02409 Multivariate Statistics

Lecture K, November 17 2025
Anders Stockmarr
anst@dtu.dk

Course developers:
Anders Stockmarr
Anders Nymark Christensen

Agenda

- Multivariate normal modeling
- Transformations to normality
- Main Topic: Linear multivariate analysis

Hotellings T² in two-sample case

Theorem 4.9

We use the same notation as given above. Now, let

$$T^2 = \frac{nm}{n+m}(\bar{X}-\bar{Y})^T \mathbf{S}^{-1}(\bar{X}-\bar{Y}).$$

Then the critical region for a test of H_0 against H_1 at level α is equal to

$$C = \{x_1, \ldots, x_n, y_1, \ldots, y_m | \frac{n+m-p-1}{(n+m-2)p} t^2 > \mathrm{F}(p, n+m-p-1)_{1-\alpha}\}$$

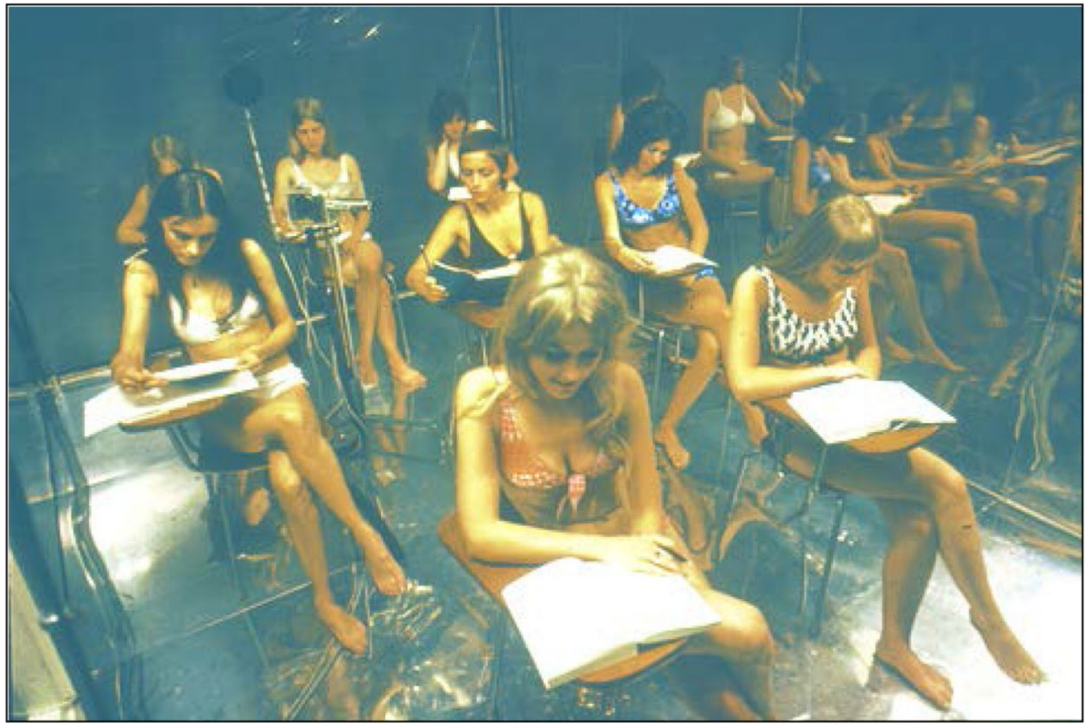
Here t^2 is the observed value of T^2.

Example: Heat Climate Data

- At the Laboratory for Heating- and Ventilation, DTU, one has measured the following on 16 men and 16 women:

  - The height in cm.
  - The evaporation loss in g/m2 skin during a 3 hour period.
  - The mean skin temperature in °C.

- The mean temperature is found by measuring the skin temperature at 14 different locations every minute for 5 minutes (same locations every time). The mean temperature is then an average of all $14 \times 5 = 70$ measurements.



Data:

`heatclimate.txt`

Hotelling's T-Square in the Two Sample Case

H_0: Evaporation and skin temperature is independent of sex

```
> T2<-(n*m/(n+m))*t(Xbar-Ybar)%*%solve(S)%*%(Xbar-Ybar)
> T2
         [,1]
[1,] 4.612128
> ((n+m-p-1)/((n+m-2)*p))*T2
         [,1]
[1,] 2.229195
> 1-pf(((n+m-p-1)/((n+m-2)*p))*T2,p,n+n-p+1)
          [,1]
[1,] 0.1245838
```

The data support that males and females have the same measurements (p=0.12).

- Did we check the model? **NO, WE DID'T!**

Multivariate Statistics
From Lecture 1:

Generally we shall distinguish between three different types of data approaches:

$$\begin{array}{c|ccc} Y_1 & X_{11} & \cdots & X_{k1} \\ \vdots & \vdots & & \vdots \\ Y_n & X_{1n} & \cdots & X_{xn} \end{array}$$

Multiple regression

$$\begin{array}{ccc|ccc} Y_{11} & \cdots & Y_{d1} & X_{11} & \cdots & X_{k1} \\ \vdots & & \vdots & \vdots & & \vdots \\ Y_{1n} & \cdots & Y_{dn} & X_{1n} & \cdots & X_{xn} \end{array}$$

Multivariate regression

$$\begin{array}{ccc} X_{11} & \cdots & X_{k1} \\ \vdots & & \vdots \\ X_{1n} & \cdots & X_{xn} \end{array}$$

Structural data analysis

Multivariate Normal Models

- Matrix observations in $\mathbb{R}^{n \times p}$!
- For the heat climate data, $n = 32$ and $p = 2$:

$$Y = (evaporation, temperature): \; Y_i \sim N_2(\mu_i, \Sigma), \; i = 1, \dots, 32,$$

and values for each individual are assumed *independent*.

