})}{\mathrm{V}(Y_i)},$$

Multiple correlation coefficient V

- Note that

$$V(Y_i|X) = \left(1 - \rho^2_{y_i|x}\right) V(Y)$$

- Thus, $\rho^2_{y_i|x}$ measures how much the variance of Y_i decreases when you know X.

- In a regression where Y_i is regressed on X, $\rho^2_{y_i|x}$ measure the degree that the regression model explains the total variation of Y_i. Since Y_i and X are implicitly given in such a situation, $\rho^2_{y_i|x}$ is usually termed without these, as $\boldsymbol{R^2}$.

Multiple correlation coefficient VI

Example

We consider a three-dimensional random variable

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

with dispersion matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

Find the squared multiple correlation between X and $(Y, Z)^T$

Theorem 1.42

We consider the situation above. Let σ_i be the i'th column in $\mathbf{\Sigma}_{xy}$, i.e. σ_i^T is the i'th row in $\mathbf{\Sigma}_{yx}$. Further, let σ_{ii} denote the i'th diagonal element, i.e. the variance of Y_i
Then

$$\rho_{y_i|x} = \frac{\sqrt{\sigma_i^T \mathbf{\Sigma}_{xx}^{-1} \sigma_i}}{\sqrt{\sigma_{ii}}}.$$

If we let

$$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{ii} & \sigma_i^T \\ \sigma_i & \mathbf{\Sigma}_{xx} \end{bmatrix},$$

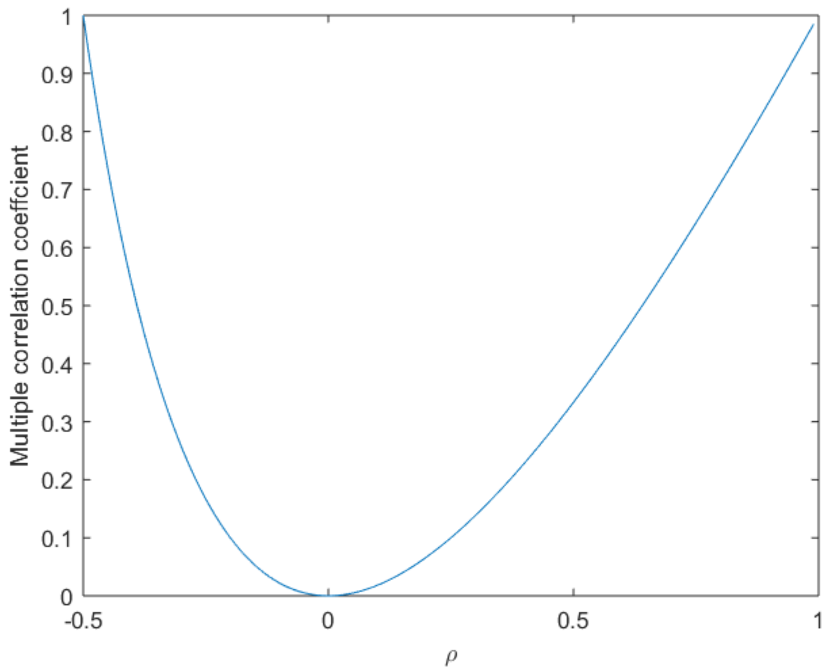
then

$$1 - \rho_{y_i|x}^2 = \frac{\det \mathbf{\Sigma}_i}{\sigma_{ii} \det \mathbf{\Sigma}_{xx}} = \frac{\mathrm{V}(Y_i|\mathbf{X})}{\mathrm{V}(Y_i)},$$

Multiple correlation coefficient VII

Example – solution, using theorem 1.40:

$$\Sigma_i = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}, \qquad \Sigma_{xx} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \qquad \sigma_{ii} = 1$$



$$\rho^2_{X|YZ} = 1 - \frac{\det \Sigma_i}{\sigma_{ii} \cdot \det \Sigma_{xx}}$$
$$= 1 - \frac{1\cdot 1\cdot 1 - \rho\cdot 1\cdot \rho + \rho\cdot\rho\cdot\rho - \rho\cdot\rho\cdot 1 + \rho\cdot\rho\cdot\rho - \rho\cdot\rho\cdot 1}{1\cdot(1\cdot 1 - \rho\cdot\rho)}$$
$$= 1 - \frac{1 - 3\rho^2 + 2\rho^3}{1-\rho^2}$$

