Lecture 10: Multiple linear regression

Pratheepa Jeganathan

10/14/2019

Recap

- What is a regression model?
- Descriptive statistics – graphical
- Descriptive statistics – numerical
- Inference about a population mean
- Difference between two population means
- Some tips on R

Recap

- Simple linear regression (covariance, correlation, estimation, geometry of least squares)
  - Inference on simple linear regression model
  - Goodness of fit of regression: analysis of variance.
  - F-statistics.
  - Residuals.
  - Diagnostic plots for simple linear regression (graphical methods).

Multiple linear regression

Outline

- Specifying the model.
- Fitting the model: least squares.
- Interpretation of the coefficients.

Model

- Given a sample $Y_i, X_{i1}, X_{i2}, \cdots, X_{ip}, i = 1, 2, \cdots, n$, the multiple linear regression model is
-
$$y_i = \underbrace{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}}_{\mathbb{E}[Y|X_1,\cdots,X_2]} + \epsilon_i,$$
  where $i = 1, 2, \cdots, n$.
  - y_i: i-th value of the response variable Y.
  - x_{ij}: j-th predictor variable for the i-th unit.
- Assumptions:
  - Errors ϵ are assumed independent $\mathrm{N}\left(0, \sigma^2\right)$, as in simple linear regression.
- Coefficients are called *partial* regression coefficients because they "allow" for the effect of other variables.

Example (Prostate data)

- For more information on the Gleason score.

Variable	Notation	Description
lpsa	Y	(log) Prostate Specific Antigen
lcavol	X_1	Cancer Volume
lweight	X_2	Weight
age	X_3	Patient age
lbph	X_4	(log) Vening Prostatic Hyperplasia
svi	X_5	Seminal Vesicle Invasion
lcp	X_6	(log) Capsular Penetration
gleason		Gleason score
pgg45	X_7	Percent of Gleason score 4 or 5
train		Label for test or training split

- We assume a linear model relating Y and seven explanatory variables.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_5 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \epsilon.$$

Example

- This data set is from the book The Elements of Statistical Learning.

```
# if not installed, install the package
if(!("ElemStatLearn" %in% installed.packages())){
  install.packages("ElemStatLearn")
}
```

```
library(xtable)
library(ElemStatLearn)
data(prostate)
```

Parameter estimation

- Just as in simple linear regression, model is fit by minimizing

$$\mathrm{SSE}(\beta_0, \ldots, \beta_p) = \sum_{i=1}^{n} \left(Y_i - \left(\beta_0 + \sum_{j=1}^{p} \beta_j \; X_{ij} \right) \right)^2$$

- BY a direct application of calculus, it can be shown that the least squares estimators $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_{0p}$ are given by the solution of a system of linear equations known as *normal equation*.
    - $\hat{\beta}_0$ is intercept (may or may not be $\bar{Y}$.)
    - $\hat{\beta}_j$ is the estimate of the partial regression coefficient of the predictor X_j.

Parameter estimation for the example

```
prostate.lm = lm(lpsa ~ lcavol + lweight +
    age + lbph + svi + lcp + pgg45,
  data = prostate)
print(xtable(prostate.lm, digits = 3))
```

% latex table generated in R 3.6.0 by xtable 1.8-4 package % Sun Oct 13 23:27:40 2019

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.494	0.874	0.566	0.573
lcavol	0.570	0.086	6.634	0.000
lweight	0.614	0.198	3.096	0.003
age	-0.021	0.011	-1.905	0.060
lbph	0.097	0.058	1.691	0.094
svi	0.752	0.238	3.159	0.002
lcp	-0.105	0.089	-1.175	0.243
pgg45	0.005	0.003	1.573	0.119

Estimating σ^2

- As in simple regression

$$\widehat{\sigma}^2 = \frac{SSE}{n-p-1}.$$

- We can show that

$$\frac{(n-p-1)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

- Why χ^2_{n-p-1}?
  - Typically, the degrees of freedom in the estimate of σ^2 is $n - \#$number of parameters in regression function.

Estimating σ^2 for the example

- The degrees of freedom

```
prostate.lm$df.resid
```

```
## [1] 89
```

- Using the formula

```
sigma.hat = sqrt(sum(resid(prostate.lm)^2)
  /df.residual(prostate.lm))
round(sigma.hat, digits = 3)
```

```
## [1] 0.696
```

- OR

```
round(sqrt(deviance(prostate.lm)/df.residual(prostate.lm)),
```

```
## [1] 0.696
```

Interpretation of regression coefficients β_j's

- Take $\beta_1 = \beta_{\text{lcavol}}$ for example. This is the amount the average `lpsa` rating increases for one "unit" of increase in `lcavol`, keeping everything else constant.
- We refer to this as the effect of `lcavol` *allowing for* or *controlling for* the other variables.

