STAT 3690 Lecture 17

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 09, 2022

Multivariate linear regression

- Interested in the relationship between random q-vector $[Y_1,\ldots,Y_p]^\top$ and random q-vector $[X_1,\ldots,X_q]^\top$
- Model
  - Population version: $[Y_1,\ldots,Y_p]^\top \mid X_1,\ldots,X_q \sim (\mathbf{B}^\top[1,X_1,\ldots,X_q]^\top,\sigma^2)$, where $\mathbf{B} = [\beta_{kj}]_{(q+1)\times p}$, i.e.,
    - $\mathrm{E}([Y_1,\ldots,Y_p]^\top \mid X_1,\ldots,X_q) = \mathbf{B}^\top[1,X_1,\ldots,X_q]^\top$
    - $\mathrm{cov}([Y_1,\ldots,Y_p]^\top \mid X_1,\ldots,X_q) = \mathbf{\Sigma} > 0$, i.e., the conditional covariance of $[Y_1,\ldots,Y_p]^\top$ does not depend on $X_1,\ldots,X_q$
  - Sample version
$$\underset{n\times p}{\mathbf{Y}} = \underset{n\times(q+1)}{\mathbf{X}}\ \underset{(q+1)\times p}{\mathbf{B}} + \underset{n\times p}{\mathbf{E}}$$
    - $\mathbf{Y} = [Y_{ij}]_{n\times p}$
    - Design matrix
$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{q1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nq} \end{bmatrix}_{n\times(q+1)}$$
      - $\mathrm{rk}(\mathbf{X}) = q+1 < p+q+1 \le n$
    - $\mathbf{E} = [\mathbf{E}_{1\cdot},\ldots,\mathbf{E}_{n\cdot}]^\top$, where $\mathbf{E}_{i\cdot}^\top$ is the ith row of $\mathbf{E}$
    - Assume the independence across i, i.e.,
      - $[Y_{i1},\ldots,Y_{ip},X_{i1},\ldots,X_{iq}]^\top \overset{\text{iid}}{\sim} [Y_1,\ldots Y_p,X_1,\ldots,X_q]^\top$
      - $\mathbf{E}_{i\cdot} \overset{\text{iid}}{\sim} (\mathbf{0}_p,\mathbf{\Sigma})$

- Relationship with univariate linear regression
  - If $\mathbf{\Sigma}$ is diagonal, the multivariate model reduces to $\mathbf{Y}_{\cdot j} = \mathbf{X}\mathbf{B}_{\cdot j} + \mathbf{E}_{\cdot j}$, $j = 1,\ldots,p$
    - $\mathbf{Y}_{\cdot j}$: the jth column of $\mathbf{Y}$
    - $\mathbf{B}_{\cdot j}$: the jth column of $\mathbf{B}$
    - $\mathbf{E}_{\cdot j} \sim (\mathbf{0}_n,\sigma_{jj}^2\mathbf{I}_n)$
      - σ_{jj}^2: (j,j)-entry of $\mathbf{\Sigma}$

- Relationship with MANOVA
  - MANOVA models can be expressed as multivariate linear regression with carefully selected dummy (explanatory) variables.

Exercise: translate the following 1-way MANOVA model

$$\mathbf{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \mathbf{E}_{ij},\quad i = 1,\ldots,m,\quad j = 1,\ldots,n_i$$

into a multivariate linear regression model, where $\mathbf{E}_{ij} \overset{\text{iid}}{\sim} MVN_p(\mathbf{0}, \mathbf{\Sigma})$ and $\sum_i \boldsymbol{\tau}_i = 0$.

$$
Y = \begin{bmatrix} Y_{11}^{\top} \\ \vdots \\ Y_{1n_1}^{\top} \\ Y_{21}^{\top} \\ \vdots \\ Y_{2n_2}^{\top} \\ \vdots \\ Y_{m-1,1}^{\top} \\ \vdots \\ Y_{m-1,n_{m-1}}^{\top} \\ Y_{m1}^{\top} \\ \vdots \\ Y_{mn_m}^{\top} \end{bmatrix}_{(\sum_i n_i)\times p}
\qquad
X = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}_{(\sum_i n_i)\times m}
\qquad
\mathcal{B} = \begin{bmatrix} \beta_{0\cdot}^{\top} \\ \vdots \\ \beta_{(m-1)\cdot}^{\top} \end{bmatrix}_{m\times p}
$$

$\underbrace{\qquad\qquad}_{m-1}$

In this case, $\beta_{0\cdot} = \mu + \tau_m$, $\beta_{i\cdot} = \tau_i - \tau_m$, $i = 1, \cdots, m-1$

- Least squares (LS) estimation (no need of (conditional) normality)
  - $\hat{\mathbf{B}}_{\text{LS}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$
    - $\text{E}(\hat{\mathbf{B}}_{\text{LS}}) = \mathbf{B}$
  - $\hat{\mathbf{\Sigma}}_{\text{LS}} = (n-q-1)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}_{\text{LS}})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}_{\text{LS}}) = (n-q-1)^{-1}\mathbf{Y}^\top(\mathbf{I} - \mathbf{H})\mathbf{Y}$
    - $\text{E}(\hat{\mathbf{\Sigma}}_{\text{LS}}) = \mathbf{\Sigma}$

