

Eléments de Biostatistiques

Jean Vaillant
Février 2011

1 Echantillonnage en population finie; Finite population sampling

\mathcal{P} est la population statistique considérée. Elle est de taille N . Ses éléments, les individus statistiques, sont répertoriés (pas forcément le cas dans la pratique) et un numéro allant de 1 à N leur est affecté. On peut donc écrire

$$\mathcal{P} = \{1, \dots, N\}.$$

On note

Y_i la valeur du caractère étudié sur l'individu i

\bar{Y} la moyenne de la population

V la variance de la population,

$$\text{avec } \bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad \text{et} \quad V = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2.$$

\bar{Y} et V sont simplement de la moyenne et de la variance de la série des N valeurs Y_1, \dots, Y_N .

Généralement, la population ne peut être observée dans sa totalité en raison du coût d'observation. Par conséquent, \bar{Y} et V ne peuvent être connues parfaitement. On choisit donc un certain nombre d'individus et l'ensemble ainsi constitué est appelé échantillon. A l'issue de l'observation du caractère sur chaque individu de l'échantillon, on a donc :

Y_i est connu si i est dans l'échantillon,

Y_i inconnu si i est dans l'échantillon.

L'échantillonnage a généralement pour but d'estimer une quantité dépendant de toutes les valeurs y_1, y_2, \dots, y_N à l'aide des seules valeurs données par l'échantillon observé. \bar{Y} et V sont de telles quantités.

Dans le cas où les Y_i sont binaires (c'est-à-dire $Y_i = 1$ si l'individu i vérifie une certaine propriété, et $Y_i = 0$ sinon), \bar{Y} est simplement la proportion notée P d'individus vérifiant cette propriété dans la population. La variance V de la population vaut alors $P(1-P)$.

1.1 Echantillonnage aléatoire simple; Simple random sampling

On considère un plan d'échantillonnage de taille n . On note A l'échantillon tiré,

$\bar{y}_A = \frac{1}{n} \sum_{i \in A} y_i$ la moyenne d'échantillon, et $s_A^2 = \frac{1}{n} \sum_{i \in A} (y_i - \bar{y}_A)^2$ la variance des $y_i, i \in A$.

Intéressons nous au Plan Aléatoire Simple Sans Remise (PASSR) et au Plan Aléatoire Simple Avec Remise (PASAR).

	PASSR	PASAR
Variance de \bar{y}_A	$(1 - \frac{n}{N}) \frac{V}{n} \frac{N}{N-1}$	$\frac{V}{n}$
Estimateur sans biais de la variance de \bar{y}_A	$(1 - \frac{n}{N}) \frac{s_A^2}{n-1}$	$\frac{s_A^2}{n-1}$

Dans le cas où le caractère étudié est binaire, on note \hat{p}_A la proportion observée dans l'échantillon A . On a alors le tableau suivant

	PASSR	PASAR
Variance de \hat{p}_A	$(1 - \frac{n}{N}) \frac{P(1-P)}{n} \frac{N}{N-1}$	$\frac{P(1-P)}{n}$
Estimateur sans biais de la variance de \hat{p}_A	$(1 - \frac{n}{N}) \frac{\hat{p}_A(1-\hat{p}_A)}{n-1}$	$\frac{\hat{p}_A(1-\hat{p}_A)}{n-1}$

Choix d'une taille d'échantillonnage; Sampling size determination

On désire déterminer une taille d'échantillon n permettant d'avoir une erreur d'évaluation de la proportion P de la population inférieure à d avec une probabilité au moins égale à $1 - \alpha$. Le fait que cette erreur soit inférieure à d avec une probabilité $1 - \alpha$ s'écrit :

$$\text{Proba}(|\hat{p}_A - P| \leq d) \geq 1 - \alpha.$$

Lorsque l'approximation d'une loi hypergéométrique ou d'une loi binomiale par une loi normale est possible, on a (avec u fractile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite)

