last change: 21/01/13@10.20

3 Cont Multiv Analysis

Summary 1 *MVN distribution properties: transformation, marginals, conditionals, concentration matrix, marginal independence, conditional independence, concentration graphs, conditional means & variance formulas, partial correlation, joint confidence intervals, one sample problem, LR test*

3.1 Multinormal Distribution

special case: **independence** $x_1, ..., x_p$ iid $\mathcal{N}(0, I) \Longrightarrow \mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, I)$

$$f(x_1, ..., x_p) = \prod_i^p f(x_i) = f(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^p} \exp\left\{-\frac{1}{2}\mathbf{x}'\mathbf{x}\right\}$$

generate general multiv density by linear transf of standard independent case

general **multiNormal density:**

we define the general mvNormal $\mathbf{y} \sim \mathcal{N}_p(\mu, \Sigma)$ as linear transformation[1] $\mathbf{y} = A\mathbf{x} + \mu$ of $\mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, I)$ (with $\det A \neq 0$):

$$f(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^p}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mu)'\left(A^{-1}\right)' A^{-1}(\mathbf{y}-\mu)\right\} \det A^{-1}$$

set

$$\Sigma = AA'$$

write

$$A = \Sigma^{1/2} \Longrightarrow \Sigma = (\Sigma)^{1/2}\left(\Sigma^{1/2}\right)'$$

then

$$\mathbf{y} \backsim \mathcal{N}_p(\mu, \Sigma)$$

with

$$f(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mu)'\Sigma^{-1}(\mathbf{y}-\mu)\right\}$$

properties. with $\det A \neq 0 \implies$

1. Σ positive definite (p.d.), i.e. $\mathbf{x} \neq \mathbf{0} \implies \mathbf{x}'\Sigma\mathbf{x} > 0$

[1] density transformation: $\mathbf{y} = \mathbf{y}(\mathbf{x})$, inverse $\mathbf{x} = \mathbf{x}(\mathbf{y}) \Longrightarrow f(\mathbf{y}) = f(\mathbf{x}(\mathbf{y}))\ |\det \partial\mathbf{x}/\partial\mathbf{y}|$

2. symmetric, i.e. $\Sigma = \Sigma'$

$\Longrightarrow$ **general case:** $\mathbf{y} \backsim \mathcal{N}_p(\mu, \Sigma)$, Σ symmetric, full rank

$$f(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mu)'\Sigma^{-1}(\mathbf{y}-\mu)\right\}$$

properties

1. $\mathbb{E}\mathbf{y} = \mu$
2. $\mathbb{V}\mathbf{y} = \Sigma$
3. $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ with $\mathbf{x} \backsim \mathcal{N}_p(\mathbf{0}, I)$

multivariate expectation & variance properties:

$$\mathbb{E}\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \mathbb{E}y_1 \\ \vdots \\ \mathbb{E}y_p \end{bmatrix}$$

lin transf

$$\mathbb{E}(B\mathbf{y} + \mathbf{c}) = B\mathbb{E}\mathbf{y} + \mathbf{c}$$

(Co-)variance (matrix): $\mathbb{V}(\mathbf{y}) = \mathbb{E}\left((\mathbf{y}-\mathbb{E}\mathbf{y})(\mathbf{y}-\mathbb{E}\mathbf{y})'\right) = \begin{bmatrix} \text{var}y_1 & \dots & \text{cov}(y_1, y_p) \\ \vdots & \ddots & \vdots \\ \text{cov}(y_p y_1) & \dots & \text{var}y_p \end{bmatrix}$

lin transf

$$\mathbb{V}(B\mathbf{y} + \mathbf{c}) = B\mathbb{V}(\mathbf{y})B'$$

B any rectang matrix that multiplies with $\mathbf{y}$.

Conclusion:

$$\mathbf{y} \sim \mathcal{N}_p(\mu, \Sigma) \implies \mu = \mathbb{E}\mathbf{y},\ \Sigma = \mathbb{V}\mathbf{y}$$

Proof: use

$$\Sigma \text{ symm, pd} \implies \Sigma = AA' \text{ for some } A$$

therefore, every multinormal can be generated by transforming a $\mathcal{N}_p(\mathbf{0}, I)$ wiith mean zero and variance identity

$$\mathbf{y} \sim \mathcal{N}(\mu, \Sigma) \implies \left\{ \begin{array}{r} \mathbf{x} \sim \mathcal{N}(0, I) \\ \mathbf{y} = A\mathbf{x} + \mu,\ |A| \neq 0 \\ \Sigma = AA' \end{array} \right\} \implies \left\{ \begin{array}{l} \mathbb{E}\mathbf{y} = \mu \\ \mathbb{V}\mathbf{y} = A\ \mathbb{V}\mathbf{x}\ A' = A\ A' = \Sigma \end{array} \right.$$

conversely, every $\mathbf{y} \sim \mathcal{N}_p(\mu, \Sigma)$ can be standardized $\mathbf{x} = A^{-1}(\mathbf{y} - \mu)$ into a $\mathcal{N}_p(\mathbf{0}, I)$.

any such matrix A is written sometimes $\Sigma^{1/2}$ with inverse $\Sigma^{-1/2}$

$$\Sigma^{-1/2}(\mathbf{y} - \mu) \sim \mathcal{N}_p(\mathbf{0}, I)$$

define **concentration** (or precision) matrix Σ^{-1}. notation:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & & \\ \vdots & & \ddots & \\ \sigma_{p1} & & & \sigma_{pp} \end{bmatrix}, \quad \Sigma^{-1} = \begin{bmatrix} \sigma^{11} & \sigma^{12} & \cdots & \sigma^{1p} \\ \sigma^{21} & \sigma^{22} & & \\ \vdots & & \ddots & \\ \sigma^{p1} & & & \sigma^{pp} \end{bmatrix}$$

