MLE under $\mathcal{N}(0, \sigma^2 I)$ is OLS.

3.5 Maximum likelihood estimation for β and σ^2 when $\varepsilon_{n\times 1} \sim N(0, \sigma^2 \boldsymbol{I})$

When the error distribution is assumed to be known, namely, $\varepsilon_{n\times 1} \sim N(0, \sigma^2 \boldsymbol{I})$ and σ^2 is unknown. The likelihood function is

$$L(\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\overbrace{(\boldsymbol{Y}-\boldsymbol{X\beta})}^{\vec{\varepsilon}}{}^\top(\boldsymbol{Y}-\boldsymbol{X\beta})}.$$

The log-likelihood function is

$$\begin{aligned}
\log L(\beta, \sigma^2) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{Y}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{X\beta}) \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{Y}^\top\boldsymbol{Y} - \boldsymbol{2Y}^\top\boldsymbol{X\beta} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X\beta});
\end{aligned}$$

score function

$$\begin{cases}
\dfrac{\partial \log L}{\partial \beta} = \dfrac{1}{\sigma^2}(\boldsymbol{X}^\top\boldsymbol{Y} - \boldsymbol{X}^\top\boldsymbol{X\beta}), \\
\dfrac{\partial \log L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}(\boldsymbol{Y}-\boldsymbol{X\beta})^\top(\boldsymbol{Y}-\boldsymbol{X\beta}).
\end{cases}$$

derive it by yourself $\square$

Setting the two equations to zero, we obtain the MLE of β and σ^2:

$$\mathrm{MLE}(\beta) = \tilde{\boldsymbol{\beta}} = \boldsymbol{(X^\top X)^{-1}X^\top Y}.$$

$$\begin{aligned}
\mathrm{MLE}(\sigma^2) = \tilde{\sigma}^2 &= \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})^\top(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) \\
&= \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X(X^\top X)^{-1}X^\top Y})^\top(\boldsymbol{Y} - \boldsymbol{X(X^\top X)^{-1}X^\top Y}) \\
&= \frac{1}{n}\boldsymbol{Y}^\top\boldsymbol{(I - X(X^\top X)^{-1}X^\top)Y} \\
&= \frac{1}{n}\boldsymbol{Y}^\top\boldsymbol{(I-H)Y} \\
&= \frac{1}{n}[\boldsymbol{Y}^\top\boldsymbol{Y} - \tilde{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{Y}].
\end{aligned}$$

It can be seen that when $\varepsilon_{n\times 1} \sim N(0, \sigma^2 \boldsymbol{I})$, the least square estimate is the MLE. In other words, when $\varepsilon_{n\times 1} \sim N(0, \sigma^2 \boldsymbol{I})$, minimizing the least-square objective function is equivalent to maximizing the likelihood function.

3.5.1 Properties of the MLE.

There are a number of properties of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\sigma}^2$.

1. The MLE of $\boldsymbol{\beta}$, $\hat{\tilde{\boldsymbol{\beta}}} = \boldsymbol{(X^\top X)^{-1}X^\top Y}$ follows $N(\beta, (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\sigma^2)$. This is because $\tilde{\boldsymbol{\beta}}$ is linear in $\boldsymbol{Y}$

and

$$
\begin{aligned}
E(\hat{\beta}) &= E[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}] \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top E(\boldsymbol{Y}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta_0} \\
&= \boldsymbol{\beta}_0, \\
Var(\hat{\beta}) &= Var((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \sigma^2.
\end{aligned}
$$

2. The sum of squares of the deviations of the observed $Y_i's$ from their estimated expected values is usually known as the *residual error sum of squares* or *sum of squares due to error*, denoted as SSE. It is given by

$$SSE = (\boldsymbol{Y} - \hat{\boldsymbol{Y}})^\top (\boldsymbol{Y} - \hat{\boldsymbol{Y}}) = \boldsymbol{Y}^\top (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{Y}.$$

Then, $\tilde{\sigma}^2 = \frac{SSE}{n}$ is the MLE. But we have shown that $\boxed{\hat{\sigma}^2 = \frac{SSE}{n - r(\boldsymbol{X})} \text{ is an unbiased estimator of } \sigma^2}$ with $r(\boldsymbol{X})$ = rank of $\boldsymbol{X}$, implying $\tilde{\sigma}^2$ is biased for finite n. However, it is consistent or asymptotically unbiased as $n \to \infty$.