- How to describe the distribution of the entire data matrix?

Tensor Products of Matrices

- Suppose that A is an $m \times n$ matrix, and B an $p \times q$ matrix.

- We define the *Kroncker tensor product* of A and B (we just call it the *tensor product*) $A \otimes B$ as the $mp \times nq$ matrix with entries

$$(A \otimes B)_{ir,js} = a_{ij} b_{rs}$$

Tensor Products of Matrices

- Two different representations:

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1j}B & \cdots & a_{1n}B \\ \vdots & & \vdots & & \vdots \\ a_{i1}B & \cdots & a_{ij}B & \cdots & a_{in}B \\ \vdots & & \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mj}B & \cdots & a_{mn}B \end{bmatrix} = \begin{bmatrix} Ab_{11} & \cdots & Ab_{1s} & \cdots & Ab_{1q} \\ \vdots & & \vdots & & \vdots \\ Ab_{r1} & \cdots & Ab_{rs} & \cdots & Ab_{rq} \\ \vdots & & \vdots & & \vdots \\ Ab_{p1} & \cdots & Ab_{ps} & \cdots & Bb_{pq} \end{bmatrix}$$

Tensor Products of Matrices

- Of special Interest: $n = m, A = I_n, B = \Sigma$:

$$I_n \otimes \Sigma = \begin{bmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{bmatrix}$$

- In the heat climate case: 16 copies of the 2×2 matrix Σ in the diagonal.

- Rows are stochastically independent, variance between columns are Σ.

Tensor Products of Matrices

- Tensor products as linear mappings:

$$A \otimes \mathrm{B}:\ \mathbb{R}^{n \times q} \to \mathbb{R}^{m \times p}$$
$$x \mapsto AxB^T$$

since

$$\left(AxB^T\right)_{ir} = \sum_{\nu} \sum_{\kappa} a_{i\nu} x_{\nu\kappa} b_{r\kappa} = \sum_{\nu,\kappa} (A \otimes B)_{ir,\nu\kappa} x_{\nu\kappa}\,.$$

Tensor Products of Matrices

- Tensor products calculus:

Remark A.53

Tensor product rules.

i) $0 \otimes A = A \otimes 0 = 0$

ii) $(A_1 + A_2) \otimes B = A_1 \otimes B + A_2 \otimes B$

iii) $A \otimes (B_1 + B_2) = A \otimes B_1 + A \otimes B_2$

iv) $\alpha A \otimes \beta B = \alpha\beta A \otimes B$

v) $A_1 A_2 \otimes B_1 B_2 = (A_1 \otimes B_1)(A_2 \otimes B_2)$

vi) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if the inverses exist

vii) $(A \otimes B)^{-} = A^{-} \otimes B^{-}$

viii) $(A \otimes B)^T = A^T \otimes B^T$

Multivariate Normal Models

$Y = (evaporation, temperature)$: $Y_i \sim N_2(\mu_i, \Sigma)$, $i = 1, \ldots, 32,$ and values for each individual are assumed *independent*.

Take $\mu \in \mathbb{R}^{32\times 2}$ $to\ be$ $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_{32} \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_{32} \end{bmatrix}, \Sigma = V(Y_1) = \cdots = V(Y_{32}).$

Then

$$\boldsymbol{Y} \sim \boldsymbol{N}_{\mathbf{32\times 2}}(\boldsymbol{\mu}, \boldsymbol{I}_{\mathbf{32}} \otimes \boldsymbol{\Sigma})$$

Example: Heat Climate Data

$$Y = (evaporation, temperature): \; Y_i \sim N_2(\mu_i, \Sigma), \; i = 1, \dots, 32,$$

and values for each individual are assumed *independent*.

Let us assume that the mean of Y depends on `sex` and `height`:

$$H: \mu_i = \begin{pmatrix} \mu_{evap,i} \\ \mu_{temp,i} \end{pmatrix} = \alpha_{sex} + \beta \cdot \mathrm{height}_i = \begin{bmatrix} \alpha_{sex,evap} \\ \alpha_{sex,temp} \end{bmatrix} + \begin{bmatrix} \beta_{evap} \cdot height_i \\ \beta_{temp} \cdot height_i \end{bmatrix}$$

Remember the criteria for multivariate normality (Lecture 1):

Y multivariate normal $\Leftrightarrow$
Y_2 is univariate normal, and $Y_1 | Y_2$ is univariate normal.

Example: Heat Climate Data

$Y = (evaporation, temperature)$: $Y_i \sim N_2(\mu_i, \Sigma)$, $i = 1, \ldots, 32$,
and values for each individual are assumed *independent*.

$$H: \mu_i = \begin{pmatrix} \mu_{evap,i} \\ \mu_{temp,i} \end{pmatrix} = \alpha_{sex} + \beta \cdot \mathrm{height}_i = \begin{bmatrix} \alpha_{sex,evap} \\ \alpha_{sex,temp} \end{bmatrix} + \begin{bmatrix} \beta_{evap} \cdot height_i \\ \beta_{temp} \cdot height_i \end{bmatrix}$$

$$Y_{1,i} | Y_{2,i} \sim N\left(\alpha_{sex,evap} + \beta_{evap} \cdot \mathrm{height}_i + \Sigma_{evap,temp} \Sigma_{temp,temp}^{-1} \left(Y_{2,i} - \left(\alpha_{sex,temp} + \beta_{temp} \cdot height_i\right)\right), \Sigma_{1\cdot 2}\right)$$