	Pas d'information sur P	$P \leq P_0 \leq 0,5$
Cas sans remise	$n \geq \frac{Nu^2}{u^2 + 4(N-1)d^2}$	$n \geq \frac{Nu^2}{u^2 + \frac{(N-1)d^2}{P_0(1-P_0)}}$
Cas avec remise	$n \geq \frac{u^2}{4d^2}$	$n \geq \frac{u^2 P_0(1-P_0)}{d^2}$

1.2 Échantillonnage aléatoire stratifié; Stratified random sampling

La population est divisée en H strates de taille N_1, \dots, N_H . La moyenne d'échantillon dans la strate h est notée \bar{y}_h . La procédure d'échantillonnage consiste à exécuter un plan aléatoire simple sans remise de taille n_h dans la strate h , indépendamment des autres strates.

La moyenne d'échantillon global n'est pas forcément sans biais pour \bar{Y} pour ce type d'échantillonnage. On utilise donc la moyenne dite stratifiée qui, elle, est sans biais :

$$\bar{y}_{st} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h$$

et de variance $Var(\bar{y}_{st}) = \sum_{h=1}^H \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{V_h}{n_h} \frac{N_h}{N_h - 1}$ où V_h est la variance de la strate h .

Le plan aléatoire stratifié est intéressant quand la variabilité intra-strate est faible.

1.3 Echantillonnage aléatoire en grappes; Cluster random sampling

La population est divisée en G grappes, pas forcément de même taille. L'échantillonnage consiste à choisir g grappes selon un plan aléatoire simple sans remise. Un estimateur sans biais de la moyenne de la population est donnée par

$$\bar{y}_{grappes} = \frac{G}{g.N} \cdot (\text{somme des valeurs observées sur les grappes échantillonnées}).$$

Le plan aléatoire en grappes est intéressant quand la variabilité inter-grappe est faible.

2 Répartitions spatiales d'individus et lois de dénombrement

L'étude de la répartition spatiale d'une population écologique peut se faire par comptage d'individus dans des unités d'échantillonnage disposées spatialement.

Une loi de dénombrement est une loi (ou distribution) de probabilité concernant une variable de comptage (donc ayant pour valeurs possibles des nombres entiers non négatifs).

Selon le mécanisme de répartition des individus, la variable aléatoire $X = \text{nombre d'individus}$ dans une unité d'échantillonnage spatiale suit une loi bien déterminée. Le rapport variance sur espérance de cette loi est notée I_d et est appelée indice de dispersion théorique de la répartition.

2.1 Répartition complètement aléatoire et loi de Poisson

Une variable aléatoire X suit une loi de Poisson de paramètre $\theta > 0$ si

$$\text{Proba}(X = k) = \frac{\theta^k}{k!} e^{-\theta} \text{ pour } k \text{ dans } \mathbb{N}.$$

Si X représente le nombre d'individus dans une unité d'échantillonnage de surface S dans le cas d'une répartition complètement aléatoire (RCA) d'intensité λ (nombre attendu d'individus par unité de surface), alors X suit la loi de Poisson de paramètre λS .

L'espérance et la variance d'une telle loi de Poisson valent toutes les deux λS . Donc l'indice de dispersion théorique d'une RCA est $I_d = 1$. Ceci explique que la valeur de référence pour l'indice de dispersion soit 1.

Si $I_d > 1$, on parle d'individus surdispersés ou de population surdispersée ou de surdispersion,

Si $I_d < 1$, on parle d'individus sousdispersés ou de population sousdispersée ou de sousdispersion.

2.2 Sousdispersion et loi binomiale

Quand $I_d < 1$, cela correspond à $Var(X) < E(X)$. La variabilité dans les unités d'échantillonnage est moindre que celle d'une RCA de même intensité.

Parmi les lois de dénombrement correspondant à une sousdispersion, on notera la loi binomiale de paramètres $N > 0$ et $p \in]0, 1[$. Si X suit cette loi, le nombre d'individus par unité est inférieur ou égal à N . Comme $E(X) = Np$ et $Var(X) = Np(1-p)$, on a $I_d = (1-p) < 1$.