3. The MLE $\hat{\boldsymbol{\beta}}$ and SSE are independent. To show this, recall that $\hat{\boldsymbol{\beta}} = \overset{B}{(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top} \boldsymbol{Y}$ and $SSE = \boldsymbol{Y}^\top \overset{A}{(\boldsymbol{I} - \boldsymbol{H})} \boldsymbol{Y}$ where $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$. Since
Linear form of Y
quadratic form of Y

$$(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{\sigma^2 I})(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{\sigma^2} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0},$$

$B \Sigma A = 0$

by Theorem 2 in section 2.4, $\hat{\boldsymbol{\beta}}$ and SSE are independent and thus $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent.

4. Distribution of $\hat{\sigma}^2$. Consider $\frac{SSE}{\sigma^2} = \frac{1}{\sigma^2} \boldsymbol{Y}^\top \underset{A}{(\boldsymbol{I} - \boldsymbol{H})} \boldsymbol{Y}$ and $\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \underset{\Sigma}{\boldsymbol{I\sigma^2}})$.
Since

$$(\frac{\boldsymbol{I} - \boldsymbol{H}}{\sigma^2})(\boldsymbol{I}\sigma^2) = \boldsymbol{I} - \boldsymbol{H}$$

which is idempotent, then

$A\Sigma$ is idempotent. then $Y^\top A Y \sim \chi^2_{r,\lambda}$

$$\frac{SSE}{\sigma^2} \sim \chi^2_{r(\frac{\boldsymbol{I}-\boldsymbol{H}}{\sigma^2}), \frac{1}{2}(\boldsymbol{X\beta})^\top (\frac{\boldsymbol{I}-\boldsymbol{H}}{\sigma^2})(\boldsymbol{X\beta})}.$$

$r = \text{rank}(A\Sigma)$
$\lambda = \frac{1}{2}\mu^\top A \mu$

However,

$$(\boldsymbol{X\beta})^\top (\frac{\boldsymbol{I} - \boldsymbol{H}}{\sigma^2})(\boldsymbol{X\beta}) = 0.$$

Therefore, the noncentrality parameter is 0. In addition,

$$r(\frac{\boldsymbol{I} - \boldsymbol{H}}{\sigma^2}) = r(\boldsymbol{I} - \boldsymbol{H}) = n - r(\boldsymbol{X}).$$

Consequently,

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n - r(\boldsymbol{X})}, \qquad \text{or}$$

$$\frac{[n - r(\boldsymbol{X})]\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n - r(\boldsymbol{X})}.$$

Key point today MLE for semiparametric model.

parametric β, nonparametric unknown $f_\varepsilon(\cdot)$.

Remark 1. When the error distribution is unknown, that is the density function $f_\varepsilon(\cdot)$ is unknown, the linear model with unknown error distribution is a semiparametric model. In semiparametric linear regression or accelerated failure time models, complications in efficient estimation arise from the multiple roots of the efficient score and density estimation. The maximum likelihood estimation or the semiparametric efficient estimation of linear models were studied by Zeng and Lin (2007, JASA) and Lin and Chen (2013, Biometrika).

Example 1. Consider $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, 2, \cdots, n$, where $\varepsilon_i \sim N(0, \sigma^2)$.

