# Utilisation de la procédure MI et MIANALYZE Pour l'imputation multiple en SAS 9.3

M.Nguile-Makao Ph.D.
&
Pr. Vincent Vradet MD. Ph.D

Unité méthodologique du Centre Clinique et Évaluative en Oncologie
Hôtel-Dieu de Québec Laval Université

3 novembre 2016

# Plan

1. Généralité sur les données manquantes
   - Problématique des données manquantes
   - Méthodes d'analyse
2. Imputation multiple MI
   - Théorie MI
   - Proc MI
   - Proc MIANALYZE
3. Problème de Norm. Ass. des estimateurs

# Données manquantes

Plus de 95% des données utilisées dans des études statistiques contiennent de l'information manquantes.
Sources :

- **Essais cliniques**
  - Sortie d'études
  - Perdus de vue
  - ...
- **Enquêtes**
  - Non réponse partielles
  - Non réponse totales
  - Erreurs de recueil de données
- **Analyse de survie**
  - Censure
  - Troncatures
- **Etc.**

# Définition DMs

- **Définition de DMs**
  - DMs d'un jeu de données $X$ est la matrice de données pour laquelle quelques valeurs réelles de $X$ sont inconnues.
  - $X = (X_{ij})_{i,j \in R^{p \times n}}$ un jeu de données, $X = (X_o, X_m)$ où $X_o$ est la matrice des observées et $X_m$ la matrice des manquantes.
  - $R = (R_{ij})_{i,i \in R^{p \times n}}$ matrice des indicatrices où $r_{ij} = 1$ si DM ou 0 sinon.
- **Distribution de DMs**
  - Mécanisme des DMs est spécifié par la distribution de R sachant X.

$$Pr(R = r|X, \theta) = p(r|x_o, x_m, \theta)$$

Notation : *DM valeurs manquantes*

# Typologie des DMs

Little & Rubin (1987) proposent une typologie des DMs

- **MAR** : Missing At Random.
  - Probabilité qu'une observation soit manquante ne dépend pas des mesures observées ou des mesures non observées.

$$p(r|x_o, x_m, \theta) = p(r, \theta).$$

- **MCAR** : Missing Completely At Random.
  - Connaissant les données observées, la probabilité qu'une observation soit manquante ne dépend pas des données non observées.

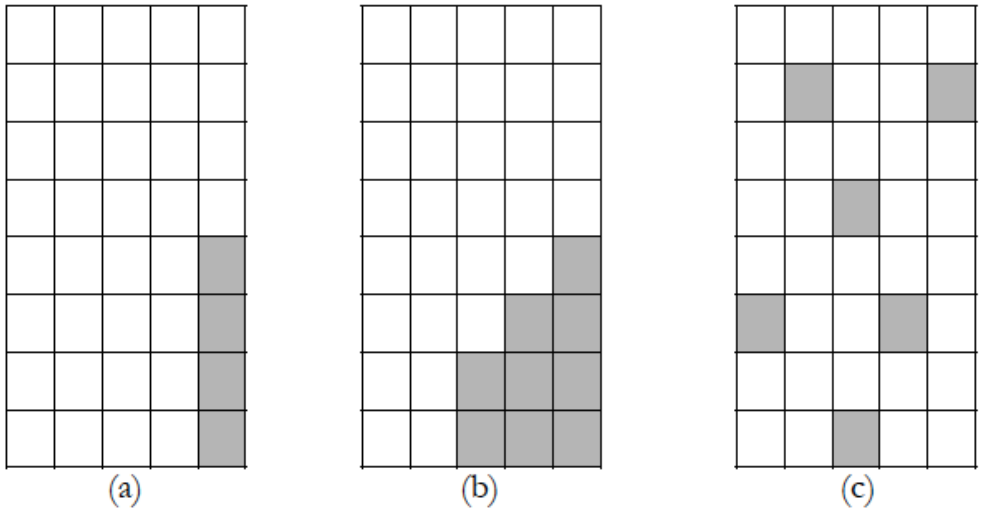
$$p(r|x_o, x_m, \theta) = p(r|x_o, \theta)$$

- **MNAR** : Missing Non At Random
  - Probabilité qu'une observation soit manquante dépend des données observés et des données non observées.

$$p(r|x_o, x_m, \theta) = p(r|x_o, x_m, \theta)$$

# Pattern des DMs

- **Différents patterns des DMs**




- **Lexique**
  $(a)$ : **DMs** univariées. $(b)$ : **DMs** monotones $(c)$ : **DMs** non monotones

# Exemples

**Exemples des DMs**
Considérons une base de données (BD) des patients atteints du cancer de la prostate (PCa). BD=(âge, Gleason,PSA).

- **MCAR**
  - La probabilité que l'âge soit NA ne dépend ni des valeurs observées et non observées du Gleason, ainsi que ceux de la PSA, de plus elle est la même pour tous les sujets.
- **MAR**
  - La probabilité que l'âge soit NA ne dépend que des valeurs observées du Gleason et la PSA (valeurs observées). Elle n'est pas la même pour tous les sujets.
- **NMAR**
  - La probabilité que l'âge soit NA dépend des non observées valeurs du Gleason et de la PSA (valeurs non observées). Elle n'est pas la même pour tous les sujets.