2.3 Répartition en milieu hétérogène et loi de Poisson composée

Considérons une population dont les individus, conditionnellement aux facteurs environnementaux, se répartissent indépendamment les uns des autres. Soit X le nombre d'individus dans une unité de surface S et λ l'intensité qui est maintenant aléatoire dans cette unité.

Conditionnellement à λ , X suit la loi de Poisson de paramètre λ .

On démontre que

$$E(X) = E(\lambda)S, \quad Var(X) = E(\lambda)S + Var(\lambda)S^2 \quad \text{et donc} \quad I_d = 1 + \frac{Var(\lambda)}{E(\lambda)}S.$$

L'aléa sur l'intensité λ crée donc de la surdispersion. Un exemple de loi pour λ est la loi dite Gamma. Dans ce cas, X suit la loi binomiale négative de paramètres $m > 0$ et $\gamma > 0$ caractérisée par

$$Proba(X = k) = \frac{(k + \gamma - 1)!}{(\gamma - 1)!k!} \left(\frac{m}{m + \gamma}\right)^k \left(\frac{\gamma}{m + \gamma}\right)^\gamma \quad \text{pour } k \text{ dans } \mathbf{N}$$

avec $E(X) = m$, $Var(X) = m(1 + \frac{m}{\gamma})$ et donc $I_d = 1 + \frac{m}{\gamma}$.

2.4 Répartition en groupes et loi de Poisson généralisée

Soit une population d'individus répartis en groupes, les groupes étant répartis indépendamment les uns des autres.

Soit Y la variable de comptage des groupes dans une unité d'échantillonnage de surface S , le nombre X d'individus dans cette unité vérifie :

$$X = g_1 + g_2 + \dots + g_Y = \sum_{i=1}^Y g_i \quad \text{où } g_i \text{ est le nombre d'individus dans le groupe } i.$$

Les tailles de groupes g_i , pour i allant de 1 à Y sont des variables aléatoires indépendantes et de même loi. On dit alors que X suit une loi de Poisson généralisée. On démontre que :

$$E(X) = E(Y)E(g_i) \quad \text{et} \quad Var(X) = Var(Y)(E(g_i))^2 + E(Y)Var(g_i)$$

d'où $I_d = E(g_i) + \frac{Var(g_i)}{E(g_i)} > 1$ dès lors que $E(g_i) > 1$ (un groupe n'est pas singleton).

2.5 Estimation de l'indice de dispersion

Soient x_1, x_2, \dots, x_n les nombres d'individus écologiques observés dans n unités d'échantillonnage. L'indice de dispersion théorique I_d est estimé par :

$$\hat{I}_d = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)\bar{x}}$$

Sous l'hypothèse de RCA, \hat{I}_d est un estimateur sans biais et convergent de I_d . De plus, $(n-1)\hat{I}_d$ suit une loi du khi-deux à $n-1$ degrés de liberté.

2.6 Tests de l'indice de dispersion

Le test de l'hypothèse de RCA peut s'effectuer contre trois types d'alternatives et s'appuie sur la loi du khi-deux à $n-1$ degrés de liberté de la statistique $T = (n-1)\hat{I}_d$.

Notons $\chi_{n-1,p}^2$ le fractile d'ordre p de la loi du khi-deux à $n-1$ degrés de liberté. Les règles de décision des tests de dispersion de niveau α sont indiquées dans ce qui suit.

2.6.1 Cas où $n-1 \leq 30$

$H_0 : Id = 1$ contre $Id \neq 1$

Rejet de H_0 si $T \notin [\chi_{n-1,\alpha/2}^2; \chi_{n-1,1-\alpha/2}^2]$

Non rejet de H_0 si $T \in [\chi_{n-1,\alpha/2}^2; \chi_{n-1,1-\alpha/2}^2]$

$H_0 : Id = 1$ contre $Id > 1$

Rejet de H_0 si $T > \chi_{n-1,1-\alpha}^2$

Non rejet de H_0 si $T \leq \chi_{n-1,1-\alpha}^2$

$H_0 : Id = 1$ contre $Id < 1$

Rejet de H_0 si $T < \chi_{n-1,\alpha}^2$

Non rejet de H_0 si $T \geq \chi_{n-1,\alpha}^2$

2.6.2 Cas où $n - 1 > 30$

Quand $n - 1 > 30$, on utilise la statistique $U = \sqrt{2(n-1)\widehat{Id}} - \sqrt{2(n-1) - 1}$ qui suit approximativement la loi normale centrée réduite. Notons u_p le fractile d'ordre p de la loi normale centrée réduite. Les règles de décision des tests de dispersion de niveau α sont indiquées dans ce qui suit.

