1

QDA 2023 03 22

MULTIPLE LINEAR REGRESSION (p regressors)

$K = p+1$

REGRESSOR

$$\Rightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots \\ \vdots & x_{21} & x_{22} & \\ \vdots & \vdots & \vdots & \\ & x_{i1} & x_{i2} & \\ & \vdots & \vdots & \\ 1 & & & \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \end{bmatrix}$$

$i = 1 \dots m$

2nd regressor x_{i2} $i=1\dots m$ (ex t^2)

1st regressor x_{i1} $i = 1 \dots m$

ex: t

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$$\underline{y} = \underline{X}\,\underline{\beta} + \underline{\varepsilon}$$

②

$$\underbrace{\underline{y}}_{m\times 1} = \underbrace{\underline{X}}_{m\times K}\,\underbrace{\underline{\beta}}_{K\times 1} + \underbrace{\underline{\varepsilon}}_{m\times 1}$$

$$\underline{\hat{\beta}} = \underbrace{(\underline{X}'\underline{X})^{-1}}_{\text{exists if}}\,\underline{X}'\underline{y} \quad \begin{cases} b_0 = \hat{\beta}_0 \\ b_1 = \hat{\beta}_1 \\ \vdots \\ b_p = \hat{\beta}_p \end{cases}$$

$$\underline{b} = \underline{\hat{\beta}}$$

the regressors are linearly INDEPENDENT

$\underline{ex}$ $\quad x_{i2} = k\, x_{i1}$ $\quad$ NOT POSSIBLE!

x_{i2}: 2nd reg

x_{i1}: 1st regressor

(3)

$$E(\hat{\underline{\beta}}) = \underline{\beta} \quad (*) \qquad \boxed{\mathrm{Cov}(\hat{\underline{\beta}}) = \sigma^2(\underline{X}'\underline{X})^{-1}} \quad (**)$$

$$\hat{\underline{y}} = \underline{X}\hat{\underline{\beta}} = \overbrace{\underline{X}\underbrace{(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}}_{\hat{\beta}}}^{H} = \underline{H}\,\underline{y}$$

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$$

$$\underline{\varepsilon} \sim MN(\underline{0}, \sigma^2\underline{I}) \Rightarrow \hat{\underline{\beta}} \sim MN(\underline{\beta}, \sigma^2(\underline{X}'\underline{X})^{-1})$$

$$\Updownarrow$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

(4) **Cov ?**

ex SIMPLE LINEAR REGRESSION

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \hat{\underline{\beta}} = \underline{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$E(\hat{\underline{\beta}}) = E(\underline{b}) = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$V(b_0)$

