Context:

linear model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$(x_1, y_1), \ldots, (x_n, y_n)$

x_i: observation for X_i (fixed)

y_i: ______ Y_i (random)

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\varepsilon_i}_{\text{noise}}$$

$$U = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \varepsilon_n \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2 I_n \right)$$

Results:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \sigma^2 V \right)$$

with $V = ({}^t X X)^{-1}$

for $i \in \{1, 2\}$,

$$\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \sim T(n-2)$$

Remember:

$$\hat{\sigma}^2_{\hat{\beta}_0} = \hat{\sigma}_n^2 \frac{\sum X_i^2}{n \sum (X_i - \bar{X}_n)^2}$$

$$\hat{\sigma}_{\hat{\beta}_1} = \hat{\sigma}_n^2 \frac{1}{\sum (X_i - \bar{X}_n)^2}$$

with $\underbrace{\hat{\sigma}_n^2 = \frac{1}{n-2} \sum (Y_i - \hat{Y}_i)^2}_{\text{unbiased estimate for } \sigma^2}$

$$\frac{(n-2)\,\hat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-2)$$

and $\hat{\beta} \perp\!\!\!\perp \hat{\sigma}_n^2$

Proof:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= (X^T X)^{-1} X^T (X\beta + U)$$

$$= \underbrace{\beta}_{\text{constant vector}} + \underbrace{(X^T X)^{-1} X^T}_{\text{deterministic matrix}} U \quad \text{gaussian vector}$$

$\Rightarrow \hat{\beta}$ is a gaussian vector

We already see that $E(\hat{\beta}) = \beta$

and $V[\hat{\beta}] = \sigma^2 ({}^t X X)^{-1}$

Since $\hat{\beta}$ is gaussian vector, we have:

$\hat{\beta}_0$ and $\hat{\beta}_1$ are gaussian random variable

$$V[\hat{\beta}] = \begin{pmatrix} V[\hat{\beta}_0] & \mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) & V[\hat{\beta}_1] \end{pmatrix}$$

$\sigma^2_{\hat{\beta}_0}$ $\qquad\qquad$ $\sigma^2_{\hat{\beta}_1}$

We see already that:

$$\sigma^2_{\hat{\beta}_0} = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x}_n)^2}$$

$$\sigma^2_{\hat{\beta}_1} = \sigma^2 \times \frac{1}{\sum (x_i - \bar{x}_n)^2}$$

$$\frac{\underbrace{\frac{\hat{\beta}_0-\beta_0}{\sigma_{\hat{\beta}_0}}}_{\sim \mathcal{N}(0,1)}}{\sqrt{\underbrace{\frac{(n-2)\,\hat{\sigma}_n^2}{\sigma^2}}_{\sim \chi^2(n-2)}\Big/(n-2)}} \sim T(n-2)$$

$$= \frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}_{\hat{\beta}_0}}$$

Construction of the confidence interval for β_j with level $100(1-\alpha)\%$

We want to determine A_n and B_n such that

$$P(\beta_j \in [A_n, B_n]) = 1 - \alpha$$

$\hat{\beta}_j$ is an estimate for β_j

$$\hat{\beta}_j \sim \mathcal{N}(\beta_j, \sigma^2_{\hat{\beta}_j})$$

$$A_n = \hat{\beta}_j - S_n$$

$$B_n = \hat{\beta}_j + P_n$$

We have to determine S_n and P_n such that:

$$P\left(\hat{\beta}_j - \beta_j \in [-P_n, S_n]\right) = 1 - \alpha$$

$$\Rightarrow P\left(\underbrace{\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}}}_{\sim T(n-2)} \in \left[-\frac{P_n}{\hat{\sigma}_{\hat{\beta}_j}}, \frac{S_n}{\hat{\sigma}_{\hat{\beta}_j}}\right]\right) = 1 - \alpha$$

By using the symmetry of the density associated to a student random variable,

we take $S_n = T_n$ and

$$\frac{S_n}{\hat{\sigma}_{\hat{\beta}_j}} = t_{n-2;\, 1-\alpha/2}$$

fractile of order $1-\alpha/2$ for a $T(n-2)$

$$P(\underbrace{T}_{T(n-2)} \leq t_{n-2,\,1-\alpha/2}) = 1-\alpha/2$$

The confidence interval for β_j with level $100(1-\alpha)\%$ is:

$$\left[\hat{\beta}_j - \hat{\sigma}_{\hat{\beta}_j} t_{n-2,\,1-\alpha/2} \,,\, \hat{\beta}_j + \hat{\sigma}_{\hat{\beta}_j} t_{n-2,\,1-\alpha/2}\right]$$

<u>Rk:</u> the main confidence interval is the one for β_1:

Indeed:

if 0 is outside your confidence interval for β_1 then you can accept with a probability equals to $1\alpha(1-\alpha)\%$ that $\beta_1 \neq 0$

↳ We can conclude that there exists an linear model between Y and X.