# Impacts des DMs

- **En présence des DMs de type MCAR**
  - Perte de précision (Puissance)
  - Aucun biais
- **En présence des DMs de type MAR**
  - Perte de précision (Puissance)
  - Aucun biais pour des méthodes d'inférence par maximum de vraisemblance

  **En présence des DMs de type MNAR**
  - Perte de précision (Puissance)
  - Biais

Remarque :

*Il est très difficile de savoir quelle est le vraie mécanisme des DMs ou au moins laquelle de typologie est le plus probable. Il y a une part d'avis subjectif dans le choix mais ce peut-être un avis éclairé !*

# Méthodes suppression/tolérance des DMs

- **Analyse des cas-concrets**
  - Suppression des DMs (problématique si NMAR))
- **Analyse des cas disponibles**
  - Suppression par paire ("Pairwise deletion")

*http ://www.ats.ucla.edu/stat/sas/modules/missing.htm*

- **Tolère les DMs**
  - Méthode CART

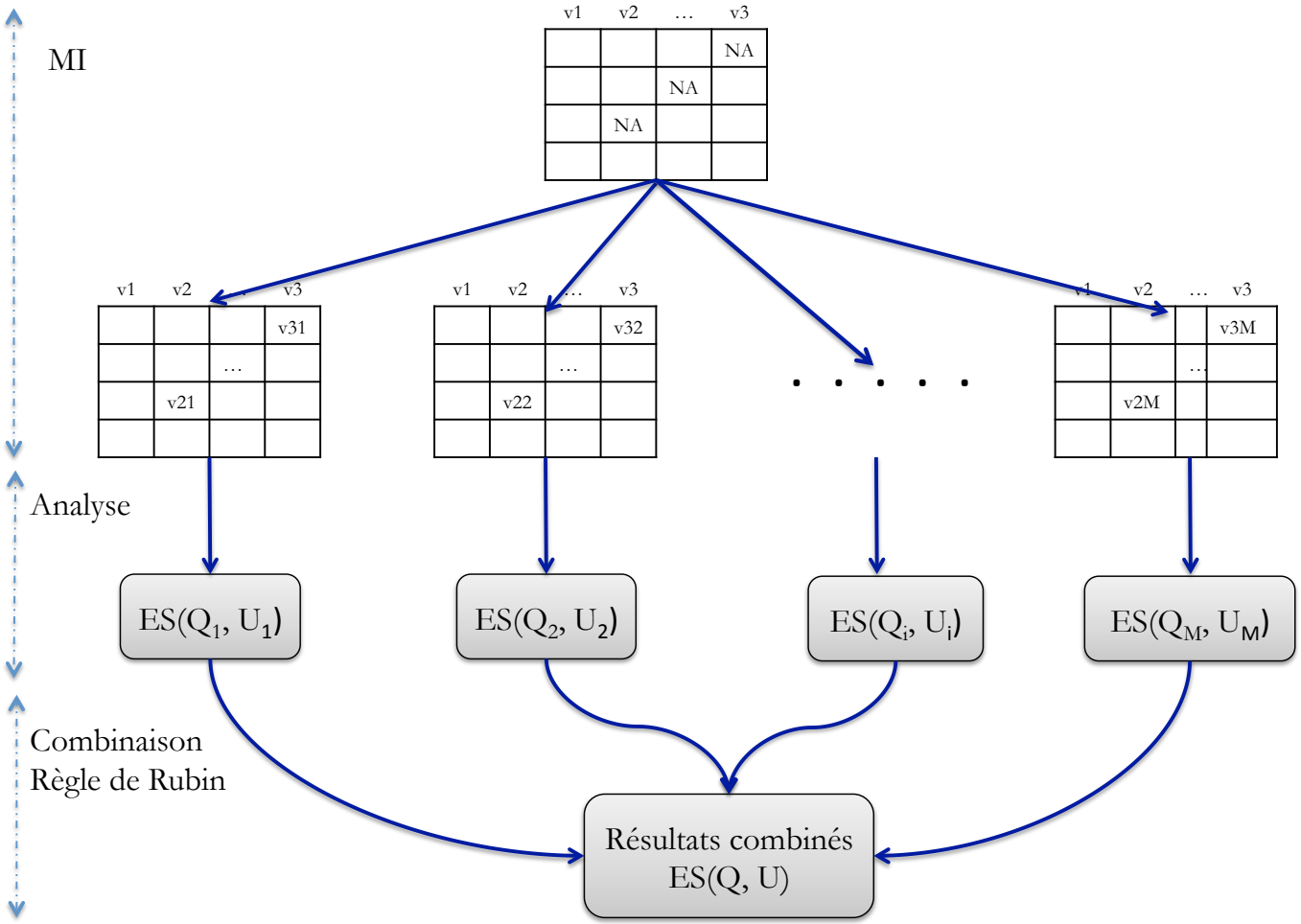
Leonard Gordon. *Using Classification and Regression Trees (CART) in SAS® Enterprise Miner™ For Applications in Public Health.* ***Data Mining and Text Analytics, SAS Global Forum (2013).***

