$$\hat{\beta}_\tau = \boxed{\underset{\beta}{\operatorname{argmin}}} \underbrace{(y-X\beta)^T(y-X\beta)+\tau\beta^T\beta}_{\text{Ridge}}$$

First order condition for a minimum is that the gradient of Ridge with respect to β should be equal to zero:

$$\nabla_\beta \text{ Ridge}=0 \Rightarrow -2X^T(y-X\beta)+2\tau\beta=0 \Rightarrow X^Ty=(\underbrace{X^TX+\tau I}_{\substack{\text{is positive definite for}\\ \text{any } \tau>0}})\beta$$

Therefore $X^TX+\tau I$ has full rank and is invertable:

$$\hat{\beta}_\tau=(X^TX+\tau I)^{-1}X^Ty \quad (1)$$

$$\mathbb{E}[\hat{\beta}_\tau]=S_\tau^{-1}S\beta^* = (X^TX+\tau I_D)^{-1}X^T\underbrace{X\beta^*}_{y-\varepsilon}$$

$$S=X^TX \,,\quad S_\tau=X^TX+\tau I_D \,,\quad y=X\beta^*+\varepsilon$$

Writting the ridge as

$$\begin{aligned}
\hat{\beta}_\tau &= (X^TX+\tau I_D)^{-1}X^Ty\\
&= (X^TX+\tau I_D)^{-1}X^T(X\beta^*+\varepsilon)\\
&= (X^TX+\tau I_D)^{-1}X^TX\beta^* + (X^TX+\tau I_D)X^T\varepsilon
\end{aligned}$$

$$\xrightarrow{\mathbb{E}[\hat{\beta}_\tau]}$$

$$\begin{aligned}
&= (X^TX+\tau I_D)^{-1}X^TX\beta^* + (X^TX+\tau I_D)X^T\cancelto{0}{\mathbb{E}[\varepsilon]}\\
&= (X^TX+\tau I_D)^{-1}X^TX\beta^* \quad \therefore \quad (2)
\end{aligned}$$

$$\operatorname{Cov}[\hat{\beta}_\tau]=S_\tau^{-1}SS_\tau^{-1}\sigma^2 = (X^TX+\tau I_D)^{-1}X^TX(X^TX+\tau I_D)^{-1}\sigma^2$$

Conditional variance of β^* $\quad Var[\beta^*|X]=\sigma^2(X^TX)^{-1}$

$$\xrightarrow{Var=\operatorname{Cov}[\hat{\beta}_\tau|X] \;+(2)}$$

$$\begin{aligned}
&= (X^TX+\tau I_D)^{-1}X^TX\, Var[\hat{\beta}_\tau|X]\left[(X^TX+\tau I_D)^{-1}X^TX\right]^T\\
&= (X^TX+\tau I_D)^{-1}X^TX\, Var[\hat{\beta}_\tau|X]\,X^TX(X^TX+\tau I_D)^{-1}\\
&= (X^TX+\tau I_D)^{-1}X^TX\,\sigma^2(X^TX)^{-1}X^TX(X^TX+\tau I_D)^{-1}\\
&= \sigma^2(X^TX+\tau I_D)^{-1}X^TX(X^TX+\tau I_D)^{-1} \quad \therefore
\end{aligned}$$