Rapport de méthodes

Enquête suisse sur la structure des salaires 2002

0

Statistische Grundlagen und Übersichten
Bases statistiques et produits généraux
Basi statistiche e presentazioni generali

La précision du salaire brut standardisé médian

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

OFS BFS UST

Neuchâtel, 2005

Statistik der Schweiz

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Informationsgesellschaft, Sport
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Kriminalität und Strafrecht
20 Wirtschaftliche und soziale Situation der Bevölkerung
21 Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene

Statistique de la Suisse

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

0 Bases statistiques et produits généraux
1 Population
2 Espace et environnement
3 Vie active et rémunération du travail
4 Economie nationale
5 Prix
6 Industrie et services
7 Agriculture et sylviculture
8 Energie
9 Construction et logement
10 Tourisme
11 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, société de l'information, sport
17 Politique
18 Administration et finances publiques
19 Criminalité et droit pénal
20 Situation économique et sociale de la population
21 Développement durable et disparités régionales et internationales

Statistik der Schweiz | **Statistique de la Suisse**

Methodenbericht | **Rapport de méthodes**

Enquête suisse sur la structure des salaires 2002

La précision du salaire brut standardisé médian

Auteurs

Monique Graf
Office fédéral de la statistique,

Alina Matei
Université de Neuchâtel

Editeur

Office fédéral de la statistique

OFS BFS UST

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

Neuchâtel, 2005

Préambule

L'enquête suisse sur la structure des salaires 2002 est la première pour laquelle les estimations au niveau des grandes régions (NUTS 2) sont systématiquement publiées. Il était donc opportun de faire le point sur la précision obtenue. Dans ce but, le Service de méthodes statistiques (Meth) a confié un mandat de recherche avec le Groupe de Statistique de l'Université de Neuchâtel (prof. Yves Tillé). Le but du mandat est l'étude systématique de la précision de la variable principale de l'enquête, le salaire mensuel brut standardisé. Ce travail a été réalisé par Alina Matei, assistante au Groupe de Statistique, sous la direction de Monique Graf, Meth.

Un grand merci à la section LOHN, en particulier à son chef Didier Froidevaux et à Jacques Méry, pour leur intérêt pour cette recherche et les fructueuses discussions tout au long de ce travail.

Résumé

Cette étude est une étude systématique de la précision des salaires mensuels médians publiés dans le cadre de l'enquête suisse sur la structure des salaires 2002. Selon la méthode développée pour l'enquête 2000, des intervalles de confiance non paramétriques à 95% pour la médiane ont été calculés, pour les secteurs privé et public, par branche économique, le niveau de qualification, le sexe, la taille de l'entreprise et la région. Les intervalles de confiance sont représentés ici sous forme graphique. Les intervalles de confiance calculés n'étant pas symétriques, on définit dans ce cas une mesure analogue celle du coefficient de variation et dénommée coefficient de variation synthétique 95%. Pour résumer la masse d'informations produites sur la précision, nous avons établi un modèle de prévision de ce coefficient de variation.

Pour le secteur privé, deux modèles linéaires ont été choisis pour estimer le coefficient de variation synthétique 95%: un modèle pour les données agrégées au niveau Suisse (où on a pris en compte la branche économique, le niveau de qualification et le sexe) et un modèle pour les données complètes (où on a pris en compte la branche économique, le sexe et la région). Les deux modèles sont utilisables pour une estimation globale de la précision.

Mots-clé

rapport de méthodes; LSE; ESS; plan d'échantillonnage; coefficient de variation; intervalle de confiance; salaire.

Complément d'information: Monique Graf, tél. 032 713 66 11
Monique.Graf@bfs.admin.ch

Réalisation: Service de méthodes statistiques, OFS

Diffusion: Office fédéral de la statistique
CH-2010 Neuchâtel
Tél. 032 713 60 60 / Fax 032 713 60 61
Order@bfs.admin.ch

Internet: http://www.statistik.admin.ch

Numéro de commande: 338-0031

Prix: gratuit

Série: Statistique de la Suisse

Domaine: 0 Bases statistiques et produits généraux

Langue du texte original: Français

Graphisme/Layout: OFS

Copyright: OFS, Neuchâtel 2005
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.

ISBN: 3-303-00317-3

Table des matières

1 Introduction **5**

2 Intervalles de confiance pour la médiane **5**

3 Modélisation du coefficient de variation synthétique 95% **6**
3.1 Suisse 6
3.2 Les régions 13

4 Discussion et conclusions **19**
4.1 Discussion 19
4.2 Conclusions 24

Annexe 1
Délimitation du fichier SAS d'entrée des différents tableaux et programmes R **25**

Annexe 2
Résultats pour la Suisse **27**

Annexe 3
Intervalles de confiance pour la médiane **28**

Annexe 4
Perte de précision de l'estimation **48**

Références **49**

1 Introduction

Cette étude est une étude systématique de la précision des salaires mensuels médians publiés dans le cadre de l'enquête suisse sur la structure des salaires 2002 (ESS 2002). L'enquête 2002 est la première pour laquelle une ventilation par régions NUTS II (grandes régions) a été introduite. Pour avoir une précision suffisante des médianes estimées, la taille de l'échantillon a été considérablement augmenté par rapport à l'enquête 2000 (environ 18 000 entreprises répondantes en 2000 pour 42 000 en 2002). Le plan d'échantillonnage est décrit dans [2]. La précision des médianes est évaluée par un intervalle de confiance à 95% (chapitre 2). De cet intervalle est déduit un coefficient de variation synthétique 95% ($CV_{syn95\%}$), qui est notre critère de précision. La procédure est entièrement décrite dans [1].

La masse d'information produite est considérable : 49 branches d'activité x 8 classifications régionales (Suisse + les 7 grandes régions) x 3 sexe (total, homme, femme) x 4 niveau de qualification (total, 1+2, 3, 4). Pour condenser les résultats, nous avons élaboré des modèles prédictifs de ces coefficients de variation (chapitre 3). On trouve des modèles prévisionnels de précision au Bureau of Labor Statistics, U.S. Department of Labor pour l'estimation de l'écart type [6], au U.S. Census 2000 pour l'estimation de la variance généralisée [7] etc. Nos modèles résument l'information et permettent également de lisser les estimations des $CV_{syn95\%}$. Lorsque nous ne prenons que les données agrégées au niveau Suisse (pas de ventilation régionale), le coefficient de détermination (R^2) du modèle est de 0.8 ; il est de 0.6 avec les données complètes. Ces modèles sont donc utilisables pour une estimation globale de la précision.

2 Intervalles de confiance pour la médiane

La méthode de calcul d'un intervalle de confiance pour la médiane est non-paramétrique et n'est donc pas basée sur l'évaluation de l'écart-type. Les intervalles de confiance calculés ne sont pas symétriques. Dans ce cas, une mesure analogue à celle du coefficient de variation et dénommée coefficient de variation synthétique 95% ($CV_{syn95\%}$) a été utilisée (voir [1], p.26). Le coefficient de variation synthétique 95% est exprimé en pourcentage et reflète la longueur d'un intervalle de confiance par rapport à un facteur qui est égal a la médiane multipliée par 1.96 :

$$CV_{syn95\%} = 100\frac{\max(BS - M, M - BI)}{1.96M}, \tag{1}$$

où BI, BS sont respectivement les bornes inférieure et supérieure de l'intervalle de confiance à 95% et M est la médiane. A partir de $CV_{syn95\%}$ on peut toujours reconstruire un intervalle de confiance conservatif, c'est-à-dire qui contient toujours l'intervalle de confiance exact. C'est le plus petit intervalle symétrique autour de la médiane qui contient l'intervalle de confiance.

La nécessité de résumer la précision nous a poussé à chercher des moyens de modéliser le $CV_{syn95\%}$ à partir de la branche économique, le niveau de qualification, le sexe et la région. Pour le secteur privé, la modélisation de $CV_{syn95\%}$ est basée sur une approche paramétrique, en utilisant

un modèle linéaire, où les variables explicatives sont les facteurs : branche économique, niveau de qualification, sexe et région.

Concernant les intervalles de confiances calculés, on a choisi une manière graphique de représentation, pour éviter la publication des grands tableaux numériques. L'Annexe 3 présente ces résultats, en spécifiant chaque fois de quel cas il s'agit.

Les intervalles de confiance pour la médiane ont été calculés pour les tableaux standard donnés en [3], [4] et [5], sauf le cas f, publié à part pour Eurostat :

a. Tableau TA1 - salaire mensuel brut selon la branche économique, le niveau de qualification requise pour le poste du travail et le sexe - Secteur privé, pour la Suisse (voir [5], p.43) ;
b. Tableau TA1 par régions - le tableau TA1 calculé pour chaque région selon la branche économique et le sexe (voir [4]) ;
c. Tableau TA2 - salaire mensuel brut selon la branche économique, le niveau de qualification requise pour le poste du travail et le sexe - Secteur public, pour la Suisse (voir [5], p.46) ;
d. Tableau TA4 - salaire mensuel brut selon la branche économique, la taille de l'entreprise et le sexe - Secteur privé, pour la Suisse (voir [5], p.48) ;
e. Tableau TA7 - salaire mensuel brut selon le domaine d'activité, le niveau de qualification requise pour le poste du travail et le sexe - Secteur privé et public ensemble, pour la Suisse (voir [3], p.16, [5], p.51-53) ;
f. Tableau Eurostat - salaire mensuel brut pour la Suisse calculé pour Eurostat - Secteur privé et public ensemble.

Les graphiques correspondants sont présentés dans l'Annexe 3. Les intervalles montrent qu'en général on a une très bonne discrimination entre les catégories considérées. On remarque cependant dans ces graphiques quelques intervalles de confiance très grands, en particulier pour la branche 61. Ce sont les cas où, en général, "l'interprétation de la valeur de la médiane est limitée, celle-ci étant relativement incertaine sur le plan statistique" (voir [5], page 42 et [4], page 36).