$\implies$

$$f(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{1}{2}\sum_{ij}(y_i - \mu_i)\,\sigma^{ij}\,(y_j - \mu_j)\right)$$

Special case: p = 2

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}\left(\sigma^{11}(x_1 - \mu_1)^2 - 2\sigma^{12}(x_1 - \mu_1)(x_2 - \mu_2) + \sigma^{22}(x_2 - \mu_2)^2\right)\right\}$$

compute Σ:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \implies \Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22}(1 - \rho^2)}\begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

$\implies$

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}\left(\frac{(x - \mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x - \mu_1)}{\sigma_1}\frac{(y - \mu_2)}{\sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2}\right)\right\}$$

where $\rho = \dfrac{\sigma_{12}}{\sigma_1\sigma_2}$, i.e.

$$= C\left\{-\frac{\sigma_1^2\sigma_2^2}{2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)}\left(\frac{(x - \mu_1)^2}{\sigma_1^2} - 2\sigma_{12}\frac{(x - \mu_1)}{\sigma_1^2}\frac{(y - \mu_2)}{\sigma_2^2} + \frac{(y - \mu_2)^2}{\sigma_2^2}\right)\right\}$$

$$= C\exp\left\{-\frac{1}{2(\sigma_1^2\sigma_2^2 - \sigma_{12}^2)}\left(\sigma_2^2(x - \mu_1)^2 - 2\sigma_{12}(x - \mu_1)(y - \mu_2) + \sigma_1^2(y - \mu_2)^2\right)\right\}$$

$$= C\exp\left\{-\frac{1}{2(\sigma_{11}\sigma_{22} - \sigma_{12}^2)}\left(\sigma_{22}(x - \mu_1)^2 - 2\sigma_{12}(x - \mu_1)(y - \mu_2) + \sigma_{11}(y - \mu_2)^2\right)\right\}$$

$\Longrightarrow$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \Longrightarrow \begin{cases} \sigma^{11} = & \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}(1-\rho^2)} & = & \frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \\ \sigma^{12} = & \sigma^{21} & = & -\frac{\sigma_{12}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \\ \sigma^{22} = & \frac{\sigma_{22}}{\sigma_{11}\sigma_{22}(1-\rho^2)} & = & \frac{\sigma_{11}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2} \end{cases}$$

Note1: general case:
bivariate case: $\text{cov}(x, y) = 0 \Leftrightarrow \sigma_{12} = 0 \Leftrightarrow \sigma^{12} = 0 \Leftrightarrow X, Y$ independent
general case: $\text{cov}(x_i, x_j) = 0 \Leftrightarrow \sigma_{ij} = 0 \Leftrightarrow X_i, X_j$ marginally independent
but: $p > 2 : \ \sigma_{ij} = 0 \not\Rightarrow \sigma^{ij} = 0$.

Note2: marginal normality$\not\Rightarrow$joint normality (simple artificial examples exist)

Theorem:

1. every multinormal is uniquely defined by its expectations, variances,covariance
2. linear transformations of a multinormal are multinormal
3. marginals of multinormals are multinormal (resp uninormal)
   $\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} \sim \mathcal{N}_{p+q} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \Longrightarrow \mathbf{y} \sim \mathcal{N}_p \left(\mu_1, \Sigma_{11} \right)$
4. (jointly) multinormal random variables are marginally independent if they are uncorrelated:
$$x_i \perp x_j \Leftrightarrow f(x_i, x_j) = f(x_i) \ f(x_i) \Leftrightarrow \sigma_{ij} = 0$$
$$y_1 \perp \ldots \perp y_p \Leftrightarrow f(\mathbf{y}) = \prod_{i=1}^{p} f(y_i) \Leftrightarrow \Sigma_{11} = \text{diag}(\Sigma_{11})$$
$$\mathbf{y} \perp \mathbf{x} \Leftrightarrow f(\mathbf{y}, \mathbf{x}) = f(\mathbf{y}) \ f(\mathbf{x}) \Leftrightarrow \Sigma_{12} = 0$$
5.
$$x_i \perp x_j, \ x_j \perp x_k \not\Rightarrow x_i \perp x_k$$
6. conditionals of multinormals are multinormal (resp uninormal)
$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix} \sim \mathcal{N}_{p+q} \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yx}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yx}} \end{bmatrix} \right) \Longrightarrow \mathbf{y}|\ \mathbf{x} \sim \mathcal{N}_p \left(\mu_{\mathbf{y}\ |\mathbf{x}}, \Sigma_{\mathbf{y}|\ \mathbf{x}} \right)$$
   where $\mu_{\mathbf{y}|\ \mathbf{x}}, \Sigma_{\mathbf{y}|\ \mathbf{x}}$ later

3.2 marginal & conditional independence:

Example: (case $p = 3$): $f(y_1, y_2, x) = \mathcal{N}_3(\mathbf{0}, \Sigma)$

$\Longrightarrow$

$$f(y_1, y_2, x) = c\ \exp\left\{-\frac{1}{2}\left(y_1\sigma^{11}y_1 + y_2\sigma^{22}y_2 + x\sigma^{33}x + 2y_1\sigma^{12}y_2 + 2y_1\sigma^{13}x + 2y_2\sigma^{23}x\right)\right\}$$

with normalizing constant

$$c = \frac{1}{\sqrt{(2\pi)^3 |\Sigma|}}$$

IDEA: set $\sigma^{12} = 0$

$$\sigma^{12} = 0 \Longrightarrow f(y_1, y_2, x) = g(y_1, x)\ \ h(y_2, x)$$

condl density $f(y_1, y_2 \mid x)$: fix x in $f(y_1, y_2, x)$ and normalize:

$$f(y_1, y_2 \mid x) = c(x)\ \ f(y_1, y_2, x) \ \text{ with } 1/c(x) = \int\int f(y_1, y_2, x)\ \ dy_1 dy_2$$

