Stat GR4315 Lecture 5

Jingchen Liu

Department of Statistics
Columbia University

Hypothesis testing

- The rationale
- Test statistic
- Test statistic and reference distribution
- p-value: the probability of observing a test statistic that is as extreme as or more extreme than the observed one.
- The null hypothesis is rejected is the p-value is less than a threshold, such as 5%.

Hypothesis testing

- Size of a test: the probability of making type I error.
- Power of a test: the probability of rejecting the null under the alternative hypothesis

Hypothesis testing

- The Z-test

$$z = \frac{\hat{\beta}_1}{\sigma\sqrt{\frac{1}{\sum(x_i - \bar{x})}}} \sim N(0,1)$$

- The t-test

$$t = \frac{\hat{\beta}_1}{\hat{\sigma}\sqrt{\frac{1}{\sum(x_i - \bar{x})}}} \sim t_{n-2}$$

Hypothesis testing – *p*-value

- Two-sided *p*-value: $H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$

$$P(|Z| \geq |z|) \quad or \quad P(|t_{n-2}| \geq |t|)$$

- One-sided *p*-value: $H_0 : \beta_1 \leq 0 \quad H_1 : \beta_1 > 0$

$$P(Z \geq z) \quad or \quad P(t_{n-2} \geq t)$$

- Duality between hypothesis testing and confidence interval

Analysis of variance

- The decomposition

$$\sum(y_i - \bar{y})^2 = \sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \bar{y})^2$$

$$SS_{total} = SS_{error} + SS_{regression}$$

- The analysis of variance table

	d.f.	Sum Sq	Mean Sq	F-value	p-value
x	$p-1$	$\sum(\hat{y}_i - \bar{y})^2$	$\frac{\sum(\hat{y}_i-\bar{y})^2}{p-1}$	$\frac{(n-p)\sum(\hat{y}_i-\bar{y})^2}{(p-1)\sum(y_i-\hat{y}_i)^2}$	$*$
Residuals	$n-p$	$\sum(y_i - \hat{y}_i)^2$	$\frac{\sum(y_i-\hat{y}_i)^2}{n-p}$		
Total	$n-1$	$\sum(y_i - \bar{y})^2$			

Analysis of variance

- The analysis of variance table of the Iris setosa data

	d.f.	Sum Sq	Mean Sq	F-value	p-value
x	1	3.9	3.2	59.0	7×10^{-10}
Residuals	48	3.2	0.066		
Total	49	7.1			

- F-test and t-test

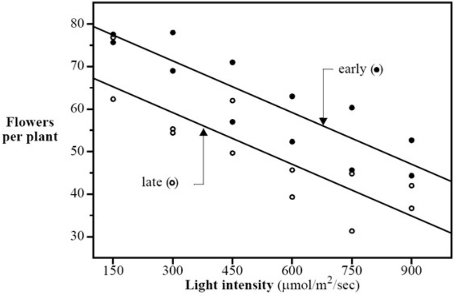
Inferential tools of simple linear model

- The four assumptions
- Point estimate
  - Understanding the least squares estimate
  - variance estimate
  - Frequentist's distribution
- Interval estimate
  - Regression coefficients
  - Prediction: conditional mean, future observation, simultaneous confidence band
- Hypothesis testing
  - Z-test
  - t-test: special case – two sample test
  - F-test: analysis of variance, R^2

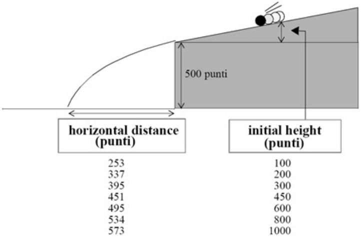
Multiple linear regression – motivation

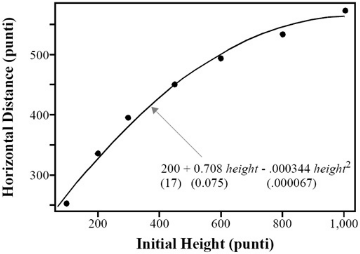
- Simple linear model in subgroups
- Nonlinear relationship
- Multiple predictors
- Variable selection

Multiple group

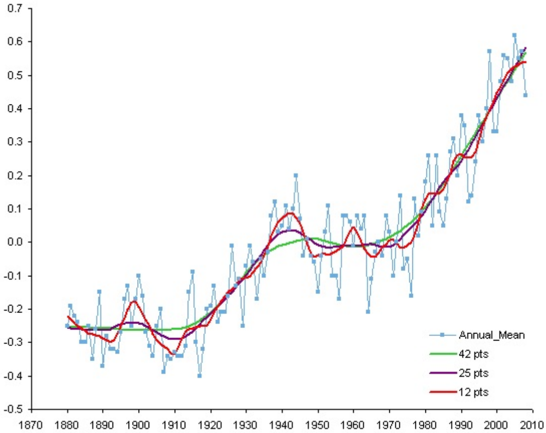


Nonlinearity





Local regression



Multiple linear regression

- Response variable y and p covariates x_1, ..., x_p
- Regression model

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p + \varepsilon$$

where $\varepsilon \sim N(0, \sigma^2)$

Matrix notation

- $$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

  ε_1, ..., ε_n are i.i.d. $N(0, \sigma^2)$.
- Write in short as

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

  where ε is multivariate normal with mean zero and covariance matrix $\sigma^2 I$.

Least squares estimate

- Least squares estimator

$$\hat{\beta} = \arg\min_{\beta}(Y - X\beta)^\top(Y - X\beta) = (X^\top X)^{-1}X^\top Y$$

- Derive it.

Frequentist distribution

- Unbiased distribution

$$E(\hat{\beta}) = \beta$$

- Variance and covariance

$$Var(\hat{\beta}) = \sigma^2 (X^\top X)^{-1}.$$

- Computation of covariance matrix
- Multivariate normal distribution

About multivariate normal distribution

- Multivariate normal distribution with mean μ and covariance matrix Σ

$$f(x) = \frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}} e^{-\frac{(x-\mu)^\top \Sigma^{-1} (x-\mu)}{2}}$$

- Generating multivariate normal random variables

Variance estimation

- An unbiased estimator

$$\hat{\sigma}^2 = \frac{(Y - \hat{Y})^\top (Y - \hat{Y})}{n - p - 1}$$

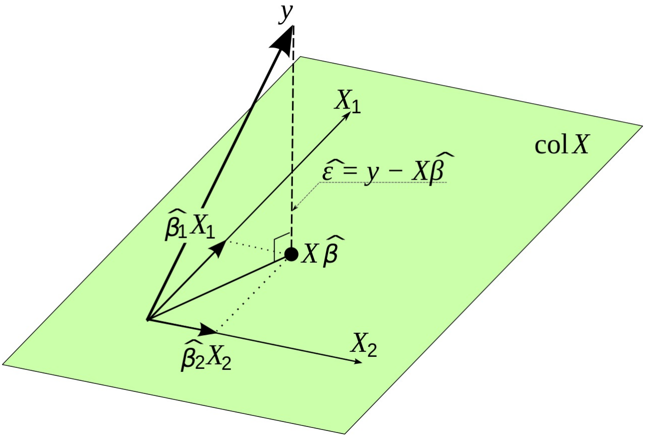
- The distribution of $\hat{\sigma}^2$
- Prediction

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

- Hat matrix

$$H = X(X^\top X)^{-1} X^\top$$

Projection


y
X_1
col X
$\widehat{\varepsilon} = y - X\widehat{\beta}$
$\widehat{\beta}_1 X_1$
$X\widehat{\beta}$
$\widehat{\beta}_2 X_2$
X_2