# Méthode d'imputation simple

- **Imputation simple : Définition**
  - Méthode basée sur la substitution à une DM une valeur choisie de manière pertinente via des algorithmes.
- **Méthodes en sas**
  - Imputation par moyenne ou médiane
  - CMCF (Concept Most Common Attribute Value Fitting)
  - LOCF (Last observation carried forward, LOCF)
  - KNN (k-nearest neighbors)
  - Regression local (LOESS)
  - L'algorithm NIPALS (Nonlinear Iterative Partial Least Squares )
  - L'algorithme EM (Expectation maximization)
- Remarque
  *L'imputation simple est dangereuse car elle ne permet pas d'évaluer l'incertitude de la valeur imputée par rapport à la vrai valeur. D'ou imputation multiple !*

# Méthode star !

- **Imputation multiple (MI) Rubin 1987**



# Phase de la MI

- **Algorithm de Rubin** (2 phases)
  - Phase 1
    $X \sim f(x|\theta), \quad f(x_m|x_o) = \int f(x_m|x_o, \theta) f(\theta|x_o) d\theta$
    Estimation de $\theta$ via la distribution à posteriori $f(\theta|x_o)$.
    $(a)^*$ : approximation de la distribution de $X$ par Algorithm MCMC.
  - Phase 2
    Imputation des DMs à partir $f(x_m|x_0)$.
- **Autres méthodes** $(a)^*$
  - Méthode : "Regression" (hypothèse normalité)
  - Méthode : "Predictive mean matching" (hypothèse normalité)
  - Méthode : "Propensity Score"
  - Méthode : "Fully Conditional Specification (FCS)"

# Pooling de resultats

- **Combinaison de l'inférence**
  Soit $Q$ la statistique, $1, \cdots, M$ imputations et $U$ la variance.
  - Estimateur M-complet data $\overline{\widehat{Q}} = \frac{1}{M} \sum_m \widehat{Q}_m$

| $Var_{Intra}$ | $Var_{Inter}$ | $Var_T$ |
|---|---|---|
| $\overline{\widehat{U}} = \frac{1}{M} \sum_m \widehat{U}_m$ | $B = \frac{1}{M-1} \sum_m \left(\widehat{Q}_i - \overline{\widehat{Q}}\right)^2$ | $T = \overline{\widehat{U}} + (1 + \frac{1}{M})B$ |

  - $\frac{(Q-\overline{\widehat{Q}}))}{\sqrt{T}} \sim t_{v_M}$ où $v_M = (M-1)\left[1 + \frac{\widehat{U}}{(1+M^{-1})B}\right]^2$
    ddl ajusté pour faible DMs $v_M^* = \left[\frac{1}{v_M} + \frac{1}{(v_{obs}}\right]^{-1}$ où $v_{obs} = \frac{v_0+1}{v_0+3} v_0(1-\gamma)$ et $\gamma = (1 + M^{-1})BT^{-1}$
- **Efficacité de la MI**
  - Variance relative accrue dû aux DMs $r = (1 + M^{-1})B\overline{\widehat{U}}$
  - Fraction d'information manquantes $\widehat{\lambda} = [2 + r(v_M + 3)]\,(r+1)^{-1}$
    Rubin ($\widehat{\lambda}$ efficace pour M=3-5 )
  - Efficacité relative $RE = (M + \lambda)^{-1}$

# Procédure MI SAS

- **proc MI sas 9.3**
  - **PROC MI** <options> ;
    **BY** variables ;
    **CLASS** variables ;
    **EM** <options> ;
    **FCS** <options> ;
    **FREQ** variable ;
    **MCMC** <options> ;
    **MONOTONE** <options> ;
    **TRANSFORM** transform ( variables </ options>) <...transform ( variables </ options>) > ;
    **VAR** variables ;
- **Définitions**
  - FCS : imputation multivariée par chaîn (supposant une distribution conjointe existe).
  - MCMC : Monte Carlo Markov Chaine .
  - MONOTONE : pour la typologie monotone des DM

# Méthodes associées à la proc MI

- **Méthodes de la procédure MI sas 9.3**

| Pattern DM | Variable | Methodes |
|---|---|---|
| Monotone | Continuous | Monotone regression |
| | | Monotone predicted mean matching |
| | | Monotone propensity score |
| Monotone | ordinal | Monotone logistic regression |
| | nominal | Monotone discriminant function |
| Arbitrary | Continuous | MCMC full-data imputation |
| | | MCMC monotone-data imputation |
| Arbitrary | Continuous | FCS regression |
| | | FCS predicted mean matching |
| Arbitrary | ordinal | FCS logistic regression |
| | nominal | FCS discriminant function |

# Exemple de données

- **Les données**

| Obs. | Length1 | Length2 | Length3 |
|---|---|---|---|
| 1 | 23.2 | 25.4 | 30.0 |
| 2 | 24.0 | 26.3 | 31.2 |
| 3 | 23.9 | 26.5 | 31.1 |
| 4 | 26.3 | 29.0 | 33.5 |
| 5 | 26.5 | 29.0 | . |
| 6 | 26.8 | 29.7 | 34.7 |
| 7 | 26.8 | . | . |
| 8 | 27.6 | 30.0 | 35.0 |
| 9 | 27.6 | 30.0 | 35.1 |
| 10 | 28.5 | 30.7 | 36.2 |
| 11 | 28.4 | 31.0 | 36.2 |
| 12 | 28.7 | . | . |
| 13 | 29.1 | 31.5 | . |
| 14 | 29.5 | 32.0 | 37.3 |

