HW -6 : Today (3-19)

Class - 7

Exam-1 Results:
Max-94, Min - 52, Mean - 79.5

>= 90 - 4, 80-89 - 7, 70-79 - 6, 60-69 - 1, =<60 - 2

CORRECTIONS (one more chance):
- Due :W (3/14) Before 4 pm,
- You get half credits back,
- Do all the corrections on separate sheets,
- Should be returned with the exam,
- No partial credits.

General Linear Regression Model (GLR model)

In general, the variables $X_1, X_2, \ldots, X_{p-1}$ in the regression model do not need to represent different predictor variables. (ie $X_1, X_2, \ldots X_{p-1}$ may not have the additive effect).

GLR model:

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

where

$\beta_0, \beta_1, \ldots \beta_{p-1}$ — parameters,

$X_{i1}, X_{i2}, \ldots X_{i\,p-1}$ - known constants,

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ (ie σ^2 — constant).

Note:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \quad \text{if } X_{i0} = 1 .$$

Since $E(\varepsilon_i) = 0$,

$$\underbrace{E(Y) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i\,p-1}}_{\text{may not be a linear surface (ie linear in parameters only)}} .$$

may not be a linear surface. (ie linear in parameters only).

The following are Examples:

1) <u>No interaction effect between the predictor variables</u>

When $X_1, X_2, \ldots, X_{p-1}$ represent different predictor variables, GLR model is same as the first order linear model with p-1 predictors.

2) <u>Qualitative Predictor Variables</u> $\qquad I_A(x) = \begin{cases} 1 : & x \in A \\ 0 : & x \notin A. \end{cases}$

GLR can also have qualitative predictors such as gender (male, female) and disability Status (not disable, Partially disable, fully disable).

We use indicator variables to identify the classes of qualitative variables.

Eg: consider the regression analysis to predict the length of hospital stay (Y) based on age (X_1) and gender (X_2) of the patient.

Here X_2 is qualitative,

$$X_{i2} = \begin{cases} 0 : & \text{if male} \\ 1 : & \text{if female} \end{cases}$$

The first order regression model

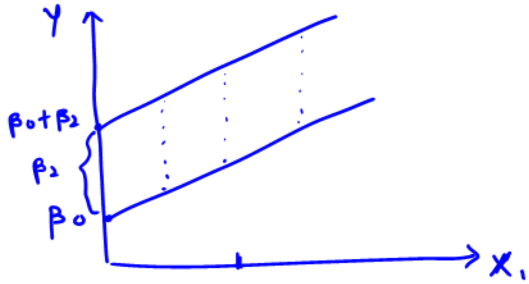
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$$

The response function:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

For male patients : $X_2 = 0$: $E(Y) = \beta_0 + \beta_1 X_1$ ←

For female patients : $X_2 = 1$: $E(Y) = (\beta_0 + \beta_2) + \beta_1 X_1$



3) Polynomial Regression

Polynomial regression moduls contains the square and higher order terms of predictor variables. Here the response function is curve linear.

Eg:- function with one predictor

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \underbrace{X_i^2} + \varepsilon_i$$

Let $X_{i1} = X_i$ and $X_{i2} = X_i^2$,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (\leftarrow \text{GLR model form})$$

4) Transformed variables

Many moduls can be transformed to the GLR model.

Eg:- $$Y_i = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i}$$

Let $Y_i' = \frac{1}{Y_i}$, then

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad (\text{GLR form}).$$

5) Interaction effects

Some times the effect of the predictor variables may not be additive.

Eg:- $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

If we let $X_{i3} = X_{i1} X_{i2}$ then,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad \text{(GLR form)}.$$

b) Combination of casses

Eg:- $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1}^2 + \beta_3 X_{i2} + \beta_4 X_{i2}^2 + \beta_5 X_{i1} X_{i2} + \varepsilon_i$

Let $Z_{i1} = X_{i1}$, $Z_{i2} = X_{i1}^2$, $Z_{i3} = X_{i2}$, $Z_{i4} = X_{i2}^2$, $Z_{i5} = X_{i1} X_{i2}$, then.

then,

$$Y_i = \beta_0 + \beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \beta_4 Z_{i4} + \beta_5 Z_{i5} + \varepsilon_i$$

(form of GLR).

