Topic 2:
Miscellaneous Topics

Outline

- Joint Estimation of β_0 and β_1
- Multiple Testing/Simultaneous CI
- Regression Through the Origin
- Measurement Error
- Inverse Predictions

Joint Inference of β_0 and β_1

- Confidence intervals are used for a single parameter
- Confidence region for two or more parameters

$$P\left((\beta_0, \beta_1) \in S \subset R^2\right) = 100(1-\alpha)\%$$

- The region for (β_0, β_1) defines a set of lines, form a band about the estimated regression line (Lecture 4)

$$\{(x, y): y = \beta_0 + \beta_1 x, (\beta_0, \beta_1) \in S\}$$

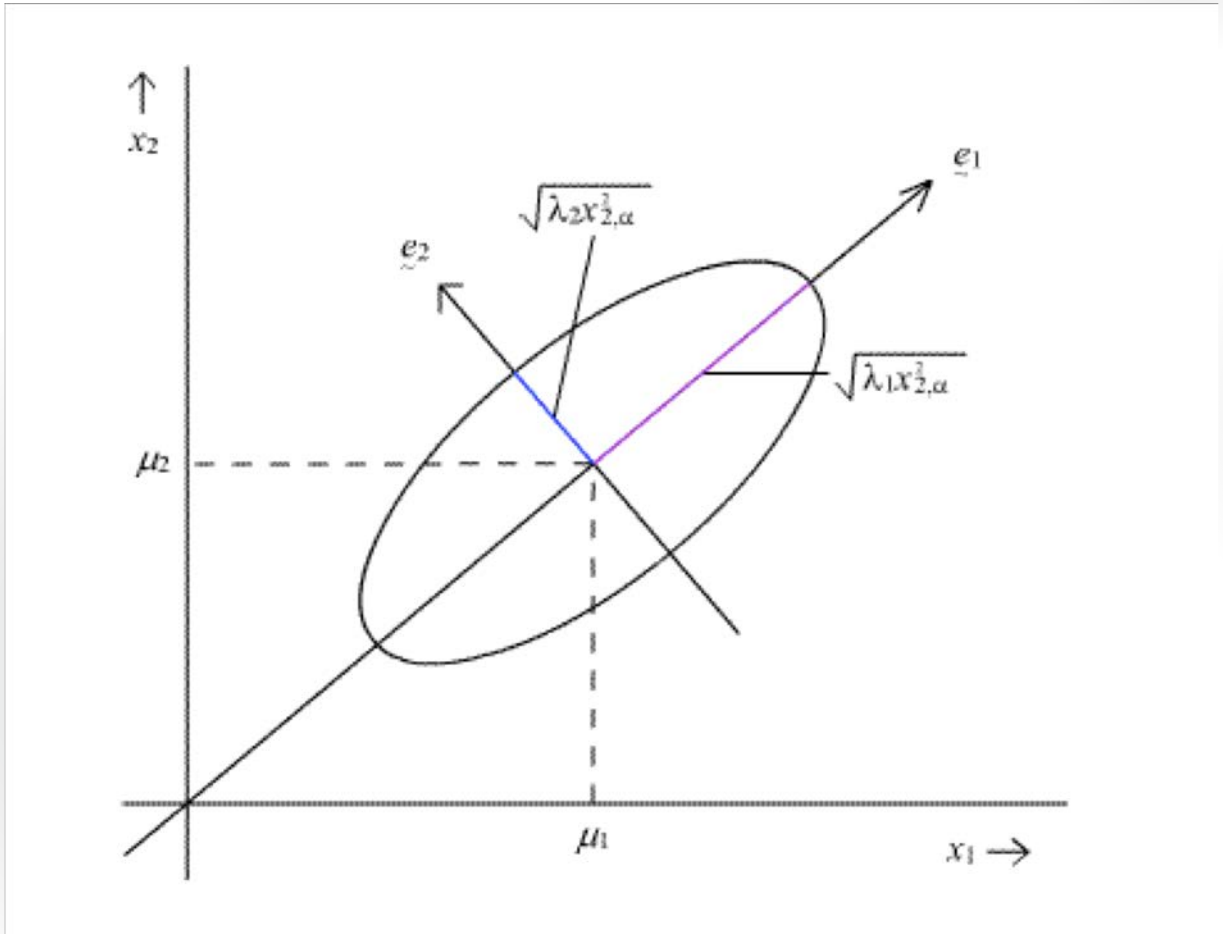
Joint Inference of β_0 and β_1

- Since $b_0(\hat{\beta}_0)$ and $b_1(\hat{\beta}_1)$ are jointly Normal,

$$(b_0, b_1)' \sim N\left((\beta_0, \beta_1)', \sigma^2 \Sigma_{2\times 2}\right)$$

the *natural* (i.e., smallest) confidence region is an ellipse

- Textbook considers rectangles (KNNL 4.1) (i.e., region formed from the product or union of two separate confidence intervals)
- Need to adjust confidence level of each CI so that the region has proper 1-α level



x_2
e_1
e_2
$\sqrt{\lambda_2 \chi^2_{2,\alpha}}$
$\sqrt{\lambda_1 \chi^2_{2,\alpha}}$
μ_2
μ_1
x_1

Bonferroni Inequality and Correction

- Individual CIs:

$$P(\beta_0 \in CI_0) = 1 - \alpha, \qquad P(\beta_1 \in CI_1) = 1 - \alpha$$

- Joint confidence region:

$$S = CI_0 \times CI_1 = \{(\beta_0, \beta_1): \beta_0 \in CI_0 \ and \ \beta_1 \in CI_1\}$$

- Confidence level of

$$P\big((\beta_0, \beta_1) \in S\big) = P(\beta_0 \in CI_0 \ and \ \beta_1 \in CI_1) = 1 - \alpha?$$

