MAST90104 - MST Revision

Weichang Yu

Room 108, Old Geology South Bldg
School of Mathematics and Statistics, University of Melbourne

Tick the option(s) that is/are always TRUE.

- ☐ (a) $\mathbf{A}$ and $\mathbf{C}$ are two symmetric square matrices of order p. Then, $(\mathbf{AC})^T = \mathbf{CA}$.
- ☐ (b) $\mathbf{A}$ is an m by n matrix with rank r. Then, $\min\{m, n\} \leq r \leq \max\{m, n\}$.
- ☐ (c) $\mathbf{A}$ are $\mathbf{C}$ are nonsingular matrices. Then, $\mathbf{AC}$ is nonsingular.
- ☐ (d) $\mathbf{A}$ are $\mathbf{C}$ are nonsingular matrices. Then, $\mathbf{A} + \mathbf{C}$ is nonsingular.
- ☐ (e) $\mathbf{A}$ is a symmetric matrix. Then, there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^T\mathbf{AP} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix with entries equal to 0 or 1.

Random distributions

Let $y_1 \sim N(-3, 1)$ and $y_2 \sim N(1, 1)$ be two independent normal random variables. We can express the distribution of $\mathbf{y} = (y_1, y_2)^T$ as:

$$\mathbf{y} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

(a) Write down the numerical forms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

(b) Let $\Omega = \begin{pmatrix} -2 & 1 \\ 0.5 & 2 \end{pmatrix}$. Does $\Omega\mathbf{y}$ follow a multivariate normal distribution? If yes, write down its mean vector and covariance matrix.

(c) Let $\mathbf{B} = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$ and $\psi = (-3.5, -7)^T$. Is $\mathbf{y}^T\mathbf{B}\mathbf{y}$ and $\psi^T\mathbf{y}$ independent? Explain your answer.

Linear model estimation theory

Consider the general linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is an n by $k+1$ full rank matrix. Consider the estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

(a) State the additional assumptions required for $\widehat{\boldsymbol{\beta}}$ to be the unique best linear unbiased estimator.

(b) Denote the i-th entry of $\mathbf{y}$ as y_i. Assume that each y_i is an independent random variable drawn from $N(\mathbf{x}_i^T\boldsymbol{\beta}, d_i)$ respectively, where $\mathbf{x}_i^T = (1, x_{i1}, \ldots, x_{ik})$ and $d_i > 0$. Is $\widehat{\boldsymbol{\beta}}$ a uniformly minimum variance unbiased estimator for $\boldsymbol{\beta}$? Justify your answer.

Hypothesis testing

Consider the squid dataset. We build a linear regression model to predict `weight` using the remaining variables in the dataset. Note that there is no missing data in the dataset.

```
> SquidDat <- read.csv("squid.csv",header=TRUE)
> names(SquidDat)
[1] "weight"      "beak"        "wing"        "beaktonotch"
[5] "notchtowing" "width"
> model.lm <- lm(weight~., data=SquidDat)
> summary(model.lm)

Call:
lm(formula = weight ~ ., data = SquidDat)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2610 -0.5373  0.1355  0.5120  0.8611

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.5122     0.9336  -6.976 3.13e-06 ***
beak          1.9994     2.5733   0.777  0.44851
wing         -3.6751     2.7737  -1.325  0.20378
beaktonotch   2.5245     6.3475   0.398  0.69610
notchtowing   5.1581     3.6603   1.409  0.17791
width        14.4012     4.8560   2.966  0.00911 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7035 on 16 degrees of freedom
Multiple R-squared:  0.9633,
```

Hypothesis testing and model selection

(a) Write down the sample size on the dataset.

(b) Consider a model relevance test for the fitted model. Write down the null distribution of the test statistic and state any assumptions you have made (You do not have to compute the numerical value of the test statistic).

(c) Use the R output to test the hypothesis $H_0 : \beta_{\text{wing}} = 0$ at 10% significance level.

(d) Describe the difference between multiple R^2 and adjusted-R^2.

(e) Compute the adjusted-R^2 of the fitted model.

(f) Consider the backward elimination procedure for model selection. Which coefficient would you drop from the model first?