| Obs. | BR | BR_time | Marge | ki67 | PSA0 | stade | agecl1 |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 1 | 2.5 | 6.7 | T3 | age<57 |
| 2 | 0 | 7 | 0 | 1.4 | 17.5 | T2 | 57-67 |
| 3 | 1 | 4 | 1 | 0.2 | 15 | T3 | age<57 |
| 4 | 1 | 13 | 1 | 1.4 | 10.8 | T3 | age<57 |
| 5 | 0 | 11 | 0 | 2.1 | 3 | T3 | age>67 |
| 6 | 0 | 8 | 0 | 0.5 | . | T2 | age>67 |
| 7 | 1 | 1 | 1 | 1.2 | . | T3 | 57-67 |
| 8 | 0 | 12 | 1 | 3.8 | 5.9 | T2 | age<57 |
| 9 | 0 | 11 | 1 | 0.2 | 8 | T3 | age>67 |
| 10 | 1 | 5 | 1 | 2 | 15.5 | T2 | age<57 |
| 11 | 0 | 9 | 0 | 0.8 | 10.2 | T2 | age>67 |
| 12 | 1 | 0 | 1 | 0.6 | . | T3 | age<57 |
| 13 | 0 | 8 | 0 | 0.6 | . | T2 | age<57 |
| 14 | 1 | 7 | 1 | . | 8 | T2 | age>67 |

FIGURE: Data Prostate cancer (hotel-Dieux Québec) et Fish (SAS)

# Exemple MI

- **Imputation par défaut (MCMC )**
  - **Programme** ($R_{1,1}$)

```
proc mi data= Fish1 seed=1396 nimpute=5 out=datimp0 round=.1 ;
var Length1 Length2 Length3;
run ;
```

- **Imputation par la méthode de régression (monotone)**
  - **Programme** ($R_{1,2}$)

```
proc mi data= Fish1 seed=1396 nimpute=5 out=datimp1 round=.1 ;
monotone reg (Length2) regpmm(Length3 = Length1 Length2);
var Length1 Length2 Length3;run ;
```

- **Imputation par la méthode de régression (arbitraire)**
  - **Programme** ($R_{1,3}$)

```
proc mi data= Fish1 seed=1396 nimpute=5 out=datimp2 ;
FCS regression ;
var Length3 Length1 Length2 ;
run ;
```

# Résultats du prog. $R_{1,1}$

- **Resultat du prog $R_{1,1}$**

**Procédure MI**

| Informations sur le modèle | |
|---|---|
| Table | WORK.FISH1 |
| Méthode | MCMC |
| Chaîne à imputation multiple | Chaîne unique |
| Valeurs estimées initiales pour | Mode a posteriori EM |
| Début | Valeur de départ |
| A priori | Jeffreys |
| Nombre d'imputations | 5 |
| Nombre d'itérations Burn-in | 200 |
| Nombre d'itérations | 100 |
| Valeur initiale pour nb aléatoi | 1396 |

**Caractéristiques des données manquantes**

| Groupe | Length1 | Length2 | Length3 | Fréq | Pourcentage | Moyennes de groupes | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Length1 | Length2 | Length3 |
| 1 | X | X | X | 30 | 85.71 | 30.603333 | 33.436667 | 38.720000 |
| 2 | X | X | . | 3 | 8.57 | 29.033333 | 31.666667 | . |
| 3 | X | . | . | 2 | 5.71 | 27.750000 | . | . |

**Estimations EM (Mode a posteriori)**

| _TYPE_ | _NAME_ | Length1 | Length2 | Length3 |
|---|---|---|---|---|
| MEAN | | 30.305714 | 33.107236 | 38.371523 |
| COV | Length1 | 11.258945 | 12.249655 | 12.974260 |
| COV | Length2 | 12.249655 | 13.383878 | 14.171168 |
| COV | Length3 | 12.974260 | 14.171168 | 15.061287 |

**Informations sur la variance (5 Imputations)**

| Variable | Variance | | | DDL | Augmentation relative dans variance | Informations manquantes fraction | Efficacité relative |
|---|---|---|---|---|---|---|---|
| | Inter | Intra | Total | | | | |
| Length2 | 0.000117 | 0.438645 | 0.438785 | 32.152 | 0.000319 | 0.000319 | 0.999936 |
| Length3 | 0.000372 | 0.491945 | 0.492392 | 32.133 | 0.000908 | 0.000908 | 0.999818 |

**Résultats estimés des paramètres (5 Imputations)**

| Variable | Moyenne | Erreur type | Intervalle de confiance à 95% | | DDL | Minimum | Maximum | Mu0 | t pour H0: Moyenne=Mu0 | Pr > \|t\| |
|---|---|---|---|---|---|---|---|---|---|---|
| Length2 | 33.106971 | 0.662409 | 31.75794 | 34.45600 | 32.152 | 33.095143 | 33.121714 | 0 | 49.98 | <.0001 |
| Length3 | 38.379200 | 0.701706 | 36.95010 | 39.80830 | 32.133 | 38.365429 | 38.412571 | 0 | 54.69 | <.0001 |