① Known: $\min Q(\mathcal{B}) := tr\{(Y - X\mathcal{B})^T(Y - X\mathcal{B})\} = \sum_{j=1}^{p} (Y_{\cdot j} - X\mathcal{B}_{\cdot j})^T(Y_{\cdot j} - X\mathcal{B}_{\cdot j})$, where $Y_{\cdot j}$ and $\mathcal{B}_{\cdot j}$ are jth column of Y and $\mathcal{B}$, respectively

$\because\ \hat{\mathcal{B}}_{\cdot j} = (X^TX)^{-1}X^TY_{\cdot j}$ minimize $(Y_{\cdot j} - X\beta_{\cdot j})^T(Y_{\cdot j} - X\beta_{\cdot j})$

$\therefore\ \hat{\mathcal{B}} = [\hat{\mathcal{B}}_{\cdot 1}, \cdots, \hat{\mathcal{B}}_{\cdot p}] = (X^TX)^{-1}X^TY = \underset{\mathcal{B}}{\arg\min}\, Q(\mathcal{B})$

② $E\{Y^T(I-H)Y \mid X\}$

$= E\{(X\mathcal{B}+E)^T(I-H)(X\mathcal{B}+E) \mid X\}$

$= E\{\mathcal{B}^T\underbrace{X^T(I-H)}_{0}X\mathcal{B} + E^T\underbrace{(I-H)X}_{0}\mathcal{B} + \mathcal{B}^T\underbrace{X^T(I-H)}_{0}E + E(I-H)E \mid X\}$

$= E\{\underbrace{(Y-X\mathcal{B})^T(I-H)(Y-X\mathcal{B})}_{A = [a_{ij}]_{p\times p}} \mid X\}$

$\because\ a_{ij} = (Y_{\cdot i} - X\mathcal{B}_{\cdot i})^T(I-H)(Y_{\cdot j} - X\mathcal{B}_{\cdot j})$

$= tr\{(I-H)(Y_{\cdot j} - X\mathcal{B}_{\cdot j})(Y_{\cdot i} - X\mathcal{B}_{\cdot i})^T\}$

$\therefore\ E(a_{ij} \mid X) = E[tr\{(I-H)E_{\cdot j}E_{\cdot i}^T\} \mid X]$

$= tr\{(I-H)E(E_{\cdot j}E_{\cdot i}^T \mid X)\}$

$= tr\{(I-H)E(E_{\cdot j}E_{\cdot i}^T)\}$

$= tr\{(I-H)E([E_{kj}E_{\ell i}]_{1\le k\le n,\ 1\le \ell\le n})\}$

$= tr\{(I-H)(\sigma_{ij}I)\}$ $\quad (\because\ E(E_{kj}E_{\ell i}) = \sigma_{ij}$ if $k=\ell$ and 0 otherwise$)$

$= \sigma_{ij}\, tr(I-H) = (n-q-1)\sigma_{ij}$

$\therefore\ E(A \mid X) = (n-q-1)\Sigma$

$\therefore\ E(\hat{\Sigma}_{LS}) = \Sigma$

- Maximum likelihood (ML) estimation (in need of (conditional) normality)
  - $\hat{\mathbf{B}}_{\mathrm{ML}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \hat{\mathbf{B}}_{\mathrm{LS}}$
  - $\hat{\mathbf{\Sigma}}_{\mathrm{ML}} = n^{-1}\mathbf{Y}^\top(\mathbf{I}-\mathbf{H})\mathbf{Y} = n^{-1}(n-q-1)\hat{\mathbf{\Sigma}}_{\mathrm{LS}}$
    * Given $\mathbf{X}$, $n\hat{\mathbf{\Sigma}}_{\mathrm{ML}} \sim W_p(\mathbf{\Sigma}, n-q-1)$

- Inference (in need of (conditional) normality)
  - Inference on $\mathbf{B}^\top\boldsymbol{a}$, given $\boldsymbol{a} \in \mathbb{R}^{q+1}$
    * Estimator $\hat{\mathbf{B}}_{\mathrm{ML}}^\top\boldsymbol{a}$
    * $100(1-\alpha)\%$ confidence region for $\mathbf{B}^\top\boldsymbol{a}$

$$\left\{\boldsymbol{u} \in \mathbb{R}^p : (\boldsymbol{u} - \widehat{\mathbf{B}}_{\mathrm{ML}}^\top\boldsymbol{a})^\top\widehat{\mathbf{\Sigma}}_{\mathrm{LS}}^{-1}(\boldsymbol{u} - \hat{\mathbf{B}}_{\mathrm{ML}}^\top\boldsymbol{a}) \le \boldsymbol{a}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\boldsymbol{a}\frac{p(n-q-1)}{n-q-p}F_{1-\alpha,p,n-p-q}\right\}$$

