Prob Stats L17

Linear Regression

April 20, 2023

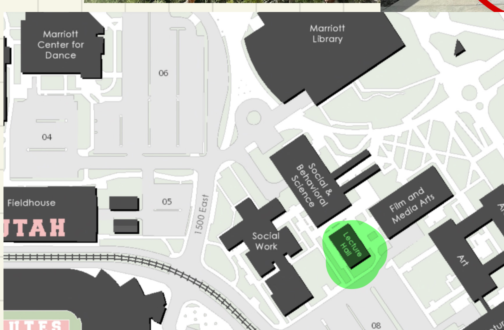
Final Exam

Monday, May 1
3:30 - 5:30 pm

location:
BEH S Aud

"the tower of terror*"

*Aud not in tower



$(1-\alpha)100\%$ **Confidence Intervals**

$$\bar{X}_n \pm t_{\alpha/2}\frac{S_n}{\sqrt{n}}$$

$$[L_n = \bar{X}_n - t_{\alpha/2}\frac{S_n}{\sqrt{n}}, \quad R_n = \bar{X}_n + t_{\alpha/2}\frac{S_n}{\sqrt{n}}$$

$$Pr(L_n \le \mu \le R_n) = 1-\alpha$$

$$\Longleftrightarrow$$

$$Pr\left(-t_{\alpha/2} \le \frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} \le t_{\alpha/2}\right) = 1-\alpha$$

$-t_{\alpha/2} =$ `qt(`$\alpha/2$`, df=20)`

Hypothesis Testing w/ critical value α

H_0: $X_i \sim N(\mu,\sigma)$, σ unknown

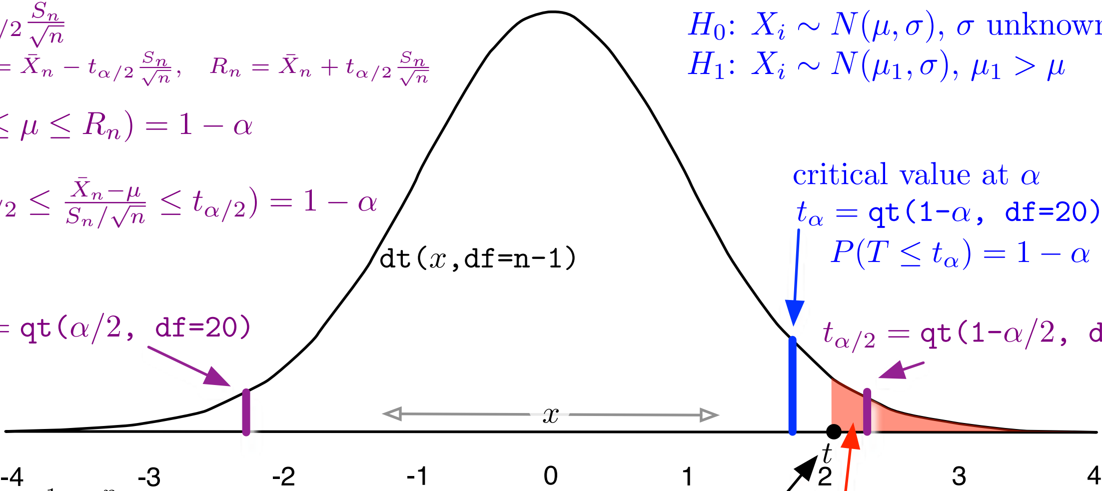
H_1: $X_i \sim N(\mu_1,\sigma)$, $\mu_1 > \mu$

critical value at α

$t_\alpha =$ `qt(1-`α`, df=20)`

$P(T \le t_\alpha) = 1-\alpha$

$t_{\alpha/2} =$ `qt(1-`$\alpha/2$`, df=20)`



random variables

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$T = \frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} \sim \text{t-dist}(\text{df} = n-1)$$