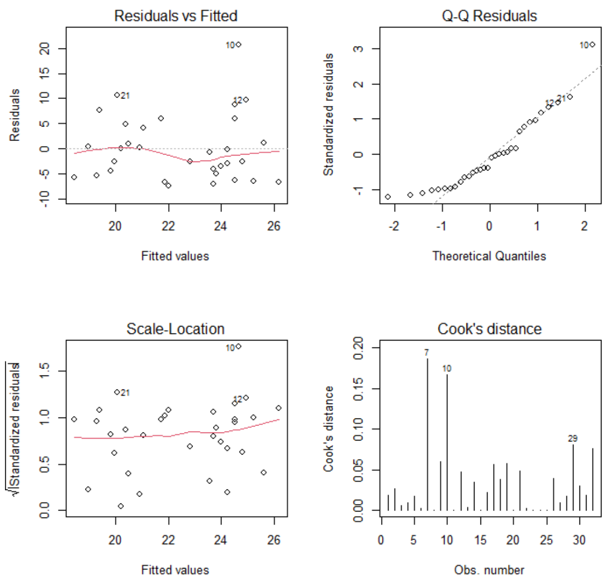
$$= N\left(\tilde{\alpha} + \tilde{\beta} \cdot height_i + \gamma Y_{2,i}, \Sigma_{1\cdot 2}\right)$$

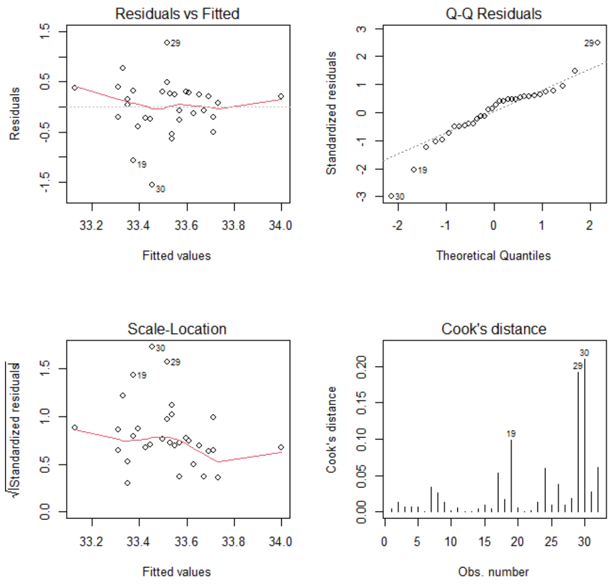
$$Y_{2,i} \sim N\left(\alpha_{sex,temp} + \beta_{temp} \cdot \mathrm{height}_i, \Sigma_{temp,temp}\right)$$

Example: Heat Climate Data

- Model control:

```
model0.1<-lm(evap~height+sex+temp,data=heatclimate)
model0.2<-lm(temp~height+sex,data=heatclimate)
```





- Not super. Tendency to increasing variances but not clear.

Transformations Towards Normality

- The Box-Cox transformation:

$$y^{(\lambda)} = \begin{cases} y^{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}$$

- Trick: Evaluate the likelihood function, and work out the value with the highest likelihood.

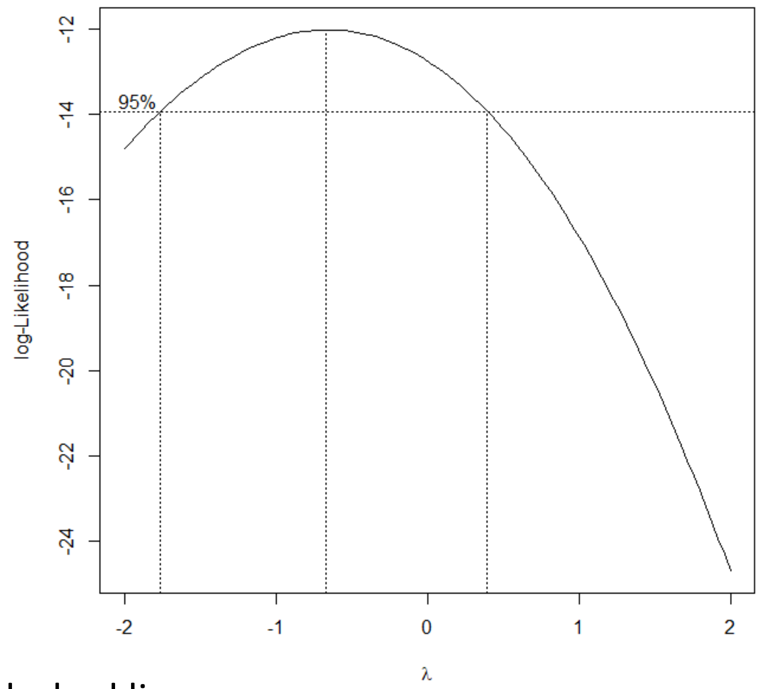
- Implementations include a confidence interval for λ. One can then select an appropriate value within this.

- $\lambda = 1$ means no transformation.

- Details: *Box GEP & Cox DR (1964).*

Transformations Towards Normality

```
boxcox(evap~height+sex+temp,data=heatclimate)
```

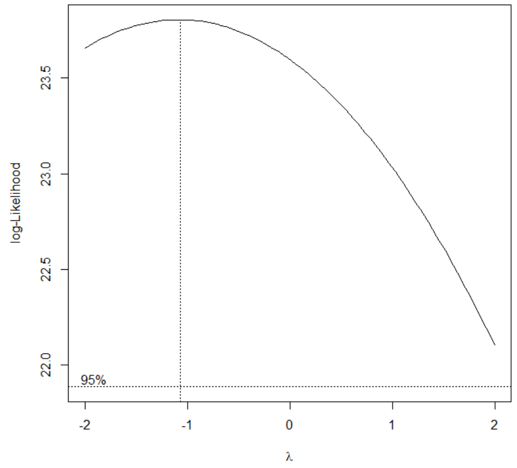


- Confidence interval by outer vertical dashed lines.
- 0 is in the confidence interval; we will use 0 (for a start).

