Lecture 1: Linear Models and Applications

Claudia Czado

TU München

Overview

- Introduction to linear models
- Exploratory data analysis (EDA)
- Estimation and testing in linear models
- Regression diagnostics

Introduction to linear models

Goal of regression models is to determine how a response variable depends on covariates. A special class of regression models are linear models. The general setup is given by

Data	$(Y_i, x_{i1}, ..., x_{ik}), i = 1, ..., n$
Responses	$\mathbf{Y} = (Y_1, ..., Y_n)^T$
Covariates	$\mathbf{x_i} = (x_{i1}, ..., x_{ik})^T$ (known)

Example: Life Expectancies from 40 countries

Source: The Economist's Book of World Statistics, 1990, Time Books, The World Almanac of Facts, 1992, World Almanac Books

LIFE.EXP	Life Expectancy at Birth
TEMP	Temperature in degrees Fahrenheit
URBAN	Percent of population living in urban areas
HOSP.POP	No. of hospitals per population
COUNTRY	The name of the country

Which is the response and which are the covariates?

Linear models (LM) under normality assumption

$$Y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \; iid, \quad i = 1, .., n,$$

where the unknown $\beta_0, ..., \beta_k$ regression parameters and the the unknown error variance σ^2 needs to be estimated. Note $E(Y_i)$ is a linear function in $\beta_0, ..., \beta_k$.

$$\begin{aligned} E(Y_i) &= \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} \\ Var(Y_i) &= Var(\epsilon_i) = \sigma^2 \text{ variance homogeneity} \end{aligned}$$

LM's in matrix notation

$$\mathbf{Y} = X\boldsymbol{\beta} + \epsilon, \quad X \in \mathbb{R}^{n \times p}, \;\; p = k+1, \;\; \boldsymbol{\beta} = (\beta_0, ..., \beta_k)^T$$
$$E(\mathbf{Y}) = X\boldsymbol{\beta} \qquad Var(\mathbf{Y}) = \sigma^2 I_n$$

Under the normal assumption we have

$$\mathbf{Y} \;\sim\; N_n(X\boldsymbol{\beta}, \sigma^2 I_n),$$

where $N_n(\boldsymbol{\mu}, \Sigma)$ is the n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. The matrix X is also called the design matrix.

Exploratory data analysis (EDA)

Consider the ranges of the responses and covariates. When covariates are discrete, group covariate levels if they are sparse.

Plot the covariates against the responses. These scatter plots should look linear. Otherwise consider transformations of the covariates.

To check if the constant variance assumption is reasonable the scatter plot of covariates against the responses should be contained in a band.

Example: Life expectancies: Data summaries

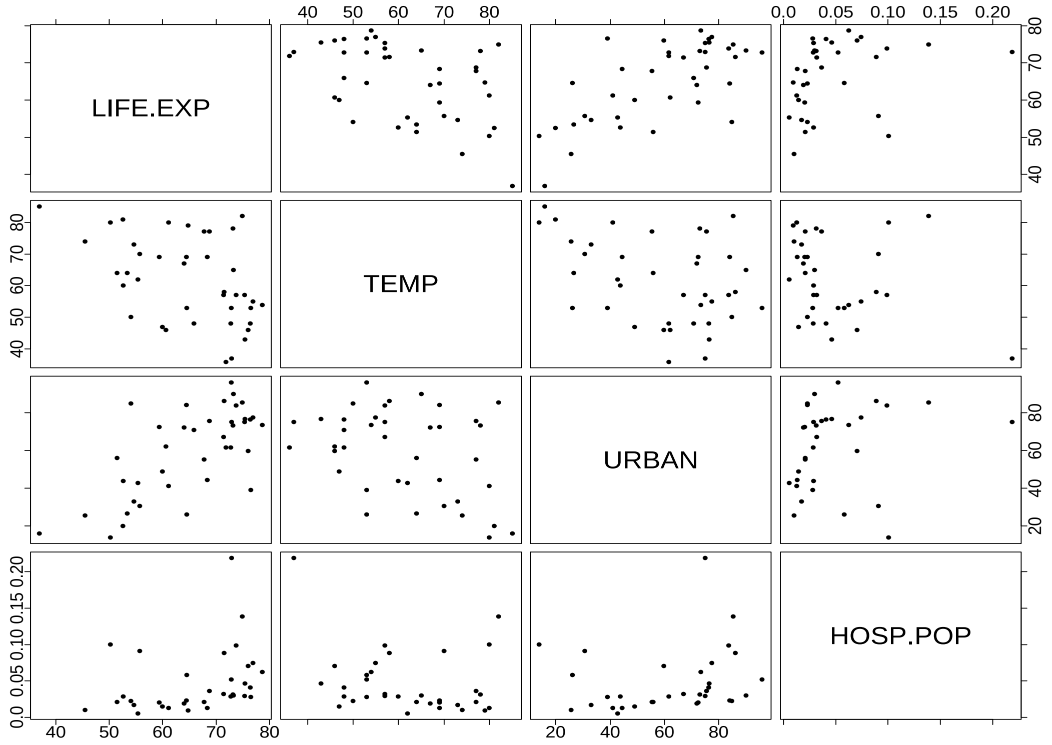
```
> summary(health.data)

   LIFE.EXP          TEMP           URBAN
 Min.   :37    Min.   :36    Min.   :14
 1st Qu.:56    1st Qu.:52    1st Qu.:42
 Median :67    Median :61    Median :62
 Mean   :65    Mean   :62    Mean   :59
 3rd Qu.:73    3rd Qu.:73    3rd Qu.:76
 Max.   :79    Max.   :85    Max.   :96
                             NA's   : 1

     HOSP.POP
   Min.   :0.0057
  1st Qu.:0.0204
  Median :0.0295
  Mean   :0.0469
  3rd Qu.:0.0614
  Max.   :0.2190
  NA's   :6.0000
```