$H_0 : Id = 1$ contre $Id \neq 1$

Rejet de H_0 si $|U| > u_{1-\alpha/2}$

Non rejet de H_0 si $|U| \leq u_{1-\alpha/2}$

$H_0 : Id = 1$ contre $Id > 1$

Rejet de H_0 si $U > u_{1-\alpha}$

Non rejet de H_0 si $U \leq u_{1-\alpha}$

$H_0 : Id = 1$ contre $Id < 1$

Rejet de H_0 si $U < -u_{1-\alpha}$

Non rejet de H_0 si $U \geq -u_{1-\alpha}$

2.7 Test du khi-deux d'ajustement

Le principe du test du khi-deux d'ajustement est de comparer les effectifs observés aux effectifs espérés sous le modèle testé. La formule du critère du khi-deux d'ajustement est :

$$C = \sum_i \frac{(n_i - np_i)^2}{np_i}$$

où n_i est l'effectif de la i ème classe d'observation, n est l'effectif global et p_i la probabilité d'appartenir à la i ème classe sous le modèle considéré.

Une formule plus simple mais moins intuitive pour C est la suivante :

$$C = \left(\sum_i \frac{n_i^2}{np_i} \right) - n.$$

La règle de décision du test d'ajustement du khi-deux de niveau α est

Rejet du modèle si $C > \chi_{\nu, 1-\alpha}^2$

Non rejet du modèle si $T \leq \chi_{\nu, 1-\alpha}^2$.

3 Etude de la régression simple

Tests concernant les paramètres de régression :

Soient \hat{a} et \hat{b} les estimateurs respectifs de la pente a de la droite de régression et de l'interception b ; soient $\hat{\sigma}^2$ l'estimateur sans biais de σ^2 , et $t_{n-2, 1-\alpha/2}$ le fractile de la loi de Student à $n - 2$ degrés de liberté.

$a = a_0$ contre $a \neq a_0$

Rejet de l'hypothèse $H_0 : a = a_0$ si $\frac{|\hat{a} - a_0|}{\sqrt{\frac{\hat{\sigma}^2}{ns_x^2}}} > t_{n-2, 1-\alpha/2}$

Acceptation de l'hypothèse H_0 sinon.

$b = b_0$ contre $b \neq b_0$

Rejet de l'hypothèse $H_0 : b = b_0$ si $\frac{|\hat{b} - b_0|}{\sqrt{\frac{\hat{\sigma}^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2}\right)}} > t_{n-2, 1-\alpha/2}$

Acceptation de l'hypothèse H_0 sinon.

Intervalles de confiance de sécurité $1 - \alpha$:

Pour a , on a

$$\left[\hat{a} \pm t_{n-2, 1-\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n s_x^2}} \right].$$

Pour b , on a

$$\left[\hat{b} \pm t_{n-2, 1-\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2}\right)} \right].$$

Pour la prédiction associée à une valeur x_0 de la variable explicative, on a

$$\left[\hat{y}_0 \pm t_{n-2, 1-\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n} \left(1 + \frac{(\bar{x} - x_0)^2}{s_x^2}\right)} \right].$$

On rappelle que $\hat{\sigma}^2 = \frac{n}{n-2} \left(s_y^2 - \frac{s_{xy}^2}{s_x^2} \right)$ où s_{xy} est la covariance empirique entre les x_i et les y_i .

4 Analyse de la variance

Analyse de la variance à un facteur

On considère I échantillons indépendants $\mathcal{E}_1, \dots, \mathcal{E}_I$.