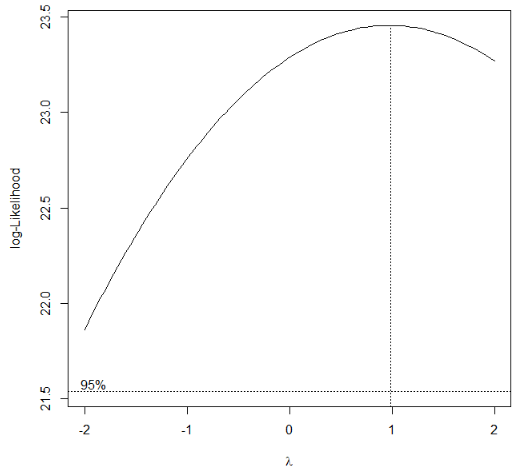
- Box-Cox Transformation of temp is indecisive; very flat likelihood function. We decide from the above.

Transformations Towards Normality

```
boxcox(log(evap)~height+sex+log(temp),data=heatclimate)
```




```
boxcox(1/log(evap)~height+sex+1/log(temp),data=heatclimate)
```



Transformations Towards Normality

- **Benefits:**
- Data closer to normality;
- Variance stabilization;
- Skewness reduction.

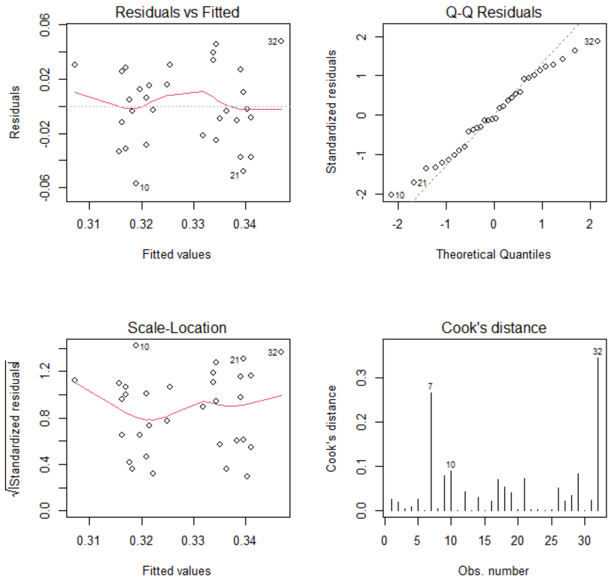
- **Drawbacks:**
- Data must be positive (workarounds exists);
- Harder to interpret coefficients;
- No guarantee that the optimal trasnformation makes the data normal.

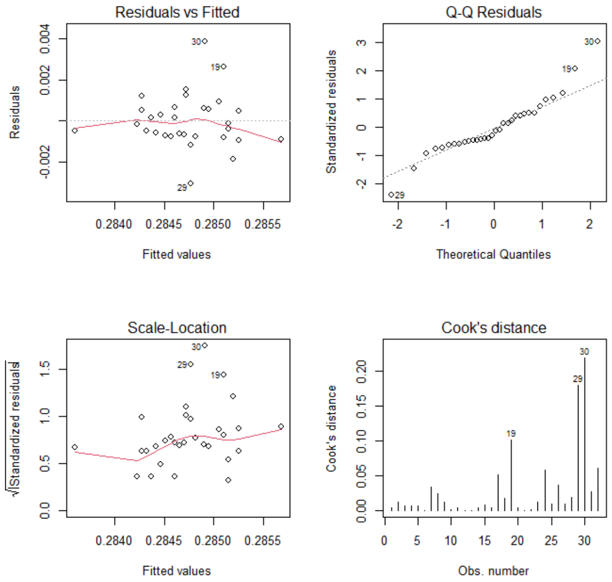
The last point is what causes us to make a **convenience choice** withing the confidence interval.

Example: Heat Climate Data

- Model control:

```
model0.1<-lm(1/log(evap)~height+sex+1/log(temp),data=heatclimate)
model0.2<-lm(1/log(temp)~height+sex,data=heatclimate)
```





- Better. Observations 30 and 32 are potentially somewhat influential but appears to be the end of a continuum. And not critically influential.

Example: Heat Climate Data

```
> summary(heatclimate)
 sex        height            evap             temp      
 f:16   Min.   :157.0   Min.   :12.60   Min.   :31.90  
 m:16   1st Qu.:166.2   1st Qu.:18.48   1st Qu.:33.20  
        Median :173.0   Median :20.75   Median :33.55  
        Mean   :172.9   Mean   :22.49   Mean   :33.52  
        3rd Qu.:180.0   3rd Qu.:25.65   3rd Qu.:33.90  
        Max.   :190.0   Max.   :45.40   Max.   :34.80  
> heatclimate[c(30,32),]
   sex height evap temp
30   f    164 20.2 31.9
32   f    180 12.6 33.5
```

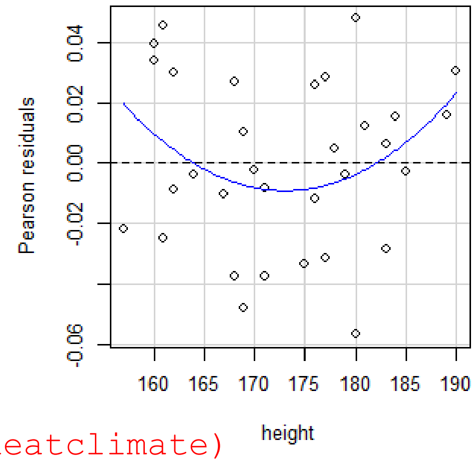
- Observation 32 is a tall female with low evaporation and median temperature, but not extreme in more than one category (evaporation). Observation 30 is similarly extreme in temperature but not evaporation. And, none of them critically influential.

