

Multiple Linear Model

designed to help answer questions like "is y related to x , after accounting for x_2, x_3, \dots, x_p ?"

response: y_1, y_2, \dots, y_n

covariates: $x_{11}, x_{21}, \dots, x_{n1}$

$x_{12}, x_{22}, \dots, x_{n2}$

$x_{13}, x_{23}, \dots, x_{n3}$

Model for y_i :

$$\varepsilon_i \stackrel{\text{ind}}{\sim} N(0, \sigma^2)$$

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \varepsilon_i \quad b_0, b_1, b_2, b_3, \sigma^2 \text{ unknown numbers}$$

why do we say "after accounting for x_2, x_3, \dots ?"

Suppose covariates for subjects 1 and 2 satisfy

$$x_{21} = x_{11} + u, \quad x_{22} = x_{12} + v, \quad x_{23} = x_{13} + w$$

$$\begin{aligned} E(Y_2 - Y_1) &= E(Y_2) - E(Y_1) & E(a+bX) &= a+bE(X) \\ &= E(b_0 + b_1 x_{21} + b_2 x_{22} + b_3 x_{23} + \varepsilon_2) \\ &\quad - E(b_0 + b_1 x_{11} + b_2 x_{12} + b_3 x_{13} + \varepsilon_1) \nearrow 0 \\ &= b_0 + b_1 x_{21} + b_2 x_{22} + b_3 x_{23} + E(\varepsilon_2) \nearrow 0 \\ &\quad - (b_0 + b_1 x_{11} + b_2 x_{12} + b_3 x_{13} + E(\varepsilon_1)) \\ &= b_1(x_{21} - x_{11}) + b_2(x_{22} - x_{12}) + b_3(x_{23} - x_{13}) \\ &= b_1 u + b_2 v + b_3 w \end{aligned}$$

The total effect is $b_1 u + b_2 v + b_3 w$, but the effect of changing x_1 after accounting for the changes due to x_2 and x_3 is $b_1 u$

The model for y_1, \dots, y_n can be written together in matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & x_{n3} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$x_{10} = 1$$

$$\underline{Y} = \underline{X} \underline{b} + \underline{\epsilon}$$

In general, we have p covariates and model y_i as

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} + \epsilon_i, \quad \epsilon_i \stackrel{\text{ind}}{\sim} N(0, \sigma^2)$$

write together in matrix form as

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{10} & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

given the data \underline{y} and \underline{X} , we estimate \underline{b} by minimizing squared residuals

$$RSS(\underline{b}^*) = \sum_{i=1}^n \left(y_i - \sum_{j=0}^p b_j^* x_{ij} \right)^2$$

$$\frac{\partial RSS(\underline{b}^*)}{\partial b_k^*} = -2 \sum_{i=1}^n \left(y_i - \sum_{j=0}^p b_j^* x_{ij} \right) x_{ik} \quad \text{set each derivative equal to 0}$$

$$\begin{aligned} \sum_{i=1}^n y_i x_{ik} &= \sum_{i=1}^n \sum_{j=0}^p \hat{b}_j x_{ij} x_{ik} \\ &= \sum_{j=0}^p \hat{b}_j \sum_{i=1}^n x_{ij} x_{ik} \end{aligned}$$

this is one equation.
We have $p+1$ of them total.

Normal Equations \rightarrow

$$\sum_{i=1}^n y_i x_{i0} = \sum_{j=0}^p \hat{b}_j \sum_{i=1}^n x_{ij} x_{i0}$$
$$\vdots$$
$$\sum_{i=1}^n y_i x_{ip} = \sum_{j=0}^p \hat{b}_j \sum_{i=1}^n x_{ij} x_{ip}$$

exercise: verify that these can be written in matrix form as

$$X^T \underline{y} = X^T X \hat{\underline{b}}$$

if $\text{rank}(X) = p+1$ (full column rank), then $(X^T X)^{-1}$ exists, and

$$(X^T X)^{-1} X^T \underline{y} = (X^T X)^{-1} (X^T X) \hat{\underline{b}} \rightarrow \hat{\underline{b}} = (X^T X)^{-1} X^T \underline{y}$$

$$\hat{\underline{b}} = (X^T X)^{-1} X^T \underline{y} \quad \text{The most important equation in all of statistics}$$

fitted values: $\hat{\underline{y}} = X \hat{\underline{b}} = X (X^T X)^{-1} X^T \underline{y} = P \underline{y}$.