- $P(\beta_0 \in CI_0 \ and \ \beta_1 \in CI_1) = 1 - P(\beta_0 \notin CI_0 \ or \ \beta_1 \notin CI_1)$

$$\geq 1 - [P(\beta_0 \notin CI_0) + P(\beta_1 \notin CI_1)] = 1 - 2\alpha$$

- In order to achieve confidence level at least α_0, set $\alpha=\alpha_0/2$, so

$$P(\beta_0 \in CI_0) = 1 - \frac{\alpha_0}{2}, P(\beta_1 \in CI_1) = 1 - \frac{\alpha_0}{2}$$

$$P(\beta_0 \in CI_0 \ and \ \beta_1 \in CI_1) \geq 1 - \alpha_0$$

Bonferroni Correction

- Recall, individually,

$$CI_0{}': b_0 \pm t_{1-\frac{\alpha}{2}, n-2} s(b_0) \ for \ \beta_0$$

$$CI_1{}': b_1 \pm t_{1-\frac{\alpha}{2}, n-2} s(b_1) \ for \ \beta_1$$

- Jointly, for β_0 and β_1,

$$CI_0: b_0 \pm t_{1-\frac{\alpha}{2\times 2}, n-2} s(b_0) \ for \ \beta_0$$

$$CI_1: b_1 \pm t_{1-\frac{\alpha}{2\times 2}, n-2} s(b_1) \ for \ \beta_1$$

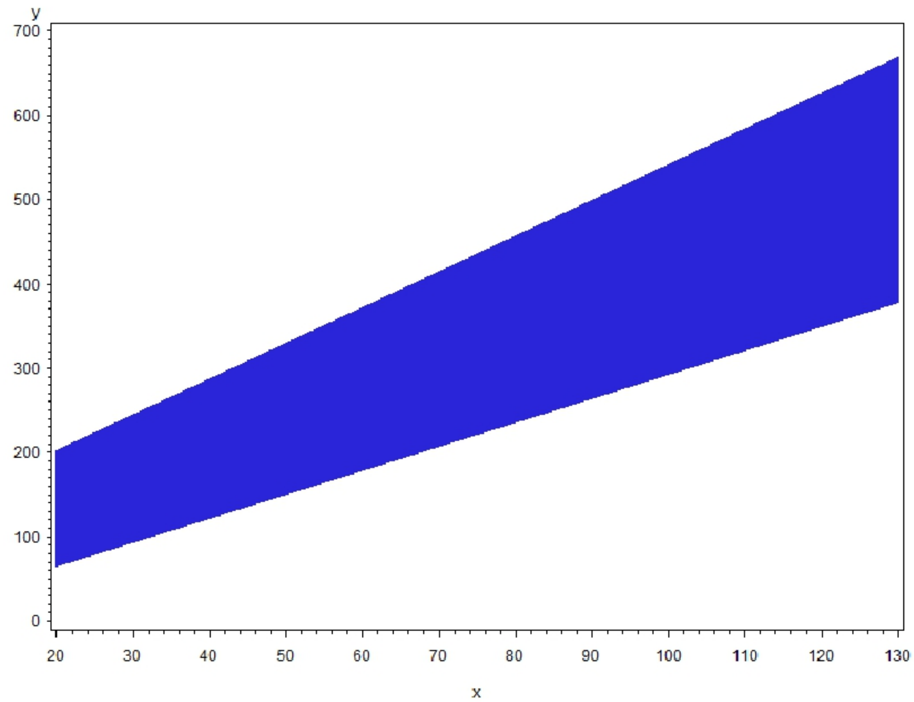
- Confidence Region for (β_0, β_1) with level at least 1-α

$$S = CI_0 \times CI_1$$

Joint Estimation of β_0 and β_1

- For Toluca example, 90% rectangular region is
  - $8.20 \leq \beta_0 \leq 116.5$
  - $2.85 \leq \beta_1 \leq 4.29$
- Region shown right…all lines when X positive between
  - $Y = 116.5 + 4.29X$
  - $Y = 8.2 + 2.85X$

Definitely not as small nor symmetric about mean X as the confidence band



Mean Response CIs

- Simultaneous estimation of μ_h at <u>*all*</u> X_h, uses Working-Hotelling (KNNL 2.6)

$$\hat{\mu}_h \pm Ws(\hat{\mu}_h)$$

where $W^2 = 2F_{\alpha,2,n-2}$

- For simultaneous estimation at <u>*few*</u> X_h, use Bonferroni. Let g = number of X_h. Then

$$\hat{\mu}_h \pm Bs(\hat{\mu}_h)$$

where $B = t_{\frac{\alpha}{2g},n-2}$

  - Use this when $B < W$, implying narrower CIs

Simultaneous Prediction Intervals

- Simultaneous prediction at *few* X_h, use Bonferroni method:

$$\hat{Y}_h \pm Bs(pred)$$

where $B = t_{\frac{\alpha}{2g}, n-2}$

- Scheffé's method

$$\hat{Y}_h \pm Ss(pred)$$

where $S^2 = gF_{\alpha, g, n-2}$

- Again choose one with narrower intervals

Regression through the Origin

- $Y_i = \beta_1 X_i + \varepsilon_i$, that is, assume β_0=0
- lm(Y ~ 0 + X) in R
- Generally *not* a good idea because of model misspecification and chance variation
- Might be forcing model to behave certain way in area with no data
- Problems with residuals, R^2 and other statistics
- See cautions, KNNL p 164

Measurement Error

- For $Y_i^* = Y_i + \tau_i$, where τ_i is extra m-error:

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i + \tau_i$$

- Not a big problem…only variance=σ_ε^2+σ_τ^2
- For X_i , X_i^*= X_i + δ_i where δ_i is m-error,

$$Y_i = \beta_0 + \beta_1 X_i^* + (\varepsilon_i - \beta_1 \delta_i) = \beta_0 + \beta_1 X_i^* + \varepsilon_i^*$$

- Because X_i^* and ε_i^* are correlated, the usual LS estimator of β is biased, and additional information or data and method are needed; see KNNL 4.5, pp165-158.
- The Berkson model: a special case where X_i^* is fixed whereas the true quantity is unknown and random, the usual LS based inference remains fine

Note:

- Proof for
  $$\beta_1^* \le \beta_1$$
  on pp 167
- Key point: observed
  $$X_i^* = X_i + \delta_i$$
  here unobservable true X_i is a r.v.