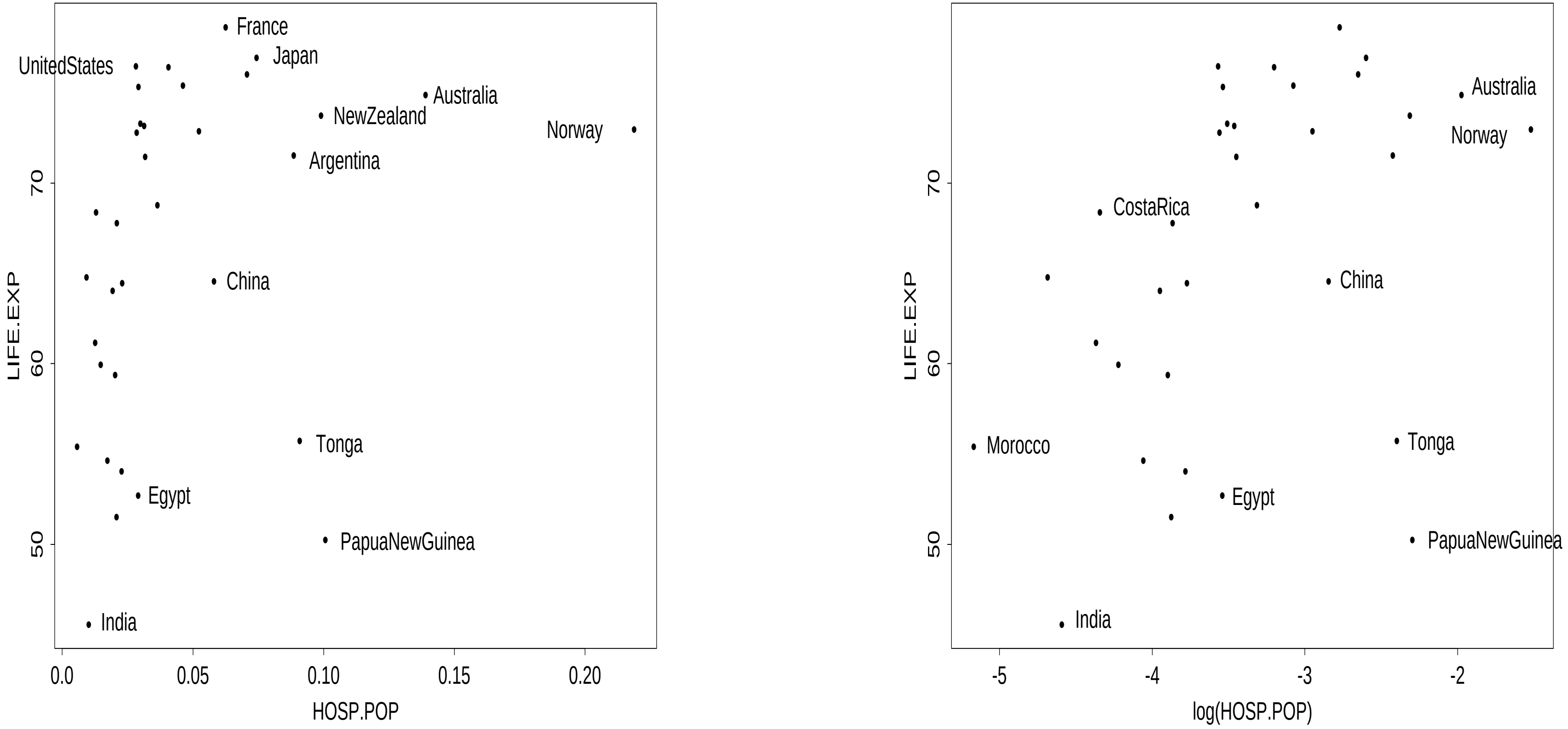
NA=not available (missing data)

Example: Life expectancies: EDA



LIFE.EXP decreases if TEMP increases
LIFE.EXP increases if URBAN increases
LIFE.EXP increases if HOSP.POP increases
Check for nonlinearities and nonconstant variances

Example: Life expectancies: Transformations



Logarithm makes relationship more linear.

Parameter estimation of β

Under the normality assumption it is enough to minimize $Q(\boldsymbol{\beta}) :=\parallel \mathbf{Y} - X\boldsymbol{\beta} \parallel^2$ for $\boldsymbol{\beta} \in \mathbb{R}^p$ to calculate the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$. An estimator which minimizes $Q(\boldsymbol{\beta})$ is also called a least square estimator (LSE). If the matrix $X \in \mathbb{R}^{n \times p}$ has full rank p the minimum of $Q(\boldsymbol{\beta})$ is taken at

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$$

$SS_{Res} := \sum_{i=1}^{n} (Y_i - \mathbf{x_i}^T \hat{\boldsymbol{\beta}})^2 = Q(\hat{\boldsymbol{\beta}})$	Residual Sum of Squares
$\hat{Y}_i := \mathbf{x_i}^T \hat{\boldsymbol{\beta}}$	fitted value for $\mu_i := E(Y_i)$
$e_i := Y_i - \hat{Y}_i$	raw residual

It follows that

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \text{ and } Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (X^T X)^{-1},$$

therefore one has under normality

$$\hat{\boldsymbol{\beta}} \ \sim \ N_p(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1}).$$

Parameter estimation of σ^2

The MLE of σ^2 is given by

$$\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbf{x_i}^T\hat{\boldsymbol{\beta}})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

One can show that the estimator is biased, in particular $E(\hat{\sigma}^2) = \frac{n-p}{n}\sigma^2$. An unbiased estimator for σ^2 is therefore given by

$$s^2 := \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \frac{n}{n-p}\hat{\sigma}^2$$

Under normality it follows that

$$\frac{(n-p)s^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sigma^2} \sim \chi^2_{n-p} \text{ is independent of } \hat{\boldsymbol{\beta}}.$$

Goodness of fit in linear models

Consider

$$SS_{Total} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2, \quad SS_{Reg} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2, \quad SS_{Res} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2,$$

where $\bar{Y} = \frac{1}{n}\sum\limits_{i=1}^{n} Y_i$. It follows that

$$SS_{Total} = SS_{Reg} + SS_{Res},$$

therefore

$$R^2 := \frac{SS_{Reg}}{SS_{Total}} = 1 - \frac{SS_{Res}}{SS_{Total}}$$

explains the proportion of variability explained by the regression. R^2 is called the multiple coefficient of determination.