residuals: $\hat{\underline{e}} = \underline{y} - X \hat{\underline{b}} = (I - X(X^T X)^{-1} X^T) \underline{y} = (I - P) \underline{y}$

P and I - P are very special matrices called projection matrices

Projections :

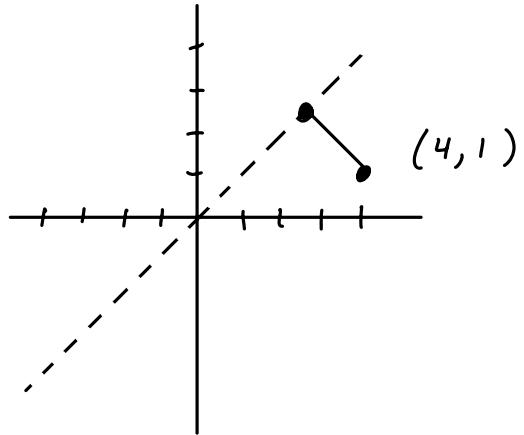
A projection is a special linear transformation. (i.e $y \rightarrow Ay$)

Example :

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \quad y = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

$$Py = \begin{bmatrix} 2 + 1/2 \\ 2 + 1/2 \end{bmatrix} = \begin{bmatrix} 5/2 \\ 5/2 \end{bmatrix}$$

What does this projection "do"?



Moves the point $(4, 1)$ onto the line described by $a(1, 1)$

- $(5/2, 5/2) = 5/2(1, 1)$.
- We say that $5/2$ is the coordinates of $(5/2, 5/2)$ in the linear space spanned by $(1, 1)$

In fact, the projection finds the closest point on $a(1, 1)$ to $(4, 1)$.

This works for any input vector y (try it!)

In general, an $n \times n$ projection matrix P maps vectors to the closest point in the linear subspace spanned by the columns of P .

Projection matrices must satisfy 2 properties

1. symmetric: $P = P^T$

2. idempotent: $PP = P$

Example :

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}, \quad P^T = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad PP = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} + \frac{1}{4} & \frac{1}{4} + \frac{1}{4} \\ \frac{1}{4} + \frac{1}{4} & \frac{1}{4} + \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

projections are the secret sauce of linear regression

Example : $P = X(X^T X)^{-1} X^T$ ✓

symmetric: $P^T = X(X^T X)^{-1} X^T = P$

↑
 $X^T X$ is symmetric, therefore so is inverse

idempotent: $PP = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = P$ ✓

Some properties of projections

① if P is a projection, so is $I - P$

symmetric: $I - P = I - P^T = I - P$ ✓

idempotent: $(I - P)(I - P) = I - P - P + PP = I - P - P + P = I - P$ ✓

Consequence: if $P = X(X^T X)^{-1} X^T y$

$$(I - P)y = y - Py = y - X\hat{b} = e$$

* The residuals are a projection onto $I - P$

$$e^T \hat{y} = \hat{y}^T (I - P)Py = \hat{y}^T (P - PP)y = \hat{y}^T (P - P)y = 0$$

* residuals are orthogonal to fitted values

② if you have a matrix M with k linearly independent columns

($k < n$, $M = [\underline{m}_1 \ \underline{m}_2 \ \dots \ \underline{m}_k]$) then $M(M^T M)^{-1} M^T$ is a projection

onto the space spanned by $\underline{m}_1, \dots, \underline{m}_k$ (onto the columns of M)

$$\underbrace{M(M^T M)^{-1} M^T}_{n \times k \quad k \times 1} y = M \underline{d} = [\underline{m}_1 \ \dots \ \underline{m}_k] \begin{bmatrix} d_1 \\ \vdots \\ d_k \end{bmatrix} = d_1 \underline{m}_1 + d_2 \underline{m}_2 + \dots + d_k \underline{m}_k$$