$$Y_i = \beta_0 + \beta_1 X_i^* + (\varepsilon_i - \beta_1 \delta_i)$$
$$Cov(X_i^*, \varepsilon_i - \beta_1 \delta_i) = -\beta_1 \sigma_\delta^2$$

$$Cov(Y_i, X_i^*)$$
$$= Cov(\beta_1 X_i^* + (\varepsilon_i - \beta_1 \delta_i), X_i^*)$$
$$= \beta_1 \sigma_{X^*}^2 - \beta_1 \sigma_\delta^2$$
$$E(Y|X^*) = \beta_0^* + \beta_1^* X^*$$

Since the Pearson coefficient
$$r_{Y,X^*} = \beta_1^* \frac{\sigma_{X^*}}{\sigma_Y}$$

By definition,
$$r_{Y,X^*} = \frac{Cov(Y, X^*)}{\sigma_Y \sigma_{X^*}}$$
$$= \frac{\beta_1 \sigma_{X^*}^2 - \beta_1 \sigma_\delta^2}{\sigma_Y \sigma_{X^*}}$$
$$\therefore \beta_1(\sigma_{X^*}^2 - \sigma_\delta^2) = \beta_1^* \sigma_{X^*}^2$$
$$\beta_1^* = \frac{\sigma_{X^*}^2 - \sigma_\delta^2}{\sigma_{X^*}^2} \beta_1 \le \beta_1$$

or:
$$\beta_1^* = \frac{\sum (X_i^* - \bar{X}^*) Y_i}{\sum (X_i^* - \bar{X}^*)^2}$$
$$= \beta_1 + \frac{\sum (X_i^* - \bar{X}^*)(\varepsilon_i - \beta_1 \delta_i)}{\sum (X_i^* - \bar{X}^*)^2}$$
$$= \beta_1 + \frac{Cov(X^*, \varepsilon - \beta_1 \delta)}{Var(X^*)}$$
$$= \beta_1 - \frac{\beta_1 \sigma_\delta^2}{\sigma_X^2 + \sigma_\delta^2}$$
$$= \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\delta^2} \beta_1 \le \beta_1$$

Inverse Predictions/Calibration

- The same model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- The fitted regression function:

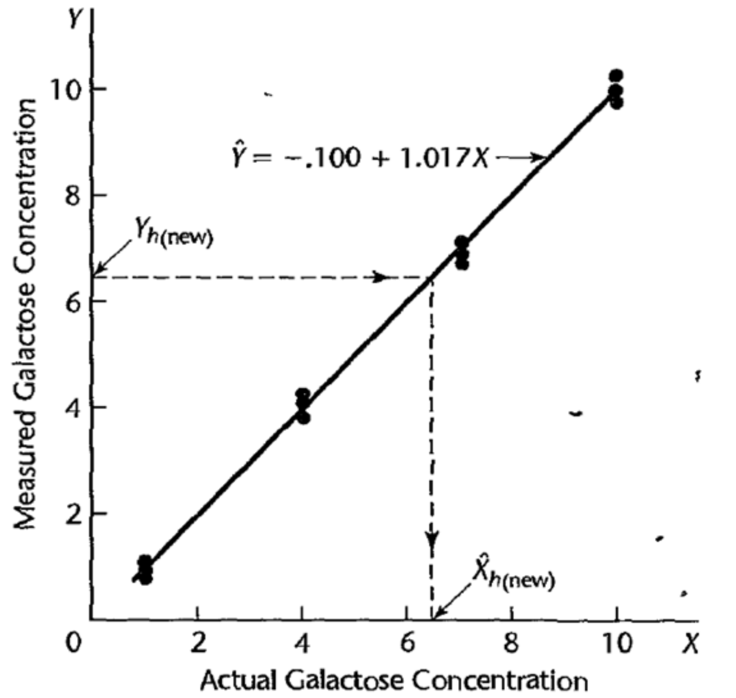
$$Y = b_0 + b_1 X$$

- Instead of predicting Y_h, predict the corresponding $X_h, \hat{X}_h$, given Y_h
- Solve the fitted equation for X_h

$$\hat{X}_h = \frac{Y_h - b_0}{b_1}, \text{ where } b_1 \neq 0$$

- Approximate CI can be obtained, see KNNL p169

- Technical applications
  - validation of new instruments
  - assessment of sample "unknowns" against a set of standard values



Background Reading

- Next class we will do simple regression with vectors and matrices so that we can generalize to multiple regression
- Scan through KNNL 5.1 to 5.7 if this is unfamiliar to you

清华大学统计学辅修课程

Linear Regression Analysis

Lecture 7-
Matrix Approach to Linear Regression & Multiple Linear Regression

周在莹

清华大学统计学研究中心

http://www.stat.tsinghua.edu.cn

Topic 1:

Matrix Approach to Linear Regression

Outline

- Linear Regression in Matrix Form
- Simple Linear Regression in another Perspective

The Model in Scalar Form

- $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
  - The ε_i's are independent Normally distributed random variables with mean 0 and variance σ^2
- Consider writing out the observations:

$$Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1$$

$$Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2$$

$$\vdots$$

$$Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n$$

The Model in Matrix Form

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_3 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

The Model in Matrix Form

$$Y = X\beta + \varepsilon$$

- Vector of responses

$$\mathbf{Y}_{n\times 1} = (Y_1, Y_2, \ldots, Y_n)^t$$

- Design Matrix

$$\mathbf{X}_{n\times 2} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$$

- Vector of parameters (coefficients)

$$\beta_{2\times 1} = (\beta_0, \beta_1)^t$$

- Vector of error terms

$$\epsilon_{n\times 1} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^t$$

Variance-Covariance Matrix

$$\sigma^2(\mathbf{Y}) = \Sigma_{\mathbf{Y}} = \begin{pmatrix} \sigma^2(Y_1) & \sigma(Y_1, Y_2) & \dots & \sigma(Y_1, Y_n) \\ \sigma(Y_2, Y_1) & \sigma^2(Y_2) & \dots & \sigma(Y_2, Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma(Y_n, Y_1) & \sigma(Y_n, Y_2) & \dots & \sigma^2(Y_n) \end{pmatrix}$$