Chaque échantillon \mathcal{E}_i est de taille n_i , de moyenne \bar{x}_i et de variance s_i^2 . On note \bar{x} la moyenne globale et n le nombre total d'observations.

Variance résiduelle (ou variance intra-groupe) : $s_R^2 = \frac{1}{n-I} \sum_{i=1}^I n_i s_i^2$.

Variance factorielle (ou variance inter-groupe) : $s_F^2 = \frac{1}{I-1} \sum_{i=1}^I n_i (\bar{x}_i - \bar{x})^2$.

Test de l'hypothèse H_0 de non influence du facteur étudié

Rejet de H_0 au niveau α si $\frac{s_F^2}{s_R^2} > f_{I-1, n-I, 1-\alpha}$

$f_{I-1, n-I, 1-\alpha}$ est le fractile approprié de la loi de Fisher-Snédecor à $(I-1, n-I)$ degrés de liberté.

Analyse de la variance à deux facteurs

On considère IJ échantillons indépendants $\mathcal{E}_{11}, \dots, \mathcal{E}_{ij}, \dots, \mathcal{E}_{IJ}$ de même taille $n_0 > 1$.

On note respectivement \bar{x}_{ij} et s_{ij}^2 la moyenne et la variance de l'échantillon \mathcal{E}_{ij} .

Variance résiduelle : $s_R^2 = \frac{n_0}{IJ(n_0 - 1)} \sum_{i=1}^I \sum_{j=1}^J s_{ij}^2$.

Variance factorielle : $s_F^2 = \frac{n_0}{IJ-1} \sum_{i=1}^I \sum_{j=1}^J (\bar{x}_{ij} - \bar{x})^2$.

Variance factorielle due au facteur A seul

$$s_A^2 = \frac{Jn_0}{I-1} \sum_{i=1}^I (\bar{x}_i - \bar{x})^2.$$

Variance factorielle due au facteur B seul

$$s_B^2 = \frac{In_0}{J-1} \sum_{j=1}^J (\bar{x}_{.j} - \bar{x})^2.$$

Variance factorielle due à l'interaction de A et B

$$s_{AB}^2 = \frac{n_0}{(I-1)(J-1)} \sum_{i=1}^I \sum_{j=1}^J (\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2.$$

On a la décomposition suivante de la variance factorielle

$$(IJ-1)s_F^2 = (I-1)s_A^2 + (J-1)s_B^2 + (I-1)(J-1)s_{AB}^2.$$

Tests de niveau α des hypothèses nulles

Ces hypothèses nulles sont

$(H_0)_A$: le facteur A n'a pas d'influence.

$(H_0)_B$: le facteur B n'a pas d'influence.

$(H_0)_{AB}$: il n'y a pas d'interaction entre les facteurs A et B .

Les règles de décision sont :

$$\text{Rejet de } (H_0)_A \text{ si } \frac{s_A^2}{s_R^2} > f_{I-1, (n_0-1)IJ, 1-\alpha}$$

où $f_{I-1, (n_0-1)IJ, 1-\alpha}$ est le fractile approprié de la loi de Fisher-Snédecor.

$$\text{Rejet de } (H_0)_B \text{ si } \frac{s_B^2}{s_R^2} > f_{J-1, (n_0-1)IJ, 1-\alpha}$$

où $f_{J-1, (n_0-1)IJ, 1-\alpha}$ est le fractile approprié de la loi de Fisher-Snédecor.

$$\text{Rejet de } (H_0)_{AB} \text{ si } \frac{s_{AB}^2}{s_R^2} > f_{(I-1)(J-1), (n_0-1)IJ, 1-\alpha}$$

où $f_{(I-1)(J-1), (n_0-1)IJ, \alpha}$ est le fractile approprié de la loi de Fisher-Snédecor.

Le **non rejet d'une hypothèse nulle** s'en déduit comme d'habitude.

5 Tableaux d'ANOVA

Plan à un facteur avec répétitions

On étudie un facteur A ayant I niveaux. On a n_i répétitions pour le niveau i .

