L20: ANOVA table

1. Space orthogonal decomposition

   (1) Definition

$$\begin{aligned}
& \mathcal{R}(L) \text{ is orthogonally decomposed by } \mathcal{R}(A) \text{ and } \mathcal{R}(B) \\
\overset{notation}{\Longleftrightarrow}\ & \mathcal{R}(L) = \mathcal{R}(A)\dot{\oplus}\mathcal{R}(B) \\
\overset{def}{\Longleftrightarrow}\ & \mathcal{R}(L) = \mathcal{R}(A) + \mathcal{R}(B) \text{ and } \mathcal{R}(A) \perp \mathcal{R}(B).
\end{aligned}$$

   (2) Properties
   Space orthogonal decomposition implies dimension, projection matrix, projection vector and squared norm of projection vector decompositions, i.e., if $\mathcal{R}(L) = \mathcal{R}(A)\dot{\oplus}\mathcal{R}(B)$, then

   (i) $\operatorname{rank}(L) = \operatorname{rank}(A) + \operatorname{rank}(B)$,
   (ii) $LL^+ = AA^+ + BB^+$,
   (iii) $\pi(y|\mathcal{R}(L)) = \pi(y \mid \mathcal{R}(A)) + \pi(y \mid \mathcal{R}(B))$,
   (iv) $\|\pi(y|\mathcal{R}(L))\|^2 = \|\pi(y \mid \mathcal{R}(A))\|^2 + \|\pi(y \mid \mathcal{R}(B))\|^2$.

   Proof (i) $\mathcal{R}(L) = \mathcal{R}(A) + \mathcal{R}(B) = \mathcal{R}[(A,\, B)]$.
   So $\operatorname{rank}(L) = \operatorname{rank}[(A,\, B)] = \operatorname{rank}(A) + \operatorname{rank}(B) - \dim[\mathcal{R}(A) \cap \mathcal{R}(B)]$.
   But $\mathcal{R}(A) \perp \mathcal{R}(B) \Longrightarrow \mathcal{R}(A) \cap \mathcal{R}(B) = \{0\}$. (i) follows.
   (ii) $\mathcal{R}(L) = \mathcal{R}(A) + \mathcal{R}(B) = \mathcal{R}[(A,\, B)]$. So $LL^+ = (A,\, B)(A,\, B)^+$.
   But $\mathcal{R}(A) \perp \mathcal{R}(B) \Longrightarrow A'B \Longrightarrow (A,\, B)^+ = \begin{pmatrix} A^+ \\ B^+ \end{pmatrix}$. (ii) follows.
   (iii) $\pi(y \mid \mathcal{R}(L)) = LL^+y = AA^+y + BB^+y = \pi(y \mid \mathcal{R}(A)) + \pi(y \mid \mathcal{R}(B))$.
   (iv) $\pi(y \mid \mathcal{R}(A)) \perp \pi(y \mid \mathcal{R}(B))$. By Pythagorean

2. Space orthogonal decomposition in linear model

   (1) Linear model
   In linear model for y, y_i, $i = 1, .., n$, are observed. So there is a $\mathbf{y} = (y_1, ..., y_n)' \in R^n$. By linear model specification, $\mathbf{y} = X\beta + e$ where $e \sim N(0,\, \sigma^2 I_n)$. So $E(y) \in \mathcal{R}(X) \subset R^n$. Suppose one can identify a known space $\mathcal{R}(D) \subset \mathcal{R}(X)$. Then

$$\mathcal{R}(D) \subset \mathcal{R}(X) \Longleftrightarrow D = XT \text{ for some } T \Longleftrightarrow \mathcal{R}^{\perp}(X) \subset \mathcal{R}^{\perp}(D)$$

   where $\mathcal{R}^{\perp}(D) = \mathcal{R}(I - DD^+)$ and $\mathcal{R}^{\perp}(X) = \mathcal{R}(I - XX^+)$.

   (2) Orthogonal decomposition of $\mathcal{R}(I - DD^+)$
   $\mathcal{R}(I - DD^+) = \mathcal{R}(I - XX^+)\dot{\oplus}\mathcal{R}(XX^+ - DD^+)$.