3 Modélisation du coefficient de variation synthétique 95%

3.1 Suisse

Nous commençons par le coefficient de variation synthétique calculé dans la Suisse entière pour le secteur privé (il s'agit des données du tableau TA1, 2a. ci-dessus). Nous prenons en compte les valeurs $\log(CV_{syn95\%})$, où $\log$ est le logarithme naturel.

3.1.1 Statistiques descriptives pour $\log(CV_{syn95\%})$ - Suisse

On est intéressé d'analyser la variable $\log(CV_{syn95\%})$ en fonction des catégories : branche économique (NOGA2), niveau de qualification (ANFORNI) et sexe (GESCHLE), où $\log$ est la fonction logarithmique. On a :

1. la branche économique (NOGA2) avec 50 classes ;

2. le niveau de qualification (ANFORNI) avec 4 modalités (0=total, 1+2=travaux les plus exigeants et tâches les plus difficiles+travail indépendant et très qualifié, 3=connaissances professionnelles spécialisées, 4=activités simples et répétitives) ;
3. le sexe (GESCHLE) avec 3 modalités (0=total, 1=homme, 2=femme).

La Fig. 1 présente l'histogramme et la boîte à moustaches de la variable $\log(CV_{syn95\%})$ pour toutes les modalités croisées des catégories. Cette variable a une faible asymétrie (le coefficient d'asymétrie est 1.26) et on remarque des valeurs extrêmes. Le coefficient d'aplatissement de 5.35 dans ce cas montre une l'allure générale de la distribution assez voûtée.

La fig. 2 présente notre variable mesurée dans chaque classe NOGA2. On peut remarquer qu'il y a une asymétrie aux niveaux : 15, 21, 22, 29, 61, 66, 70, 72, 73, 80, 5052, 6567, 7074 (dans cette figure les classes NOGA2 sont dans l'ordre numérique). Les valeurs $\log(CV95\%)$ pour la classe 61 sont plus élevées, en comparaison avec toutes les autres classes. C'est un cas spécial et des explications sont données ultérieurement dans cette étude. On observe aussi qu'il y a une grande variabilité de $\log(CV95\%)$ d'une classe NOGA2 à l'autre (voir la longueur de chaque boîte).

Pour les facteurs ANFORNI (niveau de qualification) et GESCHLE (sexe), la variable $\log(CV95\%)$ suit une distribution assez symétrique (voir la fig. 3).

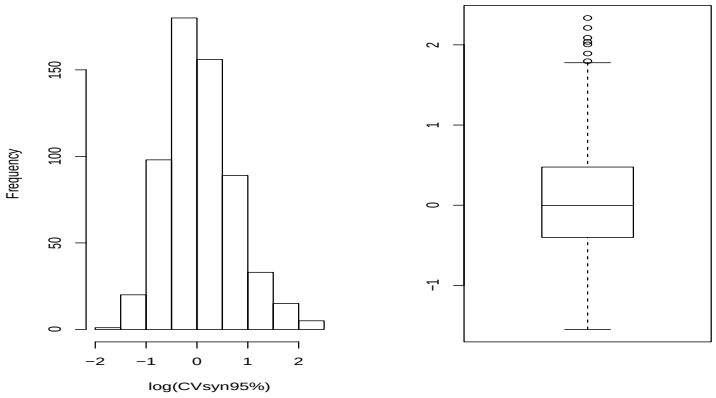


FIG. 1 – Histogramme et boîte à moustaches pour $\log(CV_{syn95\%})$ - Suisse

3.1.2 Modèle 1 - un modèle linéaire pour la Suisse

A partir d'une catégorisation par branche économique, niveau de qualification et sexe, on voudrait avoir une estimation du coefficient de variation synthétique 95%.

On considère alors la variable expliquée $CV_{syn95\%}$ et les variables explicatives : branche économique (NOGA2), niveau de qualification (ANFORNI) et sexe (GESCHLE). Les variables explicatives sont des facteurs non ordonnés à plusieurs niveaux ("levels") : la branche économique avec 50 niveaux, le niveau de qualification avec 4 niveaux et le sexe avec 3 niveaux.

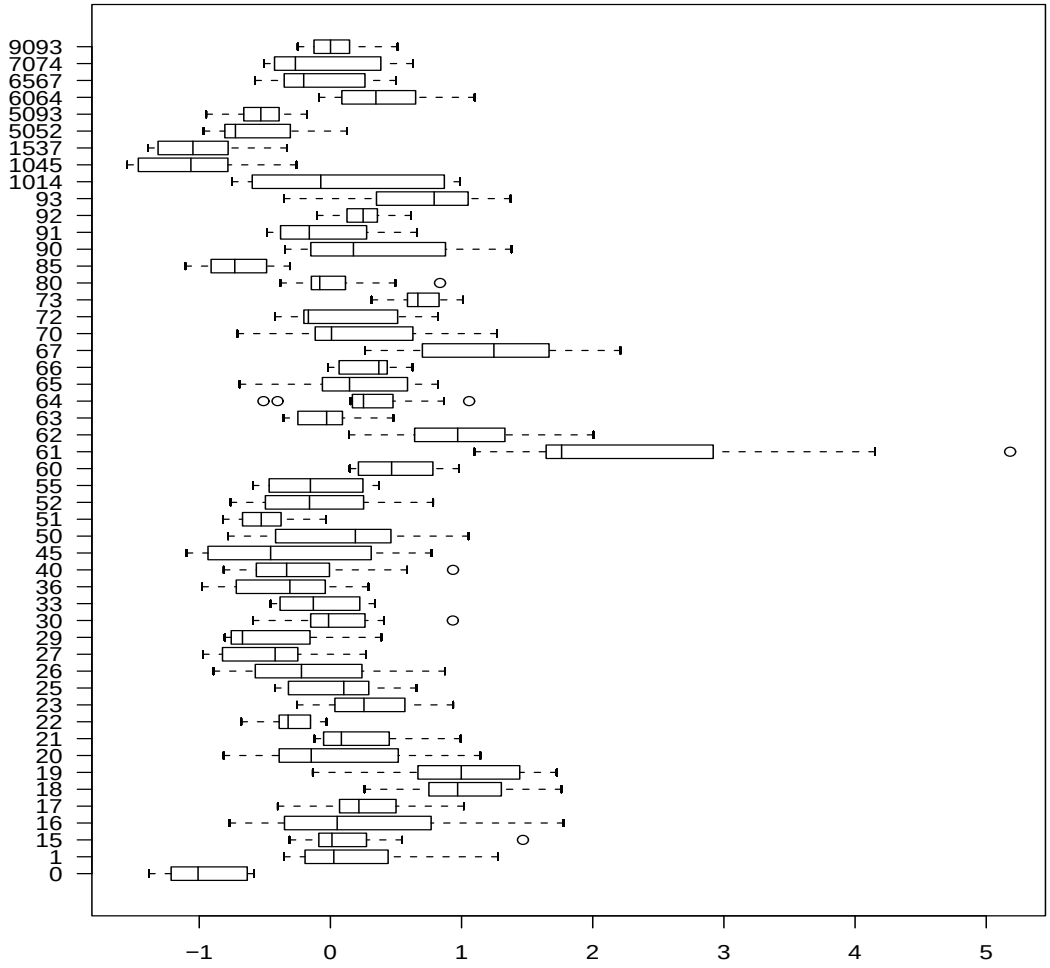


FIG. 2 – Boîte à moustaches pour $\log(CV_{syn95\%})$ pour chaque niveau NOGA2 - Suisse

Un des modèles pouvant être retenu est un modèle linéaire additif :

$$\log(CV_{syn95\%}) = \beta_0 + \beta_1(NOGA2) + \beta_2(ANFORNI) + \beta_3(GESCHLE) + \epsilon. \qquad (2)$$

Ce modèle permet de prédire les valeurs moyennes du $\log(CV_{syn95\%})$ pour une branche économique, un niveau de qualification et un sexe fixés. Trois observations avec de très grandes valeurs pour $\log(CV_{syn95\%})$ ont été éliminées. Ces trois observations correspondent à :

- NOGA2=61, ANFORNI=0, GESCHLE=2,
- NOGA2=61, ANFORNI=4, GESCHLE=0,

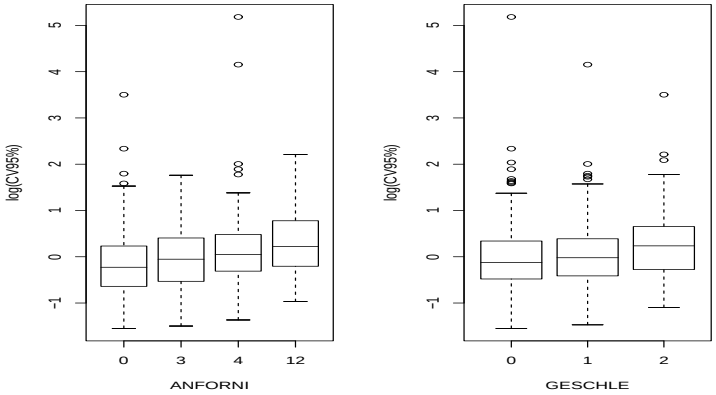


FIG. 3 – Boîte à moustaches pour $\log(CV_{syn95\%})$ pour chaque niveau ANFORNI (gauche) et chaque niveau GESCHLE (droite) - Suisse

– NOGA2=61, ANFORNI=4, GESCHLE=1.

Le fichier a 597 valeurs $CV_{syn95\%}$.