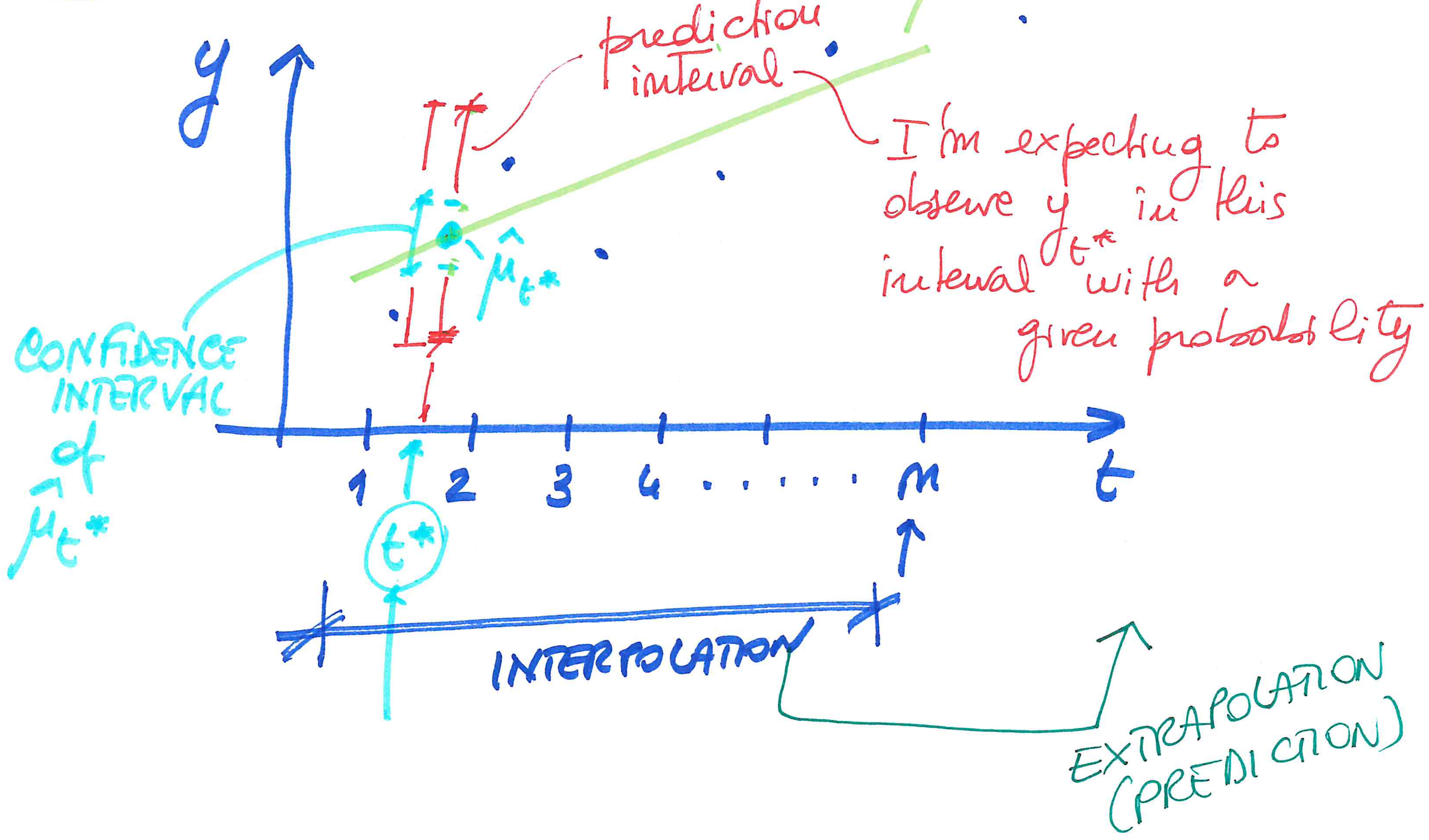
$\text{Cov}(b_0, b_1)$

$$\text{Cov}(\hat{\underline{\beta}}) = \text{Cov}(\underline{b}) = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma_{b_1}^2 \end{bmatrix} = \sigma^2 (\underline{X}'\underline{X})^{-1}$$

$V(b_1)$

$$\text{Cov}(X_1, X_2) = E\left[(X_1 - \mu_1)(X_2 - \mu_2)\right]$$

⑤ EXTRAPOLATION?