$\Longrightarrow$

$$\boxed{\sigma^{12} = 0} \Longrightarrow f(y_1, y_2, x) = g(y_1, x)\ \ h(y_2, x) \Longrightarrow f(y_1, y_2 \mid x) = \underbrace{c_1(x)\ \ g(y_1, x)}_{=f(y_1, y_2 \mid x)}\ \underbrace{c_2(x)\ \ h(y_2, x)}_{=f(y_2 \mid x)}$$

DETAILS:

$$f(y_1, y_2 | x) = c\exp\left\{-\tfrac{1}{2}\left(y_1\sigma^{11}y_1 + y_2\sigma^{22}y_2 + \underbrace{x\sigma^{33}x}_{\text{absorb into } C}\right.\right.$$

$$\left.\left. +2\underbrace{y_1\sigma^{12}y_2}_{\text{mixed}} + 2y_1\sigma^{13}\underbrace{x}_{\text{fixed}} + 2y_2\sigma^{23}\underbrace{x)}_{\text{fixed}}\right)\right\}$$

$$= c(x)\exp\left\{-\frac{1}{2}\left(\sigma^{11}y_1^2 + \sigma^{22}y_2^2 + 2\underbrace{\sigma^{12}y_1y_2}_{\text{mixed}} + 2\sigma^{13}xy_1 + 2\sigma^{23}xy_2\right)\right\}$$

$$\boxed{\sigma^{12} = 0} \Longrightarrow f(y_1, y_2 | x) = c(x)\exp\left\{-\frac{1}{2}\left(\sigma^{11}y_1^2 + \sigma^{22}y_2^2 + \sigma^{13}xy_1 + \sigma^{23}xy_2\right)\right\}$$

use quadratic complements:

$$\Longrightarrow f(y_1, y_2 | x) =$$

$$= c(x)\exp\left\{-\frac{1}{2}\left(a(y_1 - bx)^2 + c(y_2 - dx)^2\right\}\right.$$

$$= c(x)\exp\left\{-\frac{1}{2}\left(a(y_1 - bx)^2\right\}\right. \exp\left\{-\frac{1}{2}\left(c(y_2 - dx)^2\right\}\right.$$

$$= f(y_1 | x)\ \ f(y_2 | x) \qquad \Longrightarrow \text{conditional independence}$$

$\Longrightarrow$

Theorem: conditional independence

1. $\mathbf{x} \sim \mathcal{N}_p(\mu, \Sigma) \implies \Sigma = [\sigma_{ij}]$ Covariance, $\Sigma^{-1} = [\sigma^{ij}]$ Concentration

$$(Y_r \perp Y_s) \mid \mathbf{x}_{-r-s} \iff \sigma^{rs} = 0$$

   where $\mathbf{x}_{-r-s} = (y_1, ..., y_{r-1}, y_{r+1}, ..., y_{s-1}, y_{s+1}, ..., y_p)$ ("conditional on all others")

2. generalization: $[\mathbf{x}', \mathbf{y}', \mathbf{z}'] \sim \mathcal{N}_{p+q+r}(\mu, \Sigma)$

$$\implies \Sigma = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{xz}} \\ \dots & \Sigma_{\mathbf{yy}} & \Sigma_{\mathbf{yz}} \\ \dots & \dots & \Sigma_{\mathbf{zz}} \end{bmatrix} \text{ Covariance} \implies \Sigma^{-1} = \begin{bmatrix} \Sigma^{\mathbf{xx}} & \Sigma^{\mathbf{xy}} & \Sigma^{\mathbf{xz}} \\ \dots & \Sigma^{\mathbf{yy}} & \Sigma^{\mathbf{yz}} \\ \dots & \dots & \Sigma^{\mathbf{zz}} \end{bmatrix} \text{ Concentration}$$

conditional independence:

$$(\mathbf{x} \perp \mathbf{z}) \mid \mathbf{y} \iff \Sigma^{\mathbf{xz}} = 0$$

Illustration: **concentration graphs**

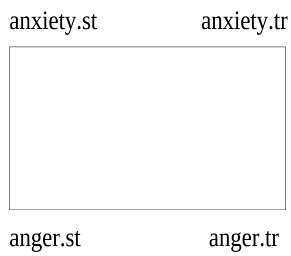
construction, separating sets

Example (Cox&Wermuth)**:**

```
library(SIN)
data(anxietyanger)
corXY<-anxietyanger$corr
stdXY<-anxietyanger$stddev
covXY<-diag(stdXY)%*%corXY%*%diag(stdXY)
concXY<-solve(covXY)
pcor<-diag(1/sqrt(diag(concXY)))%*%conXY%*%diag(1/sqrt(diag(concXY)))

options(digits=2)
corXY

           Anxiety st Anger st Anxiety tr Anger tr
Anxiety st       1.00     0.61       0.62     0.39
Anger st         0.61     1.00       0.47     0.50
Anxiety tr       0.62     0.47       1.00     0.49
Anger tr         0.39     0.50       0.49     1.00

options(digits=1)
pcor

      [,1]  [,2]  [,3]  [,4]
[1,]  1.00 -0.45 -0.47  0.04
[2,] -0.45  1.00 -0.03 -0.33
[3,] -0.47 -0.03  1.00 -0.31
[4,]  0.04 -0.33 -0.31  1.00
```

Interpretation: concentration graph:
Interpretation:

$$\texttt{angr\_st,anxi\_tr} \mid \texttt{anxi\_st,angr\_tr}$$



"a bitter person (`angr_tr`) in a state of fear (`anxi_st`) might or might not show his anger (`angr_st`), no matter if it's a coward (`anxi_tr`) or not"

$$\mathbb{P}\left(\texttt{angr\_st} \mid \texttt{anxi\_tr}, \texttt{anxi\_st}, \texttt{angr\_tr}\right) = \mathbb{P}\left(\texttt{angr\_st} \mid \texttt{anxi\_st}, \texttt{angr\_tr}\right)$$

"a bitter person in a state of fear might not show his anger even if he's brave"

Note on **Marginal & conditional independence**

1. marginal ("on all others") independence: $X^{(i)} \perp X^{(j)}$ if $\sigma_{ij} = 0$

2. conditional ("on all others") independendence $\left(X^{(i)} \perp X^{(j)}\right) \mid X^{(-i,-j)}$ if $\sigma^{ij} = 0$

3. marginal independence$\nRightarrow$conditional independence

4. marginal independence$\nLeftarrow$conditional independence

except for the case $p = 2$.