✓

Consequence: $n \times 1$

if $P = X(X^T X)^{-1} X^T$ then $\hat{y} = Py = X \hat{b} = \hat{b}_0 \underline{x}_0 + \hat{b}_1 \underline{x}_1 + \cdots + \hat{b}_p \underline{x}_p$

* regression coefficients are the coordinates of the fitted values
with respect to the basis $\underline{x}_0, \dots, \underline{x}_p$

Data (X, y) $\xrightarrow{\text{project}}$ \hat{y} $\xrightarrow{\text{get coordinates}}$ \hat{b}

③ An $n \times n$ projection matrix P with rank k can be decomposed as

$P = \underset{n \times n}{U} \underset{n \times k}{U} \underset{k \times n}{U^T}$ where the columns of U ($\underline{u}_1, \dots, \underline{u}_k$)
are an orthonormal basis for the columns of P

• $\underline{u}_i^T \underline{u}_i = 1$ and $\underline{u}_i^T \underline{u}_j = 0$ for $i \neq j$

Properties of linear transformations of random vectors

We want to know sampling dist of $\hat{\underline{B}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y} = \underline{M} \underline{Y}$

Need to know about distribution of random vectors ($\hat{\underline{B}}$ and \underline{Y})
and distributions of linear functions of random vectors ($\underline{M} \underline{Y}$)

Expectation

$$\underline{z} = (z_1, \dots, z_n)^T \quad \text{column vector}$$

$$E(\underline{z}) := (\bar{E}(z_1), \dots, \bar{E}(z_n))^T \quad \begin{matrix} \text{definition of expectation} \\ \text{of random vector} \end{matrix}$$

$$\begin{aligned} E\left(\underset{k \times 1}{a} + \underset{k \times n}{M} \underset{n \times 1}{z}\right) &= E\left[\begin{array}{c} a_1 + \sum_{j=1}^n m_{1j} z_j \\ \vdots \\ a_k + \sum_{j=1}^n m_{kj} z_j \end{array}\right] = \left[\begin{array}{c} a_1 + E\left(\sum_{j=1}^n m_{1j} z_j\right) \\ \vdots \\ a_k + E\left(\sum_{j=1}^n m_{kj} z_j\right) \end{array}\right] \\ &= \left[\begin{array}{c} a_1 + \sum_{j=1}^n m_{1j} E(z_j) \\ \vdots \\ a_k + \sum_{j=1}^n m_{kj} E(z_j) \end{array}\right] = \underset{k \times 1}{a} + \underbrace{\underset{k \times n}{M} E(\underline{z})}_{k \times 1} \quad \begin{matrix} \text{follows from the linearity} \\ \text{property of expectation} \end{matrix} \end{aligned}$$

Covariance

The covariance matrix of a vector \underline{z} is defined as

$$\underset{n \times n}{\text{Cov}}(\underline{z}) = \begin{bmatrix} \text{Cov}(z_1, z_1) & \text{Cov}(z_1, z_2) & \cdots & \text{Cov}(z_1, z_n) \\ \text{Cov}(z_2, z_1) & \text{Cov}(z_2, z_2) & \cdots & \text{Cov}(z_2, z_n) \\ \vdots & & & \\ \text{Cov}(z_n, z_1) & \text{Cov}(z_n, z_2) & \cdots & \text{Cov}(z_n, z_n) \end{bmatrix}$$

Covariance matrices are symmetric and can also be written as

$$\underset{n \times n}{\text{Cov}}(\underline{z}) = E \left[(\underline{z} - E(\underline{z})) (\underline{z} - E(\underline{z}))^T \right]$$