Nos variables explicatives sont catégorielles. Dans ce cas, on a choisi d'utiliser un contraste de type traitement. Les références sont : la branche 0 (toutes les classes NOGA2 ensemble), le niveau de qualification 0 (tous les niveaux ANFORNI ensemble) et le sexe 0 (hommes et femmes ensemble). La grande variabilité de $\log(CV_{syn95\%})$ dans les classes NOGA2 induit une hétéroscédasticité des erreurs. Pour éliminer cette hétéroscédasticité on a utilisé un ensemble de poids w_{NOGA2} défini par :

$$w_{NOGA2}[i] = var(\log(CV_{syn95\%}))/w_i, \tag{3}$$

où $var(\log(CV_{syn95\%}))$ est la variance de $\log(CV_{syn95\%})$ calculée pour toutes les observations et w_i représente la variance de $\log(CV_{syn95\%}))$ calculée dans la classe NOGA2 numéro $i, i = 1, \ldots, 50$. Le poids (3) est supposé inversement proportionnel à la variance de ϵ dans (2).

L'analyse faite à l'aide du logiciel R (voir le programme dans l'annexe 1) donne les coefficients estimés du modèle (voir TAB.1).

Statistiques générales du modèle 1 :		
Ecart-type des résidus	=	0.5985 avec 542 degrés de liberté
Coefficient de détermination multiple R^2	=	0.771
Rapport F pour la régression	=	33.8 avec 54 et 542 degrés de liberté et la p-valeur 2.2e-16

L'analyse de variance montre que tous les facteurs affectent significativement la variable dépendante. La Fig. 4 présente l'histogramme et la boîte à moustaches des résidus, le tracé QQ (Normal QQplot) et le graphique des valeurs estimées de la réponse versus les résidus standardisés. On observe que l'allure de la distribution des résidus est proche de la normale (le coefficient d'applatisment est 0.46, le coefficient d'asymétrie 0.24). Le tracé QQ des résidus a un caractère presque

TAB. 1 – Coefficients du modèle 1

Variable	Modalité	Coefficient estimé	Ecart-type estimé	Test de nullité	Significativité du test
Constante		-1.180	0.081	-14.649	< 2e-16
NOGA2	0	Référence	.	.	.
	1	1.123	0.146	7.718	5.72E-14
	15	1.106	0.144	7.653	9.07E-14
	16	1.178	0.209	5.631	2.88E-08
	17	1.209	0.120	10.072	< 2e-16
	18	1.957	0.139	14.089	< 2e-16
	19	1.919	0.158	12.163	< 2e-16
	20	0.950	0.174	5.449	7.71E-08
	21	1.159	0.118	9.842	< 2e-16
	22	0.645	0.089	7.228	1.68E-12
	23	1.221	0.118	10.375	< 2e-16
	25	0.990	0.120	8.257	1.15E-15
	26	0.816	0.166	4.924	1.13E-06
	27	0.481	0.120	4.021	6.62E-05
	29	0.487	0.127	3.832	1.42E-04
	30	0.981	0.131	7.516	2.35E-13
	33	0.861	0.110	7.829	2.59E-14
	36	0.604	0.125	4.844	1.66E-06
	40	0.739	0.154	4.800	2.06E-06
	45	0.641	0.188	3.411	6.96E-04
	50	1.038	0.164	6.345	4.70E-10
	51	0.431	0.097	4.444	1.07E-05
	52	0.826	0.150	5.502	5.80E-08
	55	0.820	0.123	6.678	6.01E-11
	60	1.448	0.106	13.661	< 2e-16
	61	2.649	0.129	20.562	< 2e-16
	62	1.943	0.169	11.481	< 2e-16
	63	0.913	0.103	8.901	< 2e-16
	64	1.221	0.137	8.920	< 2e-16
	65	1.125	0.142	7.924	1.31E-14
	66	1.236	0.093	13.346	< 2e-16
	67	2.172	0.174	12.474	< 2e-16
	70	1.151	0.172	6.674	6.15E-11
	72	1.019	0.131	7.787	3.52E-14
	73	1.625	0.094	17.202	< 2e-16
	80	0.975	0.114	8.536	< 2e-16
	85	0.229	0.099	2.311	2.12E-02
	90	1.300	0.168	7.752	4.51E-14
	91	0.898	0.128	6.996	7.82E-12
	92	1.190	0.090	13.232	< 2e-16
	93	1.625	0.147	11.028	< 2e-16
	1014	1.027	0.194	5.286	1.82E-07
	1045	-0.119	0.128	-0.932	0.352
	1537	-0.068	0.113	-0.607	0.544
	5052	0.374	0.115	3.254	1.00E-03
	5093	0.410	0.096	4.259	2.42E-05
	6064	1.324	0.118	11.219	< 2e-16
	6567	0.857	0.120	7.113	3.60E-12
	7074	0.861	0.133	6.476	2.12E-10
	9093	0.991	0.096	10.306	< 2e-16
ANFORNI	0	Référence	.	.	.
	3	0.080	0.035	2.262	0.024
	4	0.165	0.036	4.644	4.28E-06
	12	0.350	0.035	9.888	< 2e-16
GESCHLE	0	Référence	.	.	.
	1	0.064	0.031	2.095	0.037
	2	0.199	0.031	6.491	1.92E-10

linéaire ce qui donne une idée de la normalité de la distribution. La boîte à moustaches des résidus indique que la plupart des résidus ont des valeurs entre -1 et 1. Le nuage de points des résidus standardisés croisés avec les valeurs prédites présente des points aléatoirement répartis entre les bornes - 2 et + 2 et ne montre pas de formes "suspectes". Quelques points semblent hors norme, c'est-à-dire qu'il y a des valeurs aberrantes.

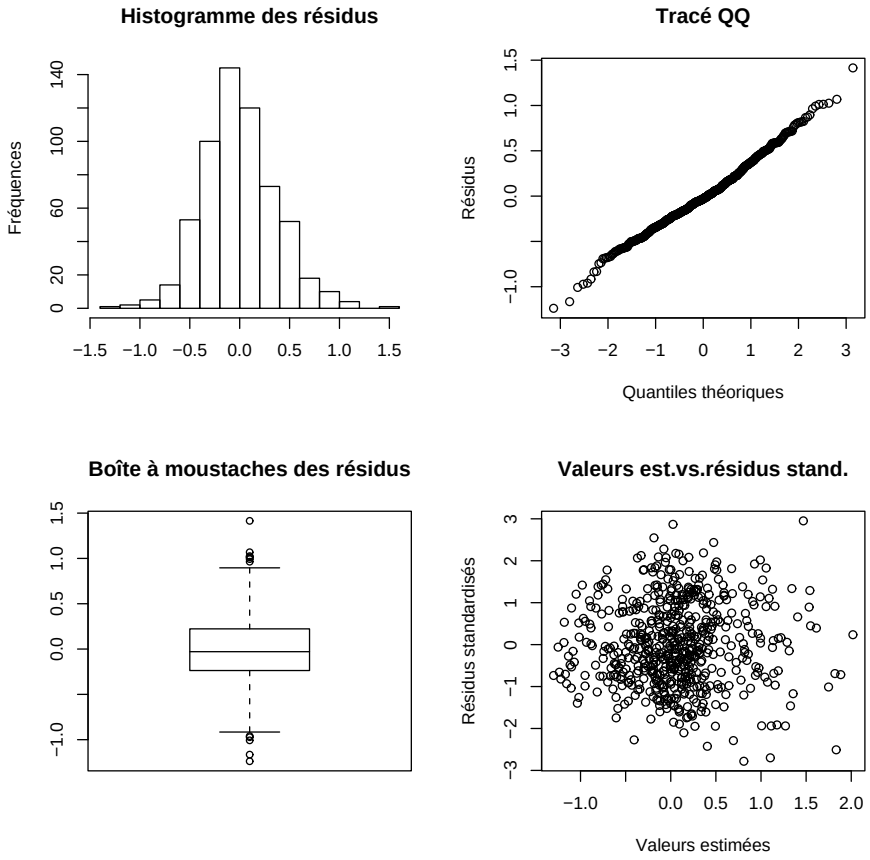


FIG. 4 – Résidus pour le modèle 1

Il y a des coefficients des niveaux qui ne sont pas significativement différents de zéro au seuil de 5% : pour les branches 1045 et 1537. Ces branches ont une précision analogue à la précision globale (toutes classes NOGA2 ensemble), appelée classe de référence. On a conservé ces branches dans le modèle pour avoir une idée sur leur comportement par rapport à la classe de référence. Le coefficient de détermination R^2 est égal a 77%, c'est-à-dire que 77% de la variance de la réponse est expliquée par le modèle.

Concernant la variable GESCHLE, on a un coefficient de 0.064 pour la modalité 1 (homme) et de 0.199 pour la modalité 2 (femme), c'est-à-dire que les salaires médians des femmes sont moins bien estimés dans notre modèle. Toutefois la différence les salaires médians des hommes et des

femmes n'est pas statistiquement significative.

3.1.3 Interprétation des coefficients

Le modèle considéré est un modèle linéaire :

$$\log(CV_{syn95\%}) = \beta_0 + \beta_1(NOGA2) + \beta_2(ANFORNI) + \beta_3(GESCHLE) + \epsilon.$$

Dans l'échelle originale de la réponse, ce modèle devient :

$$CV_{syn95\%} = \exp\left(\beta_0 + \beta_1(NOGA2) + \beta_2(ANFORNI) + \beta_3(GESCHLE) + \epsilon\right).$$

La constante $\widehat{\beta}_0$ du modèle est -1.18. Elle correspond au cas où la branche économique est 0 (NOGA2=0), le niveau de qualification est 0 (ANFORNI=0), et le sexe est 0 (GESCHLE=0), c'est-à-dire qu'on considère le cas où toutes les branches économiques, tous les niveaux de qualification et tous les sexes sont pris ensemble. On a

$$E(\log(CV_{syn95\%})|NOGA2 = 0, ANFORNI = 0, GESCHLE = 0) = -1.18,$$

et

$$\widehat{CV_{syn95\%}} = \exp(-1.18) = 0.31,$$

où E représente l'espérance et $\exp$ est la fonction exponentielle. La fonction exponentielle a dû être appliquée car le modèle choisi estime les valeurs du $\log(CV_{syn95\%})$. On a une valeur estimée $\widehat{CV_{syn95\%}} = 0.31$, alors que la valeur observée de $CV_{syn95\%}$ est 0.35. La différence entre les valeurs estimées et les vraies valeurs vient du fait que notre modèle explique 77% la variabilité du $CV_{syn95\%}$ et il y a toujours une erreur.