FIGURE: Résultat de la MI du programme $R_{1,1}$

# Données complétés prog. $R_{1,1}$

- **Résultat du prog $R_{1,1}$**

| Obs. | _Imputation_ | Length1 | Length2 | Length3 |
|---|---|---|---|---|
| 1 | 1 | 23.2 | 25.4 | 30.0 |
| 2 | 1 | 24.0 | 26.3 | 31.2 |
| 3 | 1 | 23.9 | 26.5 | 31.1 |
| 4 | 1 | 26.3 | 29.0 | 33.5 |
| 5 | 1 | 26.5 | 29.0 | 34.1 |
| 6 | 1 | 26.8 | 29.7 | 34.7 |
| 7 | 1 | 26.8 | 29.3 | 34.0 |
| 8 | 1 | 27.6 | 30.0 | 35.0 |
| 9 | 1 | 27.6 | 30.0 | 35.1 |
| 10 | 1 | 28.5 | 30.7 | 36.2 |
| 11 | 1 | 28.4 | 31.0 | 36.2 |
| 12 | 1 | 28.7 | 31.7 | 36.9 |
| 13 | 1 | 29.1 | 31.5 | 36.9 |
| 14 | 1 | 29.5 | 32.0 | 37.3 |

FIGURE: Données complétés de Fish (SAS)

# Résultats de la MI pour les prog. $R_{1,2}$ & $R_{1,3}$

- **Résultat $R_{1,2}$**

| Caractéristiques des données manquantes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Moyennes de groupes | | |
| Groupe | Length3 | Length1 | Length2 | Fréq | Pourcentage | Length3 | Length1 | Length2 |
| 1 | X | X | X | 30 | 85.71 | 38.720000 | 30.603333 | 33.436667 |
| 2 | . | X | X | 3 | 8.57 | . | 29.033333 | 31.666667 |
| 3 | . | X | . | 2 | 5.71 | . | 27.750000 | . |

| Informations sur la variance (5 Imputations) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Variance | | | | | | |
| Variable | Inter | Intra | Total | DDL | Augmentation relative dans variance | Informations manquantes fraction | Efficacité relative |
| Length3 | 0.001700 | 0.497834 | 0.499873 | 32.027 | 0.004097 | 0.004088 | 0.999183 |
| Length2 | 0.000063673 | 0.439146 | 0.439222 | 32.157 | 0.000174 | 0.000174 | 0.999965 |

FIGURE: Fish(SAS)

- **Résultat $R_{1,3}$**

| Caractéristiques des données manquantes | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Moyennes de groupes | | |
| Groupe | Length1 | Length2 | Length3 | Fréq | Pourcentage | Length1 | Length2 | Length3 |
| 1 | X | X | X | 30 | 85.71 | 30.603333 | 33.436667 | 38.720000 |
| 2 | X | X | . | 3 | 8.57 | 29.033333 | 31.666667 | . |
| 3 | X | . | . | 2 | 5.71 | 27.750000 | . | . |

| Informations sur la variance (5 Imputations) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Variance | | | | | | |
| Variable | Inter | Intra | Total | DDL | Augmentation relative dans variance | Informations manquantes fraction | Efficacité relative |
| Length2 | 0.000051429 | 0.437241 | 0.437302 | 32.158 | 0.000141 | 0.000141 | 0.999972 |
| Length3 | 0.001402 | 0.491120 | 0.492802 | 32.049 | 0.003425 | 0.003419 | 0.999317 |

FIGURE: Fish (SAS)

# Exemple 2

- **Imputation variable nominale prog.** $R_{2,1}$
  - **proc mi** data = datPca1 seed=37851 nimpute=5 out = dPcaImp0 ;
    class agecl1 stade ;
    fcs logistic ;
    var br br_time marge stade ki67 agecl1 PSA0 ;
    **run** ;
- **Imputation multi-étapes prog.** $R_{2,2}$
  - Étape 1
    **proc mi** data = datPca1 seed=155417 nimpute=5 out = dPcaImp1 ;
    mcmc impute= monotone ;
    var br br_time marge ki67 PSA0 ;
    **run** ;
  - Étape 2
    **proc mi** data= dPcaImp1 nimpute=1 seed=51343 out = datr ;
    monotone reg ;
    var br br_time marge ki67 PSA0 ;
    by _Imputation_ ; **run** ;

# Données complétés prog. $R_{2,1}$

- **Résultat du prog $R_{2,1}$**

| Obs. | _Imputation_ | BR | BR_time | Marge | ki67 | PSA0 | stade | agecl1 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 3 | 1 | 2.5 | 6.7 | T3 | age<57 |
| 2 | 1 | 0 | 7 | 0 | 1.4 | 17.5 | T2 | 57-67 |
| 3 | 1 | 1 | 4 | 1 | 0.2 | 15 | T3 | age<57 |
| 4 | 1 | 1 | 13 | 1 | 1.4 | 10.8 | T3 | age<57 |
| 5 | 1 | 0 | 11 | 0 | 2.1 | 3 | T3 | age>67 |
| 6 | 1 | 0 | 8 | 0 | 0.5 | 48.141015908 | T2 | age>67 |
| 7 | 1 | 1 | 1 | 1 | 1.2 | 17.422649028 | T3 | 57-67 |
| 8 | 1 | 0 | 12 | 1 | 3.8 | 5.9 | T2 | age<57 |
| 9 | 1 | 0 | 11 | 1 | 0.2 | 8 | T3 | age>67 |
| 10 | 1 | 1 | 5 | 1 | 2 | 15.5 | T2 | age<57 |
| 11 | 1 | 0 | 9 | 0 | 0.8 | 10.2 | T2 | age>67 |
| 12 | 1 | 1 | 0 | 1 | 0.6 | 14.410768542 | T3 | age<57 |
| 13 | 1 | 0 | 8 | 0 | 0.6 | -3.538503247 | T2 | age<57 |
| 14 | 1 | 1 | 7 | 1 | 6.2879459498 | 8 | T2 | age>67 |