$E[\text{matrix}]$ means take E of each entry (as in $E(\text{vector})$)

$$\begin{aligned} \text{Cov}\left(\underset{k \times 1}{\underline{a}} + \underset{k \times n}{M} \underline{z}\right) &= E\left(\left(\underset{k \times 1}{\underline{a}} + M\underline{z} - \underset{n \times 1}{\underline{a}} - M\underline{E}(\underline{z})\right)\left(\underset{n \times 1}{\underline{a}} + M\underline{z} - \underset{n \times 1}{\underline{a}} - M\underline{E}(\underline{z})\right)^T\right) \\ &= E\left((M\underline{z} - M\underline{E}(\underline{z})) (M\underline{z} - M\underline{E}(\underline{z}))^T\right) \\ &= E \left\{ \left[M(\underline{z} - \underline{E}(\underline{z}))\right] \left[M(\underline{z} - \underline{E}(\underline{z}))\right]^T \right\} \\ &= E \left\{ M(\underline{z} - \underline{E}(\underline{z})) (\underline{z} - \underline{E}(\underline{z}))^T M^T \right\} \\ &= M E \left[(\underline{z} - \underline{E}(\underline{z})) (\underline{z} - \underline{E}(\underline{z}))^T \right] M^T \\ &= M \text{Cov}(\underline{z}) M^T \end{aligned}$$

Multivariate normal distribution

If \underline{z} has multivariate normal distribution, we use notation

$$\underline{z} \sim N(\underline{\mu}, \Sigma)$$

$n \times 1$

where $\underline{\mu} = E(\underline{z})$ and $\Sigma = \text{Cov}(\underline{z})$

\underline{z} has probability density

$$p(\underline{z}) = \frac{1}{(2\pi)^n/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(\underline{z}-\underline{\mu})^T \Sigma^{-1} (\underline{z}-\underline{\mu})\right)$$

$$\det(\Sigma) = \text{determinant of } \Sigma \quad \exp(a) = e^a$$

Example: $E(\underline{z}) = \underline{\mu} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}$ $\text{Cov}(\underline{z}) = \begin{bmatrix} \sigma^2 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & \sigma^2 \end{bmatrix} = \sigma^2 I$

$$\begin{aligned} p(\underline{z}) &= \frac{1}{(2\pi)^n/2} \det(\sigma^2 I)^{-1/2} \exp\left(-\frac{1}{2}(\underline{z}-\underline{\mu})^T (\sigma^2 I)^{-1} (\underline{z}-\underline{\mu})\right) \\ &= \frac{1}{(2\pi)^n/2} \left((\sigma^2)^n\right)^{-1/2} \exp\left(-\frac{1}{2}(\underline{z}-\underline{\mu})^T \left(\frac{1}{\sigma^2} I\right) (\underline{z}-\underline{\mu})\right) \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2}(\underline{z}-\underline{\mu})^T (\underline{z}-\underline{\mu})\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \mu_i)^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i - \mu_i)^2}{2\sigma^2}\right) = \prod_{i=1}^n p(z_i) \end{aligned}$$

Therefore $z_1, \dots, z_n \stackrel{\text{ind}}{\sim} N(\mu, \sigma^2)$

Properties of multivariate normal

- completely determined by mean vector and covariance matrix
- if $\underline{Z} \sim N(\underline{\mu}, \Sigma)$, then $\underset{p \times 1}{\underline{a}} + M \underset{p \times n}{\underline{Z}}$ is multivariate normal

$$\underset{p \times 1}{\underline{a}} + M \underset{p \times n}{\underline{Z}} \sim N\left(\underset{p \times 1}{\underline{a}} + M \underset{n \times 1}{\underline{\mu}}, M \Sigma M^T\right)$$

exercise: verify this by using transformation
of variables

Example: distribution of average:

$$\text{Let } \underset{n \times 1}{\underline{1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\underline{Y} \sim N(b_0 \underline{1}, \sigma^2 \mathbf{I}) \longrightarrow Y_1, \dots, Y_n \stackrel{\text{ind}}{\sim} N(b_0, \sigma^2)$$