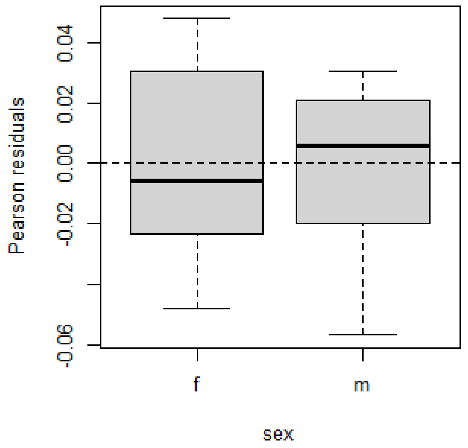
Example: Heat Climate Data

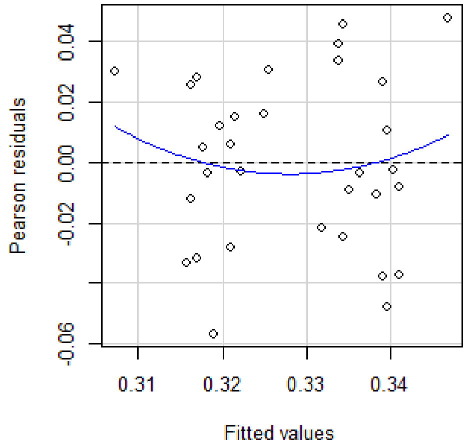
- Controlling linearity:

```
> residualPlots(model0.1)
           Test stat Pr(>|Test stat|)
height        1.8060          0.08168 .
sex
Tukey test    0.6297          0.52887
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> temp<-lm(1/log(evap)~height+sex+1/log(temp)+I(1/log(temp)^2),data=heatclimate)

> drop1(temp,test="F")
Model:
1/log(evap) ~ height + sex + 1/log(temp) + I(1/log(temp)^2)
                 Df Sum of Sq      RSS     AIC F value Pr(>F)
<none>                        0.024149 -222.06
height            1 0.0007185 0.024868 -223.12  0.8331 0.3692
sex               1 0.0032254 0.027374 -220.04  3.7398 0.0633 .
I(1/log(temp)^2)  1 0.0005192 0.024668 -223.38  0.6019 0.4443
-
```



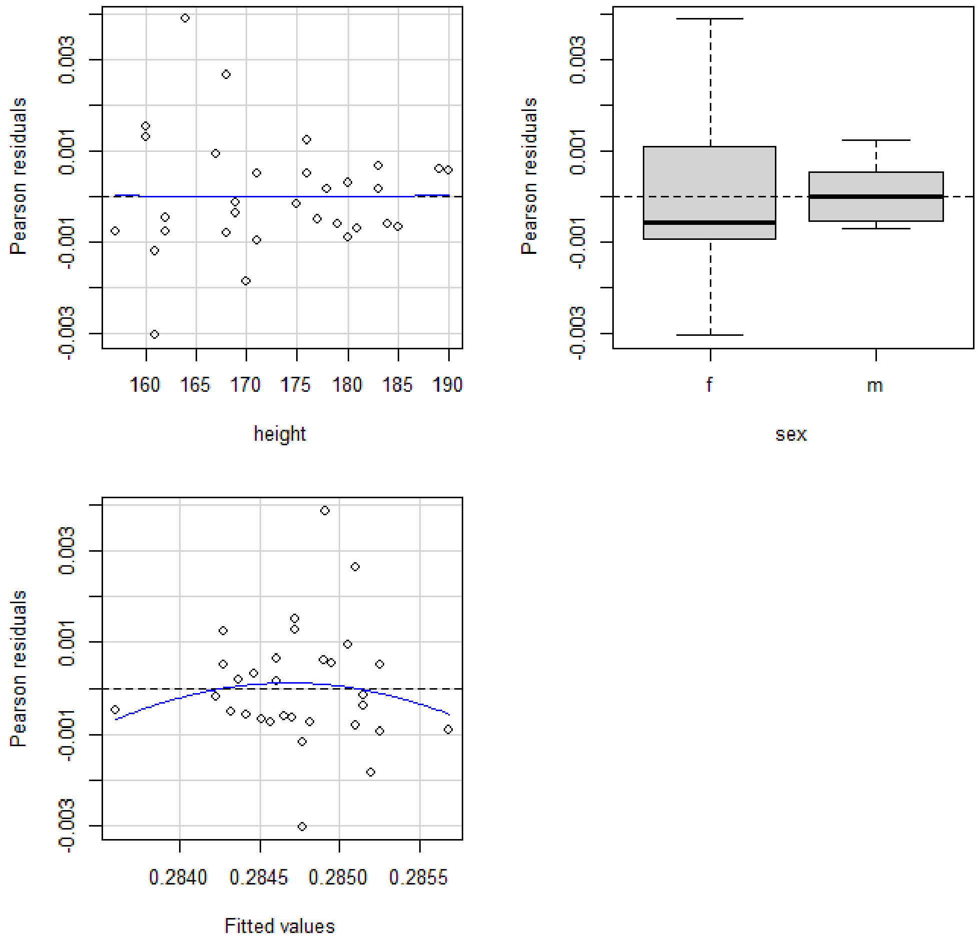




Example: Heat Climate Data

- Controlling linearity:

```
> residualPlots(model0.2)
           Test stat Pr(>|Test stat|)
height        0.0485           0.9617
sex
Tukey test   -0.7543           0.4507
```



Example: Heat Climate Data

- We accept the hypothesis of simulataneous normality of transformed variables when regressing on sex and height:

$$g(y) = \frac{1}{\log(y)};$$

$$g(Y) \sim N_{32\times 2}(\alpha_{sex} + \beta \cdot height, \Sigma)$$

- Now, how to analyze the deviation between sex, and with the potential confounder height?