FIGURE: Données complétés de Pca

# Données complétés prog. $R_{2,2}$

- **Résultat du prog $R_{2,2}$ étape1**

**Procédure MI**

| Informations sur le modèle | |
|---|---|
| Table | WORK.DATPCA1 |
| Méthode | Monotone-Données MCMC |
| Chaîne à imputation multiple | Chaîne unique |
| Valeurs estimées initiales pour | Mode a posteriori EM |
| Début | Valeur de départ |
| A priori | Jeffreys |
| Nombre d'imputations | 5 |
| Nombre d'itérations Burn-in | 200 |
| Nombre d'itérations | 100 |
| Valeur initiale pour nb aléatoi | 155417 |

| Caractéristiques des données manquantes | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Moyennes de groupes | | | | |
| Groupe | BR | BR_time | Marge | ki67 | PSA0 | Fréq | Pourcentage | BR | BR_time | Marge | ki67 | PSA0 |
| 1 | X | X | X | X | X | 178 | 79.11 | 0.483146 | 6.662921 | 0.657303 | 2.435955 | 13.376966 |
| 2 | X | X | X | X | O | 35 | 15.56 | 0.428571 | 7.657143 | 0.771429 | 2.137143 | . |
| 3 | X | X | X | . | X | 10 | 4.44 | 0.400000 | 7.200000 | 0.800000 | . | 11.830000 |
| 4 | X | X | X | O | O | 2 | 0.89 | 0.500000 | 3.500000 | 1.000000 | . | . |

FIGURE: Données complétés étape 1 PCa

# Données complétés prog. $R_{2,2}$

- **Résultat du prog $R_{2,2}$ étape 2**

**Procédure MI**

**Numéro d'imputation=1**

| Informations sur le modèle | |
|---|---|
| Table | WORK.DPCAIMP1 |
| Méthode | Monotone |
| Nombre d'imputations | 1 |
| Valeur initiale pour nb aléatoi | 51343 |

| Indication du modèle monotone | |
|---|---|
| Méthode | Variables imputées |
| Régression | BR_time Marge ki67 PSA0 |

| Caractéristiques des données manquantes | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Moyennes de groupes | | | | |
| Groupe | BR | BR_time | Marge | ki67 | PSA0 | Fréq | Pourcentage | BR | BR_time | Marge | ki67 | PSA0 |
| 1 | X | X | X | X | X | 188 | 83.56 | 0.478723 | 6.691489 | 0.664894 | 2.400199 | 13.294681 |
| 2 | X | X | X | X | . | 35 | 15.56 | 0.428571 | 7.657143 | 0.771429 | 2.137143 | . |
| 3 | X | X | X | . | . | 2 | 0.89 | 0.500000 | 3.500000 | 1.000000 | . | . |

FIGURE: Données complétés étape 2 PCa

# proc MIANALYZE

- **Syntaxe**
  - **PROC MIANALYZE** <options> ; **BY** variables ; **CLASS** variables ; **MODELEFFECTS** effects ; <label :> TEST <, ..., <> > </ > ; **STDERR** variables ;
- **Définitions**
  - MODELEFFECTS : répertorie les effets à analyser.
  - STDERR : répertorie les erreurs types associés aux effets.

# Exemple 3

- **Prog $R_{3,1}$ étape2**

```
proc reg data=fsh1imp0 outest=outreg covout noprint;
model Length1= Length3 Length2;
by _Imputation_;
run;
```

```
proc print data=outreg(obs=8);
title 'Paramètres estimés pour chaque imputation';
run;
```

```
proc mianalyze data=outreg edf=33;
modeleffects Intercept Length3 Length2;
run;
```

FIGURE: Programme

# Résultat 3

- **Prog $R_{3,1}$ étape2**

**Paramètres estimés pour chaque imputation**

| Obs. | _Imputation_ | _MODEL_ | _TYPE_ | _NAME_ | _DEPVAR_ | _RMSE_ | Intercept | Length3 | Length2 | Length1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | MODEL1 | PARMS | | Length1 | 0.23951 | -0.31639 | 0.084451 | 0.82677 | -1 |
| 2 | 1 | MODEL1 | COV | Intercept | Length1 | 0.23951 | 0.40580 | -0.086864 | 0.08849 | . |
| 3 | 1 | MODEL1 | COV | Length3 | Length1 | 0.23951 | -0.08686 | 0.026630 | -0.02825 | . |
| 4 | 1 | MODEL1 | COV | Length2 | Length1 | 0.23951 | 0.08849 | -0.028246 | 0.03007 | . |
| 5 | 2 | MODEL1 | PARMS | | Length1 | 0.24955 | 0.02017 | -0.034003 | 0.95390 | -1 |
| 6 | 2 | MODEL1 | COV | Intercept | Length1 | 0.24955 | 0.41929 | -0.087118 | 0.08833 | . |
| 7 | 2 | MODEL1 | COV | Length3 | Length1 | 0.24955 | -0.08712 | 0.026549 | -0.02813 | . |
| 8 | 2 | MODEL1 | COV | Length2 | Length1 | 0.24955 | 0.08833 | -0.028130 | 0.02993 | . |