3.3 conditional mean & variance formulas:

Theorem:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \backsim \mathcal{N}_{p+q}\left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}\right)$$

then

$$f\left(\mathbf{y} \mid \mathbf{x}\right) \sim \mathcal{N}_q\left(\mu_{\mathbf{y} \mid \mathbf{x}},\ \Sigma_{\mathbf{y} \mid \mathbf{x}}\right)$$

with

$$\mu_{\mathbf{y} \mid \mathbf{x}} = \mathbb{E}\left(\mathbf{y} \mid \mathbf{x}\right) = \mu_{\mathbf{y}} + \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}(\mathbf{x} - \mu_{\mathbf{x}})$$
$$\Sigma_{\mathbf{y} \mid \mathbf{x}} = \mathbb{V}\left(\mathbf{y} \mid \mathbf{x}\right) = \Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$$

Note: the assumption of multinormality is actually more restrictive than necessary. The theorem is just a statement of the conditional expectation of a random vector $\mathbf{y}$ conditional on $\mathbf{x}$ for which "the notion of covariance makes sense"

Proof: case $(p = 2)$

$$f(y_2 \mid y_1) = f(y_1, y_2) \Big/ \int f(y_1, y_2)\ dy_2$$

$$= c(y_1) \exp\left\{ -\frac{1}{2} \left(\underbrace{\sigma^{11} (y_1 - \mu_1)^2}_{\text{absorb}} + 2\sigma^{12} \underbrace{(y_1 - \mu_1)}_{\text{fixed}} (y_2 - \mu_2) + \sigma^{22} (y_2 - \mu_2)^2 \right) \right\}$$

$$= c(y_1) \exp\left\{ -\frac{1}{2} \sigma^{22} (y_2 - \mu_2)^2 + 2\sigma^{21} (y_2 - \mu_2)(y_1 - \mu_1) \right\}$$

$$= c(y_1) \exp\left\{ -\frac{1}{2} \left(\sigma^{22} (y_2 - \mu_2)^2 + 2\sigma^{21} \underbrace{\sigma^{22}\sigma^{-22}}_{\text{trick}} (y_2 - \mu_2)(y_1 - \mu_1) \right) \right\}$$

$$= c(y_1) \exp\left\{ -\frac{1}{2} \sigma^{22} \left((y_2 - \mu_2)^2 + 2\sigma^{21}\sigma^{-22} (y_2 - \mu_2)(y_1 - \mu_1) \right) \right\}$$

quadrat compl $\Longrightarrow$

$$= c(y_1) \exp\left\{ -\frac{1}{2} \sigma^{22} \left(y_2 - \mu_2 + \sigma^{21}\sigma^{-22} (y_1 - \mu_1) \right)^2 + \underbrace{\cdots}_{\text{depending on } y_1 \text{only}} \right\}$$

$$= c(y_1) \exp -\frac{1}{2} \frac{\left(y_2 - \underbrace{\left(\mu_2 - \sigma^{21}\sigma^{-22} (y_1 - \mu_1) \right)}_{\text{mean}} \right)^2}{\underbrace{\sigma^{-22}}_{\text{variance}}}$$

$\Longrightarrow$

$$\mathbb{V}(y_2 \mid y_1) = \sigma^{-22} = \left(\frac{\sigma_{11}}{\sigma_{11}\sigma_{22} - \sigma_{21}^2} \right)^{-1} = \sigma_{22} - \frac{\sigma_{21}^2}{\sigma_{11}}$$

and

$$\mathbb{E}(y_2 \mid y_1) = \mu_2 - \frac{\sigma^{21}}{\sigma^{22}} (y_1 - \mu_1) = \mu_2 + \frac{\sigma_{21}}{\sigma_{11}} (y_1 - \mu_1)$$

qed

proof general case:

IDEA: we see from the above that $\mathbb{V}(y_2 \mid y_1) = \sigma^{-22}$

this means that the conditional variance is obtained from the (unconditional) variance by the following steps:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \Longrightarrow \Sigma^{-1} = \begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma_{21} & \sigma^{22} \end{bmatrix} \Longrightarrow \left(\sigma^{22}\right)^{-1} = \mathbb{V}\left(y_2 \mid y_1 \right)$$

covariance matrix Σ and concentration (precision) matrix Σ^{-1}

$$\Sigma = \left[\sigma_{ij}\right] \Longrightarrow \Sigma^{-1} = \left[\sigma^{ij}\right]$$

relation between covariance and precision matrix:

$$\Longrightarrow \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \Longrightarrow \Sigma^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \Longrightarrow \left(\Sigma^{22}\right)^{-1} = \mathbb{V}\left(\mathbf{y} \mid \mathbf{x} \right)$$

this can be seen from the following argument:

mvnorm:

$f\left(\mathbf{v}\right) = C \exp((\mathbf{v} - \mu)' \Sigma^{-1} \left(\mathbf{v} - \mu\right) = C \exp(\mathbf{v}'\Sigma^{-1}\mathbf{v} + \mathbf{2}\mathbf{v}'\Sigma^{-1}\mu + \cdots)$

hence, all in $f\left(\mathbf{v}\right) = C \exp(\mathbf{v}'A\mathbf{v} + \mathbf{2}\mathbf{v}'B + \cdots)$ the variance covariance matrix of $\mathbf{v}$ is A^{-1}

equally,

$f\left(\mathbf{x}, \mathbf{y}\right) = C \exp(\mathbf{x}'\Sigma^{11}\mathbf{x} + \mathbf{x}'\Sigma^{21}\mathbf{y} + \mathbf{x}'\Sigma^{12}\mathbf{y} + \mathbf{y}'\Sigma^{22}\mathbf{y})$.