$$\bar{Y} = \left(\frac{1}{n} \underline{1}^T\right) \underline{Y} = \frac{1}{n} \sum_{i=1}^n \underline{1} Y_i = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$E(\bar{Y}) = E\left(\frac{1}{n} \underline{1}^T \underline{Y}\right) = \frac{1}{n} \underline{1}^T E(\underline{Y}) = \frac{b_0}{n} \underline{1}^T \underline{1} = \frac{b_0}{n} n = b_0$$

$$\text{cov}(\bar{Y}) = \left(\frac{1}{n} \underline{1}^T\right) (\sigma^2 \mathbf{I}) \left(\frac{1}{n} \underline{1}\right) = \frac{\sigma^2}{n^2} (\underline{1}^T \mathbf{I} \underline{1})$$

$$= \frac{\sigma^2}{n^2} (\underline{1}^T \underline{1}) = \frac{\sigma^2}{n^2} \cdot n = \frac{\sigma^2}{n}$$

$$\Rightarrow \bar{Y} \sim N\left(b_0, \frac{\sigma^2}{n}\right)$$

Properties of multiple linear regression estimators

$$\underline{Y} = \underline{X} \underline{b} + \underline{\varepsilon}, \quad \underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T \quad \varepsilon_i \stackrel{\text{ind}}{\sim} N(0, \sigma^2)$$

$$\underline{\varepsilon} \sim N(0, \sigma^2 I) \quad \text{multivariate normal}$$

$$\underline{Y} = \underline{X} \underline{b} + \underline{\varepsilon} \quad \text{is multivariate normal}$$

why?

$$E(\underline{Y}) = \underline{X} \underline{b}, \quad \text{Cov}(\underline{Y}) = \sigma^2 I$$

$$\underline{Y} \sim N(\underline{X} \underline{b}, \sigma^2 I)$$

$$\hat{\underline{B}} = (X^T X)^{-1} X^T \underline{Y} = [(X^T X)^{-1} X^T] \underline{Y}$$

$$\Rightarrow \hat{\underline{B}} \text{ is MVN}$$

$$E(\hat{\underline{B}}) = [(X^T X)^{-1} X^T] \underline{X} \underline{b} = (X^T X)^{-1} (X^T X) \underline{b} = I \underline{b} = \underline{b}$$

$$\begin{aligned} \text{Cov}(\hat{\underline{B}}) &= [(X^T X)^{-1} X^T] (\sigma^2 I) [X (X^T X)^{-1}] \\ &= \sigma^2 (X^T X)^{-1} (X^T X) (X^T X)^{-1} = \sigma^2 (X^T X)^{-1} \end{aligned}$$

$$\hat{\underline{B}} \sim N(\underline{b}, \sigma^2 (X^T X)^{-1})$$

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

$$\hat{\underline{e}} = \underline{Y} - \hat{\underline{Y}} = (\mathbf{I} - P) \underline{Y}$$

$$\hat{\underline{Y}} \sim N\left(\underline{X}\underline{b}, \sigma^2 P\right)$$

$$\hat{\underline{e}} \sim N\left(\underline{0}, \sigma^2 (\mathbf{I} - P)\right)$$

what is $\text{Cov}(\hat{\underline{Y}}, \hat{\underline{e}}) = E((\hat{\underline{Y}} - \underline{X}\underline{b})\hat{\underline{e}}^T)$?

$$= E(\hat{\underline{Y}}\hat{\underline{e}}^T - (\underline{X}\underline{b})\hat{\underline{e}}^T)$$

$$= E(P(\underline{Y} - \underline{X}\underline{b})(\underline{Y} - \underline{X}\underline{b})^T(\mathbf{I} - P)) \quad P = \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T$$

$$= P E[(\underline{Y} - \underline{X}\underline{b})(\underline{Y} - \underline{X}\underline{b})^T](\mathbf{I} - P)$$

$$= P(\sigma^2 \mathbf{I})(\mathbf{I} - P) = \sigma^2 (P \mathbf{I} - P^2) = \sigma^2 (P - P) = \underline{0}$$

residuals are uncorrelated with the fitted values

Estimating the error variance σ^2

$$\hat{e} \sim N(0, \sigma^2(I-P))$$

This is a good candidate because its distribution depends on σ^2 and no other parameters