FIGURE: Programme (SAS)

# Résultat 3

- **Prog $R_{3,1}$ étape2**

**Procédure MIANALYZE**

| Informations sur le modèle | |
|---|---|
| Table | WORK.OUTREG |
| Nombre d'imputations | 5 |

**Informations sur la variance (5 Imputations)**

| Paramètre | Variance | | | DDL | Augmentation relative dans variance | Informations manquantes fraction | Efficacité relative |
|---|---|---|---|---|---|---|---|
| | Inter | Intra | Total | | | | |
| Intercept | 0.085099 | 0.398725 | 0.500844 | 19.725 | 0.256112 | 0.219940 | 0.957865 |
| Length3 | 0.009670 | 0.024147 | 0.035751 | 13.542 | 0.480580 | 0.357563 | 0.933260 |
| Length2 | 0.011064 | 0.027105 | 0.040383 | 13.364 | 0.489842 | 0.362346 | 0.932428 |

**Résultats estimés des paramètres (5 Imputations)**

| Paramètre | Valeur estimée | Erreur type | Intervalle de confiance à 95% | | DDL | Minimum | Maximum | Thêta0 | t pour H0: Paramètre=Theta0 | Pr > \|t\| |
|---|---|---|---|---|---|---|---|---|---|---|
| Intercept | -0.255843 | 0.707703 | -1.73340 | 1.221718 | 19.725 | -0.714852 | 0.020168 | 0 | -0.36 | 0.7216 |
| Length3 | 0.073426 | 0.189080 | -0.33340 | 0.480252 | 13.542 | -0.034003 | 0.228244 | 0 | 0.39 | 0.7038 |
| Length2 | 0.837938 | 0.200954 | 0.40500 | 1.270875 | 13.364 | 0.672484 | 0.953896 | 0 | 4.17 | 0.0010 |

FIGURE: Programme (SAS)

# Problème distribution des estimateurs avec MIANALYZE

- **MiANALYSE : hypothèse des estimateurs**
  - Si $\widehat{Q}_1, \cdots, \widehat{Q}_M$ obtenue par M-imputations.
  - $\widehat{Q}_I$ suit une distribution normal pour appliquer les règles de Rubin.
  - Sinon MIANALYSE problème.
  - l'exemple l'HR ...
- **Transformation de Van Buuren (2012)**
  - 

| Statistic | Transformation |
|---|---|
| Correlation | Fisher z |
| Odds ratio | Logarithm |
| Relative risk | Logarithm |
| Hazard ratio | Logarithm |
| Explained variance R2 | Fisher z on root |
| Survival probabilities | Complementary log-log |
| Survival distribution | Logarithm |

- **Modification de programme**
  - Utilisation des petites astuces...

# Modèle logistique

- **MIANALYZE : calcul des ORs**

```
* Calcul des ORs;
proc logistic data=datPimp0;
class stade(ref='T2')agecl1(ref='age<57');
model br(event='1')= stade agecl1;
ods output parameterestimates=lgsparms ODDSRATIOS=lgsodds;
BY _Imputation_;
RUN;
proc mianalyze parms(classvar=classval)=lgsparms;
class stade agecl1;
modeleffects stade agecl1;
ods output parameterestimates=lgsparms1;
run;
DATA lgsparms2; SET lgsparms1;
Estimate_back = EXP(ESTIMATE); *Pooled odds ratio;
LCL_back=Estimate_back*EXP(-1.96*STDERR); *Pooled lower limit;
UCL_back=Estimate_back*EXP(+1.96*STDERR); *Pooled upper limit;
RUN;
```

# Résultats

- **MIANALYSE : ORs**

| Obs. | _Imputation_ | Variable | ClassVal0 | DF | Estimate | StdErr | WaldChiSq | ProbChiSq | _ESTTYPE_ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Intercept | | 1 | -0.3299 | 0.1638 | 4.0562 | 0.0440 | MLE |
| 2 | 1 | stade | T3 | 1 | 0.5936 | 0.1597 | 13.8207 | 0.0002 | MLE |
| 3 | 1 | agecl1 | 57-67 | 1 | -0.1536 | 0.1869 | 0.6755 | 0.4111 | MLE |
| 4 | 1 | agecl1 | age>67 | 1 | -0.4365 | 0.2199 | 3.9403 | 0.0471 | MLE |
| 5 | 2 | Intercept | | 1 | -0.3211 | 0.1640 | 3.8354 | 0.0502 | MLE |
| 6 | 2 | stade | T3 | 1 | 0.5848 | 0.1566 | 13.9417 | 0.0002 | MLE |
| 7 | 2 | agecl1 | 57-67 | 1 | -0.0965 | 0.1871 | 0.2659 | 0.6061 | MLE |
| 8 | 2 | agecl1 | age>67 | 1 | -0.4722 | 0.2258 | 4.3747 | 0.0365 | MLE |
| 9 | 3 | Intercept | | 1 | -0.3149 | 0.1641 | 3.6830 | 0.0550 | MLE |
| 10 | 3 | stade | T3 | 1 | 0.5645 | 0.1588 | 12.6383 | 0.0004 | MLE |
| 11 | 3 | agecl1 | 57-67 | 1 | -0.1757 | 0.1860 | 0.8920 | 0.3449 | MLE |
| 12 | 3 | agecl1 | age>67 | 1 | -0.3614 | 0.2208 | 2.6798 | 0.1016 | MLE |
| 13 | 4 | Intercept | | 1 | -0.3374 | 0.1629 | 4.2902 | 0.0383 | MLE |
| 14 | 4 | stade | T3 | 1 | 0.5695 | 0.1576 | 13.0647 | 0.0003 | MLE |