In matrix notation, $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ and $\boldsymbol{X}^\top \boldsymbol{Y} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}$.

$$
\begin{aligned}
(\boldsymbol{X}^\top \boldsymbol{X})^{-1} &= \frac{1}{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\
&= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\hat{\beta} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \\
&= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i \\ n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i \end{pmatrix} \\
&= \begin{pmatrix} \bar{y} - \hat{\beta}_1 \bar{x} \\ \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{pmatrix} \\
&= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}.
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}(\boldsymbol{Y}^\top\boldsymbol{Y} - \hat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{Y}) \\
&= \frac{1}{n-2}[\sum_{i=1}^{n} y_i^2 - \hat{\beta}_0\sum_{i=1}^{n} y_i - \hat{\beta}_1\sum_{i=1}^{n} x_i y_i] \\
&= \frac{1}{n-2}[\sum_{i=1}^{n} y_i^2 - \bar{y}\sum_{i=1}^{n} y_i + \hat{\beta}_1\bar{x}\sum_{i=1}^{n} y_i - \hat{\beta}_1\sum_{i=1}^{n} x_i y_i] \\
&= \frac{1}{n-2}[\sum_{i=1}^{n}(y_i-\bar{y})^2 - \hat{\beta}_1(\sum_{i=1}^{n} x_i y_i - n\sum_{i=1}^{n} y_i\sum_{i=1}^{n} x_i)] \\
&= \frac{1}{n-2}[\sum_{i=1}^{n}(y_i-\bar{y})^2 - \frac{[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}].
\end{aligned}
$$

Example 1(cont'd)

	parameter	UMVU Estimator
1)	β_1	$\hat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
2)	β_0	$\hat{\beta_0} = \bar{y} - \hat{\beta}_1\bar{x}$
3)	σ^2	$\hat{\sigma^2}$
4)	$2\beta_1 - 3\beta_0$	$2\hat{\beta}_1 - 3\hat{\beta}_0$
5)	$5\sigma^2 + 8\beta_1$	$5\hat{\sigma}^2 + 8\hat{\beta}_1$

6) $\beta_0 + 1.94\sigma$

Since

$$
\begin{aligned}
E[\frac{(n-2)\hat{\sigma}^2}{\sigma^2}] &= n-2 \;\text{ where}\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}, \\
\Rightarrow E[\frac{\sqrt{n-2}\hat{\sigma}}{\sigma}] &= \frac{\sqrt{2}\Gamma(\frac{n-2}{2}+\frac{1}{2})}{\Gamma(\frac{n-2}{2})} \\
\Rightarrow E[\frac{\sqrt{n-2}\Gamma(\frac{n-2}{2})}{\sqrt{2}\Gamma(\frac{n-1}{2})}\hat{\sigma}] &= \sigma \\
\Rightarrow \text{The UMVUE is } &\hat{\beta}_0 + 1.94[\frac{\sqrt{n-2}\Gamma(\frac{n-2}{2})}{\sqrt{2}\Gamma(\frac{n-1}{2})}\hat{\sigma}].
\end{aligned}
$$

7) $\frac{\beta_0}{\sigma^2}$

Since $\hat{\beta}_0$ and $\hat{\sigma}^2$ are independent and

$$
\begin{aligned}
E[\frac{\sigma^2}{(n-2)\hat{\sigma}^2}] &= \frac{\Gamma(\frac{n-2}{2}-1)}{\Gamma(\frac{n-2}{2})}2^{-1} = \frac{1}{n-4} \\
\Rightarrow E[\frac{n-4}{(n-2)\hat{\sigma}^2}] &= \frac{1}{\sigma^2} \\
\Rightarrow \text{The UMVUE is } &\hat{\beta}_0(\frac{n-4}{(n-2)\hat{\sigma}^2}).
\end{aligned}
$$

3.5.2 Deviations from Means

The following lemma is useful to find the inverse matrix of a partitioned full rank symmetric matrix.

Lemma 1.