realization of data

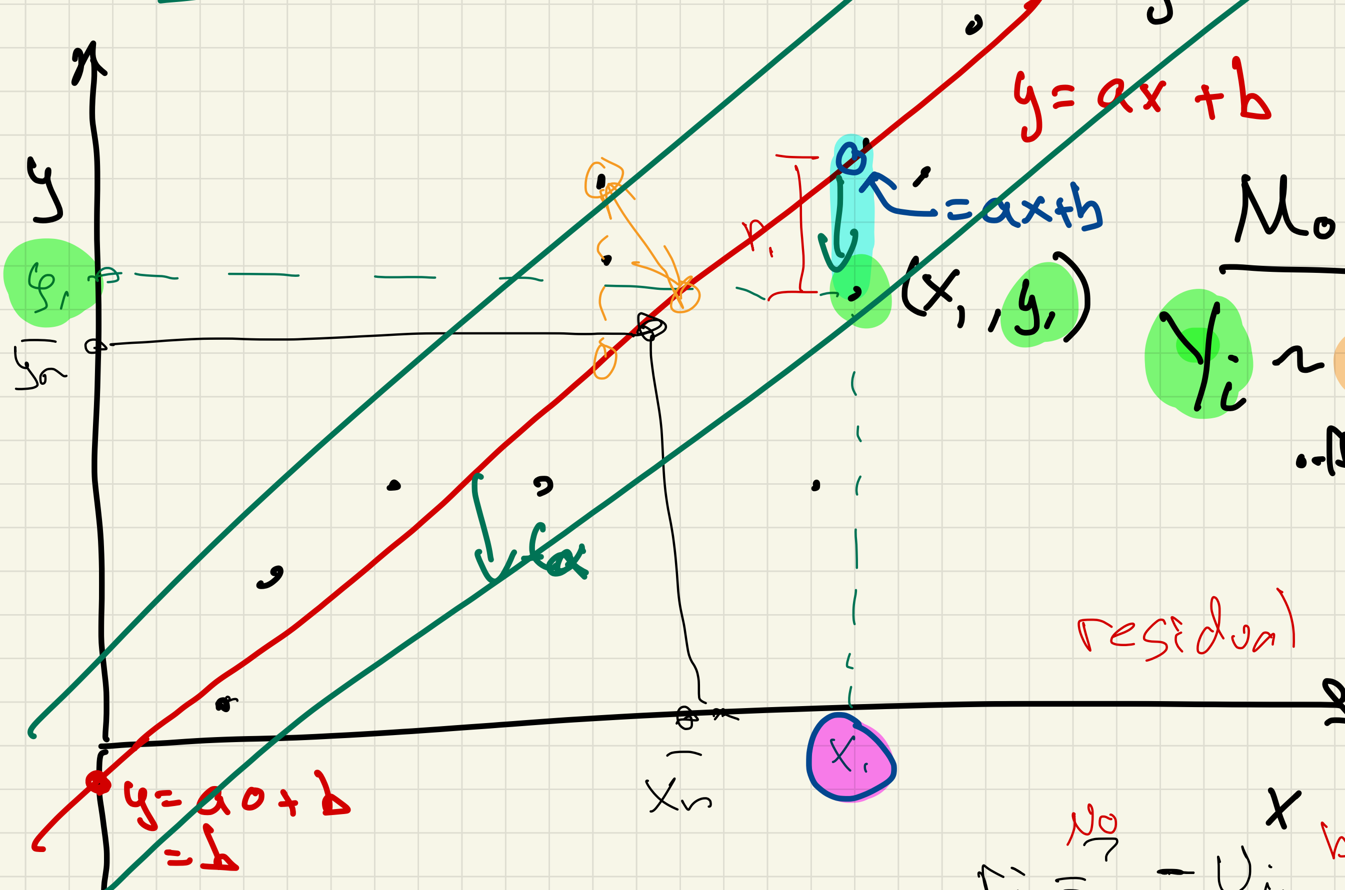
$$t = \frac{\bar{x}_n-\mu}{s_n/\sqrt{n}}$$

p-value

p = `1 - pt(`t`, df = 20)`

$$\Pr(T \le t) = 1-p$$

Linear Regression



$y \sim x$

Model

parameters

$Y_i \sim \underbrace{a X_i + b}_{\text{affine}} + U_i$

$U_i \sim N(0, \sigma^2)$

residual $= \underline{(a x_i + b)} - \underline{y_i}$

$= r_i$

$r_i \overset{?}{=} -U_i$ No

but $R_i = -U_i$

$$\underline{\text{Cost}} \quad C(a,b) = \sum_{i=1}^{n} \left(Y_i - (a x_i + b + U_i) \right)^2$$