# Résultats

- **MIANALYSE : ORs**

| Obs. | Estimate_back | LCI_back | UCI_back |
|---|---|---|---|
| 1 | 1.77957 | 1.30368 | 2.42917 |
| 2 | 0.86748 | 0.59811 | 1.25817 |
| 3 | 0.66932 | 0.42515 | 1.05371 |

# Calcul de l'effet Global

- **MIANALYSE : ORs**

```
DATA lgsodds_t; SET lgsodds(WHERE=(INDEX(EFFECT,"agecl1")));
log_or_lr_value=LOG(ODDSRATIOEST);
log_or_lr_se=(LOG(UPPERCL)-LOG(LOWERCL))/(2*1.96);
RUN;
```

```
proc mianalyze data=lgsodds_t(WHERE=(INDEX(EFFECT,'agecl1')));
ods output parameterestimates=logor1;
modeleffects log_or_lr_value;
STDERR log_or_lr_se;
run;
```

```
DATA mian_lgsodds_bt; SET logor1;
Estimate_back = EXP(ESTIMATE); *Pooled odds ratio;
LCL_back=Estimate_back*EXP(-1.96*STDERR); *Pooled lower limit;
UCL_back=Estimate_back*EXP(+1.96*STDERR); *Pooled upper limit;
RUN;
```

# Modèle de cox pour MI

- **MIANALYSE : HRs**

```
proc phreg data=datPimp0 outest=outp covout;
class stade(ref='T2') agecl1(ref='age<57');
model br_time*br(0)=marge stade agecl1 ki67 PSA0;
ods output ParameterEstimates=datpar GlobalTests=test1;
by _Imputation_;
run;
proc mianalyze parms(classvar=classval)=datpar;
class stade agecl1;
modeleffects marge stade agecl1 ki67 PSA0;
ods output ParameterEstimates=datpar1;
run;
data Hresti;set datpar1;
HR_est = EXP(ESTIMATE); *Pooled odds ratio;
LCI_HR=HR_est*EXP(-1.96*STDERR); *Pooled lower limit;
UCI_HR=HR_est*EXP(+1.96*STDERR); *Pooled upper limit;
run;
proc print data=Hresti;
var Parm stade agecl1 HR_est LCI_HR UCI_HR;
run;
```

# Resultat

- **MIANALYSE : HRs**

| Obs. | Parameter | Estimate | StdErr | HazardRatio |
|---|---|---|---|---|
| 1 | Marge | 0.33230 | 0.24508 | 1.394 |
| 2 | stade | 0.85282 | 0.25390 | 2.346 |
| 3 | agecl1 | -0.51286 | 0.23017 | 0.599 |
| 4 | agecl1 | -0.63136 | 0.28374 | 0.532 |
| 5 | ki67 | 0.17137 | 0.03816 | 1.187 |
| 6 | PSA0 | 0.02211 | 0.00403 | 1.022 |
| 7 | Marge | 0.28736 | 0.24352 | 1.333 |
| 8 | stade | 0.77297 | 0.26001 | 2.166 |
| 9 | agecl1 | -0.50194 | 0.22916 | 0.605 |
| 10 | agecl1 | -0.68766 | 0.28088 | 0.503 |
| 11 | ki67 | 0.15994 | 0.03786 | 1.173 |
| 12 | PSA0 | 0.01636 | 0.00455 | 1.016 |
| 13 | Marge | 0.34035 | 0.24107 | 1.405 |
| 14 | stade | 0.73347 | 0.25957 | 2.082 |

# Résultat

- **MIANALYSE : HRs**

| Obs. | Parm | stade | agecl1 | Estimate | StdErr |
|---|---|---|---|---|---|
| 1 | marge | | | 0.316110 | 0.244334 |
| 2 | stade | T3 | | 0.770821 | 0.261901 |
| 3 | agecl1 | | 57-67 | -0.478204 | 0.239308 |
| 4 | agecl1 | | age>67 | -0.698466 | 0.299561 |
| 5 | ki67 | | | 0.170500 | 0.041259 |
| 6 | PSA0 | | | 0.018655 | 0.004964 |

# Résultat

- **MIANALYSE : HRs**

| Obs. | Parm | stade | agecl1 | HR_est | LCI_HR | UCI_HR |
|---|---|---|---|---|---|---|
| 1 | marge | | | 1.37178 | 0.84977 | 2.21445 |
| 2 | stade | T3 | | 2.16154 | 1.29369 | 3.61159 |
| 3 | agecl1 | | 57-67 | 0.61990 | 0.38781 | 0.99088 |
| 4 | agecl1 | | age>67 | 0.49735 | 0.27648 | 0.89465 |
| 5 | ki67 | | | 1.18590 | 1.09377 | 1.28578 |
| 6 | PSA0 | | | 1.01883 | 1.00897 | 1.02879 |

**Bohdana Ratitch** et al (2013).
*Combining Analysis Results from Multiply Imputed Categorical Data*

...

Merci ...
mortderire.com