conditioning on $\mathbf{x}$

$f\left(\mathbf{y} \mid \mathbf{x}\right) = C \exp(\mathbf{x}'\Sigma^{11}\mathbf{x} + \mathbf{x}'\Sigma^{21}\mathbf{y} + \mathbf{x}'\Sigma^{12}\mathbf{y} + \mathbf{y}'\Sigma^{22}\mathbf{y}) \Longrightarrow$

$$\mathbb{V}\left(\mathbf{y} \mid \mathbf{x} \right) = \left(\Sigma^{22}\right)^{-1} = \Sigma^{-22}$$

in what follows we find an expression for $\left(\Sigma^{22}\right)^{-1}$

$$\underbrace{\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}}_{\text{covmat}} \underbrace{\begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}}_{\text{precmat}} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

find an expression for the inverse of the (2,2) block of the inverse of Σ

$$\left.\begin{array}{l} \Sigma_{21}\Sigma^{12} + \Sigma_{22}\Sigma^{22} = I \\ \Sigma_{11}\Sigma^{12} + \Sigma_{12}\Sigma^{22} = 0 \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} \Sigma^{22} = \Sigma_{22}^{-1} - \Sigma_{22}^{-1}\Sigma_{21}\Sigma^{12} \\ \Sigma^{12} = -\Sigma_{11}^{-1}\Sigma_{12}\Sigma^{22} \end{array}\right.$$

$$\Longrightarrow \Sigma^{22} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma^{22} \Longrightarrow \left(1 - \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma^{22} = \Sigma_{22}^{-1}$$

$$\Longrightarrow \Sigma_{22}\left(1 - \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)\Sigma^{22} = 1$$

$$\Longrightarrow \Sigma^{-22} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

hence

$$\mathbb{V}\left(\mathbf{y}_2 \middle| \ \mathbf{y}_1\right) = \Sigma^{-22} \Longrightarrow \mathbb{V}\left(\mathbf{y}_2 \middle| \ \mathbf{y}_1\right) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

for the expectations in the case $p=2$ we have $\mathbb{E}\left(y_2|\ y_1\right)=\mu_2+\sigma_{21}\sigma_{11}^{-1}\left(y_1-\mu_1\right)$. similarly:

$$\mathbb{E}\left(\mathbf{y}_2|\ \mathbf{y}_1\right)=\mu_2+\Sigma_{21}\Sigma_{11}^{-1}\left(\mathbf{y}_1-\mu_1\right)$$

qed

Example $(p=3)$:

$$\begin{bmatrix} x \\ y_1 \\ y_2 \end{bmatrix} \backsim \mathcal{N}_3\left(\begin{bmatrix} \mu_x \\ \mu_{y_1} \\ \mu_{y_2} \end{bmatrix}, \begin{bmatrix} \sigma_{xx} & \sigma_{xy_1} & \sigma_{xy_2} \\ \sigma_{xy_1} & \sigma_{y_1y_1} & \sigma_{y_1y_2} \\ \sigma_{xy_2} & \sigma_{y_1y_2} & \sigma_{y_2y_2} \end{bmatrix}\right)$$

$\Longrightarrow$

$$\mathbb{E}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \middle| x\right)=\begin{bmatrix}\mu_{y_1} \\ \mu_2\end{bmatrix}+\begin{bmatrix}\sigma_{xy_1}/\sigma_{xx} \\ \sigma_{xy2}/\sigma_{xx}\end{bmatrix}(x-\mu_x)$$

$$\mathbb{V}\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \middle| x\right)=\begin{bmatrix} \sigma_{y_1y_1} & \sigma_{y_1y_2} \\ \sigma_{y_1y_2} & \sigma_{y_2y_2} \end{bmatrix}-\begin{bmatrix} \sigma_{y_1x}\sigma_{xy_1}/\sigma_{xx} & \sigma_{y_1x}\sigma_{xy_2}/\sigma_{xx} \\ \sigma_{y_2x}\sigma_{xy_1}/\sigma_{xx} & \sigma_{y_2x}\sigma_{xy_2}/\sigma_{xx} \end{bmatrix}=\begin{bmatrix} \cdots & \sigma_{y_1y_2}-\frac{\sigma_{y_1x}\sigma_{xy_2}}{\sigma_{xx}} \\ \cdots & \cdots \end{bmatrix}$$

in analogy to the (unconditional) variance-covariance matrix Σ the (conditional) variance-covariance matrix $\left(\Sigma^{22}\right)^{-1}$ is called the matrix of conditional (or **partial**) variances and covariances of $\mathbf{y}$ on $\mathbf{x}$

partial correlation and regression

the following theorem gives a computational as well as conceptual interpretation of partial (co)variances

Theorem(Frisch-Waugh**):**

$$\operatorname{cov}\left(y_1,y_2\mid x\right)=\operatorname{cov}(y_1-\beta_{xy_1}x,\ y_2-\beta_{xy_2}x)$$

In words: the covariance between the residuals of the reg of y_1 on x and y_2 on x (in Econometrics: "Frisch-Waugh Theorem").