  - Inference on $\mathbf{Y}_0 = \mathbf{B}^\top\mathbf{X}_0 + \mathbf{E}_0$ with a new observation vector $\mathbf{X}_0 = [1, X_{01}, \ldots, X_{0q}]^\top \in \mathbb{R}^{q+1}$
    * Prediction $\hat{\mathbf{Y}}_0 = \mathbf{B}_{\mathrm{ML}}^\top\mathbf{X}_0$
    * $100(1-\alpha)\%$ prediction region for $\mathbf{Y}_0$

$$\left\{\boldsymbol{u} \in \mathbb{R}^p : (\boldsymbol{u} - \hat{\mathbf{Y}}_0)^\top\widehat{\mathbf{\Sigma}}_{\mathrm{LS}}^{-1}(\boldsymbol{u} - \hat{\mathbf{Y}}_0) \le \{1 + \mathbf{X}_0^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}_0\}\frac{p(n-q-1)}{n-q-p}F_{1-\alpha,p,n-p-q}\right\}$$

  - Inference on $\boldsymbol{a}^\top\mathbf{Y}_0 = \boldsymbol{a}^\top\mathbf{B}^\top\mathbf{X}_0 + \boldsymbol{a}^\top\mathbf{E}_0$, given $\boldsymbol{a} \in \mathbb{R}^p$ and a new observation vector $\mathbf{X}_0 = [1, X_{01}, \ldots, X_{0q}]^\top \in \mathbb{R}^{q+1}$
    * Prediction $\boldsymbol{a}^\top\hat{\mathbf{Y}}_0 = \boldsymbol{a}^\top\mathbf{B}_{\mathrm{ML}}^\top\mathbf{X}_0$
    * $100(1-\alpha)\%$ Scheffé's simultaneous prediction interval for $\boldsymbol{a}^\top\mathbf{Y}_0$

$$\boldsymbol{a}^\top\hat{\mathbf{Y}}_0 \pm \sqrt{\boldsymbol{a}^\top\widehat{\mathbf{\Sigma}}_{\mathrm{LS}}\boldsymbol{a}\{1 + \mathbf{X}_0^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}_0\}\frac{p(n-q-1)}{n-q-p}F_{1-\alpha,p,n-p-q}}$$

```r
install.packages(c('ellipse'))
options(digits = 4)
tear <- c(
  6.5, 6.2, 5.8, 6.5, 6.5, 6.9, 7.2, 6.9, 6.1, 6.3,
  6.7, 6.6, 7.2, 7.1, 6.8, 7.1, 7.0, 7.2, 7.5, 7.6
)
gloss <- c(
  9.5, 9.9, 9.6, 9.6, 9.2, 9.1, 10.0, 9.9, 9.5, 9.4,
  9.1, 9.3, 8.3, 8.4, 8.5, 9.2, 8.8, 9.7, 10.1, 9.2
)
(plastic <- cbind(tear, gloss))
rate <- factor(gl(2,10,length=nrow(plastic)), labels=c("Low", "High"))

# Model fitting
fit <- lm(cbind(tear, gloss) ~ rate)
summary(fit)

# Prediction
(Obs_new <- data.frame(rate = factor(c("High"), levels = c("Low", "High"))))
(prediction <- t(predict(fit, newdata = Obs_new)))

# Prediction region
n = nrow(model.matrix(fit))
```

```r
p = ncol(coef(fit))
q = ncol(model.matrix(fit))-1
(X <- model.matrix(fit))
(X0 <- t(model.matrix(~rate, Obs_new)))
(SigmaHatLS <- crossprod(resid(fit))/(n-q-1))
quad_form <- drop(t(X0) %*% solve(crossprod(X)) %*% X0)
fvalue = p*(n-q-1)/(n-p-q)*qf(0.95, p, n-p-q)

# 95% prediction region for Y0
c1 = sqrt((1 + quad_form)*fvalue)
eps1 = ellipse::ellipse(SigmaHatLS, centre = prediction, t = c1)
plot(eps1, type = "l", col='red')
points(prediction[1], prediction[2], pch = 19)

# 95% confidence region for t(B)X0
c2 = sqrt(quad_form*fvalue)
eps2 = ellipse::ellipse(SigmaHatLS, centre = prediction, t = c2)
lines(eps2, col='blue')

# 95% Scheffé's simultaneous prediction intervals for entries of Y0
a1 = c(1,0)
c(
  t(a1) %*% prediction - (t(a1) %*% SigmaHatLS %*% a1)^.5 * c1,
  t(a1) %*% prediction + (t(a1) %*% SigmaHatLS %*% a1)^.5 * c1
) # for tear
a2 = c(0,1)
c(
  t(a2) %*% prediction - (t(a2) %*% SigmaHatLS %*% a2)^.5 * c1,
  t(a2) %*% prediction + (t(a2) %*% SigmaHatLS %*% a2)^.5 * c1
) # for gloss
```