Formulae For BTRY/STSCI 4030

1 Multiple regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ip} + \epsilon_i$$

with $\epsilon_i \sim N(0, \sigma^2)$; or

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e},\ \boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 I)$$

2 Formulae

1. Estimate
$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}$$
2. Fitted values
$$\hat{\boldsymbol{y}} = X\hat{\boldsymbol{\beta}} = X(X^T X)^{-1} X^T \boldsymbol{y} = H\boldsymbol{y}$$
3. Residuals
$$\hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (I - H)\boldsymbol{y}$$

3 Sums of Squares

1. Sum of Squared Errors
$$\text{SSE} = \hat{\boldsymbol{e}}^T \hat{\boldsymbol{e}} = \boldsymbol{y}^T (I - H)\boldsymbol{y}$$
2. Sum of Squares for Regression
$$\text{SSR} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \hat{\boldsymbol{y}}^T C \hat{\boldsymbol{y}} = \boldsymbol{y}^T HCH\boldsymbol{y}$$
3. Total (corrected) sum of squares
$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \boldsymbol{y}^T C \boldsymbol{y}$$
4. Sums of squares for $\boldsymbol{x}$, or $\boldsymbol{y}$, or $\boldsymbol{xy}$ (also $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$)
$$\text{SXX} = \sum_{i=1}^{n} (x_i - \bar{x}) = \boldsymbol{x}^T C \boldsymbol{x},\ \text{SXY} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \boldsymbol{x}^T C \boldsymbol{y}$$
5. For simple linear regression $X = (\boldsymbol{1}, \boldsymbol{x})$,
$$(X^T X)^{-1} = \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{SXX} & -\frac{\bar{x}}{SXX} \\ -\frac{\bar{x}}{SXX} & \frac{1}{SXX} \end{pmatrix}$$

4 ANOVA Tables: $C = (I - H) + HCH$

1. Mean Square = (Sum of Squares)/df

2. ANOVA Table

Source	Sum of Squares	df
Regression	SSR=$\boldsymbol{y}^T HCH\boldsymbol{y}$	tr$(HCH) = p$
Error	SSE=$\boldsymbol{y}^T(I - H)\boldsymbol{y}$	tr$(I - H) = n - p - 1$
Total	SST=$\boldsymbol{y}^T C\boldsymbol{y}$	tr$(C) = n - 1$

3. Sequentially, if $X_k = [1, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k]$, $H_k = X_k(X_k^T X_k)^{-1}X_k^T$

4. In a table, note that $H_k C H_k - H_{k-1} C H_{k-1} = (H_k - \bar{J}) - (H_{k-1} - \bar{J}) = H_k - H_{k-1}$

Source	Sum of Squares	df
$\boldsymbol{x}_1$	SSR=$\boldsymbol{y}^T H_1 C H_1 \boldsymbol{y}$	tr$(H_1 C H_1) = 1$
$\boldsymbol{x}_2 \vert X_1$	SSR=$\boldsymbol{y}^T(H_2 - H_1)\boldsymbol{y}$	tr$(H_2 - H_1) = 1$
$\cdots$	$\cdots$	$\cdots$
$\boldsymbol{x}_p \vert X_{p-1}$	SSR=$\boldsymbol{y}^T(H_p - H_{p-1})\boldsymbol{y}$	tr$(H_p - H_{p-1}) = 1$
Error	SSE=$\boldsymbol{y}^T(I - H)\boldsymbol{y}$	tr$(I - H) = n - p - 1$
Total	SST=$\boldsymbol{y}^T C\boldsymbol{y}$	tr$(C) = n - 1$

5. R^2 gives relative size of fitted values versus observations

$$R^2 = \frac{\boldsymbol{y}^T HCH\boldsymbol{y}}{\boldsymbol{Y}^T C\boldsymbol{y}},\ 1 - R^2 = \frac{\boldsymbol{y}^T(I - H)\boldsymbol{y}}{\boldsymbol{y}^T C\boldsymbol{y}} = \frac{\boldsymbol{e}^T\boldsymbol{e}}{\boldsymbol{y}^T C\boldsymbol{y}}$$

6. VIF (variance inflation factors) for a covariate $\boldsymbol{x}_j$ is $1/(1 - R^2)$ for predicting $\boldsymbol{x}_j$ from X_{-j}:

$$\text{VIF}_j = \frac{1}{1 - \frac{\boldsymbol{x}_j^T H_{-j} C H_{-j} \boldsymbol{x}_j}{\boldsymbol{x}_j^T C \boldsymbol{x}_j}} = \frac{\boldsymbol{x}_j^T C \boldsymbol{x}_j}{\boldsymbol{x}_j (I - H_{-j}) \boldsymbol{x}_j}$$

5 Some Matrix Algebra

1. Eigen-decomposition
$$M_{n\times k} = V_{n\times k} D_{k\times k} U_{k\times k}^T$$
With $U^TU = V^TV = I$, orthonormal and D diagonal.

2. Special Cases

   (a) **Square and symmetric** $M = UDU^T$.

   (b) **Positive Definite** $\boldsymbol{x}^T M \boldsymbol{x} > 0$ for all $\boldsymbol{x} \Leftrightarrow d_{ii} > 0$.

   (c) **Idempotent** $M^2 = M$: then d_{ii} either 1 or 0; tr(M) = rank of M and if $X \in$ span(M) then $MX = X$.