Statistical hypothesis tests in LM's: F-test

The restriction

$$H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$$

for $\mathbf{C} \in \mathbb{R}^{r,p}, rank(\mathbf{C}) = r, \mathbf{d} \in \mathbb{R}^r$ is called general linear hypothesis with alternative $H_1: \; not \;\; H_0$. Consider the restricted least square problem

$$\begin{aligned} SS_{H_0} &= \min_{\boldsymbol{\beta}}\{\| Y - \mathbf{X}\boldsymbol{\beta} \|_2^2 \mid \underbrace{\mathbf{C}\boldsymbol{\beta} = \mathbf{d}}_{\text{under } H_0}\} \\ &= SS_{Res} + (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})^T[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d}) \end{aligned}$$

Under normality and that $H_0: \; \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ is valid we have

$$F = \frac{(SS_{H_0} - SS_{Res})/r}{SS_{Res}/(n-p)} \sim F_{r,n-p},$$

therefore the F-test is given by reject H_0 at level α if $F > F_{1-\alpha,r,n-p}$. Here $F_{n,m}$ denotes the F distribution with n numerator and m denominator degree of freedom and $F_{1-\alpha,n,m}$ is the corresponding $1-\alpha$ quantile.

Statistical hypothesis tests in LM's: t-test

Consider for each regression parameter β_j the hypothesis $H_{0j}: \beta_j = 0$, against $H_{1j}:$ not H_{0j} and use the corresponding F statistics

$$F_j := \frac{(SS_{H_{0j}} - SS_{Res})/1}{SS_{Res}/(n-p)} \sim F_{1,n-p} \text{ under } H_{0j}.$$

It follows that $F_j = T_j^2$, where

$$T_j := \frac{\hat{\beta}_j}{\hat{s}e(\hat{\beta}_j)} \sim t_{n-p} \text{ under } H_{0j},$$

where $\hat{s}e(\hat{\beta}_j)$ is the estimated standard error of $\hat{\beta}_j$ i.e.

$$\hat{s}e(\hat{\beta}_j) = s\sqrt{(X^T X)^{-1}_{jj}}.$$

Weighted Least Squares

$$\mathbf{Y} = X\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim N_n(\mathbf{0}, \sigma^2 W), \quad X \in \mathbb{R}^{n \times p}, \ p = k+1, \ \boldsymbol{\beta} = (\beta_0, ..., \beta_k)^T$$

An MLE of $\boldsymbol{\beta}$ (weighted LSE) is given by

$$\hat{\boldsymbol{\beta}} = (X^T W^{-1} X)^{-1} X^T W^{-1} \mathbf{Y}.$$

It follows that

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \text{ and } Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (X^T W^{-1} X)^{-1}.$$

The MLE of σ^2 is given by

$$\hat{\sigma}^2 := \frac{1}{n}(\mathbf{Y} - X\hat{\boldsymbol{\beta}})^T W^{-1} (\mathbf{Y} - X\hat{\boldsymbol{\beta}}) = \frac{1}{n}\mathbf{e}^T W^{-1} \mathbf{e}.$$

Example: Life expectancies: First models

Using the Original Scale for HOSP.POP

```
> attach(health.data)
> f1_LIFE.EXP ~ TEMP + URBAN + HOSP.POP
> r1_lm(f1,na.action = na.omit)
> summary(r1)
Call: lm(formula = f1, na.action = na.omit)
Residuals:
   Min     1Q Median    3Q   Max
 -18.7 -4.72   1.63 4.89 14.7
Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  60.666   9.088      6.676   0.000
       TEMP  -0.167   0.113     -1.480   0.150
      URBAN   0.232   0.063      3.710   0.001
   HOSP.POP  33.756  30.373      1.111   0.276
Residual standard error: 7.3 on 29 degrees of freedom
Multiple R-Squared: 0.46
F-statistic: 8.3 on 3 and 29 degrees of freedom, the p-value is 0.00039
Correlation of Coefficients:
         (Intercept)  TEMP URBAN
    TEMP -0.90
   URBAN -0.59        0.24
HOSP.POP -0.24        0.18 -0.14
```

TEMP and HOSP.POP are nonsignificant at the 10% level.

Example: Life expectancies: First models

Using the Logarithmic Scale for HOSP.POP

```
> f2_LIFE.EXP ~ TEMP + URBAN + log(HOSP.POP)
> r2_lm(f2,na.action = na.omit)
> summary(r2)
Call: lm(formula = f2, na.action = na.omit)
Residuals:
   Min    1Q Median   3Q  Max
 -17.6 -4.2   1.71 5.07 14.4
Coefficients:
                Value Std. Error t value Pr(>|t|)
  (Intercept)  72.964   9.917      7.357   0.000
         TEMP  -0.151   0.109     -1.387   0.176
        URBAN   0.213   0.061      3.476   0.002
log(HOSP.POP)   3.133   1.638      1.912   0.066
Residual standard error: 7 on 29 degrees of freedom
Multiple R-Squared: 0.5
F-statistic: 9.7 on 3 and 29 degrees of freedom, the p-value is 0.00013
Correlation of Coefficients:
              (Intercept)  TEMP URBAN
         TEMP -0.65
        URBAN -0.67        0.22
log(HOSP.POP)  0.52        0.19 -0.24
```

TEMP is still nonsignificant, while log(HOSP.POP) is now significant at the 10% level. 50% of the total variability is explained by the regression.