GLR model in matrix form

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

$$\Rightarrow \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1\,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2\,p-1} \\ \vdots & \vdots & & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n\,p-1} \end{bmatrix}_{n\times p} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}_{p\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

$$\varepsilon_{n\times 1} \sim N\left(\underline{0}, \sigma^2 I_{n\times n}\right)$$

$$\begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & & & \\ \cdots & & & \sigma^2 \end{bmatrix} = Var(\varepsilon)$$

$\therefore E[Y] = X\beta$ and $\sigma^2\{Y\} = \sigma^2 I$

* Leas square normal Equations: $X'Xb = X'Y$
* Estimated regression coefficients: $b = (X'X)^{-1}X'Y$
* Fitted values: $\hat{Y}_{n\times 1} = X_{n\times p} b_{p\times 1} = HY$, where $H = X(X'X)^{-1}X$

* Residuals : $\underset{n\times 1}{e} = Y - \hat{Y} = Y - Xb = (I-H)Y$

* Variance - co-variance matrix of the residuals

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

$\Rightarrow$ $S^2\{e\} = MSE\,(I-H)$ — estimated variance - covariance matrix.

Analysis of variance Results

The Sum of squares in matrix form :

$$SSTO = Y'Y - (1/n)Y'JY = Y'[I - (1/n)J]Y$$

$$SSE = e'e = Y'Y - b'X'Y = Y'(I-H)Y$$

$$SSR = b'X'Y - (1/n)Y'JY = Y'[H - (1/n)J]Y,$$

where J is $n\times n$ matrix of 1s, and H is hat matrix.

ANOVA table

Source	SS	df	MS	F
Regression	$SSR = b'X'Y - (1/n)Y'JY$	$p-1$	$MSR = \frac{SSR}{p-1}$	$F = MSR/MSE$
Error	$SSE = Y'Y - b'X'Y$	$n-p$	$MSE = \frac{SSE}{n-p}$	
Total	$SSTO = Y'Y - (1/n)Y'JY$	$n-1$		

F- test

F test is used to test whether there is a regression relationship between the response variable and predictors $X_1, X_2, X_3, \dots X_{p-1}$.

Steps:

1) Hypotheses:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \cdots \beta_{p-1} = 0$ vs

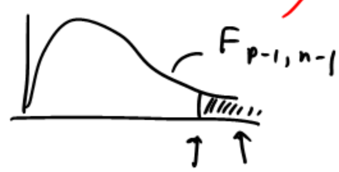
H_1 : not H_0 (ie at least on of β_ks is not zero).

2) Test Statistic:

$f^* = 55.55 \sim F_{1,39}$

$$F = \frac{MSR}{MSE} \sim F_{p-1,\,n-p}$$

3) Find $F(1-\alpha,\ p-1,\ n-p)$



OR

Calculate p- value $= P(F > f^*)$, where f^* is the observed value of F.

4) Conclusion:

If $f^* > F(1-\alpha,\ p-1,\ n-p) \Rightarrow$ reject H_0

If $f^* \leq F(1-\alpha,\ p-1,\ n-p) \Rightarrow$ conclude H_0.

(OR If p- value $< \alpha \Rightarrow$ reject H_0)

Write a non- technical Sentence about your conclusion.

Coefficient of Multiple determination

The coefficent of multiple determination (R^2) is defined as

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \quad : \quad 0 \leq R^2 \leq 1.$$

It measures the proportionate reduction of total variation in Y associated with variables $X_1, X_2, \ldots, X_{p-1}$.

* $R^2 = 0$ when all $b_K = 0$, $K = 1, 2, \dots p-1$.

* $R^2 = 1$ when $Y_i = \hat{Y}_i \ \forall \ i$
  (when all Y observations fall on the fitted regression surface.)

Note:

1) Adding more X variables to the regression model can only increase R^2 and never reduce it, because SSE can never become larger with more X-variables and SSTO is always the same for a given set of responses. $\Sigma(Y_i - \bar{Y})^2$

2) Two compare two models with different number of predictors, the adjusted coefficient of multiple determination (R^2_a) can be used.

$$R^2_a = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}$$

coefficient of multiple correlation (R)

$$R = \sqrt{R^2} \text{ — positive square root of } R^2.$$

Note:
$R = r$ for SLR model (ie when $p-1 = 1$).

Inference for normal error Regression model

Paramuter of interest is β_K.

* Point estimator $= b_K$
* Sampling distribution of b_K:

$$\frac{b_K - \beta_K}{S\{b_K\}} \sim t_{n-p}, \quad K = 0, 1, 2, \dots p-1.$$

F-test
$H_0: \beta_1 = \beta_2 = \cdots \beta_{p-1} = 0$
vs
$H_1:$ not H_0.

* Interval estimation:

$$b_k \pm t_{(1-\alpha/2 : n-p)} S\{b_k\} \quad , \quad k = 0, 1, 2, \ldots, p-1.$$

* Hypothesis test:

i) $H_0 : \beta_k = 0$ vs $H_1 : \beta_k \neq 0$

ii) $$T = \frac{b_k - 0}{S\{b_k\}} \sim t_{n-p} \quad , \quad k = 0, 1, 2, \ldots p-1.$$

iii) Calculate $t_{1-\alpha/2 : n-p}$

(OR calculate p-value $= 2P(T > |t^*|)$, (t^* is the observed value of T)).

iv) If $|t^*| \leq t_{(1-\alpha/2 : n-p)} \Rightarrow$ conclude H_0

If $|t^*| > t_{(1-\alpha/2 : n-p)} \Rightarrow$ reject H_0.