Les coefficients associés aux niveaux des variables de type facteur s'appellent des *contrastes*. Ce sont les différences entre l'effet de chaque niveau et le niveau de référence. Les coefficients des niveaux doivent être interprétés à partir de leur référence. Par exemple, le paramètre de la branche NOGA2=1 est égal a 1.12, c'est-à-dire que cette branche a une valeur moyenne estimée de 1.12 plus que la branche 0. Pour obtenir la valeur du $\widehat{CV_{syn95\%}}$ pour la branche 1, le niveau de qualification 0 et le sexe 0, il faut ajouter cette valeur au coefficient -1.18. On obtient alors :

$$\widehat{CV_{syn95\%}} = \exp(-1.18 + 1.12) = 0.95.$$

Pour la branche 1, le niveau 3 et le sexe 0, on va ajouter à la constante les coefficients correspondants à ces contrastes, c'est-à-dire les valeurs 1.12 et 0.08. Dans ce cas, on obtient

$$\widehat{CV_{syn95\%}} = \exp(-1.18 + 1.12 + 0.08) = 1.02.$$

Enfin la valeur $\widehat{CV}_{syn95\%}$ pour la branche 1, le niveau 3 et le sexe 2 est donné par :

$$\widehat{CV_{syn95\%}} = \exp(-1.18 + 1.12 + 0.08 + 0.19) = 1.25.$$

3.1.4 Une transformation des données

Selon [5], p. 42-43, il y a des cas où le calcul de la médiane n'a pas pu être publié (pas assez de données disponible) :

- NOGA2=10-14, ANFORNI=4, GESCHLE=2 ;
- NOGA2=61, ANFORNI=1+2, GESCHLE=2 ;
- NOGA2=62, ANFORNI=4, GESCHLE=2.

Il y a aussi des autres cas où l'interprétation de la médiane calculée est limitée (l'intervalle de confiance est grand ; voir [4], p.42-43) :

- NOGA2=61, ANFORNI=0, GESCHLE=0 ;
- NOGA2=61, ANFORNI=0, GESCHLE=2 ;
- NOGA2=61, ANFORNI=4, GESCHLE=0 ;
- NOGA2=61, ANFORNI=4, GESCHLE=1 ;
- NOGA2=61, ANFORNI=4, GESCHLE=2.

Trois parmi ces 8 observations ont été déjà éliminées dans le modèle 1. La fig. 5 présente notre variable mesurée dans chaque classe NOGA2. En comparaison avec la fig. 2, la classe 61 présente beaucoup moins de variabilité. Pour cette transformation des données, la fig. 3 est modifiée et le résultat est présenté dans la fig. 6.

En reprenant le modèle 1 (sans les 8 observations mentionnées) on obtient les coefficients donnés dans le tableau 2.

Statistiques générales du modèle 1 transformé :

Ecart-type des résidus	=	0.5935 avec 537 degrés de liberté ;
Coefficient de détermination multiple R^2	=	0.845 ;
Rapport F pour la régression	=	54.29 avec 54 et 537 degrés de liberté ; la p-valeur 2.2e-16.

En comparaison avec le cas précédent (voir tab.1), on observe que les coefficients du modèle changent très peu (voir tab.2), mais, par contre, le coefficient de détermination augmente ($R^2 = 0.84$ en comparaison avec 0.77). L'analyse de variance montre que tous les facteurs affectent significativement la variable dépendante. Les résidus se présentent comme à la Fig. 4. Puisque 84% de la variance de la réponse est expliquée par le modèle, on conclut que cette transformation sur les données est utilisable.

3.2 Les régions

Comme pour la Suisse entière, on prend en compte la variable $CV_{syn95\%}$ calculée pour toutes les régions (voir les tableaux TA1 par régions en [4], p.37-43), mais seulement pour le cas ANFORNI=0 (en tout). On considère la variable expliquée $\log(CV_{syn95\%})$ et les variables explicatives : branche économique (NOGA2), sexe (GESCHLE) et région (REGION). Les variables explicatives sont des facteurs :

1. branche économique : 50 niveaux (les classes NOGA2) ;
2. sexe : 3 niveaux (total=0, homme=1, femme=2) ;

TAB. 2 – Coefficients du modèle 1 sur les données modifiées

Variable	Modalité	Coefficient estimé	Ecart-type estimé	Test de nullité	Significativité du test
Constante		-1.174	0.08115	-14.46	< 2e-16
NOGA2	0	Référence	.	.	.
	1	1.12335	0.147	7.646	9.64E-14
	15	1.106	0.146	7.581	1.52E-13
	16	1.178	0.211	5.578	3.86E-08
	17	1.209	0.121	9.977	< 2e-16
	18	1.957	0.140	13.957	< 2e-16
	19	1.919	0.159	12.050	< 2e-16
	20	0.950	0.176	5.398	1.01E-07
	21	1.159	0.119	9.750	< 2e-16
	22	0.645	0.090	7.160	2.67E-12
	23	1.221	0.119	10.278	< 2e-16
	25	0.990	0.121	8.179	2.08E-15
	26	0.816	0.167	4.878	1.42E-06
	27	0.481	0.121	3.983	7.73E-05
	29	0.487	0.128	3.797	1.63E-04
	30	0.981	0.132	7.445	3.87E-13
	33	0.861	0.111	7.756	4.43E-14
	36	0.604	0.126	4.799	2.07E-06
	40	0.739	0.155	4.755	2.55E-06
	45	0.641	0.190	3.379	7.81E-04
	50	1.038	0.165	6.285	6.77E-10
	51	0.431	0.098	4.402	1.29E-05
	52	0.826	0.152	5.450	7.67E-08
	55	0.820	0.124	6.615	8.98E-11
	60	1.448	0.107	13.533	< 2e-16
	61	2.615	0.088	29.651	< 2e-16
	62	1.946	0.184	10.585	< 2e-16
	63	0.913	0.104	8.817	< 2e-16
	64	1.221	0.138	8.836	< 2e-16
	65	1.125	0.143	7.850	2.27E-14
	66	1.236	0.093	13.221	< 2e-16
	67	2.172	0.176	12.357	< 2e-16
	70	1.151	0.174	6.612	9.18E-11
	72	1.019	0.132	7.714	5.98E-14
	73	1.625	0.095	17.040	< 2e-16
	80	0.975	0.115	8.456	2.63E-16
	85	0.229	0.100	2.289	2.25E-02
	90	1.300	0.169	7.679	7.64E-14
	91	0.898	0.130	6.930	1.21E-11
	92	1.190	0.091	13.109	< 2e-16
	93	1.625	0.149	10.924	< 2e-16
	1014	1.010	0.210	4.802	2.04E-06
	1045	-0.119	0.129	-0.923	0.356457
	1537	-0.068	0.114	-0.601	0.547813
	5052	0.374	0.116	3.223	1.35E-03
	5093	0.410	0.097	4.219	2.88E-05
	6064	1.324	0.119	11.113	< 2e-16
	6567	0.857	0.122	7.047	5.64E-12
	7074	0.861	0.134	6.415	3.09E-10
	9093	0.991	0.097	10.209	< 2e-16
ANFORNI	0	Référence	.	.	.
	3	0.07406	0.03446	2.149	0.032051
	4	0.160	0.036	4.491	8.68E-06
	12	0.325	0.035	9.372	< 2e-16
GESCHLE	0	Référence	.	.	.
	1	0.077	0.030	2.600	0.00959
	2	0.1955	0.03033	6.445	2.57E-10