ANOVA Tables in LM's

ANOVA= ANalysis Of VAriance

Model	Formula	SS
null	$Y_i = \beta_0 + \epsilon_i$	$SSE_0 = \sum_{i=1}^n (Y_i - \hat{\beta}_0)^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2$
full	$\mathbf{Y} = X\boldsymbol{\beta} + \epsilon$	$SSE(X) = \parallel \mathbf{Y} - X\hat{\boldsymbol{\beta}} \parallel^2$
reduced	$\mathbf{Y} = X_1\boldsymbol{\beta}_1 + \epsilon$	$SSE(X_1) = \parallel \mathbf{Y} - X_1\hat{\boldsymbol{\beta}}_1 \parallel^2$

ANOVA table:

Source	df	Sum of Sq	MS	F
regression	$p-1$	$SS_{Reg} = SSE_0 - SSE(X)$	$MS_{Reg} = \frac{SS_{Reg}}{p-1}$	$\frac{MS_{Reg}}{s^2}$
residual	$n-p$	$SSE(X)$	$s^2 = \frac{SSE(X)}{n-p}$	
total	$n-1$	SSE_0		

Hierarchical ANOVA Tables in LM's

$$\mathbf{Y} = X\boldsymbol{\beta} + \epsilon = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \epsilon$$
$$\mathbf{X} \in \mathbb{R}^{n,p}, \mathbf{X_1} \in \mathbb{R}^{n,p_1}, \mathbf{X_2} \in \mathbb{R}^{n,p_2}, p_1 + p_2 = p$$

ANOVA table:

Source	df	Sum of Sq	MS	F
X_1	$p_1 - 1$	$SSE_0 - SSE(X_1)$	$MS(X_1) = \frac{SSE_0 - SSE(X_1)}{p_1 - 1}$	$F(X_1) = \frac{MS(X_1)}{SSE(X_1)/(n-p_1)}$
X_2 given X_1	p_2	$SSE(X_1) - SSE(X)$	$MS(X_2\|X_1) = \frac{SSE(X_1) - SSE(X)}{p_2}$	$F(X_2\|X_1) = \frac{MS(X_2\|X_1)}{s^2}$
residual	$n - p$	$SSE(X)$	$s^2 = \frac{SSE(X)}{n-p}$	
total	$n - 1$	SSE_0		

$F(X_1)$ tests $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ in $\mathbf{Y} = X_1\boldsymbol{\beta}_1 + \epsilon$ (overall F-test in reduced model).
$F(X_2|X_1)$ tests $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ in $\mathbf{Y} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \epsilon$ (partial F-test).

ANOVA Tables in Splus

```
> r <- lm(y ~x1+x2+x3)
> anova(r)
```

Source	df	Sum of Sq	MS	F
x1	1	$SS_1 = SSE_0 - SSE(x_1)$	$MS_1 = \frac{SS_1}{1}$	$\frac{MS_1}{s^2}$
x2	1	$SS_2 = SSE(x_1) - SSE(x_1, x_2)$	$MS_2 = \frac{SS_2}{1}$	$\frac{MS_2}{s^2}$
x3	1	$SS_3 = SSE(x_1, x_2) - SSE(x_1, x_2, x_3)$	$MS_3 = \frac{SS_3}{1}$	$\frac{MS_3}{s^2}$
residual	$n - p$	$SSE(x_1, x_2, x_3)$	$s^2 = \frac{SSE(X)}{n-4}$	

These F-values cannot be interpreted as partial F-values, since the denominator always assumes the full model. You need to replace it by $s_1^2 = \frac{SSE(x_1)}{n-2}$, $s_2^2 = \frac{SSE(x_1,x_2)}{n-3}$ and $s_3^2 = \frac{SSE(x_1,x_2,x_3)}{n-4} = s^2$, respectively.

Example: Life Expectancies: ANOVA Tables

ANOVA Table (Original Scale for HOSP.POP)

```
> anova(r1)
Analysis of Variance Table

Response: LIFE.EXP

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value Pr(F)
TEMP       1       455     455     8.5 0.007
URBAN      1       815     815    15.2 0.001
HOSP.POP   1        66      66     1.2 0.276
Residuals 29      1555      54
```

ANOVA Table (Log Scale for HOSP.POP)

```
> anova(r2)
Analysis of Variance Table

Response: LIFE.EXP

Terms added sequentially (first to last)
              Df Sum of Sq Mean Sq F Value  Pr(F)
TEMP           1       455     455     9.2 0.0051
URBAN          1       815     815    16.4 0.0003
log(HOSP.POP)  1       182     182     3.7 0.0658
Residuals     29      1440      50
```

Regression diagnostics for LM's

Goal is to determine

- outliers with regard to the response (y outliers)
- outliers with regard to the design space covered by the covariates (x outliers or high leverage points)
- observations which change the results greatly when removed (influential observations).

A general tool for this is to consider the hat matrix and residuals.

Residuals in LM's: hat matrix

$$\begin{aligned} H &:= X(X^TX)^{-1}X^T \text{ hat matrix} \\ \hat{\boldsymbol{Y}} &:= X\hat{\boldsymbol{\beta}} = X(X^TX)^{-1}X^T\boldsymbol{Y} = H\boldsymbol{Y} \text{ fitted values} \\ \boldsymbol{r} &:= \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (I-H)\boldsymbol{Y} \text{ residual vector} \end{aligned}$$

Properties:

- H is a projection matrix, i.e. $H^2 = H$ and symmetric.
- $E(\boldsymbol{r}) = (I-H)E(\boldsymbol{Y}) = (I-H)X\boldsymbol{\beta} = X\boldsymbol{\beta} - HX\boldsymbol{\beta} = X\boldsymbol{\beta} - X\boldsymbol{\beta} = 0$.
- $Var(\boldsymbol{r}) = \sigma^2(I-H)(I-H)^T = \sigma^2(I-H)$
- $Var(r_i) = \sigma^2(1-h_{ii})$, where h_{ii} is i-th diagonal element of H.