(OR p-value $< \alpha \Rightarrow$ reject H_0)

Make a non-technical sentence about your conclusion.

Joint Inference

The Bonferroni joint c.I.s for $g (\leq p)$ parameters with family confidence coefficient $1-\alpha$ are:

$$b_k \pm B\, S\{b_k\},$$

where $B = t_{(1-\alpha/2g : \underline{n-p})}$.

* Other methods discussed in chapter-3 can also be used.

Estimation of mean Response and prediction of new observations

* Interval Estimation of $E[Y_n]$

For given values of $X_1, X_2, \ldots, X_{p-1}$, denoted by $X_{n1}, X_{n2}, \ldots X_{np-1}$

we define the vector X_n,

$$X_n = \begin{bmatrix} 1 \\ X_{n1} \\ X_{n2} \\ \vdots \\ X_{np-1} \end{bmatrix}$$

Then mean response is denoted by

$$E[Y_n] = X_n' \beta .$$

The estimated mean response is given by

$$\hat{Y}_n = X_n' b .$$

* This estimator is unbiased

$$E[\hat{Y}_n] = X_n' \beta = E[Y_n], \text{ and}$$

$$\sigma^2\{\hat{Y}_n\} = \sigma^2 \, X_n' (X'X)^{-1} X_n$$

This can be expressed as a function of $\underbrace{\sigma^2\{b\}}$.

i.e $\sigma^2\{\hat{Y}_n\} = X_n' \, \sigma^2\{b\} X_n$

$\Rightarrow S^2\{\hat{Y}_n\} = MSE \, X_n' (X'X)^{-1} X_n = X_n' S\{b\} X_n$ — estimated Variance.

$\therefore$ $100(1-\alpha)\%$ confidence interval for $E[Y_n]$ is

$$\hat{Y}_n \pm t_{(1-\alpha/2 : n-p)} S\{\hat{Y}_n\} .$$

Simultaneous C.I.s for Several mean responses

To estimate a number of mean responses $E[Y_n]$ corresponding to different X_n vectors with family confidence coefficient $1-\alpha$, we can use two basic approaches.

① Working – Hotelling procedure

$$\hat{Y}_n \pm \underline{w}\, s\{\hat{Y}_n\} \quad \text{where} \quad w = \sqrt{2F(1-\alpha, \underline{p-1}, n-p)}$$

This confidence region covers the mean responses for all possible X_n vectors.

② Bonferroni Procedure

$$\hat{Y}_n \pm \underline{B}\, S\{\hat{Y}_n\}, \quad \text{where } B = t(1-\alpha/\underline{2g} : n-p).$$

<ins>Note:</ins>
To decide the most efficient one compare the values of w and B.

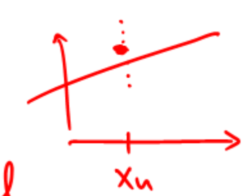
Prediction of New observation ($Y_{n(new)}$)

The $1-\alpha$ prediction limits for a new observation $Y_{n(new)}$ corresponding to X_n are

$$\hat{Y}_n \pm t_{(1-\alpha/2 : n-p)}\, S\{pred\},$$

$$\text{where } S^2\{pred\} = \underline{MSE} + \underline{s^2}\{\hat{Y}_n\} = MSE[1 + X_n'(X'X)^{-1}X_n].$$

* **Prediction of mean of m new observations**



When m new observations are to be selected at the same X level (X_n), the prediction interval for their mean $\bar{Y}_{n(new)}$ is

$$\hat{Y}_n \pm t_{(1-\alpha/2,\, n-p)}\, S\{\text{predmean}\},$$

$Y_1, Y_2, \ldots, Y_n \sim N(\mu, \sigma^2)$
$\Rightarrow \bar{y} \sim N(\mu, \sigma^2/n)$

where $$S^2\{\text{predmean}\} = \frac{MSE}{m} + s^2\{\hat{Y}_n\}$$
$$= MSE\left[\frac{1}{m} + X_n'(X'X)^{-1}X_n\right].$$

Prediction of g new observations

Simultaneous confidence intervals for g new observations at g different X-levels with family confidence interval $1-\alpha$ are:

1. Scheffe method:

$$\hat{Y}_n \pm S\, S\{\text{pred}\}, \quad \text{where } S^2 = gF(1-\alpha, g, n-p).$$

and $S^2\{\text{pred}\} = MSE\,(1 + X_n'(X'X)^{-1}X_n)$.

2. Bonferroni method:

$$\hat{Y}_n \pm B\, S\{\text{pred}\}, \quad \text{where } B = t(1-\alpha/2g : n-p).$$