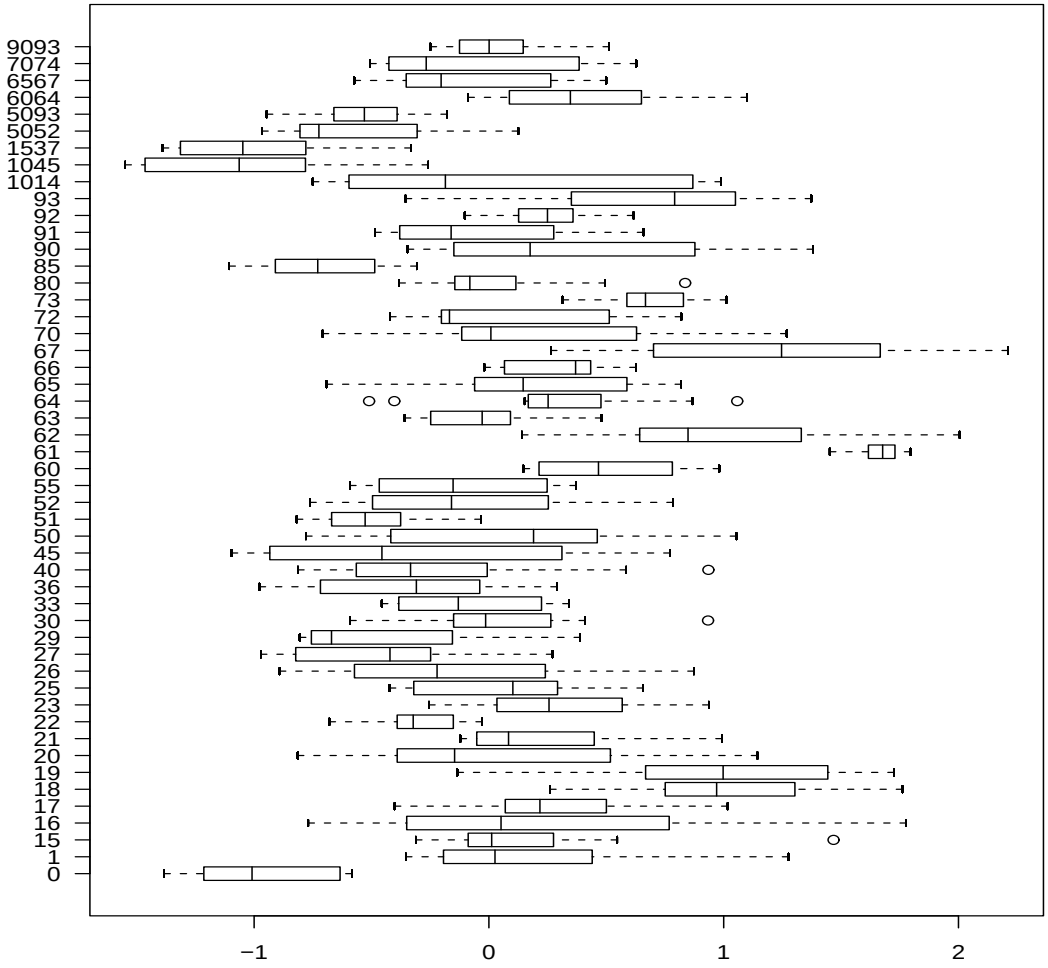


FIG. 5 – Boîte à moustaches pour $\log(CV_{syn95\%})$ pour chaque niveau NOGA2, variante II - Suisse

3. régions : 8 niveaux (les régions 0 - 7).

Dans notre notation, les régions sont les grandes régions : 1 - Région lémanique, 2 - Espace Mittelland, 3 - Suisse du Nord-Ouest, 4 - Zurich, 5 - Suisse orientale, 6 - Suisse centrale et 7 - Tessin. L'analyse a été faite par rapport à la Suisse entière, c'est-à-dire que les valeurs de $CV_{syn95\%}$ pour chaque région sont expliquées par rapport à la même variable calculée pour toute la Suisse. C'est pour ça que la Suisse entière apparaît comme la région 0.

Pour la publication, la méthode de calcul de la médiane prévoit un nombre minimal d'entreprises interrogées (5) et de salaires par entreprise (60). Les fig. 7, 8 et 9 ne utilisent pas les cas

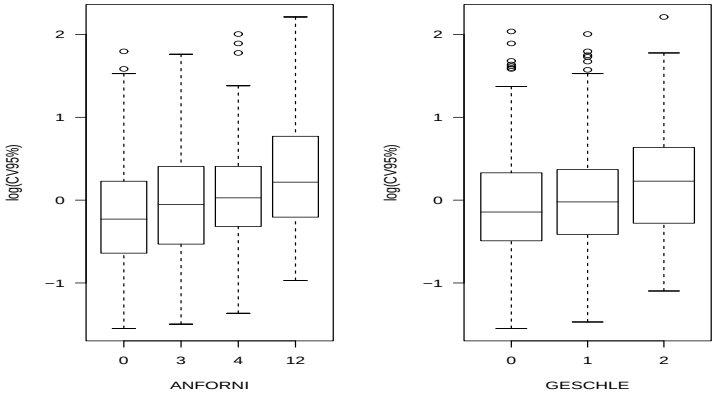


FIG. 6 – Boîte à moustaches pour $\log(CV_{syn95\%})$ pour chaque niveau ANFORNI (gauche) et chaque niveau GESCHLE (droite), variante II - Suisse

ne remplisant pas ces conditions. Encore deux cas où l'intervalle de confiance pour la médiane a une longueur égale à 0 (à cause du fait que la variance est 0) ont été éliminés : NOGA2=73, REGION=6, GESCHLE=1 et 2. Tous ces cas n'ont pas été pris en compte dans notre analyse.

3.2.1 Statistiques descriptives pour $\log(CV_{syn95\%})$ - les régions

Comme dans le premier cas, on utilise une transformation logarithmique de $CV_{syn95\%}$. La fig. 7 présente l'histogramme et la boîte à moustaches pour $\log(CV_{syn95\%})$. La distribution de cette variable est légèrement asymétrique, étalée à droite (coeff. asym. = 0.25) et il y a des valeurs extrêmes. En comparaison avec la fig. 1, les valeurs de la variable $\log(CV_{syn95\%})$ ne sont pas centrées autour de 0. La médiane est dans ce cas égale à 0.39. Le coefficient d'aplatissement de 0.11 montre une allure générale de la distribution assez aplatie.

On est intéressé d'analyser la variable $\log(CV_{syn95\%})$ en fonction des variables : branche économique (NOGA2), sexe (GESCHLE) et région (REGION). La fig. 8 présente notre variable mesurée seulement dans chaque classe NOGA2. On peut remarquer qu'il y a une asymétrie aux niveaux : 1, 20, 30, 67, 7074 (dans cette figure les classes NOGA2 sont dans l'ordre numérique). Il faut remarquer aussi que la classe NOGA2 = 1 au moment de l'analyse n'a pas été disponible pour les régions. Ce pour quoi elle est présente seulement dans les observations REGION=0. On observe aussi qu'il y a une variabilité de $\log(CV95\%)$ d'une classe NOGA2 à l'autre (voir dans la fig. 6 la longueur de chaque boîte à moustaches).

Pour les facteurs GESCHLE et REGION, notre variable suit une distribution assez symétrique (voir la fig. 9), mais il y a toujours des valeurs extrêmes. La région 7 (Tessin) présente des valeurs du $\log(CV_{syn95\%})$ plus élevées que toutes les autres régions et la Suisse entière.

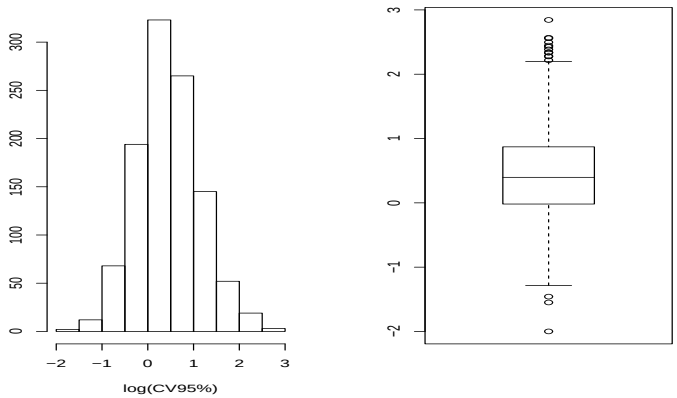


FIG. 7 – Histogramme et boîte à moustaches pour $\log(CV_{syn95\%})$ - les régions+Suisse

3.2.2 Modèle 2 - un modèle linéaire pour les régions

Un des modèles pouvant être retenu est précisé ci-dessous. Le modèle considéré est un modèle linéaire additif :

$$\log(CV_{syn95\%}) = \beta_0 + \beta_1(NOGA2) + \beta_2(GESCHLE) + \beta_3(REGION) + \epsilon. \qquad (4)$$

Ce modèle permet de prédire les valeurs moyennes de $\log(CV_{syn95\%})$ pour une branche économique, un sexe et une région fixés. L'observation extrême NOGA2=73, GESCHLE=0, REGION=6 a été éliminée. Comme dans le modèle 1, on a pris en compte un ensemble de poids calculé dans chaque classe NOGA2. Les coefficients du modèle sont donnés dans le tableau 3.

Statistiques générales du modèle 2 :

Ecart-type des résidus	=	0.5942 avec 1013 degrés de liberté ;
Coefficient de détermination multiple R^2	=	0.60 ;
Rapport F pour la régression	=	25.05 avec 58 et 1013 degrés de liberté ; et la p-valeur 2.2e-16.

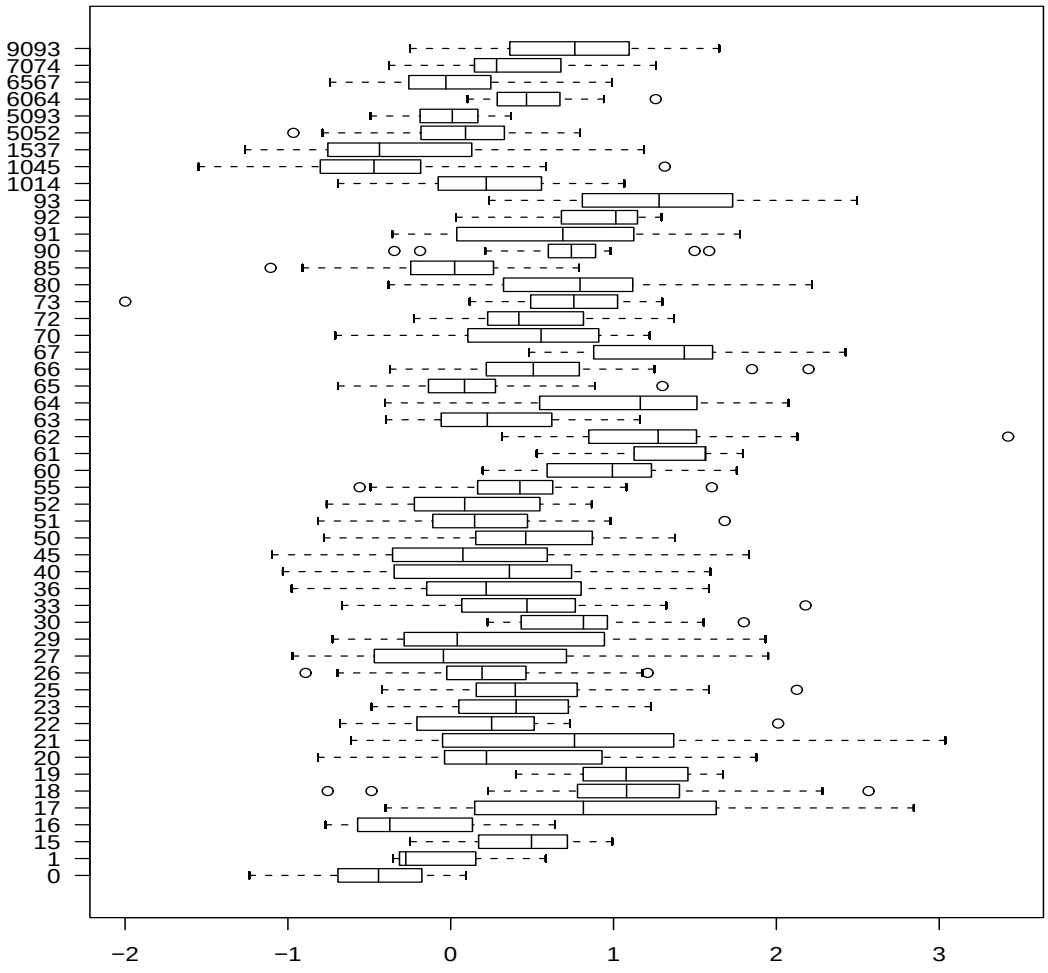


FIG. 8 – Boîte à moustaches pour $\log(CV_{syn95\%})$ pour chaque niveau NOGA2 - les régions+Suisse

