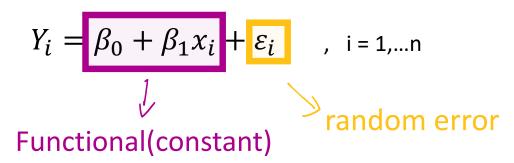
Hw 1 hints & Basic concepts of linear regression

Yuanyuan Li



- Statistical model: a model including random component with assumptions concerning the generation of sample data
- Response variables: Y variable, a random variable
- Explanatory/Predictor variables: X variables, assumed as known constants
- Parameters/regression coefficients: β_0 , β_1 , unknown constants
- Assumption about random error ε_i :
 - Model 1.1: ε_i is a random error term with mean $E\{\varepsilon_i\} = 0$ and variance $\sigma^2\{\varepsilon_i\} = \sigma^2$; ε_i and ε_j are uncorrelated so that their covariance is zero (i.e., $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$ for all $i, j; i \neq j$) $i = 1, \ldots, n$

Model 1.24:
$$\varepsilon_{i}$$
 are independent $N(0, \sigma^{2})$ $\longrightarrow E(\varepsilon_{i}) = 0$, $Var(\varepsilon_{i}) = \sigma^{2}$ $\longrightarrow E(Y_{i}) = \beta_{0} + \beta_{1}x_{i} + E(\varepsilon_{i}) = \beta_{0} + \beta_{1}x_{i}$, $Y_{i} = E(Y_{i}) + \varepsilon_{i}$ $Var(Y_{i}) = Var(\varepsilon_{i}) = \sigma^{2}$

&
$$P(-\sigma < \varepsilon_i < \sigma) = P(-1 < Z < 1) = 0.68$$

The Goals of linear regression:

• Inference: quantify the strength of the relationship between the response(Y) and the explanatory variables(X), i.e., is this effect significant?

Estimate and make inference on β_1 !

 Prediction: make a prediction of the response(Y) given additional values of the predictor variables(X), i.e., what would be the new response of an instance?

Predict *Y* and measure its uncertainty!

- $Y_i = E(Y_i) + \varepsilon_i$, $E(Y_i) = \beta_0 + \beta_1 x_i$ can be predicted value of Y_i , but they are *unknown* constants
- Estimate the *unkown* parameters, denote their estimators as $\hat{\beta}_0$, $\hat{\beta}_1$
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is estimated $E(Y_i)$ ="estimated mean"
- Overall, \hat{Y}_i is also called "predicted value" of Y_i