Proof: use that for univariate regression $y-\mu_y=\beta_{yx}\left(x-\mu_x\right) \Longrightarrow \beta_{yx}=\sigma_{yx}/\sigma_{xx} \Longrightarrow$

$$\begin{aligned}
&\operatorname{cov}(y_1-\beta_{xy_1}x,\ y_2-\beta_{xy_2}x) \\
&=\sigma_{y_1y_2}-\beta_{xy_2}\sigma_{xy_1}-\beta_{xy_1}\sigma_{xy_2}+\beta_{xy_1}\beta_{xy_1}\sigma_{xx} \\
&=\sigma_{y_1y_2}-\sigma_{xy_2}\sigma_{xy_1}/\sigma_{xx}-\sigma_{xy_1}\sigma_{xy_2}/\sigma_{xx}+\sigma_{xy_1}\sigma_{xy_2}/\sigma_{xx} \\
&=\sigma_{y_1y_2}-\frac{\sigma_{xy_1}\sigma_{xy_2}}{\sigma_{xx}}
\end{aligned}$$

qed

Applications:
1) model assessment

$$y_1 = \alpha_1 + \beta_1 x + \epsilon_1$$
$$y_2 = \alpha_2 + \beta_2 x + \epsilon_2$$

with

$$\epsilon \sim \mathcal{N}_2 (\mathbf{0}, \Sigma)$$

$\Longrightarrow$

$$\text{cov}(y_1, y_2 \mid x) = \text{cov} (y_1 - \alpha_1 + \beta_1 x,\ y_2 - \alpha_2 + \beta_2 x) = \text{cov}(\epsilon_1, \epsilon_2) \underbrace{= 0}_{\text{want}}$$

in a *good model* no further variation, covariation remains to be explained, i.e., $\mathbb{V} (\mathbf{y} \mid \mathbf{x}) = \left(\Sigma^{22}\right)^{-1}$ is diagonal

2)

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$\text{cov}(y, x_2 \mid x_1) = cov (y - \beta x_1, x_2 - \gamma x_1)$ the amount of covariation (resp corelation) between y, x_2 not explained by x_1

partial corelation:

$$(\text{diag}(\mathbb{V} (\mathbf{y}_2 | \ \mathbf{y}_1)))^{-1/2} \ \ \mathbb{V} (\mathbf{y}_2 | \ \mathbf{y}_1) \ \ (\text{diag}(\mathbb{V} (\mathbf{y}_2 | \ \mathbf{y}_1)))^{-1/2}$$

Notes:

1. $\Sigma^{-1} = \left[\sigma^{ij}\right]$ matrix of partial covariances between residuals of regressions of y_i on $\mathbf{y}_{-i-j}$ and of y_j on $\mathbf{y}_{-i-j}$. Therefore

$$\frac{\sigma^{ij}}{\sqrt{\sigma^{ii}}\sqrt{\sigma^{jj}}}$$

partial correlations of residuals of regressions of y_i on $\mathbf{y}_{-i-j}$ and of y_j on $\mathbf{y}_{-i-j}$

2. interpretation: the correlation between two variables "*after* accounting for..." or after adjusting for" or "correcting for" or "after partialling out" ....

3. Under the normality assumption a partial covariance equal zero can be interpreted in terms of *conditional independence*

4. without the normality assumption partial covariance can be interpreted as the amount covariance or "covariation" between... or of "variance shared by..." or "*common variance between...* " after adjusting for the effects of.... Similarly, partial correlation can be interpreted as a measure of colinearity betwee... after removing the effects of...

5. a zero in the conditional variance matrix implies that a regression on the conditioning variables supports the assumption of uncorrelated disturbances. the predictors not only explain all variation as well as covariation among the target variables.

6. the coefficients of the matrix $\Sigma_{yx}\Sigma_{xx}^{-1}$ are called partial regression coefficients and interpreted similarly.

Example: Simpsons paradox. There are examples of data showing
D : damand for gas
P : price
I : income

$$\begin{array}{lll} \mathbb{E}\left(D \mid p\right) \uparrow \text{ as } p \uparrow & \Longrightarrow \mathbb{E}\left(D \mid p\right) = \alpha + \beta p & \text{with } \beta > 0 \\ \mathbb{E}\left(D \mid p, I\right) \downarrow \text{ as } p \uparrow \text{ for all } I\text{, i.e.} & \mathbb{E}\left(D \mid p, I\right) = \alpha + \beta p + \gamma I & \text{with } \beta < 0 \end{array}$$

explanation: income drives up demand for gas even as prices rise

3.4 parameter estimation of multinormal distribution

general case: $\mathbf{x} \backsim \mathcal{N}\left(\mu, \Sigma\right)$

$$f(\mathbf{x}|\ \mu,\ \Sigma) = \frac{1}{\sqrt{(2\pi)^p\,|\Sigma|}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}-\mu\right)'\Sigma^{-1}\left(\mathbf{x}-\mu\right)\right\}$$

maximum likelihood estimation of $\mu,\ \Sigma$

$\mathbf{x}_1, ..., \mathbf{x}_n$ iid random sample

$$\begin{aligned} l(\mu, \Sigma|\ \mathbf{x}_1, ..., \mathbf{x}_n) &= \log \prod_{i=1}^{n} f(\mathbf{x}_i|\ \mu,\ \Sigma) = \sum_{i=1}^{n} l(\mu, \Sigma|\ \mathbf{x}_i) \\ &= -\frac{n}{2}\log|\Sigma|^{-np/2} - \frac{1}{2}\sum_{i=1}^{n}\left(\mathbf{x}_i - \mu\right)'\Sigma^{-1}\left(\mathbf{x}_i - \mu\right) \end{aligned}$$

but

$$\begin{aligned}(\mathbf{x}-\mu)'\,\Sigma^{-1}\,(\mathbf{x}-\mu) &= (\mathbf{x}-\overline{\mathbf{x}}+\overline{\mathbf{x}}-\mu)'\,\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}}+\overline{\mathbf{x}}-\mu)\\ &= (\mathbf{x}-\overline{\mathbf{x}})'\,\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}}) + (\overline{\mathbf{x}}-\mu)'\,\Sigma^{-1}\,(\overline{\mathbf{x}}-\mu)\\ &+ 2\,(\overline{\mathbf{x}}-\mu)'\,\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}})\end{aligned}$$