$$= \sum_{i=1}^{n} R_i^2$$

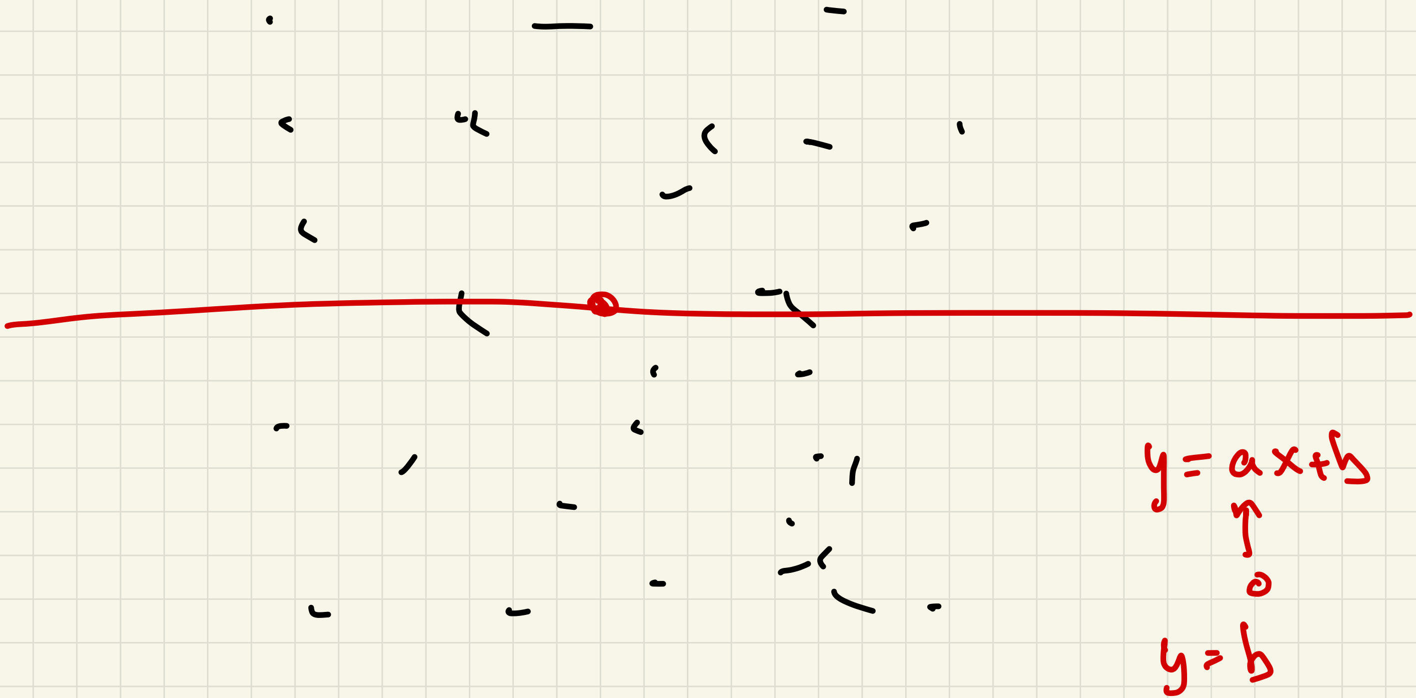
$$\frac{\partial C(a,b)}{\partial a} = 0 \iff \sum_{i=1}^{n} (Y_i - a - b x_i) X_i = 0$$

$$\hat{a} = \frac{n \sum_{i=1}^{n} x_i Y_i - \left(\sum_i x_i \right) \left(\sum_i Y_i \right)}{n \sum_i x_i^2 - \left(\sum_i x_i \right)^2} \neq \frac{\operatorname{cov}(x, Y)}{\operatorname{var}(x)}$$

$$\underbrace{\hat{b} = \bar{Y}_n - \hat{a} \bar{X}_n}$$

$$E[\hat{a}] = a$$

$$E[\hat{b}] = b$$


y = ax+b
0
y = b