Standardized Residuals in LM's

$$s_i := \frac{r_i}{\sqrt{1-h_{ii}}s}, \text{ where } s^2 := \frac{\| \boldsymbol{Y} - X\hat{\boldsymbol{\beta}} \|^2}{n-p}$$

are called standardized residuals or internally studentized residuals

h_{ii} is called the leverage of observation i.

Problem:

The estimate s^2 is biased when outliers are present in the data, therefore one wants to estimate σ^2 without the ith observation.

Jackknifed Residuals in LM's

$$\begin{aligned}
\boldsymbol{Y}_{-i} &= X_{-i}\boldsymbol{\beta}+\boldsymbol{\epsilon}_{-i} \text{ model without ith obs}\\
X_{-i} &= \text{ design matrix without ith obs}\\
\hat{\boldsymbol{\beta}}_{-i} &:= \text{ corresponding LSE of } \boldsymbol{\beta}\\
\hat{Y}_{i,-i} &:= \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_{-i}\\
&\quad \text{ ith fitted value without using ith obs}\\
r_{i,-i} &:= Y_i - \hat{Y}_{i,-i} \text{ predictive residual}\\
s^2_{-i} &:= \frac{\sum_{j=1,j\neq j}^{n} (Y_j - \boldsymbol{x}_j^T\hat{\boldsymbol{\beta}}_{-i})^2}{n-p-1} \text{ estimate of}\\
&\quad \sigma^2 \text{ in model without ith obs}\\
t_i &:= \frac{r_i}{\sqrt{1-h_{ii}}\,s_{-i}} \text{ externally studentized or jackknifed residuals}
\end{aligned}$$

Jackknifed Residuals in LM's

For a fast computation one can use linear algebra to show that

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{-i} &= \hat{\boldsymbol{\beta}} - \frac{(X^T X)^{-1}\boldsymbol{x}_i r_i}{1 - h_{ii}} \\
r_{i,-i} &= \frac{r_i}{1 - h_{ii}} \\
s^2_{-i} &= \frac{(n-p)s^2 - \frac{r_i^2}{1-h_{ii}}}{n-p-1}
\end{aligned}
$$

Residual plots are plots where the observation number i or the jth covariate x_{ij} is plotted against r_i, s_i or t_i.

There is no problem with the model if these plots look like a band. Deviations from this structure might indicate nonlinear regression effects or a violation of the variance homogeneity.

Residuals in Splus

```
r_lm(Y~x1+x2+...+xk, x=T)
```

Raw Residuals:

$r_i = Y_i - \hat{Y}_i$ `e_resid(summary(r))`

Residual Standard Error:

$s = \sqrt{\frac{||Y - X\hat{\beta}||^2}{n-p}}$ `sigma_summary(r)$sigma`

External Residual Standard Error:

s_{-i} `sigmai_lm.influence(r)$sigma`

Hat Diagonals:

h_{ii} `hi_hat(r$x)` or `hi_lm.influence(r)$hat`

Internally Studentized Residuals:

$s_i = \frac{r_i}{\sqrt{1-h_{ii}}s}$ `si_e/(sigma*((1-hi)^.5))`

Externally Studentized Residuals:

$t_i = \frac{r_i}{\sqrt{1-h_{ii}}s_{-i}}$ `ti_e/(sigmai*((1-hi)^.5))`

High leverage in LM's

Since

$$h_{ii} = \boldsymbol{x}_i (X^T X)^{-1} \boldsymbol{x}_i^T$$

one can interpret h_{ii} as standardized measure between $\boldsymbol{x}_i$ and $\overline{x}$.

Further we have

$$\sum_{i=1}^{n} h_{ii} = p,$$

therefore we call points with

$$h_{ii} > \frac{2p}{n}$$

as high leverage points or x-outliers.

Outlier detection in LM's

To detect $y - outliers$, we can use t_i, since it can be written as

$$t_i = \frac{Y_i - \hat{Y}_{-i}}{\sqrt{\widehat{Var}(Y_i - \hat{Y}_{-i})}} = \frac{\hat{\delta}_i}{\sqrt{\widehat{Var}(\delta_i)}} \text{ for } \delta_i := Y_i - \hat{Y}_{-i}.$$