   Proof: Note that $D = XT$ implies $XX^+DD^+ = DD^+XX^+ = DD^+$ since
   $XX^+DD^+ = XX^+XTD^+ = XTD^+ = DD^+$.
   $\subset$**:** If $u \in \mathcal{R}(I - DD^+)$, then $u = (I - DD^+)v$ for some v.
   Thus $u = (I - XX^+)v + (XX^+ - DD^+)v \in \mathcal{R}(I - XX^+) + \mathcal{R}(XX^+ - DD^+)$.
   $\supset$**:** $(I - DD^+)(I - XX^+) = I - XX^+ \Longrightarrow \mathcal{R}(I - XX^+) \subset \mathcal{R}(I - DD^+)$.
   $(I - DD^+)(XX^+ - HH^+) = XX^+ - HH^+ \Longrightarrow \mathcal{R}(XX^+ - HH^+) \subset \mathcal{R}(I - DD^+)$.
   $\perp$**:** $(I - XX^+)'(XX^+ - HH^+) = 0 \Longrightarrow \mathcal{R}(I - XX^+) \perp \mathcal{R}(XX^+ - HH^+)$.

(3) Rank and SS decompositions
Note that $I - DD^+$, $I - XX^+$ and $XX^+ - DD^+$ are all symmetric and idempotent. So they are projection matrices for the corresponding spaces in

$$\mathcal{R}(I - DD^+) = \mathcal{R}(I - XX^+)\dot{\oplus}\mathcal{R}(XX^+ - DD^+).$$

Let r_0 and r be the ranks of $\mathcal{R}(D)$ and $\mathcal{R}(X)$. Then
$(n - r_0) = (n - r) + (r - r_0)$ and $y'(I - DD^+)y = y'(I - XX^+)y + y'(XX^+ - HH^+)y$ are dimension (rank) and squared norm (SS) of the projection vector decompositions.

3. A general ANOVA table

(1) SS
Under $E(y) \in \mathcal{R}(D)$ and $E(y) \in \mathcal{R}(X)$, $E(y)$ is estimated by DD^+y and XX^+y with error $(I - DD^+)y$ and $(I - XX^+)y$.
The total variation in y is defined as SSTO $= y'(I - DD^+)y$.
The variation unexplained by the model due to error is SSE $= y'(I - XX^+)y$.
Hence SSM=SSTO-SSE=$y'(XX^+ - HH^+)y$ is the variation explained by the model.

(2) A general ANOVA able

Source	SS	DF	MS	F	p
Model	SSM= $y'(XX^+ - DD^+)y$	$r - r_0$	MSM	MSM/MSE	$P(F(r - r_0, n - r) > F_{ob})$
Error	SSE= $y'(I - XX^+)y$	$n - r$	MSE		
Total	SSTO= $y'(I - DD^+)y$	$n - r_0$			

With this table a LRT on the usefulness of the model can be carried out.

Ex1: For regression with intercept $y = X\beta + e$, $e \sim N(0, \sigma^2 I_n)$, the first column of $X \in R^{n\times p}$ is $1_n = Xe_1$. So $D = 1_n$ and $\mathcal{R}(1_n) \subset \mathcal{R}(X)$ with $r_0 = 1$.

Source	SS	DF	MS	F	p
Model	SSM= $y'(XX^+ - 1_n 1_n^+)y$	$r - 1$	MSM	MSM/MSE	$P(F(r - 1, n - r) > F_{ob})$
Error	SSE= $y'(I - XX^+)y$	$n - r$	MSE		
Total	SSTO= $y'(I - 1_n 1_n^+)y$	$n - 1$			

Ex2: For regression withotu intercept $y = X\beta + e$, $e \sim N(0, \sigma^2 I_n)$, $\mathcal{R}(0) \subset \mathcal{R}(X)$. Thus $D = 0$ with $r_0 = 0$. So

Source	SS	DF	MS	F	p
Model	SSM= $y'XX^+y$	r	MSM	MSM/MSE	$P(F(r, n - r) > F_{ob})$
Error	SSE= $y'(I - XX^+)y$	$n - r$	MSE		
Total	SSTO= $y'Iy$	n			

Ex3: For one-way ANOVA $y = X\mu + e$, $e \sim N(0, \sigma^2 I_n)$, with the p columns being p indicators, $1_n = X1_p$, i.e., $D = 1_n$ and $\mathcal{R}(1_n) \subset \mathcal{R}(X)$. Here $r_0 = 1$ and $r = p$. So

Source	SS	DF	MS	F	p
Model	SSM= $y'(XX^+ - 1_n 1_n^+)y$	$p - 1$	MSM	MSM/MSE	$P(F(p - 1, n - p) > F_{ob})$
Error	SSE= $y'(I - XX^+)y$	$n - p$	MSE		
Total	SSTO= $y'(I - 1_n 1_n^+)y$	$n - 1$			