Example: Heat Climate Data

Lets us redefine Y:

$$Y = [1/\log(evaporation) \quad 1/\log(temperature)]$$

$$Y = X\theta + \varepsilon, \qquad \varepsilon \sim N_{n\times p}(0, \mathrm{I_n} \otimes \Sigma)$$

with

$$X = [\mathbf{1}; 1_{male}; height]$$

$$\theta = \begin{bmatrix} \alpha_{evap} & \alpha_{temp} \\ \alpha_{evap,male} & \alpha_{temp,male} \\ \beta_{evap} & \beta_{temp} \end{bmatrix}$$

- Now, how to analyze the deviation between sex, and with the potential confounder height?

Likelihood Estimation
In the Multivariate Normal Model

- Likelihood function:

$$\mathrm{L}(\theta,\Sigma) = \frac{1}{(2\pi)^{np/2}} det(\Sigma)^{-n/2} \exp\left(-\frac{1}{2}\|Y-\mu\|^2_{I_n \otimes \Sigma^{-1}}\right)$$

- We need to project Y wrt. the inner product induced by $I_n \otimes \Sigma^{-1}$ to obtain the maximum likelihood estimator.

Likelihood Estimation
In the Multivariate Normal Model

- Note that for $R^{n \times p}$, the euclidean inner product is

$$< x, y >= \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij} y_{ij} = tr\left(x^T y\right) = tr\left(x y^T\right)$$

where tr is the $trace$, the sum of the diagonal elements.

- The trace satisfies:

1. $tr(AB) = tr(BA)$
2. $tr(\lambda A + \mu B) = \lambda tr(A) + \mu tr(B)$ for $\lambda, \mu \in \mathbb{R}$
3. $tr(A) = tr(A^T)$

Likelihood Estimation In the Multivariate Normal Model

- Note that

$$< x, y >_{I_n \otimes \Sigma^{-1}} = < x, I_n \otimes \Sigma^{-1} y > = < x, y\Sigma^{-1} > = tr\left(x\Sigma^{-1}y^T\right)$$

- Also, for a subspace $\mathrm{L} \subset \mathbb{R}^n$ we will write

$$M = L \otimes \mathbb{R}^p$$

For the subspace of $\mathbb{R}^{n \times p}$ given by

$$M = \{\mu = (\mu_1; \cdots \mu_p) | \mu_i \in L, i = 1, \dots, p\}$$

Likelihood Estimation
In the Multivariate Normal Model

- If an $n \times n$ matrix A is a projection onto a subspace $\mathrm{L} \subset \mathbb{R}^n$, then

$$\boldsymbol{A \otimes I_p} \textbf{ is a projection to } \boldsymbol{M = L \otimes \mathbb{R}^p} \textbf{ wrt. } \boldsymbol{I_n \otimes \Sigma^{-1}}\textbf{:}$$

- $A \otimes I_p$ is symmetric, because $\left(A \otimes I_p\right)^T = A^T \otimes I_p^T = A \otimes I_p$

Since A is a projection.

- $A \otimes I_p$ is idempotent: $\left(A \otimes I_p\right) A \otimes I_p = AA \otimes I_p I_p = A \otimes I_p$

- And last, for $\mu \in M$ we have

$$A \otimes I_p \mu = A \mu I_p = A\mu = \mu$$

since all the columns of μ is in L and A is a projection onto L.

Likelihood Estimation
In the Multivariate Normal Model

- Consequence: In the model

$$Y = X\theta + \varepsilon, \qquad \varepsilon \sim N_{n\times p}(0, \mathrm{I_n} \otimes \Sigma)$$

The MLE $\hat{\theta}$ for θ is found as

$$\hat{\theta} = \left(X^T X\right)^{-1} X^T Y$$

And the fitted values $\hat{Y}$ are found as

$$\hat{Y} = X\hat{\theta} = X\left(X^T X\right)^{-1} X^T Y$$

JUST AS WE ARE USED TO!

Likelihood Estimation
In the Multivariate Normal Model

Taking $A = X\left(X^T X\right)^{-1} X^T$, the MLE $\hat{\Sigma}$ for Σ is found as

$$\hat{\Sigma} = \frac{1}{n} SSD = \frac{1}{n} Y^T (I_n - A) Y:$$

The likelihood function is

$$L(\hat{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}} det(\Sigma)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \|Y - \hat{\mu}\|^2_{I_n \otimes \Sigma^{-1}}\right)$$

$$= \frac{1}{(2\pi)^{\frac{np}{2}}} det(\Sigma)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} tr\big((I_n - A) Y^T (I_n - A) Y \Sigma^{-1}\big)\right)$$

$$= \frac{1}{(2\pi)^{\frac{np}{2}}} det(\Sigma)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} tr(SSD\Sigma^{-1})\right)$$

$$= \frac{1}{(2\pi)^{\frac{np}{2}}} det(\Sigma)^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{p} \lambda_i\right)$$

Where $\lambda_1, \dots, \lambda_p$ are the eigenvalues of $SSD\Sigma^{-1}$.

By comparing with $L\left(\hat{\mu}, \frac{1}{n} SSD\right)$ (we shall not go through that) one can work out that the Likelihood is maximized when $\lambda_1 = \cdots = \lambda_p = 1$, ie. $\Sigma = \frac{1}{n} SSD$.

- As in the univariate case, we shall use the unbiased estimator $S = \frac{1}{n-k} SSD$, where k is the rank of A.

