Chap 13
Fancy Regression

—

SLR need not be so simple.

(A) log & log-log transformations.

$$\text{Model: } \mathbb{E}\{\log Y\} = \beta_0 + \beta_1 x$$

$$\mathbb{E}\{Y\} \approx e^{\beta_0} e^{\beta_1 x}$$

non-linear fit with a SLR through log transformation of Y

After you have fit a model with a transformation, it helpful to back-transform, especially for visuals.

Consider $E\{Y\} = ax^b$ a multiplicative model.

log-log model: $\log E\{Y\} = \log a + b \log x$

$\log E\{Y\}$ → $\log y$; $\log a$ → β_0; b → β_1; $\log x$ → lx

A log-log model is appropriate whenever quantities are linearly related on a multiplicative or percentage scale.

estimated slope $\hat{\beta}_1$ in a log-log SLR is known as elasticity

↓

% change in Y given 1% change in X

Ch 13

(B) Polynomial Regression

—

Taylor's theorem: any curve can be approximated locally by a low-degree polynomial

Weierstrauss' Thm: any function can be approximated globally by a high-degree polynomial

→ the sweet spot is somewhere in between.

Model: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \varepsilon_i$,

↓

Technical a MLR

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

You can choose to "power up" as much as you'd like, but be careful:

- when $n \leq d-1$ (d degree of of poly) you get a "perfet" fit, because the polynomial interpolates all points
- but it is not "ideal" since $\varepsilon_i = \emptyset$ for all i, which isn't Gaussian

→ "overfit"

you can check it you need degree d by testing it $f_d = \emptyset$ or not.

Chapter 13

MLR

© Multiple Linear Regression

extends SLR to more than one input x.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_d x_{id} + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

$$Y_i \mid x_{i1}, \ldots, x_{id} \overset{ind}{\sim} N(\underbrace{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_d x_{id}}_{\mu_i}, \sigma^2)$$

$$Y \sim N_n(X\beta, \sigma^2 \mathbb{I}_n)$$

$\leftarrow$ n-variate MVN

$Y = (Y_1, \ldots, Y_n)^T$ n-vector

$\beta = (\beta_0, \beta_1, \ldots, \beta_d)$ $p = d+1$-vector

$\mathbb{I}_n$ = $n \times n$ identity matrix

$X = n \times p$ <u>design matrix</u> with

$$x_i^T = (1, x_{i1}, \ldots, x_{id})$$

SLR is a special case where $d = 1$.

$d = 2 \rightarrow$ linear surface is a <u>plane</u>

$d \geq 3 \rightarrow$ we say <u>hyperplane</u>

$i = 1, \ldots, d$ $\quad \beta_j = \dfrac{\partial E\{Y \mid x_1, \ldots, x_d\}}{\partial x_j}$

Calculating the MLE is the same as in Ch 7 but via MVN with $X\beta \equiv \mu$ & $\Pi_n \sigma^2 \equiv \Sigma$

$$L(\beta, \sigma^2) = (2\pi \underbrace{\sigma^2}_{\Sigma})^{-n/2} \exp\left\{ \frac{-1}{2\sigma^2} (Y - \underbrace{X\beta}_{\mu})^T (Y - X\beta) \right\}$$

$$l(\beta, \sigma^2) = c - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \underbrace{(Y - X\beta)^T (Y - X\beta)}$$

$$= c - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \left(Y^T Y - 2 Y^T X \beta + \beta^T X^T X \beta \right)$$

$$\phi \overset{\text{set}}{=} \frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}(X^T y + X^T X \beta)$$

$$X^T X \beta = X^T y$$

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$$

looks like $b_1 = \frac{S_{yx}}{S_{xx}}$

Important projection result for mVNs

If $Y \sim N_n(\mu, \Sigma)$ & P is a $p \times n$ projection matrix

then $PY \sim N_p(P\mu, P\Sigma P^T)$

Consider $P = (X^T X)^{-1} X^T$ so $\hat{\beta} = PY$

observe $PX\beta = (X^T X)^{-1} X^T X \beta = \beta$

$$PP^T = (X^T X)^{-1} \cancel{X^T X (X^T X)^{-1}} = (X^T X)^{-1}$$

so $\hat{\beta} = PY \sim N_p(\beta, \sigma^2 (X^T X)^{-1})$

$\hat{\beta}$: MLE

N_p: sampling distribution

What to use for σ^2?

$$l(\hat{\beta}, \sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\hat{\beta})^T(y - X\hat{\beta})$$

$$\phi \overset{\text{set}}{=} \frac{\partial l}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\hat{\beta})^T(y - X\hat{\beta})$$

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})^T(y - X\hat{\beta})$$

Biased estimate, but you can correct that bias

$$\boxed{s^2 = \frac{1}{n-p}(y - X\hat{\beta})^T(y - X\hat{\beta})}$$

Can show that $\sigma^2 \sim (n-p)s^2 / \chi^2_{n-p}$

so you can use

Student-t for hypothesis tests

& CI.

Prediction for MLR involves another projection since

$$\hat{\underline{y}} = X\hat{\beta} = \underline{H}\,\underline{y}$$

where $H = X(X^TX)^{-1}X^T$

"hat matrix"

Details of how to use H for prediction at new x_p in book.

Chapter 13

① Dummies & Interactions

A <u>dummy variable</u> allows you to use a categorical grouping in a MLR.

if $w_i \in \{\text{Dodge}, \text{Ford}, \text{Gmc}\}$

Then $x_{i2} = \mathbb{I}_{\{w_i = \text{Ford}\}}$

$x_{i3} = \mathbb{I}_{\{w_i = \text{Gmc}\}}$

Dodge is $(x_{i2}, x_{i3}) = (0, 0)$ → in the "intercept"

ford $= (1, 0)$

Gmc $= (0, 1)$

Model: $Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon$

x_1: year

An MLR with entirely dummy variables is an ANOVA

→ see book for ice cream example

→ testing, wait to (E)

An <u>interaction effect</u> allows the contribution of one input x_j to depend on the <u>level</u> of another input x_k

Model: $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} + \cdots + \varepsilon_i$

($\beta_3 x_{i1} x_{i2}$: interaction term)

so that

$$\frac{\partial E\{Y \mid x_1, x_2\}}{\partial x_1} = \beta_1 + \beta_3 x_2$$

Chapter 13

(E) Model Selection

—

How to select from two "nested" MLR models?

→ Good stat is $R^2 \equiv cor(y, \hat{y})$

$\boxed{\Delta R^2}$

Take inspiration from ANOVA

$H_0: \beta_{Base+1} = \cdots = \beta_{full} = 0$

v. H_1: at least one of $\beta_j \neq 0$ $j > p_{base}$

"Easy" to virtually all R^2 comparison with these hypotheses.

The meta version is also similar to ANOVA: look at ratios of MSE, which can be expressed as R^2-values

$$f = \frac{(SSE_{full} - SSE_{base})/p_\Delta}{SSE_{full}/n - p_{full}} \quad \leftarrow p_b - p_f$$

$$= \frac{R_\Delta^2/p_\Delta}{(1 - R^2_{full})/n - p_\Delta} \qquad R_\Delta^2 = R^2_{full} - R^2_{base}$$

Can show that $f \sim F_{p_\Delta,\, n - p_{full}}$