We can write \hat{e} as

$$\hat{e} = \sigma(I-P)\bar{z}, \text{ where } \bar{z} \sim N(0, I_n) \text{ why?}$$

$$E(\hat{e}) = 0 \quad \checkmark$$

$$\text{Cov}(\hat{e}) = \sigma(I-P)I(I-P)^T\sigma = \sigma^2(I-P)(I-P) = \sigma^2(I-P) \quad \checkmark$$

Recall: $(I-P) = UU^T$, $U = [u_1 \dots u_k]$ u_i orthonormal

$k = \text{rank}(I-P) = n-(p+1)$ in this case.

$$\text{Consider: } \hat{e}^T \hat{e} = \sigma^2 \bar{z}^T U U^T U U^T \bar{z} = \sigma^2 \bar{z}^T U U^T \bar{z} \quad \text{ok}$$

$$V = U^T \bar{z} \sim N(0, U^T I U) = N(0, U^T U) = N(0, I_{n-p-1})$$

$$\text{Therefore } \hat{e}^T \hat{e} = \sigma^2 V^T V = \sigma^2 \sum_{i=1}^{n-p-1} V_i^2$$

A sum of $n-p-1$ squared standard normals is χ^2_{n-p-1}

$$\hat{e}^T \hat{e} = \sigma^2 W \text{ where } W \sim \chi^2_{n-p-1}$$

$$E(\hat{e}^T \hat{e}) = \sigma^2(n-p-1) \Rightarrow E\left(\frac{\hat{e}^T \hat{e}}{n-p-1}\right) = \sigma^2$$

This is why we use $\hat{\sigma}^2 = \frac{\hat{e}^T \hat{e}}{n-p-1} = \frac{1}{n-p-1} \sum_{i=1}^n \hat{e}_i^2$

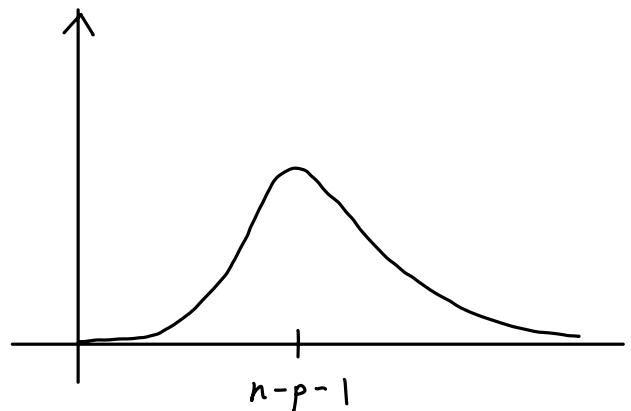
if $\hat{\sigma}^2 = \frac{\hat{e}^T \hat{e}}{n-p-1}$, what is its distribution?

Just need to re-arrange terms

$$\hat{e}^T \hat{e} \sim \sigma^2 \chi_{n-p-1}^2$$

$$\frac{\hat{e}^T \hat{e}}{n-p-1} \sim \frac{\sigma^2}{n-p-1} \chi_{n-p-1}^2$$

$$\frac{\hat{\sigma}^2}{\sigma^2} (n-p-1) \sim \chi_{n-p-1}^2$$



Can use this for hypothesis testing

$$H_0: \sigma^2 = 12.3. \text{ Your estimate is } \hat{\sigma}^2 = 9.2$$

Compare $\frac{9.2}{12.3} (n-p-1)$ to a χ_{n-p-1}^2

What about \hat{b} ?

