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 $H_k C H_k - H_{k-1}CH_{k-1} = (H_k - \bar{J}) - (H_{k-1} - \bar{J}) = H_k - H_{k-1}$

Source	Sum of Squares	df
$\boldsymbol{x}_1$	SSR=$\boldsymbol{y}^T H_1 C H_1 \boldsymbol{y}$	tr$(H_1CH_1) = 1$
$\boldsymbol{x}_2 \vert X_1$	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}:

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

5 Some Matrix Algebra

1. Eigen-decomposition
$$M_{n\times k} = V_{n\times k} D_{k\times k} U^T_{k\times k}$$
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
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}.$$

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

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 $\hat{\boldsymbol{\beta}} = (X^TX)^{-1}X^TY \sim N(\boldsymbol{\beta}, \sigma^2(X^TX)^{-1})$

   (c) Quadratic forms

$$\boldsymbol{y} \sim N(\boldsymbol{\mu}, \Sigma) \Rightarrow E\boldsymbol{y}^T A\boldsymbol{y} = \mathrm{tr}(A\Sigma) + \mu^T A\mu$$

   (d) Uncorrelated $\Leftrightarrow$ Independent:

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

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

   (f) Conditional Distributions: for

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 2 \end{pmatrix}\right)$$

   then

$$y_2|y_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(y_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

   (g) Example: $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, $\alpha_i \sim N(0, \sigma_a^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ $j = 1, \ldots, r$, then

$$\begin{pmatrix} \bar{y}i\cdot \\ \alpha_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_a^2 + \sigma_e^2/r & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 \end{pmatrix}\right)$$

   and

$$\alpha_i|\bar{y}_{i\cdot} \sim N\left(\frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2/r}(\bar{y}_{i\cdot} - \mu), \frac{\sigma_a^2\sigma_e^2/r}{\sigma_a^2 + \sigma_e^2/r}\right)$$

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

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

(b)

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

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

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

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

(a) Particularly

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

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

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

$$(x_1 \sim \chi^2_k, x_2 \sim \chi^2_l) \Rightarrow \frac{x_1/k}{x_2/l} \sim F_l^k$$

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

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

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

7 Mixed Effects Models

1. Mixed models written when all are factors in terms individual levels. For example

$$y_{ijk} = \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

   - α_i = effect for ith level of factor A
   - β_j = effect for jth level of factor B
   - $\alpha\beta_{ij}$ = effect for combination of levels of A and B
   - ϵ_{ijk} = unique effect of observation ijk.

2. Factor coding translates effects into a matrix form $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$

   (a) Mean model: $\alpha_i = \beta_{i-1}$ – indicator for each level.
   (b) Reference coding $\alpha_1 = \beta_0$, $\alpha_i = \beta_0 + \beta_{i-1}$ – intercept gives α_1
   (c) Effect coding $\alpha_a = \beta_0 - \sum_{i=1}^{a-1} \beta_i$: last level is minus sum of all the rest so that $\beta_0 = \bar{\alpha}$.

3. Contrasts apply to coefficients. Eg, $\alpha_1 = \alpha_2 + \alpha_3$ tested by $L = [1, -1, -1, \mathbf{0}]$. Can translate into reference coding by substituting for $\boldsymbol{\beta}$.

4. Random effects specify some levels to be random. Eg, $\beta_j \sim N(0, \sigma_b^2)$, $\alpha\beta_{ij} \sim N(0, \sigma_{ab}^2)$, $\epsilon_{ijk} \sim N(0, \sigma_e^2)$. *All effects are assumed independent.*

   (a) cov$(y_{ijk}, y_{i'j'k'})$ determined by shared random terms

$$\text{cov}(y_{ijk}, y_{i'j'k'}) = \begin{cases} \sigma_a^2 + \sigma_{ab}^2 + \sigma_e^2 & \text{if } i = i', j = j', k = k' \quad \text{all random terms are in common} \\ \sigma_a^2 + \sigma_{ab}^2 & \text{if } i = i', j = j' \quad \text{observations share } \alpha_i \text{ and } \alpha\beta_{ij} \\ \sigma_a^2 & \text{if } i = i' \quad \alpha_i \text{ common to both observations} \\ 0 & \text{otherwise} \quad \text{no random terms in common} \end{cases}$$

5. **Sums of Squares** used to remove lower-order effects and estimate the size of higher-order ones.

$$y_{ij} = \alpha_i + \epsilon_{ijk} \Rightarrow y_{ij} - \bar{y}_{i\cdot} = \epsilon_{ij} - \bar{\epsilon}_{i\cdot}$$

   (a) Expected sums of squares come from separating out terms in each sum

$$SSA = \sum_{i=1}^{a} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a} (\alpha_i \bar{\epsilon}_{i\cdot} - (\bar{\alpha} - \bar{\epsilon}_{\cdot\cdot})^2$$

   if $\alpha_i \sim N(0, \sigma_a^2)$ then

$$ESSA = E\sum_{i=1}^{a} (\alpha_i - \bar{\alpha})^2 + E\sum_{i=1}^{a} (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{\cdot\cdot})^2 = (a-1)(\sigma_a^2 + \sigma_e^2/r)$$

alternatively

$$ESSA = E(\boldsymbol{\alpha} + \bar{\boldsymbol{\epsilon}}_i)^T C_a(\boldsymbol{\alpha} + \bar{\boldsymbol{\epsilon}}_i) = (a-1)(\sigma_a^2 + \sigma_e^2/r)$$

(b) Test effects by finding another sum of squares that matches expectation under H_0. If α_i fixed effects then

$$ESSA = \sum(\alpha_i - \bar{\alpha})^2 + (a-1)\sigma_e^2/r$$

And $ESSA = (a-1)\sigma_e^2/r$ under $H_0 : \alpha_i = \bar{\alpha}, i = 1, \ldots, a$.

$$ESSE = E\sum_{i=1}^{a}\sum_{j=1}^{r}(y_{ij} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^{a}\sum j = 1^r(\epsilon_{ij} - \bar{\epsilon}_{i\cdot})^2 = a(r-1)\sigma^2 e$$

So that $ESSA/(a-1) = ESSE/(a(r-1)r)$. Then test using

$$F = \frac{SSA/(a-1)}{SSE/(a(r-1)r)}$$

8 Longitudinal Models

1. Also allow for continuous covariates x. For multiple measurements of subject i, we can allow each subject a different slope and intercept, but insist their *average* slopes and intercepts are fixed:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i}x_{ij} + \epsilon_{ij},\ b_{0i} \sim N(0, \sigma_{b_0}^2),\ b_{1i} \sim N(0, \sigma_{b_1}^2),\ \epsilon_{ij} \sim N(0, \sigma_e^2)$$

2. So that for the same subject

$$\text{cov}(y_{ij}, y_{ij'}) = \sigma_{b_0}^2 + x_{ij}x_{ij'}\sigma_{b_1}^2 + \sigma_e^2(j = j')$$

or

$$\text{var}(\boldsymbol{y}_i) = \sigma_{b_0}^2 \mathbf{1}\mathbf{1}^T + \sigma_{b1}^2 \boldsymbol{x}_i \boldsymbol{x}_i^T + \sigma_e^2 I$$

3. Written in vector form as

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{b} + \boldsymbol{\epsilon},\ \boldsymbol{b} \sim N(0, G),\ \boldsymbol{\epsilon} \sim N(0, \sigma_e^2 I)$$

with

$$\text{var}(\boldsymbol{y}) = ZGZ^T + \sigma_e^2 I$$

X gives fixed "average" model for subjects. Z codes subject-specific deviation from average.