summing

$$\sum (\mathbf{x}-\mu)'\,\Sigma^{-1}\,(\mathbf{x}-\mu) = \sum_i^n (\mathbf{x}-\overline{\mathbf{x}})'\,\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}}) + n\,(\overline{\mathbf{x}}-\mu)'\,\Sigma^{-1}\,(\overline{\mathbf{x}}-\mu)$$

but since[2] trABC =tr(CAB) =tr(BCA)

$$\begin{aligned}\sum_i^n (\mathbf{x}-\overline{\mathbf{x}})'\,\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}}) &= \sum_i^n \mathrm{tr}\,(\mathbf{x}-\overline{\mathbf{x}})'\,\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}})\\ &= \sum_i^n \mathrm{tr}\Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}})\,(\mathbf{x}-\overline{\mathbf{x}})'\\ &= \mathrm{tr}\sum_i^n \Sigma^{-1}\,(\mathbf{x}-\overline{\mathbf{x}})\,(\mathbf{x}-\overline{\mathbf{x}})'\\ &= \mathrm{tr}\Sigma^{-1}\sum_i^n (\mathbf{x}-\overline{\mathbf{x}})\,(\mathbf{x}-\overline{\mathbf{x}})'\end{aligned}$$

$\Longrightarrow$ with

$$S = \frac{1}{n}\sum_i^n (\mathbf{x}-\overline{\mathbf{x}})\,(\mathbf{x}-\overline{\mathbf{x}})'$$

we have

$$l(\mu,\Sigma|\ \mathbf{x}_1,...,\mathbf{x}_n) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\Sigma^{-1}nS - \frac{n}{2}\,(\overline{\mathbf{x}}-\mu)'\,\Sigma^{-1}\,(\overline{\mathbf{x}}-\mu)$$

it can be seen directly that the maximiser is

$$\widehat{\mu} = \overline{\mathbf{x}}$$

on the other hand, it is far more technical to obtain the MLE of the covariance matrix

$$\widehat{\Sigma} = S + (\overline{\mathbf{x}}-\mu)\,(\overline{\mathbf{x}}-\mu)' \quad \text{if } \mu \text{ known}$$

hence, with $\widehat{\mu} = \overline{\mathbf{x}}$

$$\widehat{\Sigma} = S \quad \text{if } \mu\ \textbf{un}\text{known}$$

[2] show (easy): $trAB = tr(BA)$

this implies $trABC = tr(CAB) = tr(BCA)$:

Note: this is a direct generalization of the univariate case

$$\mu = \overline{x}$$
$$\widehat{\sigma}^2 = \frac{1}{n}\sum (x_i - \mu)^2 = \frac{1}{n}\sum (x_i - \overline{x})^2 + (\mu - \overline{x})^2$$

Likelihood ratio tests

$\mathbf{x}_1, \ldots, \mathbf{x}_n$ iid $\sim \mathcal{N}_p(\mu, \Sigma)$

$$H_0: \ \mu = \mu_0, \ \Sigma \ \textbf{known}$$

log likelihood H_0 :

$$l(\mu, \Sigma| \ \mathbf{x}_1, ..., \mathbf{x}_n) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\Sigma^{-1}nS - \frac{n}{2}(\overline{\mathbf{x}} - \mu_0)' \Sigma^{-1} (\overline{\mathbf{x}} - \mu_0)$$

max log likelihood,

$$l(\widehat{\mu}, \Sigma| \ \mathbf{x}_1, ..., \mathbf{x}_n) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\Sigma^{-1}nS$$

$\Longrightarrow$ LR-test:

$$\lambda = -2\left(l(\mu, \Sigma| \ \mathbf{x}_1, ..., \mathbf{x}_n) - l(\widehat{\mu}, \Sigma| \ \mathbf{x}_1, ..., \mathbf{x}_n)\right) = n(\overline{\mathbf{x}} - \mu_0)' \Sigma^{-1} (\overline{\mathbf{x}} - \mu_0)$$

under H_0 :

$$\lambda \sim \chi^2 (df_1 - df_0)$$

where $df_1 = p$, $df_0 = 0$ as $n \to \infty$

Note: exact test under assumption of normality: same

general one-sample test on mean

$\mathbf{x}_1, \ldots, \mathbf{x}_n$ iid $\sim \mathcal{N}_p(\mu, \Sigma)$

$$H_0: \ \mu = \mu_0, \ \Sigma \ \textbf{un}\text{known}$$

log likelihood

$$l(\mu, \Sigma| \ \mathbf{x}_1, ..., \mathbf{x}_n) = -\frac{n}{2}\log|\Sigma| - \frac{1}{2}\mathrm{tr}\Sigma^{-1}nS - \frac{n}{2}(\overline{\mathbf{x}} - \mu)' \Sigma^{-1} (\overline{\mathbf{x}} - \mu)$$

$\Longrightarrow$ LR-test:

$$\lambda = -2\left(l(\mu_0, \widehat{\Sigma}_0 \Big| \; \mathbf{x}_1, ..., \mathbf{x}_n) - l(\widehat{\mu}, \widehat{\Sigma} \Big| \; \mathbf{x}_1, ..., \mathbf{x}_n)\right)$$
$$\widehat{\Sigma}_0 = S + (\overline{\mathbf{x}} - \mu_0)(\overline{\mathbf{x}} - \mu_0)'$$
$$\widehat{\Sigma} = S, \; \widehat{\mu} = \overline{\mathbf{x}}$$

as can be shown[3]
under $H_0: \; \mu = \mu_0$

$$\lambda \sim \chi^2 (df_1 - df_0)$$

where $df_1 = p$, $df_0 = 0$ as $n \to \infty$

Note:

1. (it can be shown:) simplification$\lambda = n \log(1 + (\overline{\mathbf{x}} - \mu_0)' S^{-1} (\overline{\mathbf{x}} - \mu_0))$
2. exact test (under strict assumption of normality):

   $$\frac{n-p}{p} (\overline{\mathbf{x}} - \mu_0)' S^{-1} (\overline{\mathbf{x}} - \mu_0) \sim F_{p,n-p}$$