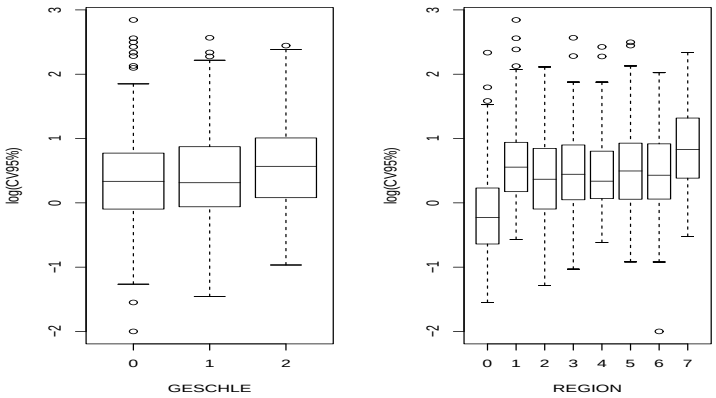


FIG. 9 – Boîte à moustaches pour $\log(CV_{syn95\%})$ pour chaque niveau GESCHLE (gauche) et chaque niveau REGION (droite) - les régions+Suisse

Pour la variable GESCHLE, le coefficient pour la modalité 2 est 0.134 et 0.077 pour la modalité 1, ce qui montre que les salaires médians des femmes sont moins bien estimés que les salaires médians des hommes, mais mieux par rapport au modèle 1. Pour les régions, le coefficient de la modalité 7 (Tessin) égal à 0.915 est plus grand que tous les autres coefficients de la même variable. La région 7 est moins bien estimée dans notre modèle. La région 2 a une meilleure estimation, avec le coefficient égal à 0.53, qui est le plus petit. Cela s'explique par le fait que deux petits cantons de la région 2 (Neuchâtel et Fribourg) ont bénéficié d'une enquête exhaustive.

Dans ce cas, le coefficient de détermination est $R^2 = 0.60$. L'analyse de variance montre que tous les facteurs affectent significativement la variable dépendante.

L'interprétation des coefficients peut être faite de la même manière que dans le modèle 1 (voir 3.1.3). La fig. 10 présente l'histogramme et la boîte à moustaches des résidus, le tracé QQ (Normal QQplot) et le graphique des valeurs estimées de la réponse versus les résidus standardisés. En comparaison avec le modèle 1, on observe que l'allure de la distribution des résidus est moins proche de la normale. Concernant le nuage de points des résidus standardisés croisés avec les valeurs prédites, on tire la même conclusion qu'avant : les points sont aléatoirement répartis entre les bornes - 2 et + 2 et il n'y a pas de formes "suspectes". Quelques points semblent hors norme, c'est-à-dire qu'il y a des valeurs aberrantes.

4 Discussion et conclusions

4.1 Discussion

Dans le tableau TA1 (voir p.6), Suisse et Grandes Régions (gr), le plus grand niveau de détail publié est la combinaison de 41 branches d'activité et des 7 grandes régions, croisée avec le sexe et le niveau de qualification (3 classes). Que peut-on dire globalement sur la relation entre la précision

TAB. 3 – Coefficients du modèle 2

Variable	Modalité	Coefficient estimé	Ecart-type estimé	Test de nullité	Significativité du test
Constante		-1.091	0.073	-15.043	< 2e-16
NOGA2	0	Référence	0	.	.
	1	1.005	0.265	3.794	0.000157
	15	0.896	0.086	10.409	< 2e-16
	16	0.853	0.364	2.343	0.019312
	17	1.418	0.172	8.241	5.21E-16
	18	1.548	0.159	9.758	< 2e-16
	19	1.623	0.114	14.275	< 2e-16
	20	0.927	0.145	6.387	2.57E-10
	21	1.308	0.187	7.012	4.27E-12
	22	0.659	0.114	5.774	1.03E-08
	23	0.848	0.103	8.268	4.22E-16
	25	0.947	0.118	7.996	3.46E-15
	26	0.715	0.116	6.182	9.16E-10
	27	0.537	0.141	3.806	0.00015
	29	0.685	0.140	4.898	1.12E-06
	30	1.224	0.096	12.775	< 2e-16
	33	0.887	0.126	7.026	3.88E-12
	36	0.760	0.132	5.754	1.15E-08
	40	0.672	0.140	4.788	1.94E-06
	45	0.630	0.146	4.319	1.72E-05
	50	0.929	0.112	8.274	4.00E-16
	51	0.646	0.116	5.566	3.32E-08
	52	0.569	0.104	5.490	5.08E-08
	55	0.870	0.102	8.540	< 2e-16
	60	1.421	0.099	14.389	< 2e-16
	61	1.915	0.202	9.469	< 2e-16
	62	1.935	0.199	9.701	< 2e-16
	63	0.734	0.102	7.175	1.38E-12
	64	1.540	0.145	10.644	< 2e-16
	65	0.606	0.098	6.197	8.34E-10
	66	1.044	0.124	8.413	< 2e-16
	67	1.765	0.115	15.311	< 2e-16
	70	0.942	0.109	8.672	< 2e-16
	72	0.924	0.095	9.750	< 2e-16
	73	1.300	0.095	13.716	< 2e-16
	80	1.185	0.121	9.799	< 2e-16
	85	0.424	0.102	4.144	3.69E-05
	90	1.193	0.110	10.855	< 2e-16
	91	1.100	0.127	8.691	< 2e-16
	92	1.347	0.090	15.025	< 2e-16
	93	1.747	0.133	13.146	< 2e-16
	1014	0.742	0.123	6.007	2.62E-09
	1045	-0.008	0.129	-0.059	0.953158
	1537	0.122	0.127	0.964	0.335059
	5052	0.454	0.097	4.699	2.98E-06
	5093	0.426	0.075	5.670	1.86E-08
	6064	0.957	0.078	12.288	< 2e-16
	6567	0.438	0.089	4.916	1.03E-06
	7074	0.826	0.094	8.751	< 2e-16
	9093	1.184	0.107	11.073	< 2e-16
GESCHLE	0	Référence	0	.	.
	1	0.077	0.030	2.571	0.010267
	2	0.134	0.030	4.410	1.15E-05
REGION	0	Référence	0	.	.
	1	0.684	0.049	14.086	< 2e-16
	2	0.530	0.048	11.017	< 2e-16
	3	0.616	0.049	12.593	< 2e-16
	4	0.584	0.049	11.980	< 2e-16
	5	0.614	0.049	12.640	< 2e-16
	6	0.627	0.050	12.660	< 2e-16
	7	0.915	0.050	18.272	< 2e-16

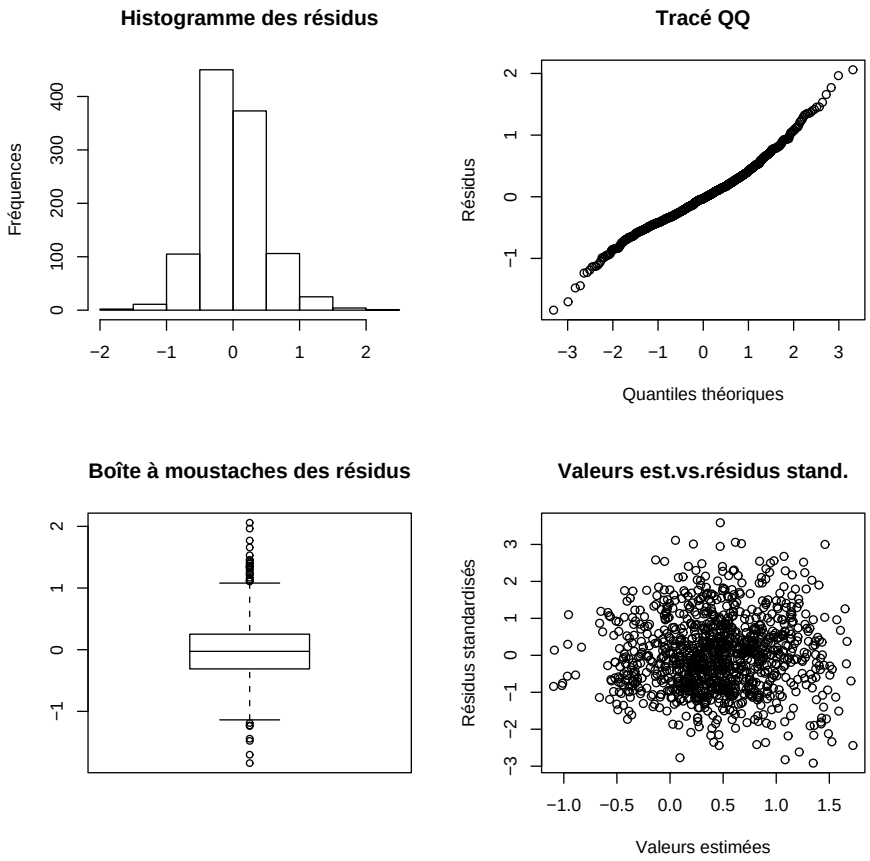


FIG. 10 – Résidus pour le modèle 2

de l'enquête sur la structure des salaires et le niveau de détail ?