- Diagonal entries are variances, and off-diagonal entries are covariances
- When $Y_1, Y_2, \ldots, Y_n$ are independent, the covariances are equal to zero

Covariance Matrix of ϵ

$$\sigma^2(\epsilon)_{n\times n} = \mathrm{Cov}\begin{pmatrix}\epsilon_1\\ \epsilon_2\\ \vdots\\ \epsilon_n\end{pmatrix} = \sigma^2 I_{n\times n}$$

where $I_{n\times n}$ is the $n\times n$ identity matrix with diagonal 1 and off-diagonal 0

- Because the error terms are independent, the covariance between any two error terms is zero
- The error terms have common variance, therefore, the diagonal are equal to σ^2

Distributional Model

- Covariance Matrix of ***Y***
- The covariance matrices of ***Y*** and ε are the same, because the design matrix ***X*** is fixed

$$\sigma^2(\mathbf{Y})_{n\times n} = \mathrm{Cov}\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \sigma^2 I_{n\times n}$$

- The distributional Model:

$$Y \sim N(X\beta, \sigma^2 I_n)$$

Least Squares Estimation

- Analytic approach:

$$Q(\beta) = \|Y - X\beta\|^2 = (Y - X\beta)^t(Y - X\beta)$$

$$\hat{\beta} = \text{argmin}_{\beta \in R^2} Q(\beta)$$

$$\frac{\partial Q}{\partial \beta} = -2X^t(Y - X\beta) = 0$$ **(Normal Equation)**

$$\frac{\partial^2 Q}{\partial \beta^2} = 2X^t X$$

- Under the condition rank(X) = 2, the minimizer of $Q(\beta)$ exists and is unique:

$$\boxed{b = \hat{\beta} = (X'X)^{-1}X'Y}$$

Least Squares Estimation

- Vector of predicted/fitted responses

$$\hat{Y} = Xb = X(X'X)^{-1}X'Y$$

$$\hat{Y} = HY$$

- Hat matrix (Projection matrix)

$$H = X(X'X)^{-1}X'$$

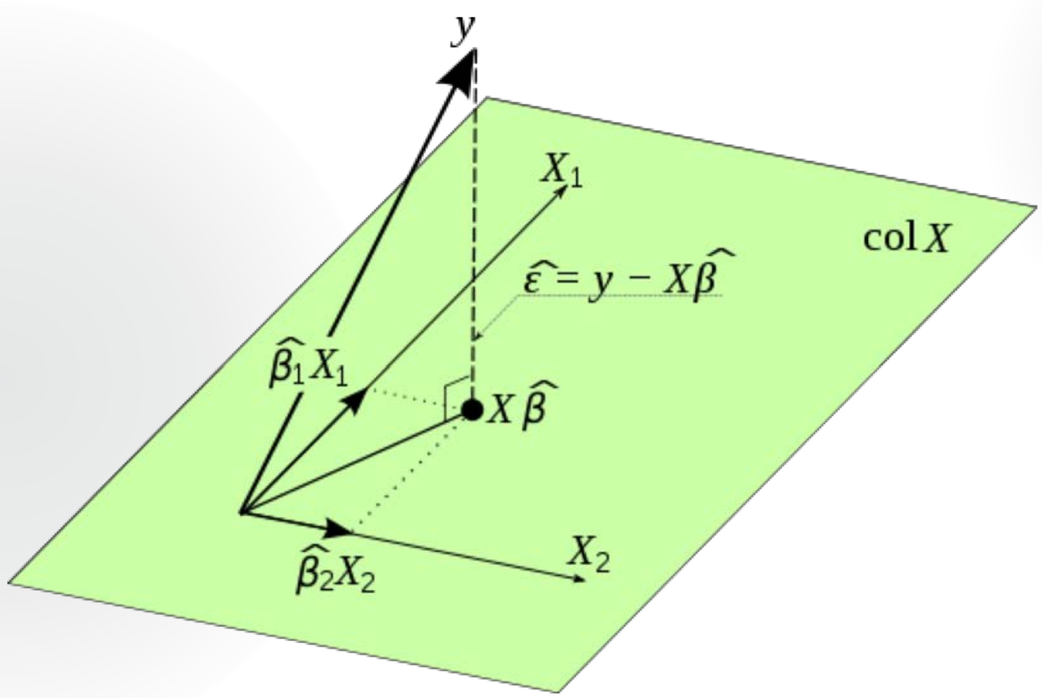
- Vector of residuals

$$e = Y - \hat{Y} = Y - HY = (I - H)Y$$

Projective Geometry Approach for LSE

- Column space:

  col(X) = linear space spanned by column vectors of X



Projective Geometry Approach for LSE

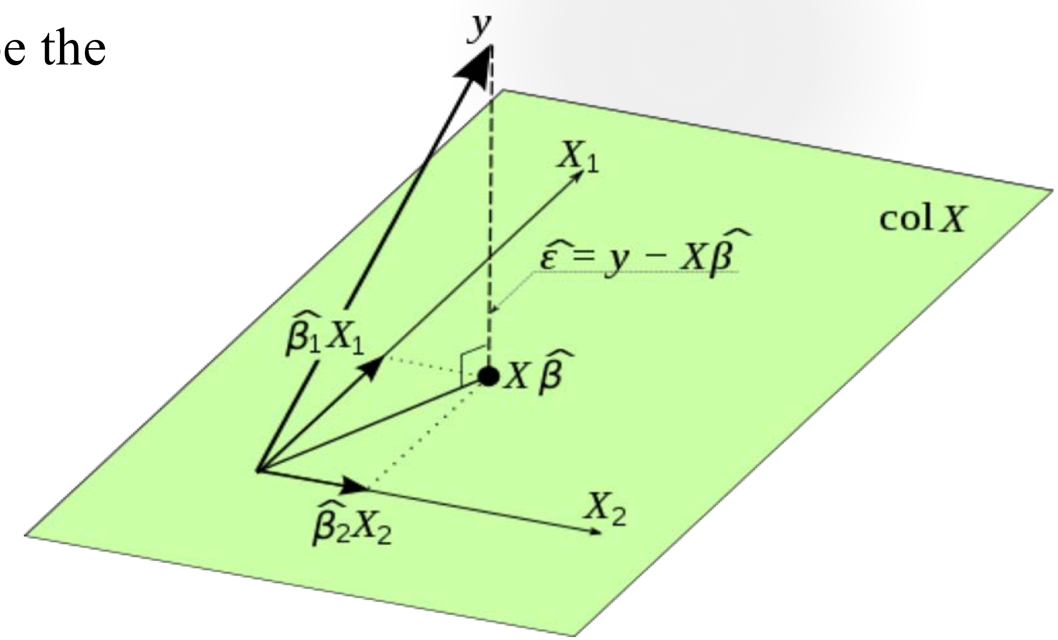
- Reformulate the problem

$$\min_{\beta\in R^2}\|Y-X\beta\|^2=\min_{y=X\beta}\|Y-y\|^2$$
$$=\min_{y\in\mathrm{col}(X)}\|Y-y\|^2$$

- The solution: the minimizer must be the orthogonal projection of Y onto the column space of X:

$$\hat{Y} = PY$$

where P is the orthogonal projection matrix/operator (Pythagoras)



Projection Matrix and LS Estimates

- It can be shown that P is symmetric and idempotent ($P^2=P$)
- It turns out that ?

$$P = H = X(X'X)^{-1}X' \quad ?$$

- From

$$PY = X(X'X)^{-1}X'Y = Xb$$

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

- Predicted responses:

$$\hat{Y} = HY$$

- Residuals:

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

- Residuals and predicted responses are orthogonal:

$$e'\hat{Y} = 0$$

Note: Orthogonal Projection

- Idempotent matrix has complementary range(image) and kernel(null space)

$$P^2 = P \Rightarrow R^n = Ker(P) \oplus Im(P)$$

- P is an orthogonal projection $\Rightarrow$ $Ker(P)$ and $Im(P)$ are orthogonal

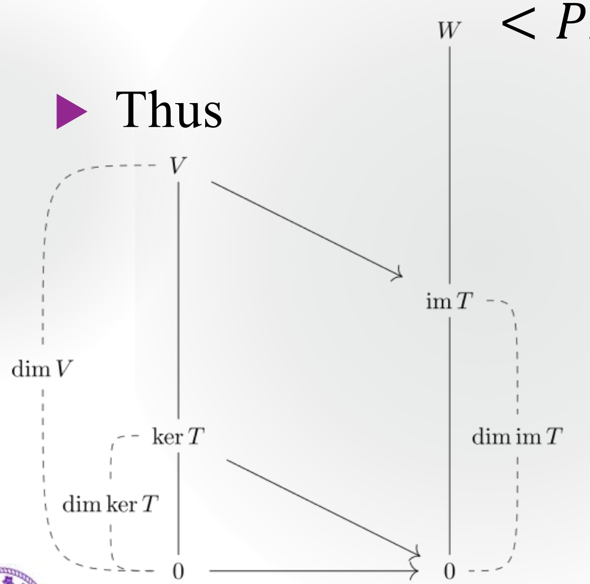
- $\forall x, y \in R^n$,

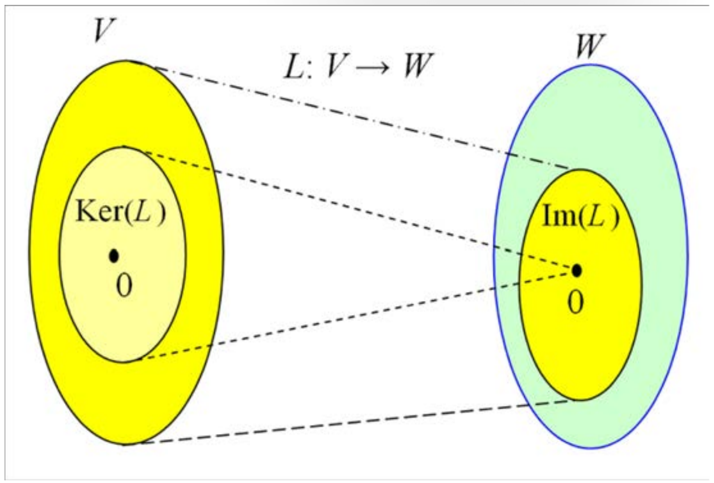
$$< Px, y - Py > = < x - Px, Py > = 0$$

- Thus

$$< Px, y > = < Px, Py >$$
$$=< x, Py > =< P^* x, y >$$

$$\underline{P = P^*}$$





Kernel and image of a map L

Sampling Distribution of b

- Theorem: Suppose $U_{m\times 1} \sim N(\mu, \Sigma)$. Let $V_{d\times 1} = c + D_{d\times m}U$. Then,

$$V \sim N(c + D\mu, D\Sigma D')$$

- Recall $Y \sim N(X\beta, \sigma^2 I_n)$, and $b = (X'X)^{-1}X'Y$. Applying the theorem,

$$b \sim N(\beta, \sigma^2(X'X)^{-1})$$

- The estimated covariance matrix of b is

$$s^2(b) = s^2(X'X)^{-1}$$

where

$$s^2 = \frac{e'e}{n-2} = \frac{Y'(I-H)Y}{n-2}$$

Background Reading

- We will use this framework to do multiple regression → we have more than one explanatory variable
- Adding another explanatory variable is to add another column in the design matrix X
- See Chapter 6

Topic 2:

Multiple Linear Regression

Outline

- Multiple Regression
  - Data and notation
  - Model
  - Inference
- Recall notes from simple linear regression regarding differences

Data for Multiple Regression

- Cases still denoted by $i = 1$ to n

$$\{(Y_i\,;\, X_{i,1}, X_{i,2}\,,\ldots, X_{i,p\text{-}1})\}_{1 \le i \le n}$$

  - Y_i = response variable for the i^{th} case
  - $X_{i,1}, X_{i,2}, \ldots, X_{i,\,p\text{-}1}$ are the p-1 explanatory (or predictor) variables for the i^{th} case