   This is the one-sample **Hotelling** test with

   $$T^2 = (\overline{\mathbf{x}} - \mu_0)' S^{-1} (\overline{\mathbf{x}} - \mu_0)$$

   remember[4]: $pF_{p,n-p} \to \chi^2(p)$ as $n \to \infty$

3. The corresponding 2-sample Hotelling test statistic is for $H_0: \; \mu_1 = \mu_2$ is

   $$T^2 = (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_1)' S^{-1} (\overline{\mathbf{x}} - \overline{\mathbf{x}}_2)$$

   This holds under the *assumption of equal variances*

4. The general problem $\mathbf{x}_1, \ldots, \mathbf{x}_m$ iid$\sim \mathcal{N}(\mu_1, \Sigma_1)$ independent $\mathbf{y}_1, \ldots, \mathbf{y}_n$ iid$\sim \mathcal{N}(\mu_2, \Sigma_2)$ needs special care. (This is called the *Behrends-Fisher problem).*

Example: swiss bank notes.

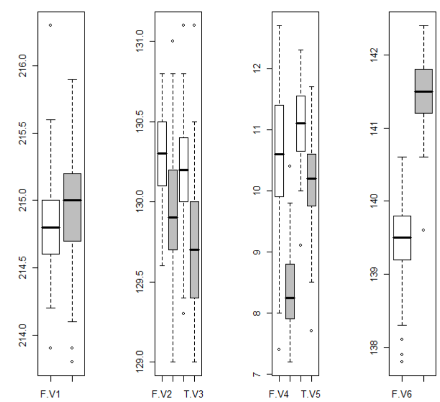
```
Six variables measured on 1000 genuine and 100 counterfeit old Swiss
1000-franc bank notes.  The data stem from Flury and Riedwyl (1988).
The columns correspond to the following 6 variables.
   $X_{1}$:      Length of the bank note,
```

[3] same as in the 1-dim case:
$\frac{1}{n}\sum(x_i - \mu_0)^2 = \frac{1}{n}\sum(x_i - \overline{x} + \overline{x} - \mu_0)^2 = \frac{1}{n}\sum\left((x_i - \overline{x})^2 + (\overline{x} - \mu_0)^2\right)$

[4] $\left.\begin{array}{l} F = \frac{\chi_1^2(p)/p}{\chi_1^2(n-p)/(n-p)} \\ \frac{\chi_1^2(n-p)}{n-p} \to 1 \end{array}\right\} \Longrightarrow pF \to \chi_1^2(p)$

```
$X_{2}$:      Height of the bank note, measured on the left,
$X_{3}$:      Height of the bank note, measured on the right,
$X_{4}$:      Distance of inner frame to the lower border,
$X_{5}$:      Distance of inner frame to the upper border,
$X_{6}$:      Length of the diagonal.
```

Observations 1-100 are the genuine bank notes and the other 100 observations are the counterfeit bank notes.




```
     xbar     mu
V1 214.823 214.969
V2 130.300 129.943
V3 130.193 129.720
V4  10.530   8.305
V5  11.133  10.168
V6 139.450 141.517

> dd<-t(xbar-mu)%*%solve(S)%*%(xbar-mu)
> chiq<-n*(log(1+dd))
> round(chiq,3)
[1,] 430.254
> chiqExact<-(n-p)/p*dd
> round(chiqExact,3)
[1,] 1141.895
```

Notes:

1. Note: this is a one-sample LR-test of the (not very interesting) hypothesis that the population mean of falsified bank notes equals the (sample)mean of true bank notes.

2. the test-statistic is asymptotic in nature, its apprication is fully justified in a sample of 100 falsified bank notes

3. the huge size of the LR-test statistic from its distribution under the null hypothesis is to be expected since the differences between both samples are clearly visible to the naked eye.

4. the assumption of normality is reasonably satisfied as density plots from the sample clearly suggest. even under clear deviations from the normality assumption the results could be interpreted reasonable, see the article of H.White

5. the more practical problems of two samples and few (normally distributed) observations are much more technically to be solved (Hotellings T-distribution, Behrends-Fisher problem)

3.5 Appendix:

prop: $\Sigma = AA' \implies \Sigma^{-1} = \left(A^{-1}\right)' A^{-1}$

proof:$\Sigma = AA' \implies \Sigma^{-1} = (AA')^{-1} = (A')^{-1} A^{-1} = \left(A^{-1}\right)' A^{-1}$
last equation because $(A')^{-1} = \left(A^{-1}\right)'$ because
$I = (A')^{-1} A' = \left((A')^{-1} A'\right)' = A\left((A')^{-1}\right)' = I \implies \left((A')^{-1}\right)' = A^{-1} \implies (A')^{-1} = \left(A^{-1}\right)'$

qed

prop: $\Sigma = AA' \implies \det A^{-1} = (\det \Sigma)^{-1/2}$

proof: $\det A^{-1} = (\det A)^{-1} = (\det A \det A')^{-1/2} = (\det AA')^{-1/2} = (\det \Sigma)^{-1/2}$

qed

literature

1. `SIN package R`

2. Cox, D.R. & Wermuth, N. (1993) Linear Dependencies Represented by Chain Graphs. Statistical Science 8(3): 204-283. (See Table

3. Spielberger, C.D., Gorsuch, R.L. and Luschene, R.E. (1970) Manual for the State Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto, CA

4. Spielberger, C.D., Russell, S. and Crane, R. (1983). Assessment of anger. In Advances in Personality Assessment, (J.N. Butcher and C.D. Spielberger, eds.), 2 159-187, Erlbaum Hillsdale, NJ.

5. Mardia et al "Multivariate Analysis, Academic Press

6. Härdle, W & Simar, L "Applied Multivariate Analysis", Springer