6

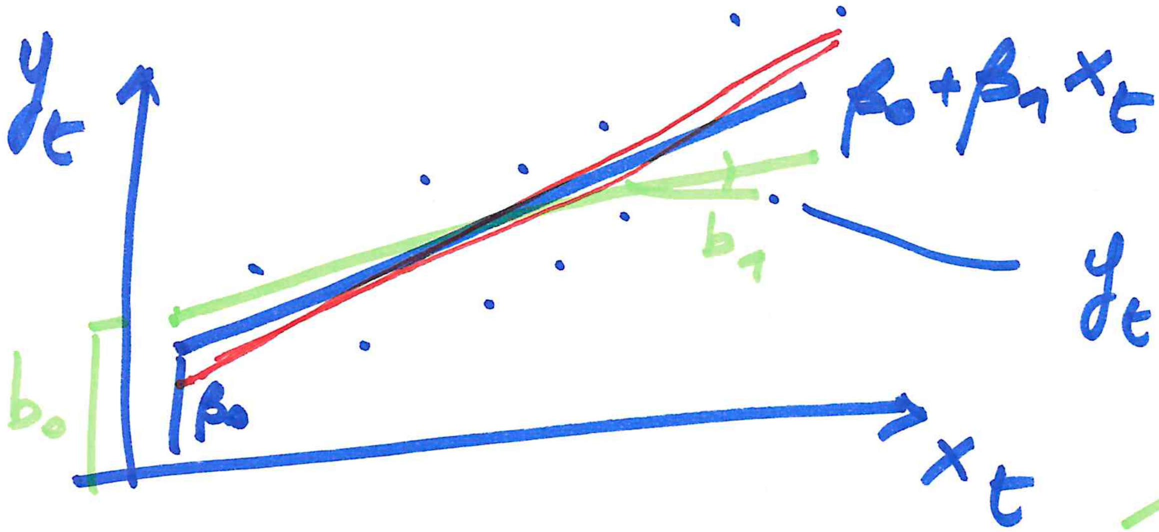
$Cov(\underline{b})$

$$\sigma_{b_0}^2 = V(b_0) = \sigma_\varepsilon^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]$$

SIMPLE LINEAR MODEL

$$\sigma_{b_1}^2 = V(b_1) = \frac{\sigma_\varepsilon^2}{S_{xx}}$$

$$\sigma_{b_0 b_1} = Cov(b_0, b_1) = -\sigma_\varepsilon^2 \frac{\bar{x}}{S_{xx}}$$



$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

over β_0 under β_1

~~over~~ under β_0
over β_1

⑦ CONFIDENCE INTERVAL $\hat{\mu}_{x_0}$

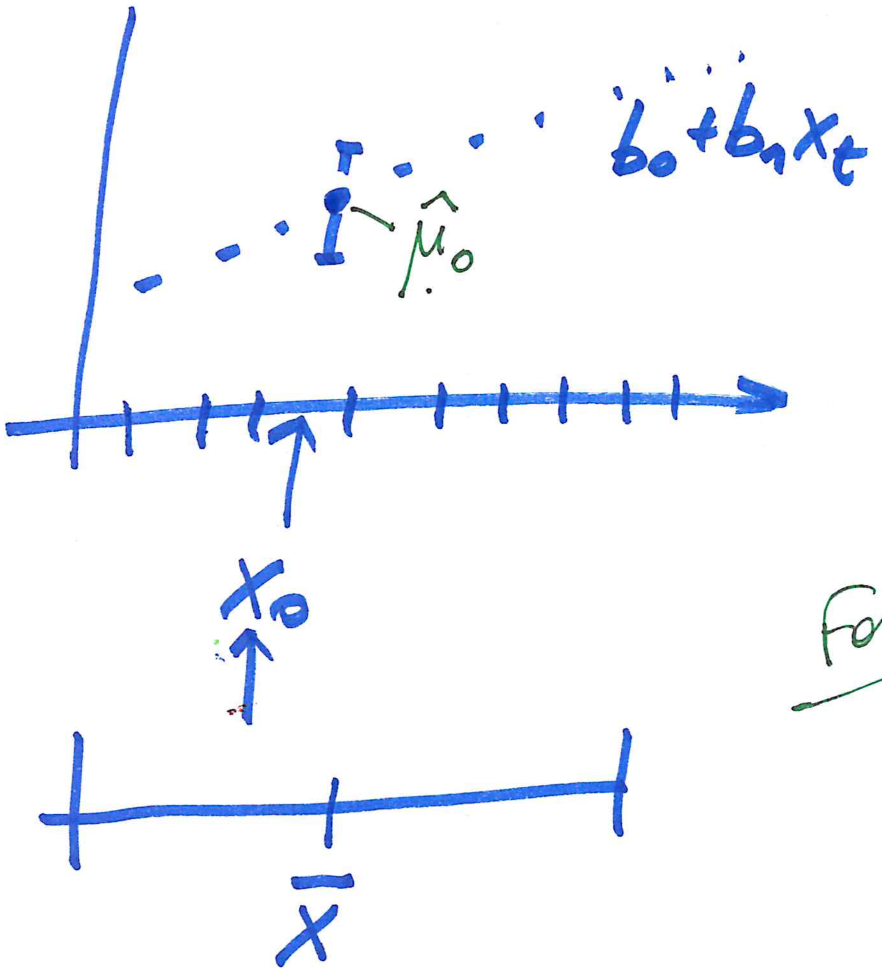
$$\mu_0 = E(Y \mid X_t = x_0) = E(Y / x_0)$$