While, if 0 is inside the confidence interval for β_1, then we do not have a linear connection between Y and X

Those confidence intervals are not the ones we use in practice when we consider simple linear model.

We prefer confidence interval for $\underbrace{E(Y)}_{=\beta_0+\beta_1 X}$

and for Y_{n+1}.

Confidence interval for $E(Y)$ at level $100(1-\alpha)\%$

To make such confidence interval, we have to fix a value for X. We denote it x.

for this value of x, we get:

$$E[Y] = \beta_0 + \beta_1 x$$

We have to find A_n and B_n such that

$$P\left(E[Y] \in [A_n, B_n]\right) = 1 - \alpha$$

An estimate of $\beta_0 + \beta_1 x$ is $\hat{\beta}_0 + \hat{\beta}_1 x$

We denote $\hat{Y}_x = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{\underline{Y}}_x$ is a gaussian random variable.

Indeed, $\hat{Y}_x$ is a linear combination of $\hat{\beta}_0$ and $\hat{\beta}_1$ and we know that

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$ is a gaussian random vector

We obtain :

$$\hat{Y}_x \sim \mathcal{N}\left(\beta_0 + \beta_1 x , \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{X}_n)^2}{\sum (X_i - \bar{X}_n)^2}\right)\right)$$

We take $A_n = \hat{Y}_x - S_n$

$$B_n = \hat{Y}_x + P_n$$

$\Rightarrow$ we just need to determine S_n and P_n such that:

$$P\Big(\underbrace{\hat{Y}_x - \overbrace{(\beta_0 + \beta_1 x)}^{E(\hat{Y}_x)}}_{\text{centered gaussian variable}} \in [-P_n \, ; \, S_n]\Big) = 1 - \alpha$$

$$\mathbb{P}\left(\frac{\hat{Y}_x - (\beta_0 + \beta_1 x)}{\sigma\sqrt{\frac{1}{n} + \frac{(x-\bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2}}} \in \left[-\frac{T_n}{c}, \frac{S_n}{c}\right]\right)$$

$$= 1 - \alpha$$

$$c =$$

Pb: c depends on σ which is unknown

$$\frac{\overbrace{\dfrac{\hat{Y}_x - (\beta_0+\beta_1 x)}{\cancel{\sigma}\sqrt{\dfrac{1}{n}+\dfrac{(x-\bar{X}_n)^2}{\sum (X_i-\bar{X}_n)^2}}}}^{\sim \mathcal{N}(0,1)}}{\sqrt{\underbrace{\cancel{(n-2)}\,\hat{\sigma}_n^2 \big/ \cancel{\sigma^2}}_{\chi^2(n-2)} \Big/ \cancel{(n-2)}}} \sim T(n-2)$$

$$\frac{\hat{Y}_x - (\beta_0 + \beta_1 x)}{\hat{\sigma}_n \sqrt{\frac{1}{n} + \frac{(x - \bar{x}_n)^2}{\sum (x_i - \bar{x}_n)^2}}} \sim T(n-2)$$

We have to find P_n and S_n such that

$$P\left(\underbrace{\frac{\hat{Y}_x - (\beta_0 + \beta_1 x)}{\underbrace{\hat{\sigma}_n \sqrt{\frac{1}{n} + \frac{(x - \bar{X}_n)^2}{\sum (X_i - \bar{X}_n)^2}}}_{= \hat{\tau}_n}}}_{\sim T(n-2)} \in \left[-\frac{P_n}{\hat{\tau}_n} ; \frac{S_n}{\hat{\tau}_n} \right] \right) = 1 - \alpha$$

By using the symmetry of a student density, we choose:

$$P_n = S_n = \hat{C}_n \quad t_{n-2,\,1-\alpha/2}$$

The confidence interval for $E(Y) = \beta_0 + \beta_1 x$ at level $100(1-\alpha)\%$ is:

$$\left[\hat{Y}_x \pm \hat{\sigma}_n \sqrt{\frac{1}{n} + \frac{(x - \bar{X}_n)^2}{\sum (X_i - \bar{X}_n)^2}} \; t_{n-2,\, 1-\alpha/2} \right]$$

