Approximation of mean and variance for functions of random variables.

Example. Measure voltage, U, and current, I, and want to calculate the resistance $R = \frac{U}{I}$. What about $E[R]$ and $Var[R]$.

Let $E[U] = \mu_1$ and $E[I] = \mu_2$

From a Taylor-expansion of $f(x,y) = \frac{x}{y}$ around (μ_1, μ_2) we get.

$$f(x,y) = f(\mu_1, \mu_2) + \frac{\partial f}{\partial x}\Big|_{(\mu_1,\mu_2)}(x-\mu_1) + \frac{\partial f}{\partial y}\Big|_{(\mu_1,\mu_2)}(y-\mu_2) + \varepsilon$$

$$= \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2}(x-\mu_1) - \frac{\mu_1}{\mu_2^2}(y-\mu_2) + \varepsilon$$

i.e. if ε is small, R can be approximated by

$$R = \frac{U}{I} \approx \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2}(U-\mu_1) - \frac{\mu_1}{\mu_2^2}(I-\mu_2)$$

such that.

$$E[R] \approx \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2}E[U-\mu_1] - \frac{\mu_1}{\mu_2^2}E[I-\mu_2] = \frac{\mu_1}{\mu_2},$$

and

$$Var[R] \approx \left[\frac{1}{\mu_2}\right]^2 Var[U] + \frac{\mu_1^2}{\mu_2^4} Var[I] - \frac{2\mu_1}{\mu_2^3} Cov[U,I]$$

If the uncertainty in U and I ~~is~~ are measurement-errors it is reasonable that $Cov[U,I] = 0$

In general if $Y = f(x_1, \ldots, x_n)$, we have:

$$f(x_1, \ldots, x_n) = f(\mu_1, \ldots, \mu_n) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_i}\Big|_{(\mu_1,\ldots,\mu_n)}(x_j - \mu_j) + \varepsilon$$

Approximation of mean and variance for functions of random variables.

Example. Measure voltage, U, and current, I, and want to calculate the resistance $R = \frac{U}{I}$. What about $E[R]$ and $Var[R]$.

Let $E[U] = \mu_1$ and $E[I] = \mu_2$

From a Taylor-expansion of $f(x,y) = \frac{x}{y}$ around (μ_1, μ_2) we get.

$$f(x,y) = f(\mu_1,\mu_2) + \frac{\partial f}{\partial x}\Big|_{(\mu_1,\mu_2)}(x-\mu_1) + \frac{\partial f}{\partial y}\Big|_{(\mu_1,\mu_2)}(y-\mu_2) + \varepsilon$$

$$= \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2}(x-\mu_1) - \frac{\mu_1}{\mu_2^2}(y-\mu_2) + \varepsilon$$

i.e. if ε is small, R can be approximated by

$$R = \frac{U}{I} \approx \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2}(U-\mu_1) - \frac{\mu_1}{\mu_2^2}(I-\mu_2)$$

such that.

$$E[R] \approx \frac{\mu_1}{\mu_2} + \frac{1}{\mu_2}E[U-\mu_1] - \frac{\mu_1}{\mu_2^2}E[I-\mu_2] = \frac{\mu_1}{\mu_2}.$$

and

$$Var[R] \approx \left[\frac{1}{\mu_2}\right]^2 Var[U] + \frac{\mu_1^2}{\mu_2^4}Var[I] - \frac{2\mu_1}{\mu_2^3}Cov[U,I]$$

If the uncertainty in U and I ~~is~~ are measurement-errors it is reasonable that $Cov[U,I] = 0$

In general if $Y = f(x_1,\dots,x_n)$, we have:

$$f(x_1,\dots,x_n) = f(\mu_1,\dots,\mu_n) + \sum_{j=1}^{n}\frac{\partial f}{\partial x_i}\Big|_{(\mu_1,\dots,\mu_n)}(x_j-\mu_j) + \varepsilon$$

and if higher order terms are small, we have

$$E[Y] \approx f(\mu_1, \dots, \mu_n)$$

$$Var[Y] \approx \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mu_1, \dots, \mu_n)^2 \, Var[X_i] + 2 \sum_{j<k} \frac{\partial f}{\partial x_j}(\mu_1, \dots, \mu_n) \frac{\partial f}{\partial x_k}(\mu_1, \dots, \mu_n) \, Cov(X_j, X_k)$$

Example $U \sim N(100, 2^2)$ $\qquad R = \frac{U}{I}$

$I \sim N(20, 1)$

We get. $E[R] \approx 5$ and $Var[R] \approx \left(\frac{1}{20}\right)^2 \cdot 4 + \left(\frac{100}{400}\right)^2 \cdot 1$

$$= \frac{4}{400} + \frac{1}{16} = \frac{1}{100} + \frac{1}{16} = 0.077.$$

$U \sim N(100, 5^2)$
$I \sim N(20, 2^2)$ $\Rightarrow$ $E[R] \approx 5$

$Var[R] \approx \left(\frac{1}{20}\right)^2 \cdot 25 + \left(\frac{100}{400}\right)^2 \cdot 4 \approx \underline{0.31}$

Chap 11 Regression Analysis

Purpose: Find relationships between one responsevariable and one or more explanatory variables. Regression analysis is a modelling tool.