$$\hat{\mu} \sim N(\; , \;)$$

$$\hat{\mu}_0 = b_0 + b_1 x_0$$

$$\hat{V}(\hat{\mu}_0) = \hat{\sigma}_\varepsilon^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

$$\hat{\sigma}_\varepsilon^2 = MS_E$$



For given n

$\hat{V}(\hat{\mu}_0)$ MIN $x_0 = \bar{x}$

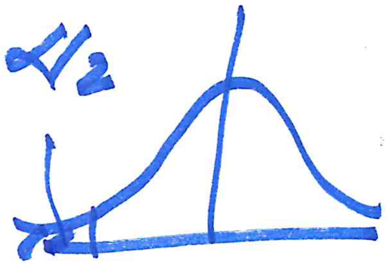
↑ x_0 moves far from $\bar{x}$

8

$$\frac{\hat{\mu}_0 - \mu_0}{\sqrt{V(\hat{\mu}_0)}} \sim N(0,1) \qquad \frac{\hat{\mu}_0 - \mu_0}{\sqrt{\hat{V}(\hat{\mu}_0)}} \sim t_{m-\bar{K}}$$

Confidence interval on μ_0:

$$P\Big(L \leq \underbrace{\frac{\hat{\mu}_0 - \mu_0}{\sqrt{\hat{V}}}}_{t_{m-\bar{K}}} \leq U\Big) = 1-\alpha$$