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$n - 1$	$CMT = \frac{SCT}{n - 1}$	
Facteur A	$SCFA$	$I - 1$	$CMFA = \frac{SCFA}{I - 1}$	$F_A = \frac{CMFA}{CMR}$
Résiduelle	SCR	$n - I$	$CMR = \frac{SCR}{n - I}$	

avec $SCT = ns^2$; $SCFA = \sum_{i=1}^I n_i(\bar{x}_i - \bar{x})^2$; et $SCR = \sum_{i=1}^I (n_i - 1)s_i'^2$.

Equation d'ANOVA : $SCT = SCFA + SCR$

Plan à deux facteurs sans répétitions

On étudie deux facteurs A et B ayant respectivement I et J niveaux. Le nombre de répétitions n_0 est tel que $n_0 = 1$.

Données :

x_{ij} est la valeur observée pour le traitement (i, j)

\bar{x}_i est la moyenne observée pour le niveau i de A ,

\bar{x}_j est la moyenne observée pour le niveau j de B .

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$IJ - 1$	$CMT = \frac{SCT}{IJ - 1}$	
Facteur A	SCF_A	$I - 1$	$CMF_A = \frac{SCF_A}{I - 1}$	$F_A = \frac{CMF_A}{CMR}$
Facteur B	SCF_B	$J - 1$	$CMF_B = \frac{SCF_B}{J - 1}$	$F_B = \frac{CMF_B}{CMR}$
Résiduelle	SCR	$(I - 1)(J - 1)$	$CMR = \frac{SCR}{(I - 1)(J - 1)}$	

avec $SCT = \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x})^2$; $SCF_A = \sum_{i=1}^I J(\bar{x}_i - \bar{x})^2$; $SCF_B = \sum_{j=1}^J I(\bar{x}_j - \bar{x})^2$

Equation d'ANOVA : $SCT = SCF_A + SCF_B + SCR$

Plan à deux facteurs avec répétitions

On étudie deux facteurs A et B ayant respectivement I et J niveaux.

Données :

x_{ijk} est la valeur observée pour la k 'eme répétition du traitement (i, j)

$\bar{x}_{i..}$ est la moyenne observée pour le niveau i de A ,

$\bar{x}_{.j.}$ est la moyenne observée pour le niveau j de B ,

$\bar{x}_{ij.}$ est la moyenne observée pour le traitement (i, j)

n_{ij} est le nombre de répétitions du traitement (i, j) ,

n_{i+} est le nombre de répétitions du niveau i de A ,

n_{+j} est le nombre de répétitions du niveau j de B .

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	F de Fisher
Totale	SCT	$n - 1$	$CMT = \frac{SCT}{n - 1}$	
Facteur A	SCF_A	$I - 1$	$CMF_A = \frac{SCF_A}{I - 1}$	$F_A = \frac{CMF_A}{CMR}$
Facteur B	SCF_B	$J - 1$	$CMF_B = \frac{SCF_B}{J - 1}$	$F_B = \frac{CMF_B}{CMR}$
Interaction entre A et B	SCF_{AB}	$(I - 1)(J - 1)$	$CMF_{AB} = \frac{SCF_{AB}}{(I - 1)(J - 1)}$	$F_{AB} = \frac{CMF_{AB}}{CMR}$
Résiduelle	SCR	$n - IJ$	$CMR = \frac{SCR}{n - IJ}$	

$$\text{avec } SCT = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \bar{x})^2; SCF_A = \sum_{i=1}^I n_{i+} (\bar{x}_{i..} - \bar{x})^2$$

$$SCF_B = \sum_{j=1}^J n_{+j} (\bar{x}_{.j.} - \bar{x})^2; SCF_{AB} = \sum_{i=1}^I \sum_{j=1}^J n_{ij} (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})^2$$

$$\text{et } SCR = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \bar{x}_{ij.})^2.$$

Equation d'ANOVA : $SCT = SCF_A + SCF_B + SCF_{AB} + SCR$