Multiple Regression Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- β_0 is the intercept
- $\beta_1, \beta_2, \ldots, \beta_{p\text{-}1}$ are the regression coefficients for the explanatory variables
- Notice switch from slope to regression coefficient. More on this soon
- X_{ik} is the value of the k^{th} explanatory variable for the i^{th} case
- ε_i's are independent Normally distributed random errors with mean 0 and variance σ^2

Multiple Regression Parameters

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- β_0, the intercept
- $\beta_1, \beta_2, \ldots, \beta_{p-1}$, the regression coefficients for the explanatory variables
- σ^2, the variance of the error term

Some Special Cases

- Polynomial Model of order p-1

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- Some explanatory variables X can be indicator or dummy variables taking values of 0 and 1 or other distinct numbers
- Interactions between explanatory variables can be represented as products of X's and included in the model (crossed terms)

An Example: Children's Weight Growth

- The response variable is Weight in metric of standard units (Y), and two explanatory variables Age in Months X_1 and Gender X_2
- Consider the following model:
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$
where X_2 is a dummy variable, that is, $X_2 = 0$ if the child is a girl; and $X_2 = 1$ if the child is a boy
- If Child i is a girl:
$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i, \ (X_{i2} = 0)$$
- If Child i is a boy:
$$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_{i1} + \varepsilon_i, \ (X_{i2} = 1)$$
- One model represents two different regression lines of Y versus X_1 for girls and boys, respectively

Model in Matrix Form

- Response vector: $Y = (Y_1, Y_2, \cdots, Y_n)'$
- Design matrix:

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{pmatrix}$$

- Regression coefficients: $\beta = (\beta_0, \beta_1, \cdots, \beta_{p-1})'$
- Error terms: $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)'$
- The model:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$$

$$\varepsilon \sim N(0, \sigma^2 I_n),\ Y \sim N(X\beta, \sigma^2 I_n)$$

Least Squares Estimation

- Analytic approach:

$$Q(\beta) = \|Y - X\beta\|^2 = (Y - X\beta)^t(Y - X\beta)$$

$$\hat{\beta} = \text{argmin}_{\beta \in R^p} Q(\beta)$$

$$\frac{\partial Q}{\partial \beta} = -2X^t(Y - X\beta) = 0 \qquad \textbf{(Normal Equation)}$$

$$\frac{\partial^2 Q}{\partial \beta^2} = 2X^tX$$

- Under the condition rank(X) = p, the minimizer of $Q(\beta)$ exists and is unique:

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

Least Squares Estimation

- Vector of predicted/fitted responses

$$\hat{Y} = Xb = X(X'X)^{-1}X'Y$$

$$\hat{Y} = HY$$

- Hat matrix (Projection matrix)

$$H = X(X'X)^{-1}X'$$

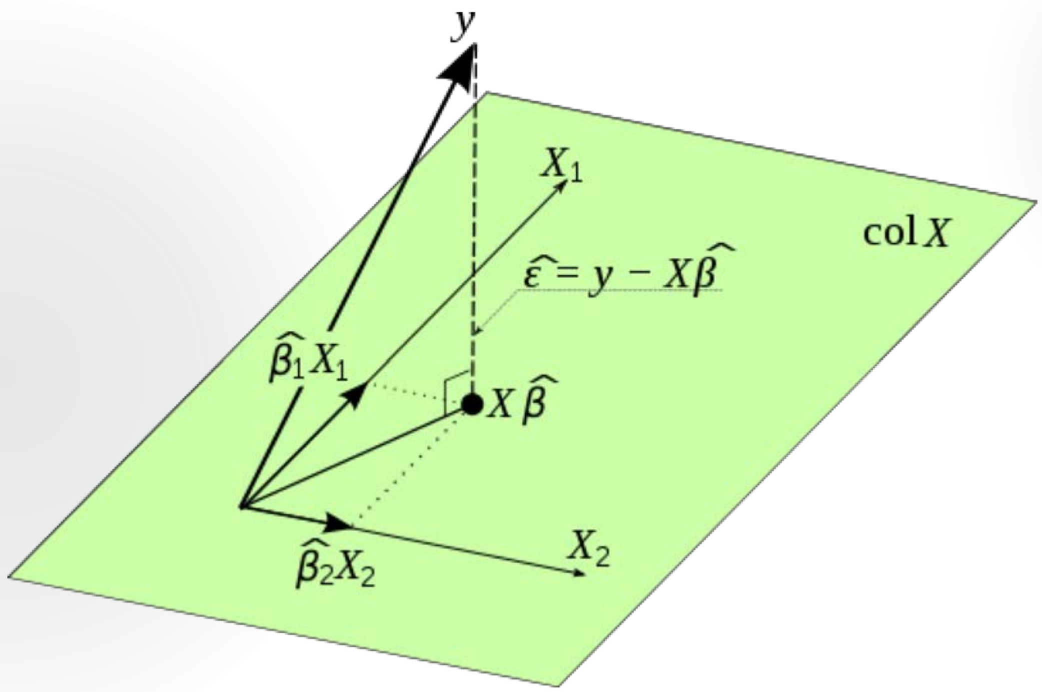
- Vector of residuals

$$e = Y - \hat{Y} = Y - HY = (I - H)Y$$

Projective Geometry Approach for LSE

- Column space:

  col(X) = linear space spanned by column vectors of X



Projective Geometry Approach for LSE

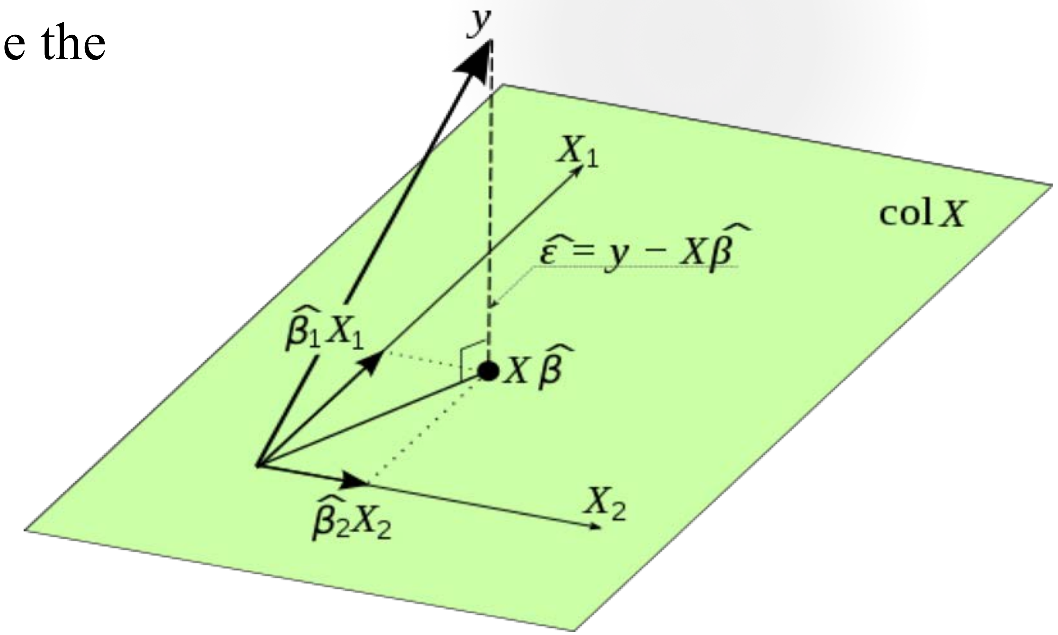
- Reformulate the problem

$$\min_{\beta\in R^2}\|Y-X\beta\|^2=\min_{y=X\beta}\|Y-y\|^2$$
$$=\min_{y\in \mathrm{col}(X)}\|Y-y\|^2$$

- The solution: the minimizer must be the orthogonal projection of Y onto the column space of X:

$$\hat{Y}=PY$$

where P is the orthogonal projection matrix/operator (Pythagoras)



Projection Matrix and LS Estimates

- It can be shown that P is symmetric and idempotent ($P^2=P$)
- It turns out that

$$P = H = X(X'X)^{-1}X'$$

- From

$$PY = X(X'X)^{-1}X'Y = Xb$$
$$b = \hat{\beta} = (X'X)^{-1}X'Y$$

- Predicted responses: $\hat{Y} = HY$
- Residuals: $e = Y - \hat{Y} = (I - H)Y$
- Residuals and predicted responses are orthogonal: $e'\hat{Y} = 0$
- Degree of Freedom of residuals

$$df = rank(I - H) = n - p$$

Note on $df = rank(I\text{-}H) = n\text{-}p$

- df = the rank of a quadratic form
- For any idempotent matrix A, tr(A) = rank(A)
- A_n is idempotent $\Leftrightarrow$ rank(A) + rank($I - A$) = n
- With $X_{n,p}$ and idempotent H, $I–H$ is also idempotent, that is, $(I–H)^2 = I–H$,

$$\mathrm{tr}(H) = \mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}(I_p) = p$$

- For Residuals $e = (I - H)Y$,

$$\text{df of } e = \mathrm{rank}(I–H) = \mathrm{tr}(I–H) = n - p$$

- http://www.jerrydallal.com/LHSP/dof.htm

Note on Covariance Matrix of Residual $e = (I\text{-}H)Y$

- Eigenvalues of $I–H$ are either 1 or 0
- Covariance Matrix of residuals e

$$\mathrm{Cov}(e) = \sigma^2(I - H)^2 = \sigma^2(I - H)$$

Denote $H = (h_{ij}) = X(X'X)^{-1}X'$

$$\mathrm{Var}(e_i) = \sigma^2(1 - h_{ii})$$

$$\mathrm{Cov}(e_i, e_j) = -\sigma^2 h_{ij}$$

Let $X_{i.} = (1, X_{i1}, X_{i2}, \ldots, X_{i,p-1})$ for $i = 1, \ldots, n$

$$h_{ii} = X_{i.}(X'X)^{-1}X'_{i.}; \;\; h_{ij} = X_{i.}(X'X)^{-1}X'_{j.}$$

Estimation of σ^2

- Sum of Squared Errors (Residuals) SSE:

$$SSE = e'e = (Y - Xb)'(Y - Xb) = Y'(I - H)Y$$

- df of SSE:

$$df_E = df \text{ of } e = n - p$$

- Mean of Squared Errors MSE

$$MSE = \frac{SSE}{df_E} = \frac{Y'(I - H)Y}{n - p}$$

- Estimator of σ^2

$$s^2 = MSE$$

$$s = \sqrt{MSE} \quad \text{Root MSE}$$

Sampling Distribution of b

- Recall $Y \sim N(X\beta, \sigma^2 I_n)$, and $b = (X'X)^{-1}X'Y$. Applying the theorem,

$$b \sim N(\beta, \sigma^2 (X'X)^{-1})$$

- Mean of b: $E(b) = \beta$, unbiased
- Covariance matrix of b: $Cov(b) = \sigma^2 (X'X)^{-1}$, optimal in some sense
- The estimated covariance matrix of b is

$$s^2(b) = s^2 (X'X)^{-1}$$

where

$$s^2 = \frac{e'e}{n-p} = \frac{Y'(I-H)Y}{n-p}$$

ANOVA Table

- Sources of variation include
  - Model (SAS) or Regression (KNNL)
  - Error (SAS, KNNL) or Residual (R)
  - Total
- *SS* and *df* add/decompos as before
- $SSM + SSE$ (RSS in R) $= SST$
- $df_M + df_E = df_T$

Sum of Squares

- Sum of Squares due to Model- SSM:

$$\text{SSM} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (\hat{Y} - \bar{Y}1_n)'(\hat{Y} - \bar{Y}1_n)$$

$$1_n = (1, 1, \ldots, 1)'_{n\times 1}$$

- Sum of Squares due to Error- SSE:

$$\text{SSE} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = (Y - \hat{Y})'(Y - \hat{Y})$$

- Sum of Squares in Total- SST:

$$\text{SST} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = (Y - \bar{Y}1_n)'(Y - \bar{Y}1_n)$$