In the mean shift outlier model for obs l given by

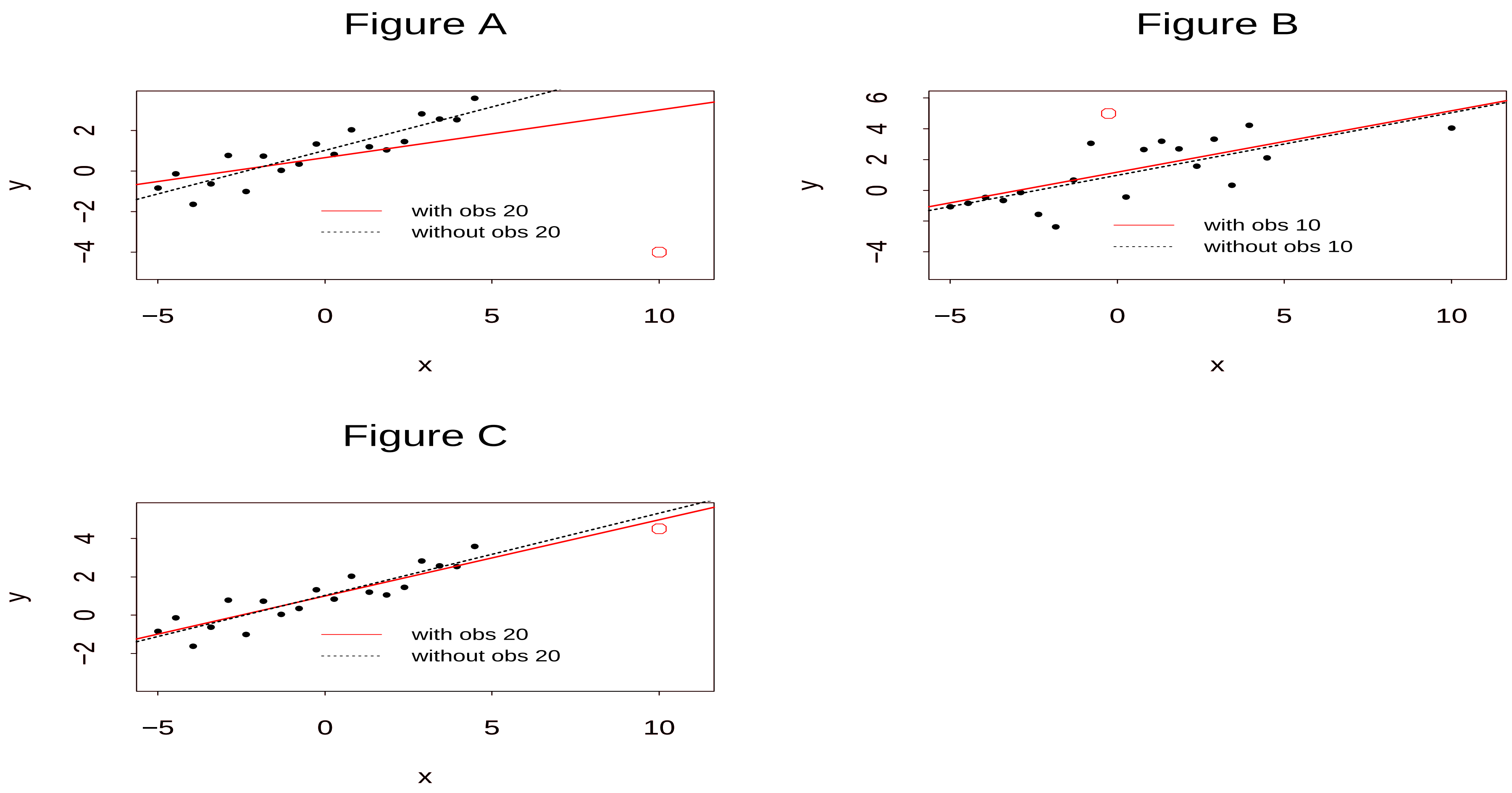
$$Y_i = \boldsymbol{x}_i^t \boldsymbol{\beta} + \gamma D_{li} + \epsilon_i \text{ with } D_{li} = \begin{cases} 1 & l = i \\ 0 & l \neq i \end{cases} .$$

Therefore reject $H : \gamma = 0$ versus $K : \gamma \neq 0$ at level α iff $|t_l| > t_{n-p-1,1-\frac{\alpha}{2}}$, where $t_{m,\alpha}$ is the α quantil of a t_m distribution.

Problem:

If one uses this outlier test for every obs. l, one has the problem of multiple testing, since one needs to substitute α by $\frac{\alpha}{n}$.

Leverage and Influence



Leverage and Influence

- Figure A shows that obs 20 is a high leverage point , which is also influential.
- Figure B shows that obs 10 is not a high leverage point and not influential.
- Figure C shows that obs 20 is a high leverage point, which is not influential.

A real valued measure for influential obs. is the Cook's distance, which is defined as

$$D_i := \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})^T (X^T X)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})}{ps^2},$$

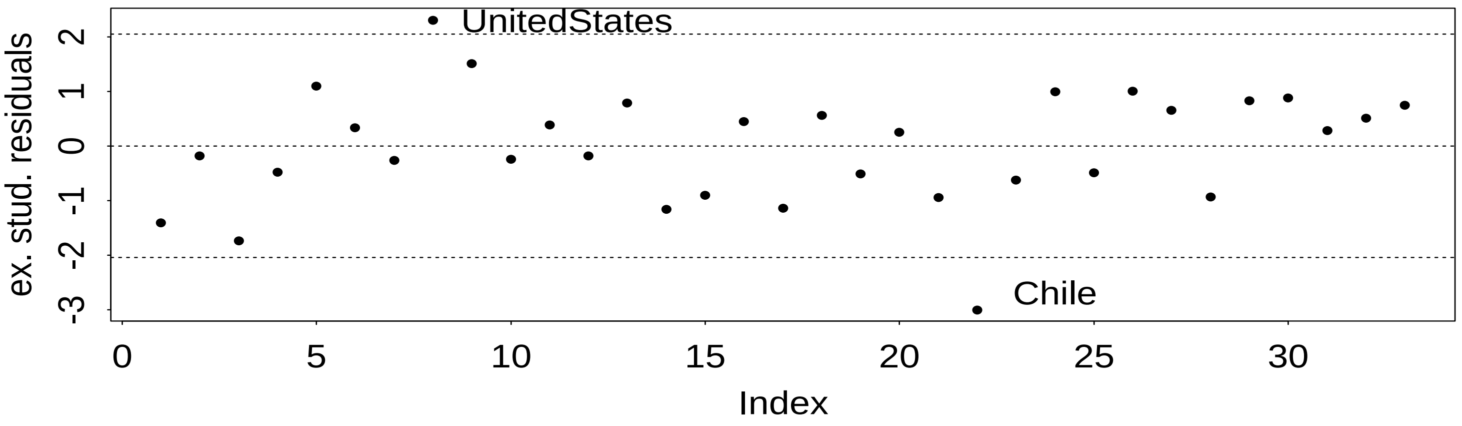
which can be interpreted as the shift in the confidence region when the ith obs. is deleted. Therefore obs. with