If

Don't spend time memorizing such complex formulas.

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{X}^\top \\ \boldsymbol{Z}^\top \end{bmatrix} [\boldsymbol{X} \quad \boldsymbol{Z}] = \begin{bmatrix} \boldsymbol{X}^\top\boldsymbol{X} & \boldsymbol{X}^\top\boldsymbol{Z} \\ \boldsymbol{Z}^\top\boldsymbol{X} & \boldsymbol{Z}^\top\boldsymbol{Z} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\top & \boldsymbol{D} \end{bmatrix},$$

and put

$$\begin{aligned} \boldsymbol{W} &= (\boldsymbol{D} - \boldsymbol{B}^\top\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ &= [\boldsymbol{Z}^\top\boldsymbol{Z} - \boldsymbol{Z}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Z}]^{-1}, \end{aligned}$$

then,

$$\begin{aligned} \boldsymbol{M}^{-1} &= \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{W}\boldsymbol{B}^\top\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{W} \\ \boldsymbol{W}\boldsymbol{B}^\top\boldsymbol{A}^{-1} & \boldsymbol{W} \end{bmatrix} \leftarrow \text{often used formula.} \\ &= \begin{bmatrix} -\boldsymbol{A}^{-1}\boldsymbol{B} \\ \boldsymbol{I} \end{bmatrix} \boldsymbol{W}[-\boldsymbol{B}^\top\boldsymbol{A}^{-1} \quad \boldsymbol{I}] + \begin{bmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}. \end{aligned}$$

In this section, we still consider the linear model in matrix form

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}\boldsymbol{\beta}_{(k+1)\times 1} + \underbrace{\varepsilon_{n\times 1}} \sim \mathcal{N}(0, \sigma^2 I_n).$$

Recall that the least square estimate $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$ and $var(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\sigma^2$.

Write $\boldsymbol{X} = (\boldsymbol{1}, \quad \boldsymbol{X_1})$ and $\boldsymbol{\beta}^\top = (\beta_0, \boldsymbol{b}^\top)$, where $\boldsymbol{X}_1 = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$ and $\boldsymbol{b}^\top = (\beta_1, \beta_2, \cdots, \beta_k)$.

Let $\bar{\boldsymbol{X}}^\top = (\bar{x}_{.1}, \bar{x}_{.2}, \cdots, \bar{x}_{.k})$ with $\bar{x}_{.i} = (1/n)\sum_{j=1}^n x_{j,i}, i = 1, \ldots, n$.

Note that

$$\begin{aligned} \boldsymbol{1}^\top\boldsymbol{1} &= n \\ \boldsymbol{1}^\top\boldsymbol{Y} &= n\bar{y} \\ \boldsymbol{1}^\top\boldsymbol{X_1} &= n\bar{\boldsymbol{X}}^\top, \end{aligned}$$

where $\mathbf{1}_{n\times 1} = (1, 1, \ldots, 1)^\top$. We then rewrite $\hat{\boldsymbol{\beta}}$ as follows:

$$
\begin{aligned}
\hat{\beta} &= \boldsymbol{(X^\top X)^{-1} X^\top Y} \\
&= \left[\begin{pmatrix} \mathbf{1}^\top \\ \boldsymbol{X}_\mathbf{1}^\top \end{pmatrix} (\mathbf{1} \quad \boldsymbol{X}_\mathbf{1}) \right]^{-1} \begin{pmatrix} \mathbf{1}^\top \\ \boldsymbol{X}_\mathbf{1}^\top \end{pmatrix} \boldsymbol{Y} \\
&= \begin{bmatrix} n & n\bar{\boldsymbol{X}}^\top \\ n\bar{\boldsymbol{X}} & \boldsymbol{X}_\mathbf{1}^\top \boldsymbol{X}_\mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} n\bar{y} \\ \boldsymbol{X}_\mathbf{1}^\top \boldsymbol{Y} \end{bmatrix}.
\end{aligned}
$$

