Multiple Linear Model

designed to help answer questions like "is y related to x_1 after accounting for $x_2, x_3, \ldots, x_p$?"

response: $y_1, y_2, \ldots, y_n$

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

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

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

Model for y_i

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

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \varepsilon_i$$

$b_0, b_1, b_2, b_3, \sigma^2$ unknown numbers

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

Suppose covariates for subjects 1 and 2 satisfy

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

$$E(a + bX) = a + bE(X)$$

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

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 $y_1, \ldots, y_n$ can be written together in matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{10} & x_{11} & x_{12} & x_{13} \\ x_{20} & x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n0} & x_{n1} & x_{n2} & x_{n3} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$x_{i0} = 1$

$$\underset{n \times 1}{\underline{Y}} = \underset{n \times 4}{X} \; \underset{4 \times 1}{\underline{b}} + \underset{n \times 1}{\underline{\varepsilon}}$$

In general, we have p covariates and model y_i as

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + x_{ip} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

write together in matrix form as

$$\underset{n \times 1}{\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}} = \underset{n \times (p+1)}{\begin{pmatrix} x_{10} & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & \cdots & x_{np} \end{pmatrix}} \underset{(p+1) \times 1}{\begin{bmatrix} b_0 \\ \vdots \\ b_p \end{bmatrix}} + \underset{n \times 1}{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

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

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

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

set each derivative equal to 0

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

$$= \sum_{j=0}^{p} \hat{b}_j \sum_{i=1}^{n} x_{ij} x_{ik}$$

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

Normal Equations $\longrightarrow$

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

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

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

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

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

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

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

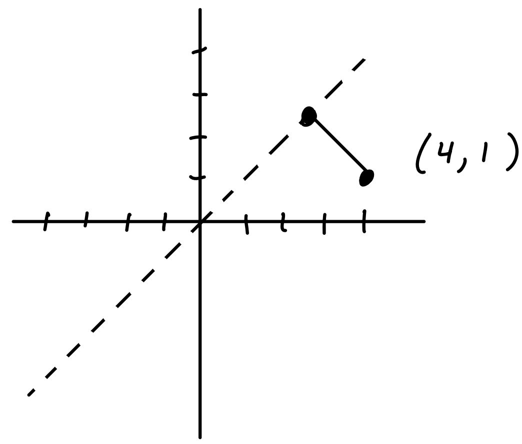
residuals: $\hat{e} = \underline{y} - X\hat{b} = (I - X(X^T X)^{-1} X^T) \underline{y} = (I - P) \underline{y}$

P and $I - P$ are very special matrices called <u>projection matrices</u>

Projections:

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

Example:



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

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

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 closet 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:

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

projections are the secret sauce of linear regression

Example: $P = X(X^TX)^{-1}X^T$ ✓

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

X^TX is symmetric, therefore so is inverse

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

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 $P = X(X^TX)^{-1}X^T\underline{y}$

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

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

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

* residuals are orthogonal to fitted values

(2) if you have a matrix M with k linearly independent columns $(k < n,\ M = [\underline{m}_1\ \underline{m}_2 \cdots \underline{m}_k])$ then $M(M^TM)^{-1}M^T$ is a projection onto the space spanned by $\underline{m}_1, \ldots, \underline{m}_k$ (onto the columns of M)

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

✓

Consequence:

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

* regression coefficents are the coordinates of the fitted values

  with respect to the basis $\underline{x}_0, \ldots, \underline{x}_p$

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

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

$\underset{n\times n}{P} = \underset{n\times k}{U}\,\underset{k\times n}{U^T}$ where the columns of U $(\underline{u}_1, \ldots, \underline{u}_k)$

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

Need to know about distribution of random vectors ($\hat{\underline{\beta}}$ and $\underline{Y}$)

and distributions of linear functions of random vectors ($M\underline{Y}$)

Expectation

$\underline{Z} = (Z_1, \ldots, Z_n)^T$ column vector

$E(\underline{Z}) := (E(Z_1), \ldots, E(Z_n))^T$ definition of expectation of random vector

$$E(\underset{k\times 1}{a} + \underset{k\times n}{M}\,\underset{n\times 1}{\underline{Z}}) = E\begin{bmatrix} a_1 + \sum_{j=1}^{n} m_{1j} Z_j \\ \vdots \\ a_k + \sum_{j=1}^{n} m_{kj} Z_j \end{bmatrix} = \begin{bmatrix} a_1 + E\left(\sum_{j=1}^{n} m_{1j} Z_j\right) \\ \vdots \\ a_k + E\left(\sum_{j=1}^{n} m_{kj} Z_j\right) \end{bmatrix}$$

$$= \begin{bmatrix} a_1 + \sum_{j=1}^{n} m_{1j} E(Z_j) \\ \vdots \\ a_k + \sum_{j=1}^{n} m_{kj} E(Z_j) \end{bmatrix} = \underset{k\times 1}{\underline{a}} + \underbrace{\underset{k\times n}{M}\, \underset{n\times 1}{E(\underline{Z})}}_{k\times 1}$$

follows from the linearity property of expectation

Covariance

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

$$\underset{n\times n}{Cov(\underline{Z})} = \begin{bmatrix} Cov(Z_1,Z_1) & Cov(Z_1,Z_2) & \cdots & Cov(Z_1,Z_n) \\ Cov(Z_2,Z_1) & Cov(Z_2,Z_2) & \cdots & Cov(Z_2,Z_n) \\ \vdots & & & \\ Cov(Z_n,Z_1) & Cov(Z_n,Z_2) & \cdots & Cov(Z_n,Z_n) \end{bmatrix}$$