$$D_i > 1$$

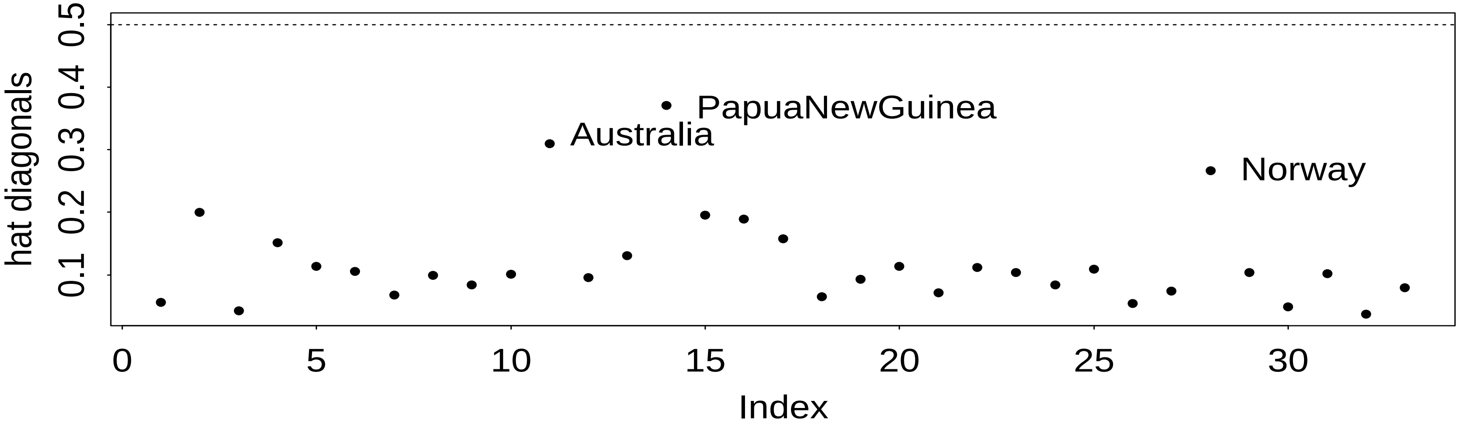
are considered influential. The cutpoint 1 corresponds that the ith obs moves $\hat{\boldsymbol{\beta}}$ to the edge of a 50 % confidence region.

Ex:Life expectancies: Residuals and hat values

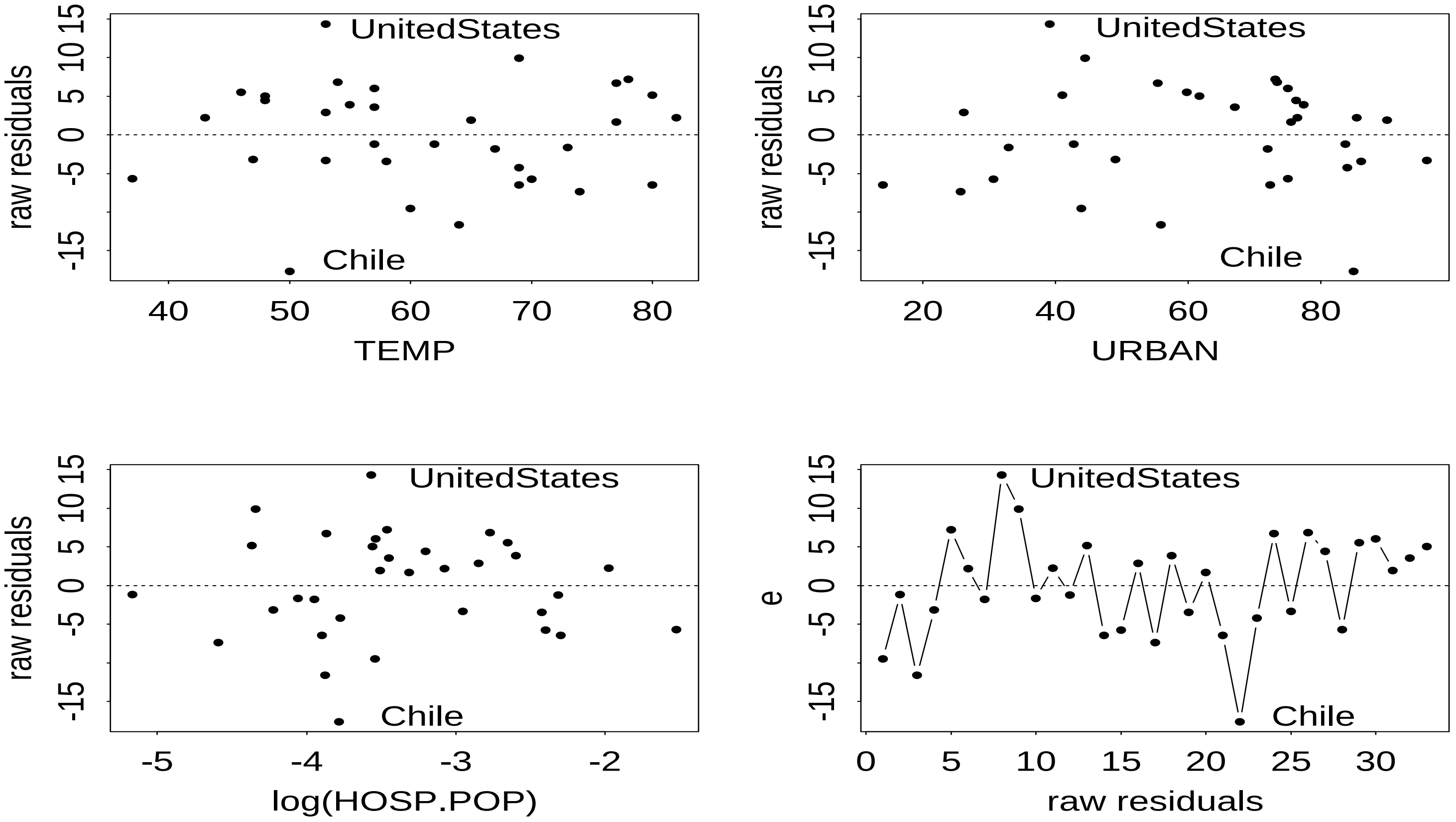
Externally Studentized Residuals



Hat Diagonals



Ex:Life expectancies: Residual plots



Bottom right plot plots r_i versus r_{i-1} to detect correlated errors.