There's a theorem that says if you define R.V. T as

$$T = \frac{Z}{\sqrt{W/k}}$$

where $Z \sim N(0,1)$ and $W \perp\!\!\!\perp Z$ (independent)
 and $W \sim \chi_k^2$

Then $T \sim t_k$ (t distribution with k degrees of freedom)

Let's see if we can use this to our advantage given what we have

$$\hat{B}_j \sim N(b_j, \sigma^2 (X^T X)_{j+1, j+1}^{-1})$$

$$\text{Var}(\hat{B}_j) = \sigma^2 (X^T X)_{j+1, j+1}^{-1}$$

$$\hat{\text{Var}}(\hat{B}_j) = \hat{\sigma}^2 (X^T X)_{j+1, j+1}^{-1} \quad \hat{\sigma}^2 \text{ is P.V. version here}$$

Hypothesis: $b_j = 7.2$. Let's look at the following.

$$\frac{\hat{B}_j - 7.2}{\sqrt{\hat{\text{Var}}(\hat{B}_j)}} \rightarrow \text{looks like a t statistic. Let's verify}$$

Rewrite as

$$\frac{(\hat{B}_j - 7.2) / \sqrt{\text{Var}(\hat{B}_j)}}{\sqrt{\hat{\text{Var}}(\hat{B}_j) / \text{Var}(\hat{B}_j)}}$$

$$\frac{z}{\sqrt{\hat{\sigma}^2 (X^T X)_{j+1, j+1}^{-1} / \sigma^2 (X^T X)_{j+1, j+1}^{-1}}}$$

$$\frac{z}{\sqrt{\hat{\sigma}^2 / \sigma^2}} = \frac{z}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2} (n-p-1) / (n-p-1)}}$$

$$= \frac{z}{\sqrt{W/(n-p-1)}} \quad \text{this is a } N(0,1) \text{ divided by the square root of a } \chi^2 \text{ divided by its degrees of freedom}$$

Independence:

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

$$\hat{\underline{\epsilon}} = (\mathbf{I} - \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T) \underline{Y}$$

Both are MVN. In fact $\begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{\epsilon}} \end{bmatrix} \sim MVN$

what is $Cov(\hat{\underline{\beta}}, \hat{\underline{\epsilon}})$?

$$\begin{aligned} Cov(\hat{\underline{\beta}}, \hat{\underline{\epsilon}}) &= Cov((\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y}, (\mathbf{I} - \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T) \underline{Y}) \\ &= (\underline{X}^T \underline{X})^{-1} \underline{X}^T Cov(\underline{Y})(\mathbf{I} - \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T) \\ &= \sigma^2 (\underline{X}^T \underline{X})^{-1} \underline{X}^T (\mathbf{I} - \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T) \\ &= \sigma^2 ((\underline{X}^T \underline{X})^{-1} \underline{X}^T - (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{X} (\underline{X}^T \underline{X})^{-1} \underline{X}^T) \\ &= \sigma^2 ((\underline{X}^T \underline{X})^{-1} \underline{X}^T - (\underline{X}^T \underline{X})^{-1} \underline{X}^T) \\ &= 0 \end{aligned}$$

Because $\begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{\epsilon}} \end{bmatrix} \sim MVN$, $Cov = 0 \Rightarrow$ independent

\bar{z} is a function of $\hat{\underline{\beta}}$, but not $\hat{\underline{\epsilon}}$

W is a function of $\hat{\underline{\beta}}$, which depends on $\hat{\underline{\epsilon}}$ but not $\hat{\underline{\beta}}$

Therefore \bar{z} is independent of W , and

Therefore,

$$\frac{\hat{\beta}_j - b_j^*}{\sqrt{\hat{\text{Var}}(\hat{\beta}_j)}} \quad \text{is } t_{n-p-1} \quad \text{when } H_0: b_j = \hat{b}_j$$

$$\text{same as } \frac{\hat{\beta}_j - b_j^*}{SE(\hat{\beta}_j)} \sim t_{n-p-1}$$