Likelihood Estimation
In the Multivariate Normal Model

The distribution of the estimators:

Since $\hat{\theta} = (X^T X)^{-1} X^T Y = \left((X^T X)^{-1} X^T \otimes I_p\right) Y$, it is immediate that $\hat{\theta}$ is normal with

$$E(\hat{\theta}) = (X^T X)^{-1} X^T E(Y) = (X^T X)^{-1} X^T X \theta = \theta,$$

$$\begin{aligned} V(\hat{\theta}) &= \left((X^T X)^{-1} X^T \otimes I_p\right) V(Y) \left((X^T X)^{-1} X^T \otimes I_p\right)^T \\ &= \left((X^T X)^{-1} X^T \otimes I_p\right) (I_n \otimes \Sigma) \left((X^T X)^{-1} X^T \otimes I_p\right)^T \\ &= (X^T X)^{-1} \otimes \Sigma \end{aligned}$$

$$\hat{\Sigma} = \frac{1}{n-k} SSD \sim W_p(n-k, \Sigma)$$

Hvor k is the rank of the projection matrix A.

Likelihood Estimation
In the Multivariate Normal Model

Theorem 4.14

We consider the above mentioned situation. If the observations Y_i are normally distributed the maximum likelihood estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{x}^T\mathbf{x})^{-1}\mathbf{x}^T\mathbf{Y}.$$

Theorem 4.19

We consider the situation from theorems 4.14 and 4.18 and we introduce the usual notations

$$\tilde{\boldsymbol{\theta}} = \mathrm{vc}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\theta}_{|1} \\ \vdots \\ \boldsymbol{\theta}_{|p} \end{bmatrix}$$

$$\hat{\tilde{\boldsymbol{\theta}}} = \mathrm{vc}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \hat{\boldsymbol{\theta}}_{|1} \\ \vdots \\ \hat{\boldsymbol{\theta}}_{|p} \end{bmatrix}.$$

Then we have that $\hat{\tilde{\boldsymbol{\theta}}}$ is normally distributed

$$\hat{\tilde{\boldsymbol{\theta}}} = \mathrm{vc}(\hat{\boldsymbol{\theta}}) \sim \mathrm{N}_{pk}(\tilde{\theta}, \Sigma \otimes (\mathbf{x}^T\mathbf{x})^{-1}),$$

and $n\hat{\Sigma}^*$ is Wishart distributed

$$n\hat{\Sigma}^* \sim \mathrm{W}(n-k, \Sigma).$$

Finally $\hat{\Sigma}^*$ and $\hat{\tilde{\boldsymbol{\theta}}}$ and therefore also $\hat{\Sigma}^*$ and $\hat{\boldsymbol{\theta}}$ are stochastically independent.

- Note that the book (here) considers the observations as stacked in a columns and not as rows in a matrix. Hence the change in tensors…

Likelihood Estimation
In the Multivariate Normal Model

- Procedure for estimation:

1. Estimate the mean for each coloumn separately, using the univariate techniques;
2. Subtract the estimate of the mean from the observations, to form the matrix R of residuals;
3. Estimate the unknown covariance matrix Σ with the unbiased estimator

$$\hat{\Sigma} = \frac{1}{f} SSD = \frac{1}{f} R^T R = \frac{1}{f} Y(I - A)Y, \qquad R = (I - A)Y$$

where the degrees of freedom follows from the number of observations n and the rank of the design matrix X, JUST AS IF IT WAS A UNIVARIATE MODEL.

Testing in the Multivariate Normal Model

- Suppose that $Y \sim N_{n\times p}(\mu, I_n \otimes \Sigma)$, where $\mu \in L_0 \otimes \mathbb{R}^p, \dim(\mathrm{L}_0) = \mathrm{d}_0$.

$$\hat{\mu} = P_0 Y, n\hat{\Sigma} = SSD_0 = Y^T(I_n - P_0)Y$$

$$L(\hat{\mu}, \hat{\Sigma}) = c \cdot (\det(SSD_0))^{-n/2} \exp\left(-\frac{np}{2}\right)$$

Suppose that we want to test the hypothesis

$$H_1: \mu \in L_1 \otimes \mathbb{R}^p, \dim(L_1) = d_1, L_1 \subseteq L_0$$

Testing in the Multivariate Normal Model

$$H_1 : \mu \in L_1 \otimes \mathbb{R}^p, \dim(L_1) = d_1, L_1 \subseteq L_0$$

$$\hat{\hat{\mu}} = P_1 Y, n\hat{\hat{\Sigma}} = SSD_{01} = Y^T(I_n - P_1)Y$$

$$L\left(\hat{\hat{\mu}}, \hat{\hat{\Sigma}}\right) = c \cdot (\det(SSD_{01}))^{-n/2} \exp\left(-\frac{np}{2}\right)$$

$$Q = \frac{L\left(\hat{\hat{\mu}}, \hat{\hat{\Sigma}}\right)}{L(\hat{\mu}, \hat{\Sigma})} = \frac{\left(det(SSD_0)\right)^{n/2}}{\left(det(SSD_{01})\right)^{n/2}}$$

Testing in the Multivariate Normal Model

$$H_1 : \mu \in L_1 \otimes \mathbb{R}^p, \dim(L_1) = d_1$$

We note that

$$I_n - P_1 = (I - P_0) + (P_0 - P_1)$$

Where

$$(I - P_0)(P_0 - P_1) = 0$$

because $L_1 \otimes \mathbb{R}^p \subseteq L_0 \otimes \mathbb{R}^p$, so that with $SSD_1 = Y^T(P_0 - P_1)Y$ it holds that SSD_0 and SSD_1 are independent, and

$$SSD_{01} = SSD_0 + SSD_1$$

and thus

$$R_1 = Q^{2/n} = \frac{det(SSD_0)}{det(SSD_0 + SSD_1)}$$

Testing in the Multivariate Normal Model

$$R_1 = Q^{2/n} = \frac{det(SSD_0)}{det(SSD_0 + SSD_1)}$$

Here, under H_1, SSD_0 and SSD_1 are independent and

$$SSD_0 \sim W_p(n - d_0, \Sigma), \qquad SSD_1 \sim W_p(d_0 - d_1, \Sigma)$$

Thus per definition, R_1 is Wilks distributed, $R_1 \sim \wedge (p, n - d_0, d_0 - d_1)$.

