Example using the skin cream example

y : sales in district

x_1 : size of district

x_2 : per capita income of district

one can compute:

$p = 3$ $\quad B = (X^t X)^{-1} X^t Y = \begin{pmatrix} 3.4526 \\ 0.4960 \\ 0.0092 \end{pmatrix}$

$n = 15$

$$MSR = \frac{1}{p-1}\left(Y^t Y - B^t X^t Y\right) = 26922.4$$

$$MSE = 4.74$$

$$F^* = \frac{MSR}{MSE} = 5680$$

Assuming α at 0.05 and assuming the ε_i are independent $N(0, \sigma^2)$, we require

$$F(0.95, 2, 12) = 3.89$$

Since F^* exceeds 3.89 we conclude

C_2: Sales are related to population and income.

But is this relation useful for predictions,

well

$$R^2 = \frac{SSR}{SSTO} = 0.9989$$

so when the independent variables x_1 and x_2 are considered, the variation in sales is "99.9% explained".

The F-test for deciding if

$\beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ or not

in the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i\,p-1} + \varepsilon_i \quad (\dagger)$$

assumes that the ε_i are independent $N(0, \sigma^2)$.

To test these assumptions, in the skin cream example, we can plot

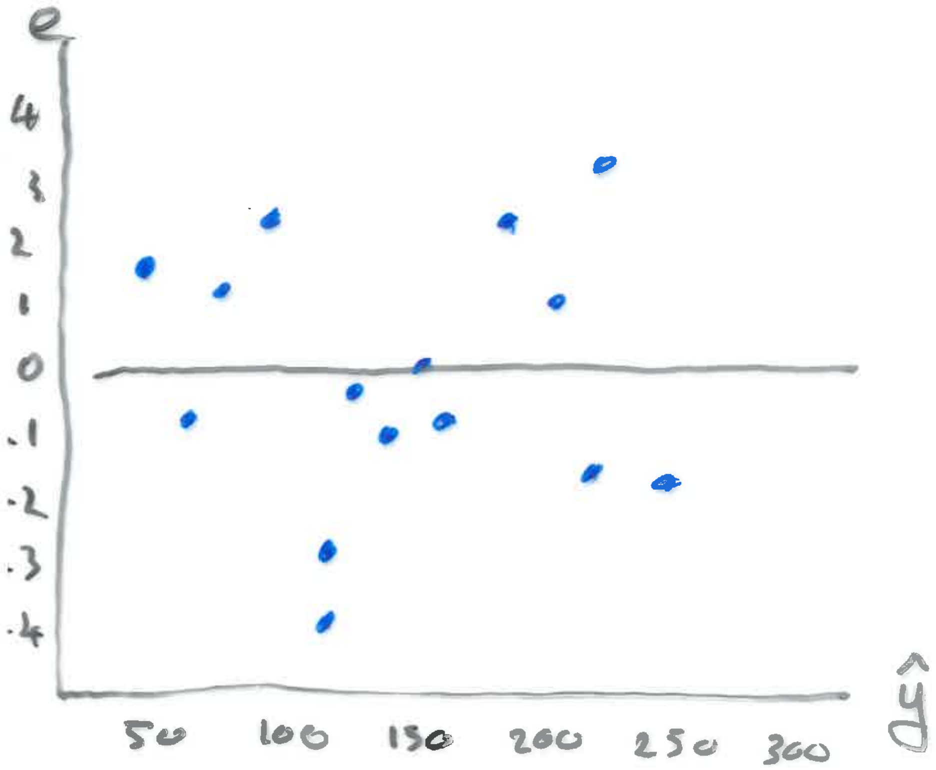
i) $\hat{y}_i$ against e_i

ii) x_{i1} against e_i

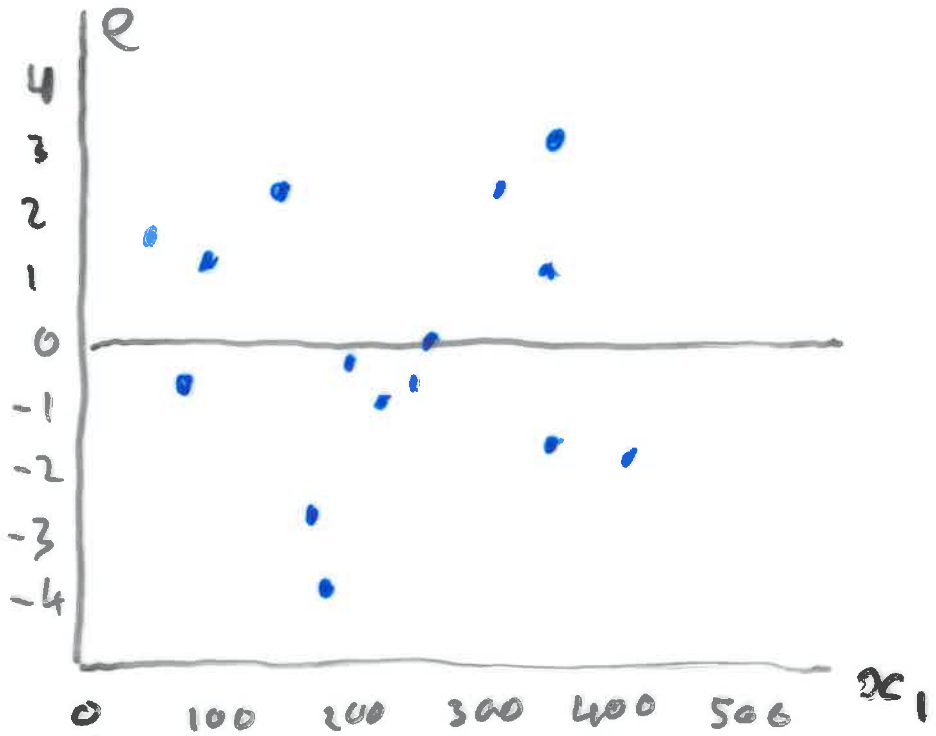
iii) x_{i2} against e_i.