Covariance matrices are symmetric and can also be written as

$$\underset{n\times n}{Cov(\underline{Z})} = E\left[\underset{n\times 1}{(\underline{Z} - E(\underline{Z}))} \underset{1\times n}{(\underline{Z} - E(\underline{Z}))^T} \right]$$

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

$$Cov(\underset{k\times 1}{\underline{a}} + \underset{k\times n}{M}\underset{n\times 1}{Z}) = E\left(\left(\underline{a} + MZ - \underline{a} - ME(Z)\right)\left(\underline{a} + MZ - \underline{a} - ME(Z)\right)^T\right)$$

$$= E\left(\left(MZ - ME(Z)\right)\left(MZ - ME(Z)\right)^T\right)$$

$$= E\left\{\left[M\left(Z - E(Z)\right)\right]\left[M\left(Z - E(Z)\right)\right]^T\right\}$$

$$= E\left\{M\left(Z - E(Z)\right)\left(Z - E(Z)\right)^T M^T\right\}$$

$$= M\,E\left[\left(Z - E(Z)\right)\left(Z - E(Z)\right)^T\right]M^T$$

$$= M\,Cov(\underline{Z})\,M^T$$

Multivariate normal distribution

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

$$\underset{n\times 1}{\underline{z}} \sim N(\underline{\mu}, \Sigma)$$

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

$\underline{z}$ has probability density

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

$\det(\Sigma)$ = determinant of Σ $\qquad \exp(a) = e^a$

<u>Example</u>: $E(\underline{z}) = \underline{\mu} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}$ $\quad Cov(\underline{z}) = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma^2 & \\ 0 & & \ddots \;\; \sigma^2 \end{bmatrix} = \sigma^2 I$

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

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

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

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (z_i - \mu_i)^2\right)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z_i-\mu_i)^2}{2\sigma^2}\right) = \prod_{i=1}^{n} p(z_i)$$

Therefore $z_1, \ldots, z_n \overset{ind}{\sim} N(\mu, \sigma^2)$

Properties of multivariate normal

- completely determined by mean vector and covariance matrix

- if $\underset{\sim}{Z} \sim N(\underline{\mu}, \Sigma)$, then $\underset{p\times 1}{\underline{a}} + \underset{p\times n}{M} \underset{n\times 1}{\underset{\sim}{Z}}$ is multivariate normal

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

exercise: verify this by using transformation of variables

<u>Example</u>: distribution of average:

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

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

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

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

$$Cov(\bar{Y}) = \left(\frac{1}{n}\mathbb{1}^T\right)(\sigma^2 I)\left(\frac{1}{n}\mathbb{1}\right) = \frac{\sigma^2}{n^2}\left(\mathbb{1}^T I\, \mathbb{1}\right)$$

$$= \frac{\sigma^2}{n^2}\left(\mathbb{1}^T\mathbb{1}\right) = \frac{\sigma^2}{n^2}\cdot n = \frac{\sigma^2}{n}$$

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

Properties of multiple linear regression estimators

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

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

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

why?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\boxed{\text{residuals are uncorrelated with the fitted values}}$$

Estimating the error variance σ^2

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

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

We can write $\hat{e}$ as

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

$E(\hat{e}) = 0$ ✓

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

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

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

Consider: $\hat{e}^T\hat{e} = \sigma^2 Z^T UU^TUU^T Z = \sigma^2 Z^T UU^T Z$ ok

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

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}

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

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

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

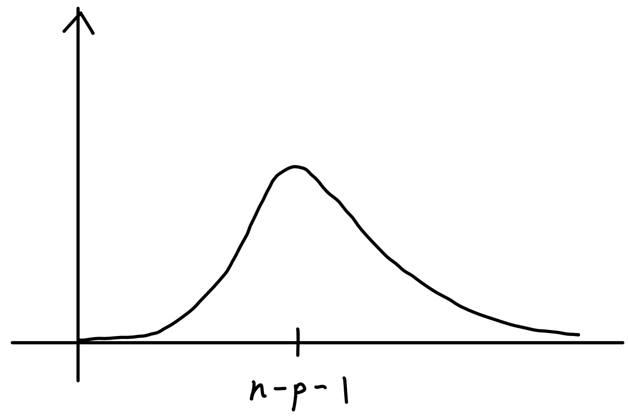
if $\hat{\sigma}^2 = \dfrac{\underline{\hat{e}}^T \underline{e}}{n-p-1}$, what is its distribution?

Just need to re-arrange terms

$\underline{\hat{e}}^T \underline{\hat{e}} \sim \sigma^2 \chi^2_{n-p-1}$

$\dfrac{\underline{\hat{e}}^T \underline{\hat{e}}}{n-p-1} \sim \dfrac{\sigma^2}{n-p-1} \chi^2_{n-p-1}$

$\dfrac{\hat{\sigma}^2}{\sigma^2}(n-p-1) \sim \chi^2_{n-p-1}$



Can use this for hypothesis testing

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

Compare $\dfrac{9.2}{12.3}(n-p-1)$ to a χ^2_{n-p-1}

<u>What about $\hat{\underline{b}}$?</u>

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

$$T = \frac{Z}{\sqrt{W/k}} \quad \text{where } Z \sim N(0,1) \text{ and } W \sim \chi^2_k$$

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

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

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

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

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

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

Rewrite as

$$\frac{(\hat{B}_j - 7.2)/\sqrt{Var(\hat{B}_j)}}{\sqrt{\hat{Var}(\hat{B}_j)/Var(\hat{B}_j)}}$$ top part is $N(0,1)$

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

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

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

Therefore,

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

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