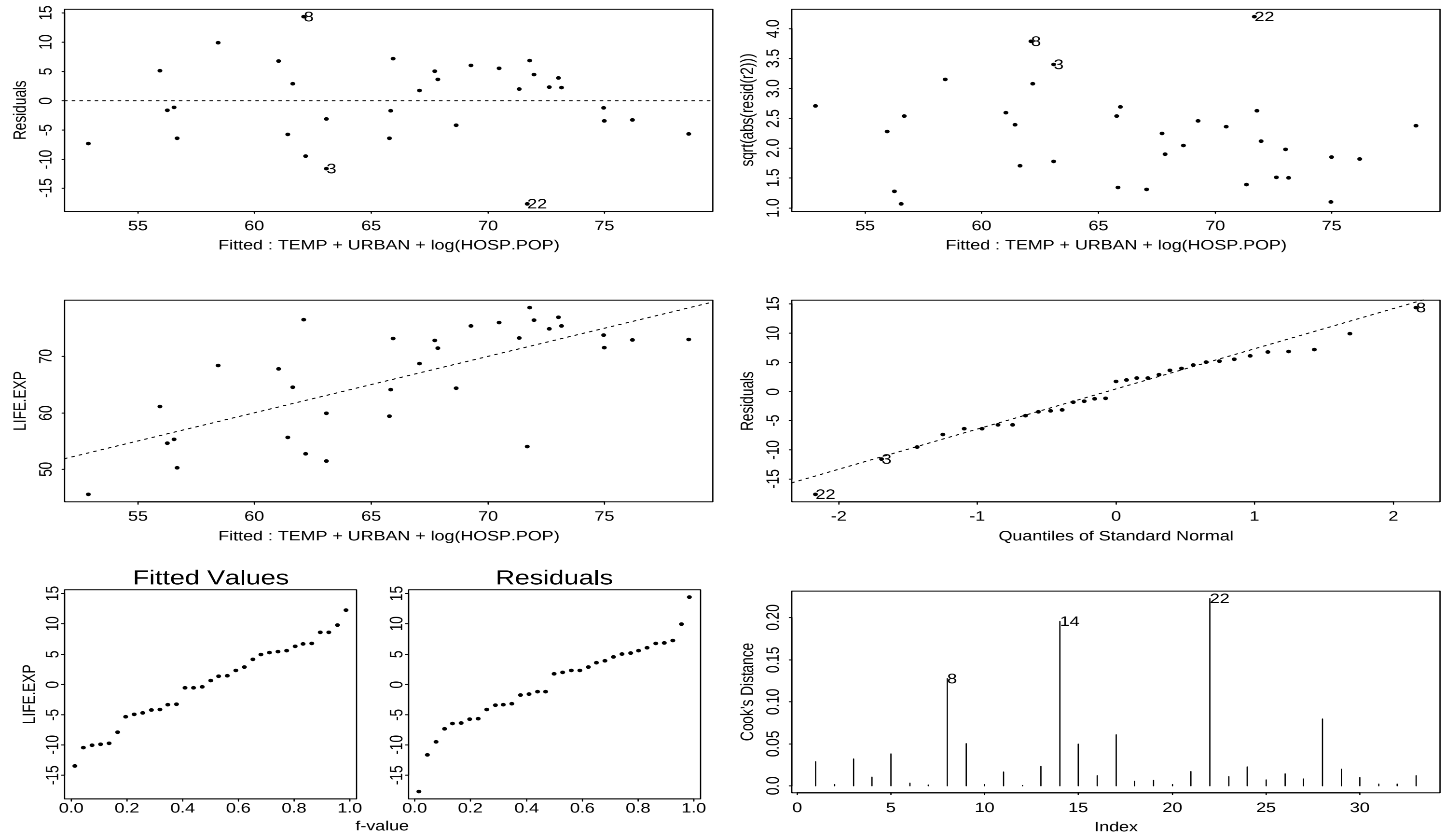
Splus function plot() for LM's

```
> r_lm(y~x)
> plot(r)
```

- Plot 1: $\hat{Y}_i$ versus r_i to check for linearity and variance homogeneity.
- Plot 2: $\hat{Y}_i$ versus $\sqrt{|r_i|}$ to check for large residuals.
- Plot 3: $\hat{Y}_i$ versus Y_i should cluster around $x = y$ line if fit is good.
- Plot 4: qq-plot of r_i to check normality of errors
- Plot 5: Residual-Fit (r-f) spread plot: f-values $(:= \frac{(1:n)-.5}{n}$ against ordered fitted values (left panel) and ordered residuals (right panel). When a good fit is present the spread of the residuals should be much smaller than that of the fitted values.
- Plot 6: i versus D_i to check for influential observations.

Ex:Life expectancies: Output from plot()

```
> plot(r2)
```


Residuals
Fitted : TEMP + URBAN + log(HOSP.POP)
sqrt(abs(resid(r2)))
LIFE.EXP
Quantiles of Standard Normal
Fitted Values
Residuals
f-value
Cook's Distance
Index