Confidence interval for y_{n+1} with level $100(1-\alpha)\%$

Let consider a new point x_{n+1}

y_{n+1} is a realization of $Y_{n+1} = \beta_0 + \beta_1 x_{n+1} + \varepsilon_{n+1}$

We want to construct a confidence interval, it means $[A_n, B_n]$ such that

$$\mathbb{P}\left(Y_{n+1} \in [A_n, B_n] \right) = 1 - \alpha$$

An estimate of Y_{n+1} is $\underbrace{\hat{Y}^{(P)}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}}_{\text{a gaussian variable}}$

We consider $A_n = \hat{y}^{(T)}_{n+1} - S_n$

$$B_n = \hat{y}^{(T)}_{n+1} + T_n$$

with :

$$P\Big(\underbrace{Y_{n+1} - \hat{y}^{(T)}_{n+1}}_{= \hat{\varepsilon}^{(T)}_{n+1}} \in [-S_n, T_n] \Big) = 1 - \alpha$$

$$\hat{\varepsilon}_{n+1}^{(T)} = \underbrace{y_{n+1}}_{\text{gaussian variable that depends on } \varepsilon_{n+1}} - \underbrace{\underbrace{\hat{y}_{n+1}^{(T)}}_{= (1 \;\; x_{n+1}) \underbrace{\hat{\beta}}_{\text{gaussian vector}}}}_{\text{gaussian variable just depends on } \varepsilon_1, \dots, \varepsilon_n}$$

we deduce that

$$\hat{\varepsilon}_{n+1}^{(r)} \sim \mathcal{N}\left(0, \sigma^2\left(1+\frac{1}{n}+\frac{(x_{n+1}-\bar{X}_n)^2}{\sum (X_i-\bar{X}_n)^2}\right)\right)$$

$$\frac{\hat{\varepsilon}_{n+1}^{(r)}}{\hat{\sigma}_n\sqrt{1+\frac{1}{n}+(x_{n+1}-\bar{X}_n)^2/\sum(X_i-\bar{X}_n)^2}} \sim T(n-2)$$

A confidence interval for y_{n+1} at level $100(1-\alpha)\%$ is:

$$\left[\hat{Y}_{n+1}^{(T)} \pm \hat{\sigma}_n \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{X}_n)^2}{\sum (X_i - \bar{X}_n)^2}} \; t_{n-2;\,1-\alpha/2} \right]$$

Confidence region for $\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$ at level $100(1-\alpha)\%$

The scope is to identify a region of $\mathbb{R}^2$ such that

denoted R_n.

$$P\left(\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} \in R_n\right) = 1-\alpha$$

An estimate of $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ is $\hat{\beta}$.

We know that

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 \left({}^t\mathbb{X}\mathbb{X}\right)^{-1}\right)$$

Result:

$$\frac{1}{2\hat{\sigma}_n^2}\,{}^t(\hat{\beta}-\beta)\times({}^t\mathbb{X}\mathbb{X})\times(\hat{\beta}-\beta)$$

$$\sim F(2, n-2)$$

Why? More generally

$$U \sim \mathcal{N}(A, \Sigma)$$

$$^t(U-A)\,\Sigma^{-1}(U-A) \sim \chi^2(p)$$

p depends on the dimension of U

Idea of the construction

We want to identify a region R_n' such that

$$P\left(\hat{\beta}-\beta \in R_n'\right) = 1-\alpha$$

↳ we identify A_n and B_n such that

$$P\left(\underbrace{\frac{1}{\ell \hat{\sigma}_n^2} {}^t(\hat{\beta}-\beta)({}^t\mathbb{X}\mathbb{X})(\hat{\beta}-\beta)}_{} \in [A_n, B_n] \right)$$

$= 1 - \alpha$

$\|\hat{\beta}-\beta\|^2$ with a special metric we take $A_n = 0$

The region of confidence for β is

$$\frac{1}{2\hat{\sigma}_n^2}\left(n(\hat{\beta}_0-\beta_0)^2 + 2n\bar{X}_n(\hat{\beta}_0-\beta_0)(\hat{\beta}_1-\beta_1) + \sum X_i^2(\hat{\beta}_1-\beta_1)^2 \right) \leq f_{(2,n-2)}(1-\alpha)$$