Example (interpretation of β_j)

- For example, let's take the 10th case in our data and change `lcavol` by 1 unit.

```
case10 = prostate[10,]
print(xtable(case10))
```

% latex table generated in R 3.6.0 by xtable 1.8-4 package % Sun Oct 13 23:52:37 2019

	lcavol	lweight	age	lbph	svi	lcp	gleason	pgg45	lpsa
10	0.22	3.24	63	-1.39	0	-1.39	6	0	1.05

```
case10.temp = case10
case10.temp['lcavol'] =
  case10.temp['lcavol'] + 1
```

Example (interpretation of β_j)

```r
Yhat = predict(prostate.lm, rbind(case10, case10.temp))
names(Yhat) = c("lcavol_10", "lcavol_10+1")
Yhat
```

```
##   lcavol_10 lcavol_10+1
##    1.307754    1.877300
```

- Our regression model says that this difference should be $\hat{\beta}_{\texttt{lcavol}}$.

```r
c(Yhat[2]-Yhat[1], coef(prostate.lm)['lcavol'])
```

```
## lcavol_10+1      lcavol
##    0.569546    0.569546
```

Partial regression coefficients

- The term *partial* refers to the fact that the coefficient β_j represent the partial effect of X_j on Y, i.e. after the effect of all other variables have been removed.
  - Specifically,

$$Y_i - \sum_{l=1, l \neq j}^{k} X_{il}\beta_l = \beta_0 + \beta_j X_{ij} + \varepsilon_i.$$

- Let $e_{i,(j)}$ be the residuals from regressing Y onto all $X_.$'s except X_j, and let $X_{i,(j)}$ be the residuals from regressing X_j onto all $X_.$'s except X_j.
  - $e_{i,(j)} = Y - (\hat{b}_0 + \hat{b}_1 X_1 + \cdots + \hat{b}_{j-1} X_{j-1} + \hat{b}_{j+1} X_{j+1} + \cdots + \hat{b}_p X_p)$.
  - $X_{i,(j)} = X_j - (\hat{a}_0 + \hat{a}_1 X_1 + \cdots + \hat{a}_{j-1} X_{j-1} + \hat{a}_{j+1} X_{j+1} + \cdots + \hat{a}_p X_p)$.
- If we regress $e_{i,(j)}$ against $X_{i,(j)}$, the coefficient is *exactly* the same as in the original model.
  - $e_{i,(j)} = C_0 + \beta_j X_{i,(j)} + \epsilon_i$.

Example

- Let's verify this interpretation of regression coefficients.

```
partial_resid_lcavol = resid(lm(lcavol ~  lweight +
    age + lbph + svi +
    lcp + pgg45, data=prostate))
partial_resid_lpsa = resid(lm(lpsa ~  lweight +
    age + lbph + svi +
    lcp + pgg45, data=prostate))
```

```r
print(xtable(summary(lm(partial_resid_lpsa ~
    partial_resid_lcavol))))
```

% latex table generated in R 3.6.0 by xtable 1.8-4 package % Mon Oct 14 00:13:19 2019

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0000	0.0684	-0.00	1.0000
partial_resid_lcavol	0.5695	0.0831	6.85	0.0000

Centering and Scaling

- Regression coefficient magnitude depends on the unit of measurements of the variable.
  - If regression coefficient of income is 5.123 when measured in dollars, but 5123 when measured in $1000.
- To make regression coefficient unitless
  - Can *center* and *scale* the variables.

Centering and scaling for the intercept model

- Centering $X_j - \bar{x}_j$ and $Y - \bar{y}$.
- Scaling
  - Unit-length scaling:

$$\tilde{Z}_y = \frac{Y - \bar{y}}{L_y}, L_y = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

    and

$$\tilde{X}_j = \frac{X_j - \bar{x}_j}{L_j}, L_j = \sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}, j = 1, 2, \cdots, p.$$

  - Standardizing:

$$\tilde{Z}_y = \frac{Y - \bar{y}}{s_y}, s_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}}$$

    and

$$\tilde{X}_j = \frac{X_j - \bar{x}_j}{s_j}, s_j = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n-1}}, j = 1, 2, \cdots, p.$$

Centering and scaling for the intercept model

- Let $\hat{\theta}_j$ be the estimated regression coefficient for the standardized data.
  - The estimated regression coefficient for the original data are

$$\hat{\beta}_j = \frac{s_y}{s_j}\hat{\theta}_j, j = 1, 2, \cdots, p,$$

    and

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{p} \hat{\beta}_j \bar{x}_j.$$

- θ_j measures the change in standardized units of Y corresponding to an increase of one standard deviation in X_j.

Scaling for no-intercept model

- Centering has an effect of including an intercept in the model.
- Let

$$Y = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

  be the no-intercept model.
- Scaled variables are

$$\tilde{Z}_y = \frac{Y}{L_y}, L_y = \sqrt{\sum_{i=1}^{n} y_i^2},$$

  and

$$\tilde{X}_j = \frac{X_j}{L_j}, L_j = \sqrt{\sum_{i=1}^{n} x_{ij}^2}, j = 1, 2, \cdots, p.$$

References for this lecture

- **CH** Chapter 3.1-3.6
- Lecture notes of Jonathan Taylor .