   (d) In particular, if M, M_1 idempotent and span(M_1) contained in span(M) then $(M - M_1)^2 = M - M_1$ and tr$(M - M_1) =$ tr$(M) -$ tr(M_1).

   (e) Examples: I, $\bar{J}$, C, H, $HCH = H - \bar{J}$ and note that $HX = X$.

3. Inverses. Note that $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

   (a) 2×2 matrices
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -b & a \end{bmatrix}.$$

   (b) A specialist block matrix if $X = [\boldsymbol{x}_1, X_{-1}]$ then
$$\begin{aligned}
(X^TX)^{-1} &= \begin{bmatrix} \boldsymbol{x}_1^T\boldsymbol{x}_1 & \boldsymbol{x}_1^T X_{-1} \\ X_{-1}^T\boldsymbol{x}_1 & X_{-1}^T X_{-1} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \frac{1}{r} & -\frac{1}{r}(X_{-1}^T X_{-1})^{-1} X_{-1}^T \boldsymbol{x}_1 \\ -\frac{1}{r}\boldsymbol{x}_1^T X_{-1}(X_{-1}^T X_{-1})^{-1} & \left(X_{-1}^T X_{-1} - \frac{X_{-1}^T\boldsymbol{x}_1\boldsymbol{x}_1^T X_{-1}}{\boldsymbol{x}_1^T\boldsymbol{x}_1}\right)^{-1} \end{bmatrix}
\end{aligned}$$
   with $r = \boldsymbol{x}_1^T\boldsymbol{x}_1 - \boldsymbol{x}_1^T X_{-1}(X_{-1}^T X_{-1})^{-1} X_{-1}^T \boldsymbol{x}_1$.

6 Distributions

1. **Normal/Gaussian**: (μ, Σ)

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma) \Rightarrow f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

   (a) Linear transforms

$$\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma) \Rightarrow A\boldsymbol{x} + b \sim N(A\mu + b, A\Sigma A^T)$$

   (b) In particular, in linear regression $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T Y \sim N(\boldsymbol{\beta}, \sigma^2 (X^T X)^{-1})$

   (c) Uncorrelated $\Leftrightarrow$ Independent:

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right), \ \Sigma_{12} = 0 \Leftrightarrow \boldsymbol{x}_1 \perp \boldsymbol{x}_2$$

   (d) In particular, $\mathrm{cor}(\hat{\boldsymbol{e}}, \hat{\boldsymbol{y}}) = \sigma^2 (I - H) H = 0$.

2. $\boldsymbol{\chi_k^2}$: $(k, 2k)$

$$\boldsymbol{z} \sim N(0, I_{k \times k}) \Rightarrow x = \sum_{i=1}^{k} z_i^2 = \boldsymbol{z}^T \boldsymbol{z} \sim \chi_k^2$$

   (a) Let $(I - H) = UDU^T$ then $\boldsymbol{u} = U^T \boldsymbol{e} \sim N(0, I)$ and

$$\mathrm{SSE} = \boldsymbol{e}^T (I - H) \boldsymbol{e} = \boldsymbol{u}^T D \boldsymbol{u} = \sum_{i=1}^{n} d_{ii} u_i^2 = \sum_{i=1}^{n-p-1} u_i^2 \sim \sigma^2 \chi^2_{n-p-1}$$

   (b)

$$E\hat{\sigma}^2 = E\mathrm{MSE} = E\left(\frac{\mathrm{SSE}}{n - p - 1} \right) = \sigma^2.$$

   (c) Noncentral if $\boldsymbol{z} \sim N(\boldsymbol{\mu}, I)$ then $x \sim \chi_k^2(\boldsymbol{\mu}^T \boldsymbol{\mu}/2)$.

3. $\boldsymbol{t_k}$: $(0, k/(k-2))$

$$(z \sim N(0,1),\ x \sim \chi_k^2) \to \frac{z}{\sqrt{X/k}} \sim t_k$$

(a) Particularly

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 (X^T X)^{-1}_{jj}}} \sqrt{\frac{1}{\sigma^2} \frac{\text{SSE}}{n-p-1}}^{-1} \sim t_{n-p-1}$$

(b) Noncentral if $z \sim N(\mu, 1)$ then $z/\sqrt{x/k} \sim t_k(\mu)$.

4. $\boldsymbol{F_l^k}$: $(l/(l-2), 2l^2(k+l-2)/[k(l-2)^2(l-4)])$

$$(x_1 \sim \chi_k^2, x_2 \sim \chi_l^2) \Rightarrow \frac{x_1/k}{x_2/l} \sim F_l^k$$

(a) In particular, if $\boldsymbol{\beta} = \mathbf{0}$ then

$$\text{SSR} = \boldsymbol{e}^T HCH\boldsymbol{e} \sim \chi_p^2 \text{ and } \frac{\text{MSR}}{\text{MSE}} = \frac{\frac{1}{\sigma^2}\frac{\text{SSR}}{p}}{\frac{1}{\sigma^2}\frac{\text{SSE}}{n-p-1}} \sim F_{n-p-1}^p.$$

(b) Noncentral: if $x_1 \sim \chi_k^2(\lambda)$, then $(x_1/k)/(x_2/l) \sim F_l^k(\lambda)$.