Stat873 **Exam 3** **April 10, 2024**

Name:

1. For Model $Y \sim N(X\beta, \sigma^2\Sigma)$ where $X \in R^{n\times p}$ with $\text{rank}(X) = p$ and $1_n \notin \mathcal{R}(X)$, $R^2 = \frac{SSM}{SSTO}$, called the coefficient of determination, is the proportion of the variation in Y explained by the model.

   (1) Define F in ANOVA table via R^2. (10 points)

$$F = \frac{MSM}{MSE} = \frac{SSM}{SSE} \cdot \frac{n-p}{p} = \frac{SSM}{SSTO-SSM} \cdot \frac{n-p}{p} = \frac{R^2}{1-R^2} \cdot \frac{n-p}{p} = \frac{R^2/p}{(1-R^2)/(n-p)}.$$

   (2) Replacing SSE and SSTO in $R^2 = \frac{SSM}{SSTO} = \frac{SSTO-SSE}{SSTO}$ by MSE and MSTO we obtain R^2_{adj}, the adjusted R^2. Express F via R^2_{adj}. (20 points)

$R^2_{adj} = \frac{MSTO-MSE}{MSTO} = 1 - \frac{MSE}{MSTO}$. So $\frac{MSE}{MSTO} = 1 - R^2_{adj}$.

$$\begin{aligned} F &= \frac{MSM}{MSE} = \frac{SSM}{MSE} \cdot \frac{1}{p} = \frac{U.SSTO-SSE}{MSE} \cdot \frac{1}{p} = \frac{n\cdot MSTO-(n-p)\cdot MSE}{p\cdot MSE} = \frac{n-(n-p)\frac{MSE}{MSTO}}{p\cdot\frac{MSE}{MSTO}} \\ &= \frac{n-(n-p)(1-R^2_{adj})}{p(1-R^2_{adj})} = 1 - \frac{n}{p} + \frac{n}{p} \cdot \frac{1}{1-R^2_{adj}} = 1 + \frac{R^2_{adj}/p}{(1-R^2_{adj})/n}. \end{aligned}$$

   (3) R^2_{adj} could assume negative values. Find the condition on R^2 for $R^2_{adj} < 0$. (20 points)

$$\begin{aligned} R^2_{adj} < 0 \quad &\Longleftrightarrow \quad \frac{MSTO-MSE}{MSE} < 0 \Longleftrightarrow MSTO < MSE \Longleftrightarrow \frac{SSTO}{n} < \frac{SSE}{n-p} \\ &\Longleftrightarrow \quad \frac{n-p}{n} < \frac{SSE}{SSTO} = \frac{SSTO-SSM}{SSTO} = 1 - R^2 \Longleftrightarrow -\frac{p}{n} < -R^2 \\ &\Longleftrightarrow \quad R^2 < \frac{p}{n}. \end{aligned}$$

So $R^2_{adj} < 0$ if and only if $R^2 < \frac{p}{n}$.

2. Model M: $Y \sim N(X\beta, \sigma^2 I_n)$ where $X = (X_I, X_{II}) \in R^{n\times p}$ has full column rank and $X_I \in R^{n\times p_1}$. With $\beta = \begin{pmatrix} \beta_I \\ \beta_{II} \end{pmatrix} \in R^p$ where $\beta_I \in R^{p_1}$, $H_0 : \beta_{II} = 0$ reduces M to Model M$_*$. Let SSE and SSE$_*$ be from M and M$_*$. Define SSD=SSE$_*$−SSE.

(1) Fill out the form below. Write SS as quadratic forms. (15 points)

Source	SS	DF
Difference	SSD= $Y'(XX^+ - X_I X_I^+)Y$	$p - p_1$
Error M	SSE= $Y'(I - XX^+)Y$	$n - p$
Error M$_*$	SSE$_*$= $Y'(I - X_I X_I^+)Y$	$n - p_1$

(2) Derive the distribution of $\frac{SSD}{\sigma^2}$ under H_0. (20 points)

Under H_0, $Y \sim N(X_I\beta_I, \sigma^2 I_n)$.
$\frac{SSD}{\sigma^2} = Y'AY$ where $A = \frac{XX^+ - X_I X_I^+}{\sigma^2}$.
Note that $XX^+X_IX_I^+ = X_IX_I^+ = X_IX_I^+XX^+$. So
(i) $A\sigma^2 I_n A = A = A'$.
(ii) $(X_I\beta_I)'A(X_I\beta_I) = (X_I\beta_I)'A(X_IX_I^+)(X_I\beta_I) = 0$.
(iii) $\text{tr}(A\sigma^2 I_n) = p - p_1$.
Thus $\frac{SSD}{\sigma^2} \overset{H_0}{\sim} \chi^2(p - p_1)$.

(3) Show that SSD and SSE are independent. (15 points)

$Y \sim N(X\beta, \sigma^2 I_n)$, $SSD = Y'(XX^+ - X_IX_I^+)Y$ and $SSE = Y'(I - XX^+)Y$.

$(XX^+ - X_IX_I^+)\sigma^2 I_n(I - XX^+) = \sigma^2(XX^+ - X_IX_I^+ - XX^+ + X_IX_I^+) = 0$.

So SSD and SSE are independent.