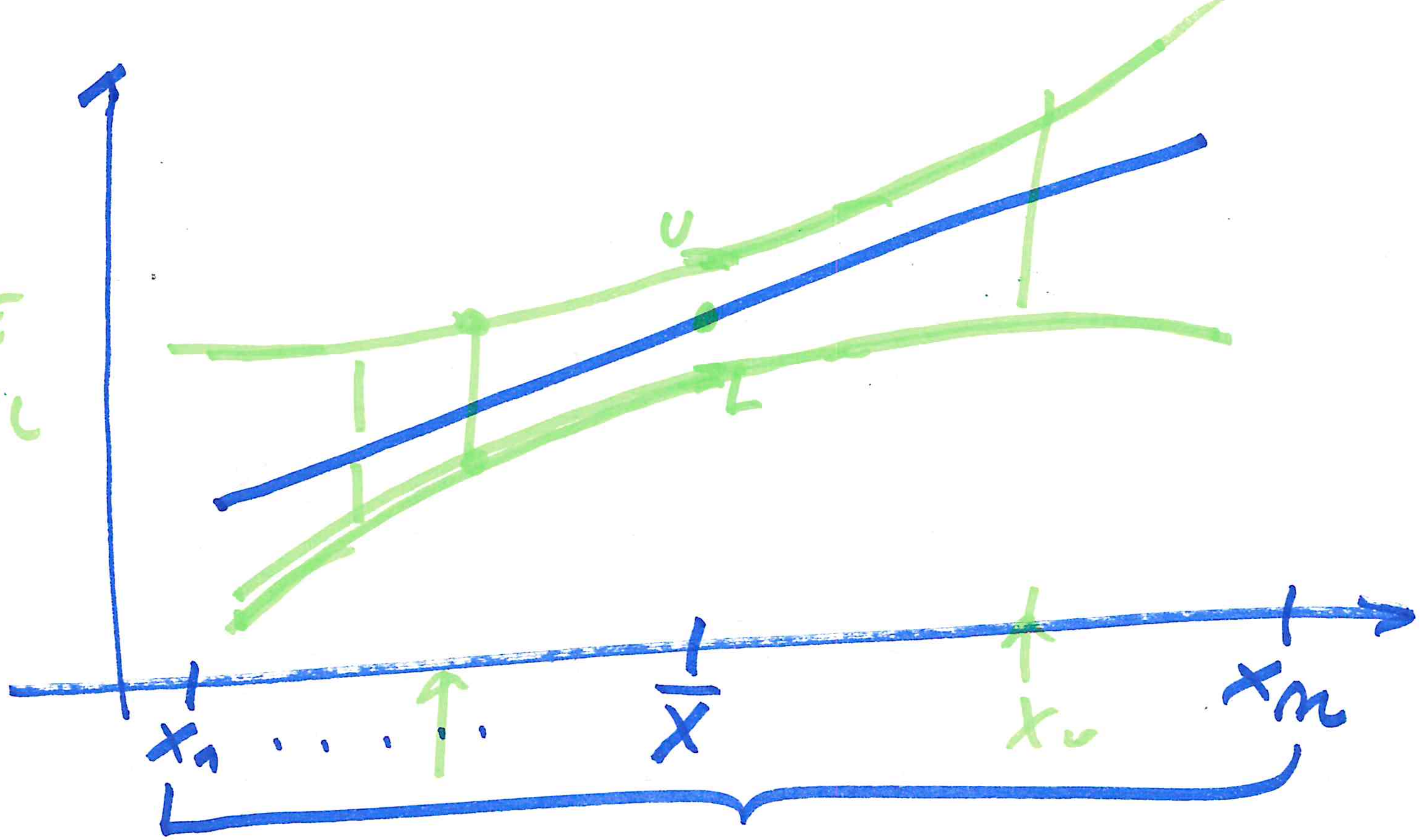


$$L = \hat{\mu}_0 - t_{\alpha/2, m-\bar{K}} \sqrt{\hat{V}(\hat{\mu}_0)}$$

$$U = \hat{\mu}_0 + t_{\alpha/2, m-\bar{K}} \sqrt{\hat{V}(\hat{\mu}_0)}$$

9

CONFIDENCE INTERVAL ON THE MEAN $\hat{\mu}_0$



$1-\alpha$

$$\hat{V}(\hat{\mu}_0) = \hat{\sigma}_\varepsilon^2 \left[\frac{1}{m} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

$$V(y_0 - \cancel{\quad}) = V(\hat{\mu}_0) + \sigma_\varepsilon^2 \qquad Y = \mu_0 + \varepsilon$$

