

STAT 3690 Lecture 18

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 11, 2022

Testing for nested models

- $H_0 : E(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}_{(0)}\mathbf{B}_{(0)}$ (nested model) vs. $H_1 : E(\mathbf{Y} \mid \mathbf{X}) = \mathbf{X}_{(0)}\mathbf{B}_{(0)} + \mathbf{X}_{(1)}\mathbf{B}_{(1)}$ (full model)
 - When $\mathbf{X}_{(0)}$ has only the column of ones, we are testing the empty model (i.e., only the intercept) against the full model.
 - When $\mathbf{X}_{(1)}$ only contains one column, we are testing for the significance of that variable.

- Likelihood ratio

$$\lambda = \left(\frac{\det \hat{\Sigma}_{\text{ML}, H_0}}{\det \hat{\Sigma}_{\text{ML}}} \right)^{-n/2} = \left[\det \left\{ (\hat{\Sigma}_{\text{ML}, H_0} - \hat{\Sigma}_{\text{ML}}) \hat{\Sigma}_{\text{ML}}^{-1} + \mathbf{I} \right\} \right]^{-n/2}$$

- Alternatives to likelihood ration

- Suppose $\eta_1 \geq \dots \geq \eta_p$ are eigenvalues of $(\hat{\Sigma}_{\text{ML}, H_0} - \hat{\Sigma}_{\text{ML}}) \hat{\Sigma}_{\text{ML}}^{-1}$
- Wilks' lambda: $\prod_i (1 + \eta_i)^{-1}$
- Pillai's trace: $\sum_i \{\eta_i (1 + \eta_i)^{-1}\}$
- Hotelling-Lawley trace: $\sum_i \eta_i$
- Roy's largest root: $\eta_1 (1 + \eta_1)^{-1}$
- When $\mathbf{X}_{(1)}$ has only one column, all four tests are equivalent; as n increases, all four tests give similar results.

```
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
)
opacity <- c(
  4.4, 6.4, 3.0, 4.1, 0.8, 5.7, 2.0, 3.9, 1.9, 5.7,
  2.8, 4.1, 3.8, 1.6, 3.4, 8.4, 5.2, 6.9, 2.7, 1.9
)
rate <- factor(gl(2,10,length=n), labels=c("Low", "High"))
additive <- factor(gl(2,5,length=nrow(X)), labels=c("Low", "High"))

# Testing the necessity of interaction
fit0 <- lm(cbind(tear, gloss, opacity) ~ rate+additive)
```

```
fit1 = lm(cbind(tear, gloss, opacity) ~ rate*additive)
anova(fit1, fit0, test='Wilks')
anova(fit1, fit0, test='Pillai')
anova(fit1, fit0, test='Hotelling')
anova(fit1, fit0, test='Roy')
```

Information criteria

- Akaike's information criterion (AIC)

$$- \ln \text{Likelihood} + 2 \times \text{number of parameters to estimate}$$

- Number of parameters to estimate in \mathbf{B} and $\mathbf{\Sigma}$: $p(q+1) + p(p+1)/2$
- Smaller is better

- Bayesian information criterion (BIC)

$$- \ln \text{Likelihood} + \ln n \times \text{number of parameters to estimate}$$

- Model selection using information criteria proceeds as follows
 - Select models of interest M_1, \dots, M_K . They do not need to be nested.
 - * Candidate models should be selected using domain-specific expertise, if possible. Or, you can go through all possible models.
 - Compute the specific information criterion for each model.
 - Select the model with the smallest value of the information criterion.

```
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
)
opacity <- c(
  4.4, 6.4, 3.0, 4.1, 0.8, 5.7, 2.0, 3.9, 1.9, 5.7,
  2.8, 4.1, 3.8, 1.6, 3.4, 8.4, 5.2, 6.9, 2.7, 1.9
)
n = length(opacity)
rate <- factor(gl(2,10,length=n), labels=c("Low", "High"))
additive <- factor(gl(2,5,length=n), labels=c("Low", "High"))

fit0 <- lm(cbind(tear, gloss, opacity) ~ rate)
logLik(fit0)
AIC(fit0)
BIC(fit0)

logLik.mlm <- function(object, ...) {
  resid <- residuals(object)
  Sigma_ML <- crossprod(resid)/nrow(resid)
  ans <- sum(mvtnorm::dmvnorm(resid, sigma = Sigma_ML, log = TRUE))
}
```

```

    df <- prod(dim(coef(object))) + choose(ncol(Sigma_ML) + 1, 2)
    attr(ans, "df") <- df
    class(ans) <- "logLik"
    return(ans)
}
logLik(fit0)
AIC(fit0)
BIC(fit0)

fit1 <- lm(cbind(tear, gloss, opacity) ~ additive)
AIC(fit1)
BIC(fit1)

```