4.1.1 Précision actuelle

En l'absence de standards reconnus pour la LSE, nous définissons notre critère de précision en fonction des résultats atteints, et en supposant que la taille actuelle de l'enquête (environ 55'000 entreprises interrogées) est la taille maximale réalisable. Un critère réaliste est alors de demander que, pour un niveau de détail donné,

$$\text{le 75e percentile des } CV_{syn95\%} \text{ soit } < 5\%.$$

Toute l'analyse présentée dans ce rapport est basée sur $\log(CV_{syn95\%})$. Nous reformulons notre critère dans cette échelle : un niveau de détail est acceptable, si le 75e percentile des $\log(CV_{syn95\%})$ est inférieur à $\log_e(5) = 1.61$. En prenant en compte nos modèles, on remarque :

1) La fig.2 ne prend en compte que la branche d'activité. On considère pour chaque branche

la médiane du salaire. On voit qu'au niveau suisse deux branches : 61, et 67 en partie, sont telles que le 75e percentile sorte de la limite de $\log_e(5) = 1.61$. Les regroupements dont ces branches sont des constituants (respectivement 60-64 et 65-67) sont eux bien mieux estimés. Au niveau des regroupements, aucun $CV_{syn95\%}$ ne dépasse la valeur de 5%. Les $\log(CV_{syn95\%})$ des branches 18 et 19 sortent parfois de la limite, mais très rarement.

2) Lorsque l'on prend séparément les classifications sexe et niveau de qualification, le critère est largement rempli (voir fig. 3).

3) La fig. 8 montre la variabilité des $\log(CV_{syn95\%})$ lorsque l'on considère pour chaque branche la médiane du salaire par Grandes Régions. Pratiquement toutes les branches qui sont des classes NOGA2 sans regroupement ont une fois ou l'autre un $CV_{syn95\%}$ qui dépasse la valeur de 5%. On voit que, malgré la taille importante de l'échantillon, plusieurs percentiles 75 dépassent aussi la valeur limite de $\log_e(5) = 1.61$. Il s'agit des branches 17, 61, 67 et 93.

4) La même constatation peut se faire à l'aide du tableau 3. Dans ce tableau, pour garantir que la limite de 5% n'est pas dépassée, il faudrait que

$$\log(CV_{syn95\%}) = -1.09 + \beta_1(NOGA2) + \beta_2(GESCHLE) + \beta_3(REGION) < 1.61, \quad (5)$$

donc

$$\begin{aligned} \beta_1(NOGA2) &< 1.61 + 1.09 - \beta_2(GESCHLE) - \beta_3(REGION) \\ &= 2.70 - \beta_2(GESCHLE) - \beta_3(REGION). \end{aligned}$$

Pour garantir la précision, il faut prendre le maximum des effets région et sexe (0.92 pour région 7 et 0.13 pour les femmes) :

$$\beta_1(NOGA2) < 2.70 - 0.92 - 0.13 = 1.65. \quad (6)$$

Pour tenir compte de la variabilité dans l'estimation de $\beta_1(NOGA2)$, on peut encore soustraire à la valeur limite l'écart type multiplié par le 75e percentile de la loi normale :

$$\beta_1(NOGA2) < 1.65 - 0.67 s(\beta_1(NOGA2)), \quad (7)$$

où $s(\beta_1(NOGA2))$ est l'écart-type estimé de $\beta_1(NOGA2)$. Par cette approche, les branches 19, 61, 62, 67 et 93 ne remplissent pas les critères. Ainsi, même avec une taille d'échantillon d'environ 55000 entreprises, il n'est pas possible d'estimer correctement toutes les combinaisons Grandes Régions $\times$ sexe $\times$ branche pour ces branches-là. A fortiori cela n'est pas garanti pour un canton.

Remarquons que ces branches appartiennent à des regroupements (15-37, 60-64, 65-67 et 90-93) qui sont d'une précision acceptable à très bonne. Les regroupements de branche demandés par Eurostat (voir fig.29) sont tous très bien estimés.

4.1.2 Effet sur la précision d'une diminution de la taille de l'échantillon

1) Avec une taille d'échantillon si grande, il est légitime de se demander quel serait l'effet sur la précision d'une diminution. Voyons l'effet d'une diminution d'environ 10000 entreprises,

ce qui correspond à un échantillon à 80% environ de la taille 2002. L'étude de la précision par regroupements Eurostat et par grandes régions montre que l'on rencontre la situation la plus critique pour la branche 90-93 et la grande région 6. L'effet sur le $CV_{syn95\%}$ d'une diminution nominale à 80% de la taille de l'échantillon 2002 (cela correspond à une diminution réelle à 81%, à cause des petites strates (<10 entreprises) qui sont exhaustives) est au pire une perte sur la précision de 7% par région et de 10% par branche Eurostat (voir les figures en Annxe 4). Ainsi, pour pouvoir garantir un $CV_{syn95\%} < 5\%$ dans 75% des cas au moins, il faut ajouter à la valeur critique (7) $\log_e(1-0.07) + \log_e(1-0.10)$, pour tenir compte de la perte sur la région et le regroupement Eurostat . On obtient donc :

$$\begin{aligned}\beta_1(NOGA2) &< 1.65 - 0.67s(\beta_1(NOGA2)) + \log_e(0.93) + \log_e(0.90) \\ &= 1.47 - 0.67s(\beta_1(NOGA2)).\end{aligned}$$

Pour la branche 90-93, on obtient :

$$\beta_1(9093) = 1.18 < 1.47 - 0.67s(\beta_1(9093)) = 1.40. \quad (8)$$

La perte de précision sur l'estimation par sexe est $< 10\%$, donc la médiane du salaire brut standardisé de la branche 90-93 (par gr $\times$ sexe) pourrait être estimée avec une précision suffisante, si on diminuait la taille nominale de l'échantillon de 20%, ce qui correspond à une diminution effective de 19%.

2) Le tableau 2 (voir 3.1.4) donne les coefficients du modèle de prévision pour les combinaisons branche x niveau de qualification (3 classes) x sexe. On voit que :
   a. si l'on additionne la constante à la valeur de branche, on obtient des valeurs très proches de celles du tableau 3,
   b. les valeurs pour le sexe ne sont pas significativement différentes entre les deux tableaux.

   Nous supposons donc un modèle additif pour le $\log(CV_{syn95\%})$ des combinaisons publiées au niveau de détail le plus grand, c'est-à-dire région x branche x niveau de qualification x sexe. L'effet du niveau de qualification (ANFORNI) est au pire de 0.33 pour la catégorie 1+2 (tableau 2). Augmentons aussi cet effet de $\log_e(1/0.9)$ pour tenir compte de la diminution de taille. On arrive à :

$$\begin{aligned}\beta_1(NOGA2) < 2.7 &- \beta_2(GESCHLE) - \beta_3(REGION) - \beta_4(ANFORNI) \\ &+3\log_e(0.9) - 0.67s(\beta_1(NOGA2)).\end{aligned}$$

   Alors

$$\beta_1(NOGA2) < 0.92 - 0.67s(\beta_1(NOGA2)).$$

   Pour la branche 90-93, il faudrait que :

$$\beta_1(9093) < 0.92 - 0.67s(\beta_1(90-93)) = 0.85.$$

3) La valeur du tableau 3 est de 1.18, donc supérieure à la valeur critique de 0.85 pour ce niveau de détail. Cela implique que plus de la moitié des $CV_{syn95\%}$ du salaire brut standardisé médian des combinaisons les plus détaillées impliquant la branche 90-93 dépassera la valeur de 5%. On fait la même constatation pour le regroupement 60-64. Pour les autres regroupements Eurostat, la valeur critique n'est pas dépassée.

4.2 Conclusions

Selon notre critère ($3/4$ des $CV_{syn95\%} < 5\%$), l'enquête 2002 est conforme :

- pour les combinaisons {Suisse x NOGA2 x sexe x niveau de qualification} : critère rempli pour 95% des branches ;
- pour les combinaisons {Grandes Régions x NOGA2 x sexe} : critère rempli pour 85% des branches ;
- au niveau des regroupements de branches, tels qu'ils sont demandés par Eurostat (voir fig. 29), les résultats détaillés sont conformes au critère.

D'après nos modèles, il n'y a que 70% des branches d'activité pour lesquelles les combinaisons les plus détaillées {Grandes Régions x NOGA2 x sexe x niveau de qualification} ont $3/4$ des $CV_{syn95\%} < 5\%$. Une diminution de la taille de 20% ferait descendre cette proportion à 51%. Nous voyons donc que si les spécifications de l'enquête sont conservées, il ne faudrait pas diminuer la taille.

Annexe 1
Délimitation du fichier SAS d'entrée des différents tableaux et programmes R

Le fichier de données utilisé est DATDYN.P02STD_neu.