Similar:

$$\Sigma_{yy} = 1,\ \Sigma_{y,xz} = [\rho \quad \rho] = \rho[1 \quad 1],\ \Sigma_{xz,y} = \begin{bmatrix} \rho \\ \rho \end{bmatrix} = \rho \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

Since $V(Y) = 1$ and $\Sigma^{-1}_{xz,xz} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$, it follows that

$$\rho^2_{x|yz} = \Sigma_{\mathrm{y,xz}}\Sigma^{-1}_{xz,xz}\Sigma_{xz,y} = \frac{\rho^2}{1-\rho^2}[1 \quad 1]\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{\rho^2}{1-\rho^2}\big(2(1-\rho)\big) = \frac{2\rho^2}{1+\rho}$$

Interpretation: The more correlated the variables are, the more we know about X given Y and Z.

Multiple correlation coefficient VIII

Theorem 1.45

Let $R = \hat{\rho}_{y_i|x}$ be the empirical multiple correlation coefficient between Y_i and $\boldsymbol{X} = (Z_{m+1}, \ldots, Z_p)$ based upon n observations. Then

$$\frac{R^2}{1-R^2} \cdot \frac{n-(p-m)-1}{p-m} \sim \mathrm{F}(p-m, n-(p-m)-1),$$

if $\rho_{y_i|x} = \rho_{y_i|z_{m+1,\ldots,p}} = 0$.

Cement strength–R program

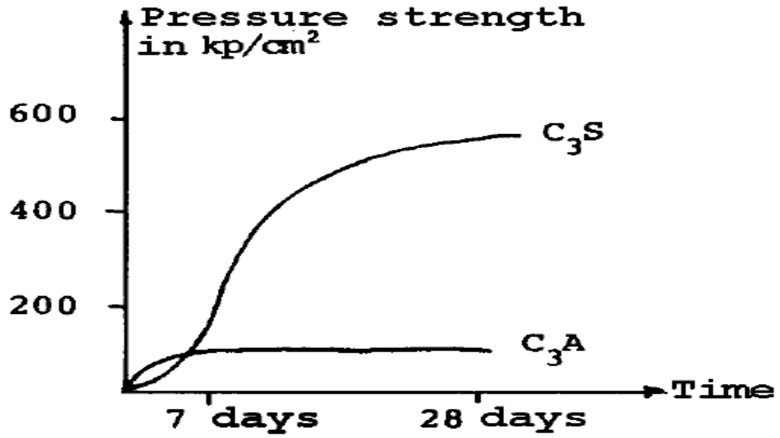
```
my.corr<-diag(rep(1,5))
my.corr[lower.tri(my.corr)]<-c(-0.309,
                                0.091,0.192,
                                0.158,0.120,0.745,
                                0.344,-0.166,0.320,0.464)
my.corr<-my.corr+t(my.corr)-diag(rep(1,5))

colnames(my.corr)<-
c("C3A","C3S","Blaine","Strgth3","Strgth28")
row.names(my.corr)<-colnames(my.corr)

#principal components and values:
eigen(my.corr)

#partial covariance:

Sigma.11<-my.corr[-3,-3]
Sigma.12<-my.corr[-3,3]
Sigma.21<-my.corr[3,-3]
Sigma.22<-my.corr[3,3]
Sigma.1.2<-Sigma.11-Sigma.12%*%solve(Sigma.22)%*%Sigma.21

# partial correlation:
cov2cor(Sigma.1.2)
#principal components and values:
eigen(Sigma.1.2)
```

If only the correlation matrix and not the full dataset is available, we must use input statements as shown.