Given that the the test is accepted, one could proceed to a subspace $L_2 \otimes \mathbb{R}^p \subseteq L_1 \otimes \mathbb{R}^p$ and so on...

Testing in the Multivariate Normal Model

Hypothesis	SSD	df	Test
$L_0 \otimes \mathbb{R}^p$	SSD_0	$n - d_0$	
$H_1\colon\ L_1 \otimes \mathbb{R}^p$	SSD_1	$d_0 - d_1$	$R_1 = \dfrac{det(SSD_0)}{det(SSD_0 + SSD_1)}$
$H_2\colon\ L_2 \otimes \mathbb{R}^p$	SSD_2	$d_1 - d_2$	$R_2 = \dfrac{det(SSD_{01})}{det(SSD_{01} + SSD_2)}$
Sum	SSD_{02}	$n - d_2$	

Testing in the Multivariate Normal Model

- Take $A = Q_{01}^T X$, where the columns in Q_{01} form a orthonormal basis for $L_0 \cap L_1^{\perp}$, $B = I_p$ $C = 0$:

- $E = SSD_0$
- $\Delta = Q_{01}^T X\hat{\theta} = Q_{01}^T P_0 Y$
  $= Q_{01}^T Y$

- $H = Y^T Q_{01} \left(Q_{01}^T X (X^T X)^{-1} X^T Q_{01} \right)^{-1} Q_{01}^T Y$
  $= Y^T Q_{01} I_{d_0 - d_1} Q_{01}^T Y$
  $= Y^T (P_0 - P_1) Y$
  $= SSD_1$

Theorem 4.21

We consider the above mentioned situation including the assumption of the normality of the observations. Furthermore we consider the hypothesis

$$H_0 : \mathbf{A}\,\boldsymbol{\theta}\,\mathbf{B}^T = \mathbf{C} \quad \text{against} \quad H_1 : \mathbf{A}\,\boldsymbol{\theta}\,\mathbf{B}^T \neq \mathbf{C},$$

where $\mathbf{A}(r \times k)$, $\mathbf{B}(s \times p)$ and $\mathbf{C}(r \times s)$ are given (known) matrices. We introduce

$$\begin{aligned} \boldsymbol{\Delta} &= \mathbf{A}\,\hat{\boldsymbol{\theta}}\,\mathbf{B}^T - \mathbf{C} \\ \mathbf{R} &= n\hat{\boldsymbol{\Sigma}}^* = (\mathbf{Y} - \mathbf{x}\hat{\boldsymbol{\theta}})^T(\mathbf{Y} - \mathbf{x}\hat{\boldsymbol{\theta}}) = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\theta}}^T(\mathbf{x}^T\mathbf{x})\hat{\boldsymbol{\theta}} \end{aligned}$$

and

$$\begin{aligned} \mathbf{E} &= \mathbf{B}\,\mathbf{R}\,\mathbf{B}^T \\ \mathbf{H} &= \boldsymbol{\Delta}^T[\mathbf{A}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{A}^T]^{-1}\boldsymbol{\Delta}. \end{aligned}$$

The likelihood ratio test for testing H_0 against H_1 is equivalent to the test given by the critical region

$$\{\mathbf{y} | \frac{\det(\mathbf{e})}{\det(\mathbf{e}+\mathbf{h})} \leq \mathrm{U}(s, r, n-k)_\alpha\},$$

where $\mathrm{U}(s, r, n-k)_\alpha$ is the α quantile in the null-hypothesis distribution of the test statistic (see below).

F approximations to Wilks Distribution

Theorem 4.22

Let U be U(s,r,n-k)-distributed and let

$$
\begin{aligned}
t &= \begin{cases} 1 & s^2+r^2=5 \\ \sqrt{\frac{s^2r^2-4}{s^2+r^2-5}} & s^2+r^2\neq 5 \end{cases} \\
v &= \frac{1}{2}(2(n-k)+r-s-1).
\end{aligned}
$$

Then

$$
\mathrm{F} = \frac{1-U^{\frac{1}{t}}}{U^{\frac{1}{t}}} \cdot \frac{vt+1-\frac{1}{2}sr}{sr}
$$

is approximately distributed as

$$
\mathrm{F}(sr, vt+1-\frac{1}{2}sr).
$$

If either s or r are equal to 1 or 2, then the approximation is exact.

- Compare with the table in Lecture F.

Example: Heat Climate Data

$$H_0: Y = \alpha_{sex} + \beta \cdot height + \varepsilon$$

$$H_1: Y = \alpha_{sex} + \varepsilon$$

```
> analysis<-manova(cbind(1/log(evap),1/log(temp))~sex+height,data=heatclimate)
> summary(analysis,test="Wilks")
          Df   Wilks approx F num Df den Df  Pr(>F)
sex        1 0.83721   2.7222      2     28 0.08312 .
height     1 0.92171   1.1892      2     28 0.31937
Residuals 29
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$p = 0.32$

Example: Heat Climate Data

$$H_1: Y = \alpha_{sex} + \varepsilon$$

$$H_2: Y = \alpha + \varepsilon$$

```
> analysis<-manova(cbind(1/log(evap),1/log(temp))~sex,data=heatclimate)
> summary(analysis,test="Wilks")
          Df   Wilks approx F num Df den Df  Pr(>F)
sex        1 0.84651   2.6292      2     29 0.08926 .
Residuals 30
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$p = 0.09$

Exercises

- Problem 8 ex. 10/12 2001
- Problem 2 ex. 5/1 2001
- Problem 5 ex 7/12 2010 Find the test statistics by using R
- Problem 6 ex 8/12 2009
- Problem 4 ex 9/12 2008
- Problem 3 ex 9/12 2011.