$\boxed{\text{See next slide}}$

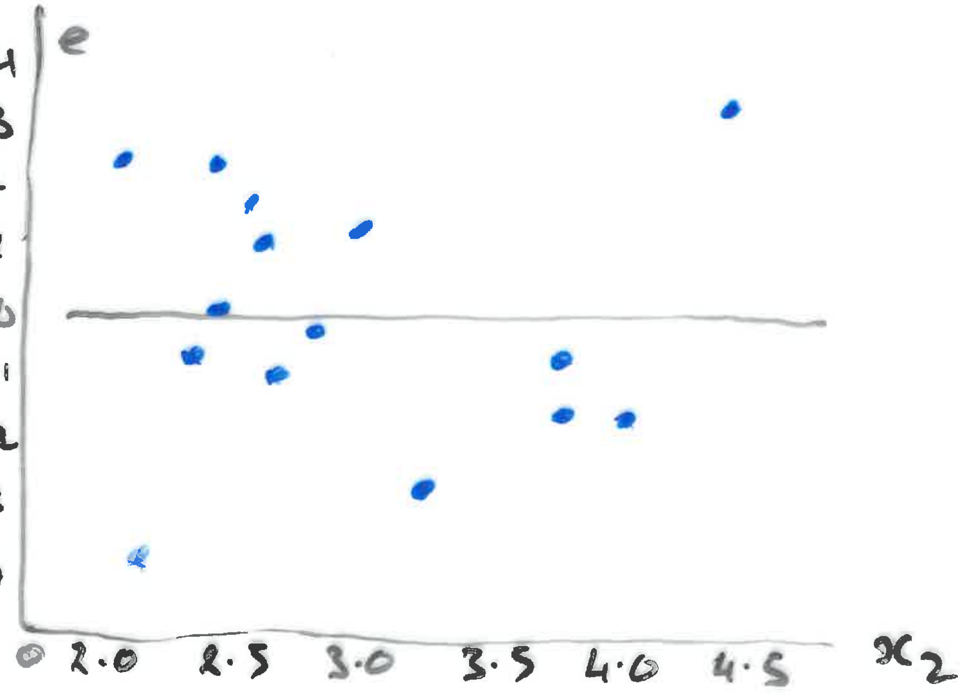
There appears to be no systematic deviation, the residuals seem to be independent and not depend on the level of $\hat{y}$ or the values of x_{i1}, x_{i2}. So it seems ok to accept that ε_i are indep and $N(0, \sigma^2)$.



y vs e



x_1 vs e



x_2 vs e

Defn The estimated covariance matrix for (T) is

$$s^2(B) = MSE\,(X^t X)^{-1}$$

$$= \begin{pmatrix} s^2(b_0) & s(b_0, b_1) & \cdots & s(b_0, b_{p-1}) \\ s(b_1, b_0) & s^2(b_1) & \cdots & \\ & & \ddots & \\ & & & s^2(b_{p-1}) \end{pmatrix}.$$

we only need $s^2(b_0), s^2(b_1), \cdots$.

Theorem Assume ε_i are independent $N(0, \sigma^2)$ the quantity

$$\frac{b_k - \beta_k}{s(b_k)}$$

follows a t-distribution with $n-p$ degrees of freedom.

So, if $q \leq p$ parameters β_k are to be estimated jointly, the confidence intervals with family coefficient $1-\alpha$ are:

$$b_k - T\, s(b_k) \leq \beta_k \leq b_k + T\, s(b_k)$$

where

$$T = t\left(1 - \frac{\alpha}{2q},\ n-p\right).$$

Example Continuing with skin cream sales, it is desired to estimate β_1 and β_2 jointly with a family confidence coefficient of 0.90.

$$s^2(B) = MSE(X^tX)^{-1} = \begin{pmatrix} 5.9081 & * & * \\ * & .000036656 & * \\ * & * & .000000937 \end{pmatrix}$$

$s^2(b_1) = .000036656$, $\quad s(b_1) = .006054$

$s^2(b_2) = .000000937$, $\quad s(b_2) = .0009681$

$$T = t\left(1 - \frac{0.10}{2 \times 2}, 12\right) = t(0.975, 12) = 2.179$$

So

$$0.4961 - (2.179)(.006054) \le \beta_1 \le 0.4961 + (2.179)(.006054)$$

or

$$0.483 \le \beta_1 \le 0.509$$

and similarly

$$0.0071 \le \beta_2 \le 0.0113$$