MA 214 Lecture 11-12

- Another study: Multiple Regression.
  - Right now, only one predictor.
- Multiple covariates
  - Identify the subset of the covariates that provide a simple and useful interpretation of a model.

General structure of data in multiple regression.

id	Y	X_1	X_2	$\dots$	X_{p-1}
1	y_1	x_{11}	x_{12}	$\dots$	$x_{1,p-1}$
2	y_2	x_{21}	x_{22}	$\dots$	$x_{2,p-1}$
.	.	.	.	.	.
.	.	.	.	.	.
.					
n	y_n	x_{n1}	x_{n2}	$\dots$	$x_{n,p-1}$

Parameter Estimation

$$f(\beta_0, \beta_1 \dots \beta_{p-1}) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_{p-1} x_{i,p-1})^2$$

Example: $\hat{y} = 3 + 2x_1 + 6x_2$

Interpretation of the slope x_1

- Variables are highly related
  - Let's say, as x_1 change, x_2 might change or stay the same.

Degrees of Freedom of the SSE.

$(n-p)$

Estimate of σ^2

$$\hat{\sigma}^2 = \frac{SSE}{n-p} = \frac{\sum_{i=1}^{n} e_i^2}{n-p} = MSE.$$

Multiple regression model.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i, \quad i = 1, 2, 3, \dots, n.$$

Note: The model implies that all $(p-1)$ covariates are linearly related to the dependent variable Y.

Legend:

y_i: ith observation on Y

x_{ij}: ith observation on covariate.

ANOVA table.

Source	SS	DF	MS	F
Regression	SSR	$p-1$	MSR	$F = \frac{MSR}{MSE}$
Error	SSE	$n-p$	MSE	
SSTO	SSTO	$n-1$		

$$SSTO = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2 = SSTO - SSE.$$

$$MSE = \sigma^2$$

Coefficient of Determination: $R^2 = \frac{SSR}{SSTO}$

meas. total variability jointly explained by the covariates $(x_1, x_2, \dots, x_{p-1})$.

Correlation coefficient r is not defined in MRM.

F Test in Multiple Regression Model.

$H_0: \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$

H_a: Not all β_i are equal to 0.

At least some of the covariates in the model are linearly related to the dependent variable y.

- Can be useful predictors of the dependent variable.

Test statistic:

$$F = \frac{MSR}{MSE}$$

Reject H_0 if $F > F_{\alpha}$

Test for individual regression parameter β.

$H_0: \beta_1 = 0$ and $H_a: \beta_1 \neq 0$

Test statistic: $t = \frac{\hat{\beta}_1}{s\{\hat{\beta}_1\}}$

Decision rule: Reject H_0 if $t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$

Correlation plots of the covariates

- Scatterplot Matrix

- Covariates are highly related to each other.
  - Predictors are providing overlapping info
    - Some info are common between them.
  - So we want to find the subset of them that has the least amount of info that are common.

- Multicollinearity
  - The estimated regression coefficients tend to have large variability.
    - F test shows a significant overall regression relation Y and X.
    - But then the t-tests for individual variables show no significance.

Model selection techniques.

- R^2 procedure
- Stepwise procedure.
- AIC (Akaike Information Criterion)

Adjusted R^2

$$R^2_{adjusted} = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \frac{n-1}{n-p} \cdot \frac{SSE}{SSTO}$$

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

p = how many parameters we have.

Example:

$$= 1 - \frac{13-1}{13-3} \times (1-0.75)$$

$p = 3$ $n = 13$

- X
- Y

(X_1, X_2, X_3) with Y

So: $p = 3 + 1$
$= 4$

- Fitting a model with lots of data and they are highly correlated to each other
  - Inefficiency.

$\downarrow$

Solution: Variable Selection.

R^2 procedure.

- Carrying out regression of the dependent variable Y on all possible subsets of the covariates $(X_1, X_2, \ldots, X_{p-1})$, and computing R^2 in each case.
- Objective: find the subset of the covariates such that adding more covariates ...
  - Find the subset that "used" the least possible amount of the covariates with the highest score for R^2.
  - The more parameters we estimate, the more degrees of freedom we will losing.

Check out previous: Other Model Selection Criterion: Adjusted R^2.

Stepwise Regression

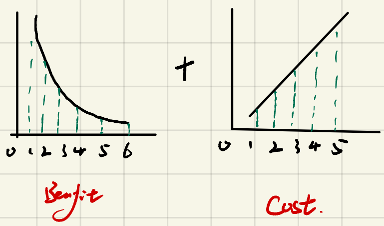
- Backward Elimination.
- Stepwise
- Forward Elimination.

- Cost and model performance.

AIC (Akaike Information Criterion).

- Based on Maximum Likelihood and a penalty for total number of parameters in the model.
- In multiple regression:
  - $AIC = n\log\left(\frac{SSE}{n}\right) + 2p + c$
    where p is the # of parameters and c is the constant.
  - Choose model with smallest AIC.

AIC Models



BIC: Schwartz's Bayesian Information Criterion.

- Assigns a greater penalty for larger models

$$BIC = n \cdot \log\left(\frac{SSE}{n}\right) + (\log n)\, p + c$$

All other things are similar to AIC.