STAT 525

Chapter 6 (Optional reading materials) Multiple Regression - Distribution Theory

Dr. Qifan Song

Matrix Algebra

- Eigendecomposition: for any symetric matrix A, there always exist some orthogonal matrix P and diagnal matrix Λ such that

$$A = P\Lambda P'.$$

- $PP' = I$, and the diagnal values in Λ are called eigenvalues of A. Number of non-zero eigenvalues is exactly the rank of matrix A.

- If A is symmetric and idempotent, then

$$P\Lambda P' = A = A^2 = P\Lambda P' P\Lambda P' = P\Lambda^2 P',$$

  thus all its eigenvalues are either 1 or 0, and the number of 1's is the rank of matrix A.

Quadratic form

If matrix $A = [a_{ij}]$, then

$$f(x) = \sum_{1 \le i,j \le p} a_{ij} x_i x_j = x'Ax$$

- If $x \sim N(\mu, \sigma^2 I)$, then what is the distribution of $x'Ax$?
- If A is positive definite, then $x'Ax$ belongs to the family of Chi-square distributions.
- Example:

$$\begin{aligned} s^2 &= \frac{\mathbf{e}'\mathbf{e}}{n-p} = \frac{Y'(I-H)(I-H)Y}{n-p} \\ &= \frac{Y'(I-H)Y}{n-p}, \end{aligned}$$

where $Y \sim N(X\beta, \sigma^2 I)$.

Quadratic form

If $x \sim N(\mu, \sigma^2 I)$, A is symmetric and idempotent, $r = rank(A)$ and $A\mu = 0$, then $x'Ax/\sigma^2 \sim \chi^2_r$.

Proof:

$$\begin{aligned} x'Ax &= x'A'Ax = (x' - \mu')A'A(x - \mu) = (x' - \mu')A(x - \mu) \\ &= (x' - \mu')P'\Lambda P(x - \mu) = N(0, \sigma^2 I)'\Lambda N(0, \sigma^2 I). \end{aligned}$$

where $P(x - \mu) \sim N(0, \sigma^2 I)$.

Example

- SSE:

$$\begin{aligned} SSE &= Y'(I-H)Y \\ (I-H)EY &= (I-H)X\beta = X\beta - X(X'X)^{-1}X'X\beta = 0 \\ rank(I-H) &= n-p, \end{aligned}$$

  if all columns of X are linearly independent.

Example

- SSR (under null hypothesis $\beta_1 = \ldots = \beta_{p-1} = 0$):

$$\begin{aligned} SSR &= \sum(\hat{Y}_i - \bar{Y})^2 = Y'(H-K)'(H-K)Y \\ (H-K)'(H-K) &= HH - 2HK + K^2 = H - 2K + K = H - K \\ (H-K)E(Y) &= E(Y) - K\mathbf{1}\beta_0 = 0 \\ rank(H-K) &= p - 1, \end{aligned}$$

where K is a matrix of all $1/n$ and $\mathbf{1}$ is a vector of all 1's. We can use the fact that $K = XK_0$ and K_0's first row is $1/n$'s and the rest are 0.

Independence between Quadratic form

If A and B are symmetric and idempotent and x is a multivariate normal, then if $AB = 0$, $x'Ax$ and $x'Bx$ are independent. Furthermore, $x'Ax/x'Bx \sim F_{r_1,r_2}$ with $r_1 = rank(A)$ and $r_2 = rank(B)$.

Proof:
Ax and Bx are both normal, and their covariance is $Cov(Ax, Bx) = A(\sigma^2 I)B' = \sigma^2 AB = 0$. Thus, Ax and Bx are independent, which implies $(Ax)'Ax$ and $(Bx)'Bx$ are independent as well.

Example

- ANOVA F-test (under null hypothesis $\beta_1 = \ldots = \beta_{p-1} = 0$).

  To show that MSR/MSE follows F-test, by existence results in previous slides, it is sufficient to show $(I-H)(H-K) = 0$, as

  $$(I-H)(H-K) \; = \; H - K - H^2 + HK = 0$$

- Proof of independence between b and s^2 (required for t-test for β_i):

  $$\begin{aligned} b &= (X'X)^{-1}X'Y \\ s^2 &= Y'(I-H)Y/(n-p) \\ (X'X)^{-1}X'(I-H) &= (X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X' = 0 \end{aligned}$$

  Thus, b and s^2 are completely independent.

Example

- General Linear F test (includes lack-of-fit F test):
  - $H_0 : \mathbf{Y} = \mathbf{X}_2\beta_1$ and $H_1 : \mathbf{Y} = \mathbf{X}_1\beta_2$, where $\mathbf{X}_2 = \mathbf{X}_1 * \mathbf{Z}$ with non-full-rank $\mathbf{Z}$.
  - $p_1 = rank(X_1) > p_2 = rank(X_2)$.
  - Hat matrics $H_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T$ for $i = 1, 2$.
  - $F = [(SSE(R) - SSE(F))/(p_1 - p_2)]/[SSE(F)/(n - p_1)]$

$$\begin{aligned}
SSE(F) &= Y'(I - H_1)Y \\
SSE(R) - SSE(F) &= Y'(I - H_2)Y - Y'(I - H_1)Y \\
&= Y'(H_1 - H_2)Y \\
(I - H_1)(H_1 - H_2) &= H_1 - H_1 - H_2 + H_1H_2 \\
&= H_1H_2 - H_2 = 0
\end{aligned}$$