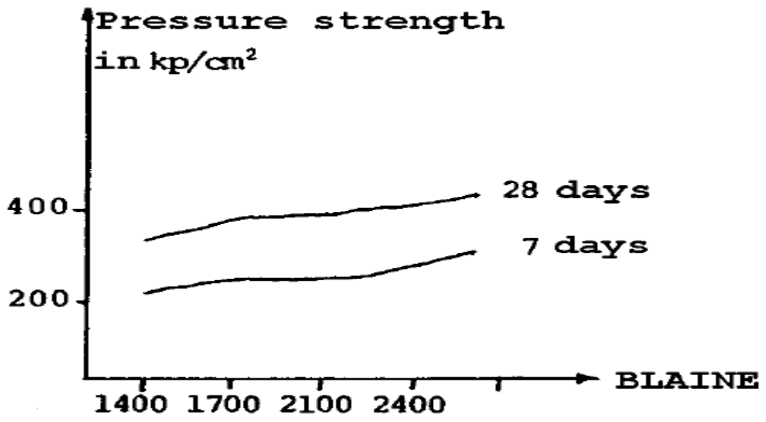
Finding partial covariances from matrix formula; converting to correlations directly.

Using eigen to find eigenvalues and eigenvectors of the correlation matrix.

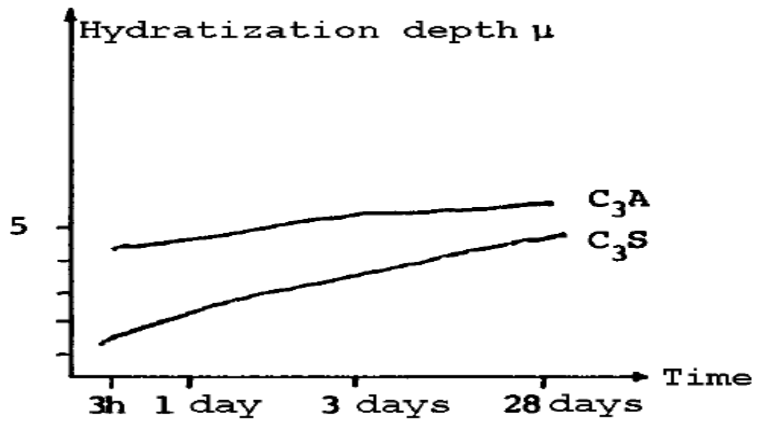
Cement strength-Conventionel Wisdom



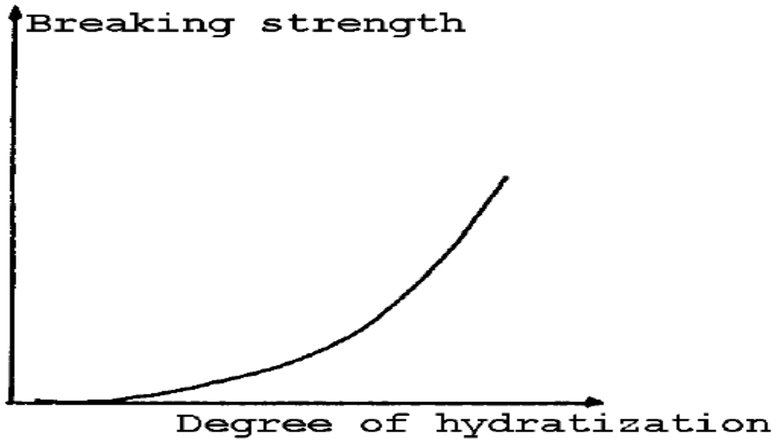
(a) Strength by pressure test at ordinary temperature of paste of C_3S and C_3A seasoned for different amounts of time. (from [13]).



(b) Pressure strengths for different finegrainedness of the cement. (from [13]).



(c) Degree of hydratisation for cement minerals and their dependence on time (from [13]).



(d) Relationship between degree of hydratisation and strength (from [13]).

Cement strength

Correlation Matrix					
	C3S	**C3A**	**BLAINE**	**Strgth3**	**Strgth28**
C3S	1.000	.	.	.	.
C3A	-0.309	1.000	.	.	.
BLAINE	0.091	0.192	1.000	.	.
Strgth3	0.158	0.120	0.745	1.000	.
Strgth28	0.344	-0.166	0.320	0.464	1.000

Partial Correlation Matrix					
	C3S	**C3A**	**BLAINE**	**Strgth3**	**Strgth28**
C3S	1.0000				
C3A	-.3340	1.0000			
BLAINE					
Strgth3	0.1358	-.0352		1.0000	
Strgth28	0.3337	-.2446		0.3570	1.0000

Exercises

1.3 , 2.4, 2.5.

1.3 is a 'pen and paper' exercise to explore multiple and partial correlation

2.4 is a real data case on hay production, that explores partial correlation.

2.5 is a real data case on fitness data.