10

Pay <u>attention</u> to <u>check</u> <u>the</u> <u>assumptions</u> at the end

$$\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$$

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

ERRORS $e_t = \hat{\varepsilon}_t$

RUNS TEST

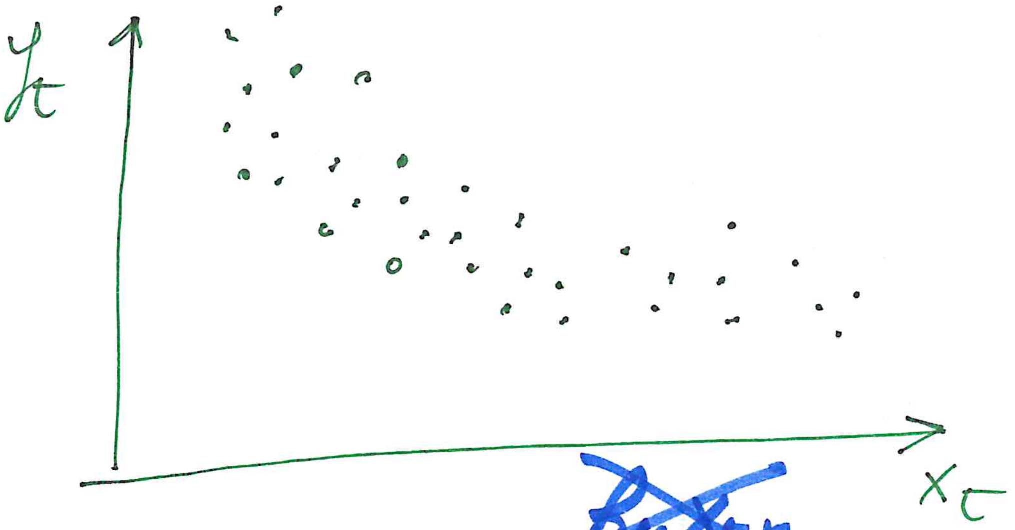
BARTLETT'S $H_0: \rho_1 = 0$

NORMALITY

$$\hat{\varepsilon}_t = e_t = y_t - \underbrace{(b_0 + b_1 x_t)}_{\hat{\mu}_t}$$

ERRORS

11 SELECTION OF REGRESSORS ?



1) DATA SNOOPING

? $x_{t1} = t$

? $x_{t2} = t^2$

? $x_{t3} = 1/t$

⋮

x_{tp}

$\cancel{\beta_k x_{tk}}$

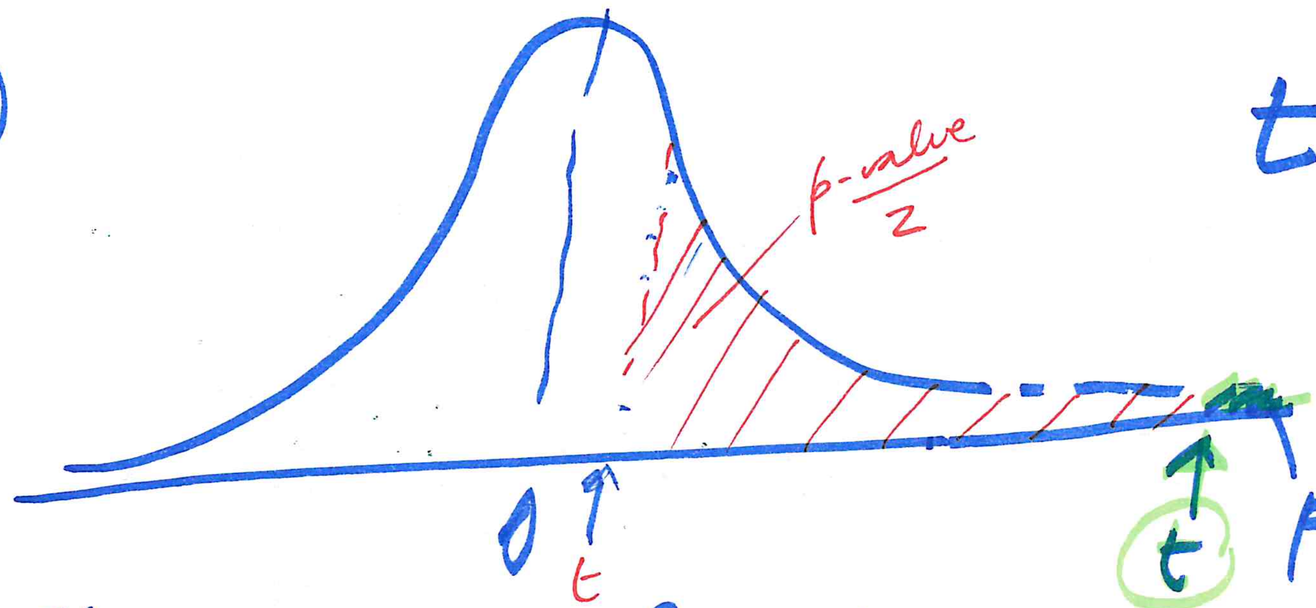
$$Y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_p x_{tp} + \varepsilon_t$$

$H_0: \beta_k = 0 \rightarrow$ "ACCEPT" H_0 (CANNOT REJECT)

$H_1: \beta_k \neq 0$

(12)

$$t = \frac{b_k - 0^{\text{ if } H_0}}{S_{b_k}}$$



p-value/2

|t| large

p-value "small" $(\leq \alpha)$ → REJECT H_0 → X_K IMPORTANT (INCLUDE)

p-value "large" $(\geq \alpha)$ → "ACCEPT" H_0 → X_K MEANINGLESS (ERASE it)

$|t| \approx 0$ (small)