By Lemma 1, we have

$$
\hat{\beta} = \begin{bmatrix} \frac{1}{n} + \bar{\boldsymbol{X}}^\top \boldsymbol{S}^{-\mathbf{1}} \bar{\boldsymbol{X}} & -\bar{\boldsymbol{X}}^\top \boldsymbol{S}^{-\mathbf{1}} \\ -\boldsymbol{S}^{-\mathbf{1}} \bar{\boldsymbol{X}} & \boldsymbol{S}^{-\mathbf{1}} \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \boldsymbol{X}_\mathbf{1}^\top \boldsymbol{Y} \end{bmatrix},
$$

where $\boldsymbol{S} = \boldsymbol{X}_\mathbf{1}^\top \boldsymbol{X}_\mathbf{1} - n\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^\top = \boldsymbol{Z}^\top \boldsymbol{Z}$ and $\underline{\boldsymbol{Z} = \boldsymbol{X}_\mathbf{1} - \mathbf{1}\bar{\boldsymbol{X}}^\top}$. This implies

$$
\begin{aligned}
\Rightarrow \begin{bmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{b}} \end{bmatrix} &= \begin{bmatrix} \bar{y} - \bar{\boldsymbol{X}}^\top \underline{\boldsymbol{S}^{-\mathbf{1}}(\boldsymbol{X}_\mathbf{1}^\top \boldsymbol{Y} - n\bar{y}\bar{\boldsymbol{X}})} \\ \boldsymbol{S}^{-\mathbf{1}}(\boldsymbol{X}_\mathbf{1}^\top \boldsymbol{Y} - n\bar{y}\bar{\boldsymbol{X}}) \end{bmatrix} \hat{b} \\
\Rightarrow \hat{\beta}_0 &= \bar{y} - \bar{\boldsymbol{X}}^\top \hat{\boldsymbol{b}} = \frac{1}{n} \mathbf{1}^\top \boldsymbol{Y} - \bar{\boldsymbol{X}}^\top \hat{\boldsymbol{b}} \\
\hat{\boldsymbol{b}} &= \boldsymbol{S}^{-\mathbf{1}}(\boldsymbol{X}_\mathbf{1}^\top \boldsymbol{Y} - n\bar{y}\bar{\boldsymbol{X}}) \\
&= (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{Y}.
\end{aligned}
$$

Similarly, it follows directly that

$$
\begin{aligned}
var(\hat{\boldsymbol{\beta}}) = var \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{b}} \end{pmatrix} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \sigma^2 \\
&= \begin{bmatrix} \frac{1}{n} + \bar{\boldsymbol{X}}^\top \boldsymbol{(Z^\top Z)^{-1}} \bar{\boldsymbol{X}} & -\bar{\boldsymbol{X}}^\top (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \\ -\boldsymbol{(Z^\top Z)^{-1}} \bar{\boldsymbol{X}} & (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \end{bmatrix} \sigma^2,
\end{aligned}
$$

$$
\Rightarrow var(\hat{\boldsymbol{b}}) = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \sigma^2,
$$

$$
\begin{aligned}
var(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \bar{\boldsymbol{X}}^\top \boldsymbol{(Z^\top Z)^{-1}} \bar{\boldsymbol{X}} \sigma^2 \\
&= \frac{\sigma^2}{n} + \bar{\boldsymbol{X}}^\top var(\hat{\boldsymbol{b}}) \bar{\boldsymbol{X}},
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
& cov(\hat{\beta}_0, \hat{\boldsymbol{b}}^\top) \\
&= -\bar{\boldsymbol{X}}^\top \boldsymbol{(Z^\top Z)^{-1}} \sigma^2 \\
&= -\bar{\boldsymbol{X}}^\top var(\hat{\boldsymbol{b}}).
\end{aligned}
}
$$

intercept & coefficients are not independent! Take care.