UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Lecture 4 Linear Regression II

STAT 441/505: Applied Statistical Methods in Data Mining

Linglong Kong

Department of Mathematical and Statistical Sciences
University of Alberta

Winter, 2016

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Outline

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Multiple Linear Regression

- Multiple Linear Regression has more than one covariates,

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon,$$

  where usually $\varepsilon \sim N(0, \sigma^2)$.
- We interpret β_j as the average effect on Y of a one unit increase in X_j, while holding all the other covariates fixed.
- In the advertising example, the model becomes

$$\text{Sales} = \beta_0 + \beta_1 \times \text{TV} + \beta_2 \times \text{Radio} + \beta_3 \times \text{Newspaper} + \varepsilon.$$

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Coefficient Interpretation

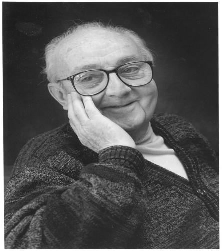
- The ideal scenario is when the predictors are uncorrelated — a balanced design.
  - Each coefficient can be estimated and tested separately.
  - Interpretations such as a unit change in X_j is associated with a β_j change in Y , while all the other variables stay fixed, are possible.
- Correlations amongst predictors cause problems.
  - The variance of all coefficient tends to increase, sometimes dramatically.
  - Interpretations become hazardous — when X_j changes, everything else changes.
- Claims of causality should be avoided for observational data.

The woes of regression coefficients

Data Analysis and Regression, Mosteller and Tukey 1977

- A regression coefficient β_j estimates the expected change in Y per unit change in X_j, with all other predictors held fixed. But predictors usually change together!
- Example: Y total amount of change in your pocket; $X_1 = \#$ of coins; $X_2 = \#$ of pennies, nickels and dimes. By itself, regression coefficient of Y on X_2 will be > 0. But how about with X_1 in model?
- $Y =$ number of tackles by a football player in a season; W and H are his weight and height. Fitted regression model is $Y = \beta_0 + 0.50W - 0.10H$. How do we interpret $\hat{\beta}_2 < 0$?

Two quotes by famous Statisticians



1919 - 2013 (aged 93)

- Essentially, all models are wrong, but some are useful.
  George Box
- The only way to find out what will happen when a complex system is disturbed is to disturb the system, not merely to observe it passively.
  Fred Mosteller and John Tukey, paraphrasing George Box

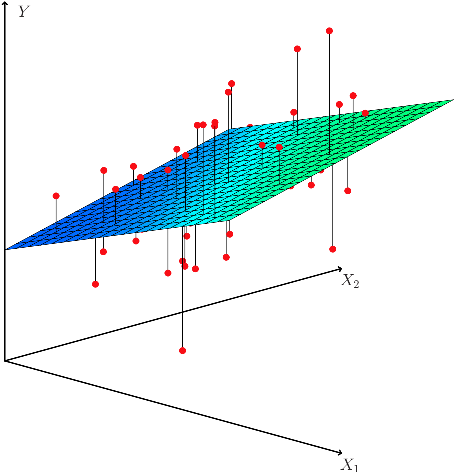
UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Coefficient estimation

- Given the estimates $\hat{\beta}_0, \hat{\beta}_1, \cdots$, and $\hat{\beta}_p$, the estimated regression line is
$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p.$$
- We estimate all the coefficients β_i, $i = 0, 1, \cdots, p$ as the values that minimize the sum of squared residuals
$$\text{RSS} = \sum_i (y_i - \hat{y}_i)^2,$$
 where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$ is the predicted values.
- This is done using standard statistical software. The values $\hat{\beta}_0$, $\hat{\beta}_1, \cdots$, and $\hat{\beta}_p$ that minimize RSS are the multiple least squares regression coefficient estimates.

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Estimation Example



Inference

- Is at least one predictor useful?

$$F = \frac{(\text{TSS} - \text{RSS})/p}{\text{RSS}/(n-p-1)} \sim F_{p,n-p-1}.$$

- What about an individual coefficient, say if β_i useful?

$$t = \frac{\hat{\beta}_i - 0}{\text{SE}\left(\hat{\beta}_i\right)} \sim t_{n-p-1}.$$

- For given $x_1, \cdots, x_p$, what is the prediction interval (PI) of the corresponding y?
- What about the estimation interval (CI) of y?
- What is the difference — PI, individual and CI, average, PI wider than CI.

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Advertising example

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.938889   0.311908   9.422   <2e-16 ***
TV           0.045765   0.001395  32.809   <2e-16 ***
Radio        0.188530   0.008611  21.893   <2e-16 ***
Newspaper   -0.001037   0.005871  -0.177     0.86
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 1.686 on 196 degrees of freedom
Multiple R-squared:  0.8972,Adjusted R-squared:  0.8956
F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16

> predict(TVadlm, newdata, interval="c", level=0.95)
       fit      lwr      upr
1 20.52397 19.99627 21.05168
> predict(TVadlm, newdata, interval="p", level=0.95)
       fit      lwr      upr
1 20.52397 17.15828 23.88967
```

Indicator Variables

- Some predictors are not quantitative but are qualitative, taking a discrete set of values.
- These are also called categorical predictors or factor variables.
- Example: investigate difference in credit card balance between males and females, ignoring the other variables. We create a new variable,

$$x_i = \begin{cases} 1 & \text{if } i\text{-th person is female,} \\ 0 & \text{if } i\text{-th person is male} \end{cases}.$$

- Resulting model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & \text{if } i\text{-th person is female,} \\ \beta_0 + \varepsilon_i & \text{if } i\text{-th person is male} \end{cases}.$$

- Interpretation and more than two levels (categories)?

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Indicator Variables

- In general, if we have k levels, we need $(k-1)$ indicator variables.
- For example, we have 3 levels — $A, B,$ and C for a covariate x,

$$x_A = \begin{cases} 1 & \text{if } x \text{ is A,} \\ 0 & \text{if } x \text{ is not A} \end{cases}; \quad x_B = \begin{cases} 1 & \text{if } x \text{ is B,} \\ 0 & \text{if } x \text{ is not B} \end{cases}.$$

- If x is C, then $x_A = x_B = 0$. We call C as baseline.
- β_A is the contrast between A and C and β_B is the contrast between B and C.

UNIVERSITY OF ALBERTA
EDMONTON·ALBERTA·CANADA

Summary and Remark

- Multiple linear regression
- Estimation and inference
- Indicator variables
- Read textbook Chapter 3
- Do R lab