Sum of Squares and Mean Squares

- Can show that

$$(Y-\bar{Y}1_n)'(Y-\bar{Y}1_n) = (Y-\hat{Y}+\hat{Y}-\bar{Y}1_n)'(Y-\hat{Y}+\hat{Y}-\bar{Y}1_n)$$

$$= (\hat{Y}-\bar{Y}1_n)'(\hat{Y}-\bar{Y}1_n) + (Y-\hat{Y})'(Y-\hat{Y})$$

- That is, SST = SSM + SSE
- Can show that

$$df_T = n-1;\ df_M = p-1;\ df_E = n-p$$

- Mean Squares:

$$\mathrm{MSM} = \mathrm{SSM}/df_M$$

$$\mathrm{MSE} = \mathrm{SSE}/df_E$$

$$\mathrm{MST} = \mathrm{SST}/df_T$$

ANOVA Table

Source	SS	df	MS	F
Model	SSM	df	MSM	MSM/MSE
Error	SSE	df_E	MSE	
Total	SST	df_T	MST	

Expected Mean Squares

- Formula for expected MS

$$\text{MSE} = \frac{Y'(I-H)Y}{(n-p)} = \frac{\epsilon'(I-H)\epsilon}{(n-p)}$$

$$E(\text{MSE}) = \frac{E(\epsilon'(I-H)\epsilon)}{(n-p)} = \sigma^2$$

- Can show that

$$E(\text{MSM}) = \sigma^2 + \text{Function}(\beta_1, \ldots, \beta_{p-1}, X)$$

- Under good design rank(X) = p,

$$\text{Function}(\beta_1, \ldots, \beta_{p-1}, X) = 0 \quad \text{if and only if}$$

$$\beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$

$$\text{Function}(\beta_1, \ldots, \beta_{p-1}, X) > 0 \quad \text{when at least one } \beta_i \neq 0$$

Note on df of SST

- $\because \bar{Y} = \frac{1}{n} 1_n' Y = \frac{1}{n} Y' 1_n$
- $SST = (Y - \bar{Y}1)'(Y - \bar{Y}1) = Y'Y - 2Y'\bar{Y}1 + 1'1\bar{Y}^2$
- $-2Y'1\bar{Y} + 1'1\bar{Y}^2 = -\frac{2}{n} Y'11'Y + n\bar{Y}^2$

$$= -\frac{2}{n} Y'11'Y + \frac{1}{n} Y'11'Y = -\frac{1}{n} Y'JY$$

- $\therefore SST = Y'\left(I - \frac{1}{n}J\right)Y, \quad J = \begin{pmatrix} 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & 1 \end{pmatrix}$
- Notice that $\frac{1}{n}J$ is idempotent, rank(J)=1, so rank$\left(I - \frac{1}{n}J\right) = n - 1$

Note on $E(MSM)$ $E(\mathrm{MSM}) = \sigma^2 + \mathrm{Function}(\beta_1, \dots, \beta_{p-1}, X)$

- Theorem: $X \sim N(\mu, \Sigma) \Rightarrow E(X'AX) = tr(A\Sigma) + \mu' A \mu$
- $SSE = Y'(I - H)Y,\ SSM = Y'\left(H - \frac{1}{n}J\right)Y$
- $E(MSM) = \frac{1}{p-1} E\left[Y'\left(H - \frac{1}{n}J\right)Y\right]$
- $\because Y \sim N(X\beta, \sigma^2 I_n)$

$$E\left[Y'\left(H - \frac{1}{n}J\right)Y\right] = (X\beta)'\left(H - \frac{1}{n}J\right)(X\beta) + \sigma^2 tr\left(H - \frac{1}{n}J\right)$$

- $tr\left(H - \frac{1}{n}J\right) = tr(H) - tr\left(\frac{1}{n}J\right) = rank(H) - rank\left(\frac{1}{n}J\right) = p - 1$
- $\left(X\beta\right)'\left(H - \frac{1}{n}J\right)(X\beta) = \beta'\left(X'HX - \frac{1}{n}X'JX\right)\beta$

$$= \beta'\left(X'X - \frac{1}{n}X'JX\right)\beta = (X\beta)'\left(I - \frac{1}{n}J\right)X\beta$$

- Since $\left(I - \frac{1}{n}J\right)$ is idempotent, the eigenvalues can only be 0 or 1, thus definite positive

ANOVA F Test

- Hypotheses:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$$
$$H_1 : \beta_k \neq k \text{ for at least one } k \text{ in } 1, 2, \ldots, p-1.$$

- Test statistic: $$F^* = \frac{MSM}{MSE}$$

- Sampling distribution under H_0:

$$F^* \sim F_{p-1,n-p}$$

- Decision rule at α
  - Reject H_0 if the calculated $F_0 > F_{p\text{-}1,n\text{-}p,\alpha}$
  - Reject H_0 if the P-value $P(F^* > F_0 | H_0) < \alpha$

Interpret Test Results

- Reject H_0:

  There exists evidence suggesting that one or more of the explanatory variables in the linear model is potentially useful for predicting (explaining) the response variable.

- Fail to reject H_0:

  There does not exit evidence to conclude that any of the explanatory variables can help model/predict/explain the response variable using the linear model

Coefficient of Multiple Determination R^2

- Correlation between responses Y and predicted responses $\hat{Y}$

$$r = \frac{\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\sum(Y_i - \bar{Y})^2}\sqrt{\sum(\hat{Y}_i - \bar{Y})^2}} = \frac{\sqrt{\sum(\hat{Y}_i - \bar{Y})^2}}{\sqrt{\sum(Y_i - \bar{Y})^2}}$$

- CMD R^2:

$$R^2 = r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\mathrm{SSM}}{\mathrm{SST}}$$

- Interpretation:

  R^2 gives the proportion of variation in the response variable, which can be explained by the model or all the explanatory variables in the model

R^2 & Adjusted R^2

- In terms of SSE:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \text{proportion not explained}$$

- Relation to F test statistics:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n-p}{p-1}$$

- Adjusted CMD R_a^2:

$$R_a^2 = 1 - \frac{MSE}{MST} = 1 - \frac{n-1}{n-p} \cdot \frac{SSE}{SST}$$

- When adding more explanatory variables, R^2 always increases, but R_a^2 can decrease

Background Reading

- KNNL 6.1 - 6.5