11.2. Simple linear regression (linear in the coefficients)

Model: $Y_i = \alpha + \beta x_i + \varepsilon_i$ } independent $E[\varepsilon_i] = 0$, $Var[\varepsilon_i] = \sigma^2$

response
dependent variable

regression variable
independent variable
explanatory variable

x_i may be a transformation of other variables.

i.e. $x_i = \ln(t_i)$, $x_i = \sin(t_i)$ and so on.

Observe: $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ which ~~are~~ have to fulfill.

$$y_1 = \alpha + \beta x_1 + \varepsilon_1$$
$$y_2 = \alpha + \beta x_2 + \varepsilon_2$$
$$\vdots$$
$$y_n = \alpha + \beta x_n + \varepsilon_n$$

or

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

We need to find estimate, (a, b), for (α, β).

Method of Least Squares

Find the a and b that minimize

$$Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - b x_i)$$

$$\frac{\partial Q}{\partial a} = 0 \iff -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0$$

$$\frac{\partial Q}{\partial b} = 0 \iff -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i = 0$$

This gives the 2 normal equations

$$\sum_{i=1}^{n} a + \sum_{i=1}^{n} bx_i = \sum_{i=1}^{n} y_i \quad (1)$$

$$\sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} bx_i^2 = \sum_{i=1}^{n} x_i y_i \quad (2)$$

From (1) we get:

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \text{ which implies } a = \bar{y} - b\bar{x}$$

and substitution of a with $\bar{y} - b\bar{x}$ in (2) gives

$$(\bar{y} - b\bar{x})\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\Rightarrow b\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \sum_{i=1}^{n} x_i(y_i - \bar{y})$$

$$\Rightarrow b = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Here is used: $\sum_{i=1}^{n}\bar{x}(y_i - \bar{y}) = \bar{x}\sum_{i=1}^{n}(y_i - \bar{y}) = 0$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$\sum_{i=1}^{n}\bar{y}(x_i - \bar{x}) = \bar{y}\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

The corresponding estimators are: $\hat{\beta} = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

and $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$.

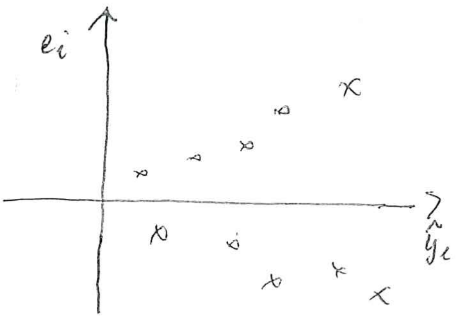
The estimated expected regression line (estimated model) is:

$$\hat{y} = a + b x$$

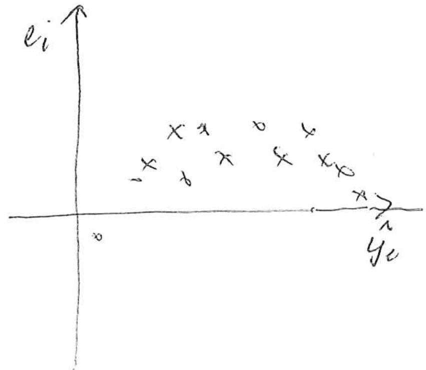
Residuals: $e_i = y_i - \hat{y}_i = y_i - a - b x_i$

These should be plotted against $\hat{y}_i$, against x_i and eventually other regression variables. Normal-plot should be used to check for normal distribution

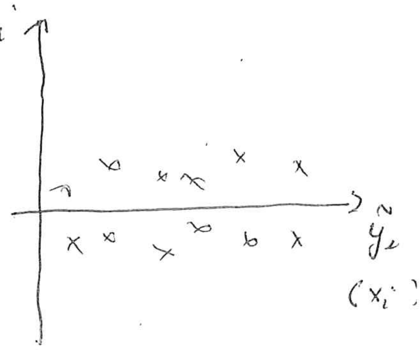
Pattern in the residuals indicates that the model does not fit.



Variance not constant.
Transform y



Should consider second order terms
$\text{e}: E[Y_i] = \alpha + \beta_1 x_i + \beta_2 x_i^2$



No pattern (ok.)