– Tableau TA1 - prog_TA1.sas

```
th=1/initgew1;th=min(th,1);    * taux de réponse corrigé des entreprises;
thi=1/initgew2;thi=min(thi,1); * taux de réponse corrigé des salaires;
gewibgrs=gewicht*ibgrs;        * poids adapté au salaire brut standardisé;
if anforni ne .;               * choix LOHN pour la publication;
if geschle ne .;               * choix LOHN pour la publication;
if reform lt 20;               * suppression Bund;
keep   burnr gr grs nog_2 stragrs mbls anforni
       geschle gewicht  ibgrs  gewibgrs
       anzlohn  nrep  thi  th;
```

– Tableau TA1 par régions - prog_TA1_reg.sas

```
th=1/initgew1;th=min(th,1);    * taux de réponse corrigé des entreprises;
thi=1/initgew2;thi=min(thi,1); * taux de réponse corrigé des salaires;
gewibgrs=gewicht*ibgrs;        * poids adapté au salaire brut standardisé;
if anforni ne .;               * choix LOHN pour la publication ;
if geschle ne .;               * choix LOHN pour la publication ;
if reform lt 20;               * suppression Bund ;
keep   burnr gr grs nog_2 stragrs mbls anforni
       geschle gewicht  ibgrs  gewibgrs
       anzlohn  nrep  thi  th;
```

– Tableau TA2 - prog_TA2.sas

```
th=1/initgew1;th=min(th,1);    * taux de réponse corrigé des entreprises;
thi=1/initgew2;thi=min(thi,1); * taux de réponse corrigé des salaires;
gewibgrs=gewicht*ibgrs;        * poids adapté au salaire brut standardisé;
if anforni ne .;               * choix LOHN pour la publication ;
if geschle ne .;               * choix LOHN pour la publication ;
if reform in ( 20, 30) ;       *  Bund ;
keep   burnr gr grs nog_2 stragrs mbls anforni
       geschle gewicht  ibgrs  gewibgrs
       anzlohn  nrep  thi  th  initgew2  anzbe;
```

– Tableau TA4 - prog_TA4.sas

```
th=1/initgew1;th=min(th,1);    * taux de réponse corrigé des entreprises;
thi=1/initgew2;thi=min(thi,1); * taux de réponse corrigé des salaires;
gewibgrs=gewicht*ibgrs;        * poids adapté au salaire brut standardisé;
if geschle ne .;               * choix LOHN pour la publication;
if reform lt 20;               * suppression Bund ;
keep   burnr gr grs nog_2 stragrs mbls anforni
       geschle gewicht  ibgrs  gewibgrs  ta3
       anzlohn  nrep  thi  th reform;
```

– Tableau TA7 - prog_TA7.sas

```
th=1/initgew1;th=min(th,1);    * taux de réponse corrigé des entreprises;
thi=1/initgew2;thi=min(thi,1); * taux de réponse corrigé des salaires;
gewibgrs=gewicht*ibgrs;        * poids adapté au salaire brut standardisé;
if anforni ne .;               * choix LOHN pour la publication;
if geschle ne .;               * choix LOHN pour la publication;
if taetigk ne .;               * choix LOHN pour la publication;
PB=0;
if reform lt 20 then PB=1;        * privé ;
if reform in (20, 30) then PB=1;  * Bund  ;
if PB=1;
keep   burnr gr grs nog_2 stragrs mbls anforni
       geschle gewicht  ibgrs  gewibgrs  ta3
       anzlohn  nrep  thi  th reform taetigk;
```

– Tableau Eurostat - eurostat.sas

```
th=1/initgew1;th=min(th,1);    * taux de réponse corrigé des entreprises;
thi=1/initgew2;thi=min(thi,1); * taux de réponse corrigé des salaires;
if reform < 20 or reform in (20,30);
if untgroe >=10;
if ibgr >= 0.9;
keep   burnr gr grs nog_2 stragrs mbliu anforni
       geschle gewicht ibgr
       anzlohn  nrep  thi  th;
```

Programme R pour le modèle 1

Le fichier utilisé (nommé ci-dessous 'cvsuisse.txt') a 4 colonnes : $CV_{syn95\%}$, NOGA2, ANFORNI et GESCHLE.

```
> a = read.table ("h:/cvsuisse.txt")
> data = a[,1:4]
> data = cbind(log(data[,1]),data[,2:4])
> data[,2]=as.factor(data[,2])
> data[,3]=as.factor(data[,3])
> data[,4]=as.factor(data[,4])
> data=data[-c(132,482,532),]
> var_classe=numeric(nrow(data))
> for(i in 1:nlevels(data[,2]))
+ { + classe=subset(data,data[,2]==levels(data[,2])[i]) +
v=var(classe[,1]) + for(j in 1:nrow(data)) +
if(data[j,2]==levels(data[,2])[i]) var_classe[j]=v + }
> w=var(data[,1])/var_classe
> modele_1 = lm(data[,1] ~ data[,2]+data[,3]+data[,4],weight=w)
> summary(modele_1)

Call: lm(formula = data[, 1] ~ data[, 2] + data[, 3] + data[, 4],
weights = w)
```

Programme R pour le modèle 2

Le fichier utilisé (nommé ci-dessous 'cvregion.txt') a 4 colonnes : $CV_{syn95\%}$, NOGA2, GESCHLE et REGION.

```
> a = read.table ("h:/cvregion.txt")
> a = a[-c(32,132,180,225,459,506,551,924,966,1090),]
> data = a[,1:4]
> data = cbind(log(data[,1]),data[,2:4])
> data[,2]=as.factor(data[,2])
> data[,3]=as.factor(data[,3])
> data[,4]=as.factor(data[,4])
> var_classe=numeric(nrow(data))
> for(i in 1:nlevels(data[,2]))
+ { + classe=subset(data,data[,2]==levels(data[,2])[i]) +
v=var(classe[,1]) + for(j in 1:nrow(data)) +
if(data[j,2]==levels(data[,2])[i]) var_classe[j]=v + }
> w=var(data[,1])/var_classe
> modele_2 = lm(data[,1] ~ data[,2]+data[,3]+data[,4],weight=w)
> summary(modele_2)

Call: lm(formula = data[, 1] ~ data[, 2] + data[, 3] + data[, 4],
weights = w)
```

Annexe 2
Résultats pour la Suisse

NOGA2	n_e	n_s	*Moyenne*	CV_{moy}	*Médiane*	*BI*	*BS*	$CV_{syn95\%}$	CV_{syn}	CV_{perc}
0	42032	1031538	6157	0.41	5379	5344	5416	0.35	0.35	0.75
1	696	5016	4494	0.94	4302	4238	4334	0.76	0.72	2.23
10-45	13803	365830	6046	0.27	5523	5500	5543	0.21	0.24	0.64
10-14	165	2500	5933	0.68	5469	5413	5518	0.52	0.57	2.37
15-37	11269	311223	6155	0.34	5602	5571	5633	0.28	0.32	0.75
15	731	33707	5407	0.95	4938	4850	5039	1.04	1.01	2.48
16	11	2003	7974	0.67	7119	7054	7215	0.69	0.58	1.31
17	311	6557	4992	0.92	4542	4457	4643	1.13	1.10	2.64
18	202	2189	4216	1.84	3673	3570	3786	1.57	1.47	3.30
19	73	860	4845	4.03	4089	3936	4300	2.63	2.20	4.36
20	754	8030	5238	0.62	5069	5025	5108	0.44	0.39	2.05
21	159	7653	5900	0.75	5526	5423	5609	0.95	0.90	2.59
22	1441	23154	6509	0.57	6088	6034	6167	0.66	0.59	1.39
23-24	495	39132	7625	1.16	6965	6820	7110	1.06	1.13	2.68
25	457	9934	5572	0.77	5078	5007	5137	0.71	0.55	2.04
26	506	8470	5819	0.49	5349	5307	5392	0.41	0.39	1.61
27-28	1438	33063	5627	0.38	5258	5220	5297	0.38	0.42	1.38
29,34,35	1630	58122	6465	0.44	6016	5959	6069	0.48	0.52	1.35
30	861	27731	6704	1.13	6161	5987	6343	1.51	1.59	2.98
33	1209	39327	6182	0.65	5571	5482	5648	0.82	0.83	1.81
36-37	991	11291	5468	0.46	5100	5070	5147	0.47	0.51	1.33
40-41	182	7281	7485	0.56	6904	6844	6954	0.44	0.62	1.50
45	2187	44826	5665	0.36	5345	5311	5380	0.33	0.39	1.32
50-93	27533	660692	6241	0.63	5262	5200	5325	0.61	0.63	1.13
50-52	7011	209812	5517	0.68	4694	4642	4755	0.66	0.70	1.48
50	1323	16874	5313	0.49	4894	4850	4930	0.46	0.67	1.45
51	3236	68588	6888	0.63	5762	5714	5812	0.44	0.40	0.98
52	2452	124350	4674	0.54	4152	4117	4190	0.47	0.53	1.39
55	2415	35366	3968	0.47	3660	3619	3700	0.57	0.57	1.70
60-64	2246	66887	5964	1.03	5345	5223	5469	1.18	1.18	2.69
60	868	17807	5238	1.33	4864	4767	4980	1.22	1.15	3.37
61	44	1303	5193	11.59	5145	4103	5895	10.33	9.00	16.73
62	72	1113	6564	1.45	5877	5706	6146	2.34	2.01	3.26
63	1063	26334	5762	0.77	5192	5111	5268	0.80	0.67	1.96
64	199	20330	7566	1.18	6936	6745	7122	1.40	1.49	2.91
65-67	1988	89445	8816	1.06	7380	7273	7488	0.75	0.75	1.35
65	993	49118	9196	1.49	7642	7539	7770	0.85	0.80	1.55
66	428	34855	7908	1.40	6883	6684	7092	1.55	1.70	2.88
67	567	5472	9989	3.51	7367	6766	8071	4.88	4.79	6.55
70-74	6970	109514	7522	0.72	6397	6305	6496	0.79	0.81	1.26
70-71	1211	10938	6636	0.74	5907	5850	5958	0.49	0.61	1.57
72,74	5586	93232	7543	0.79	6400	6303	6500	0.80	0.86	1.38
73	173	5344	8655	1.70	7525	7250	7777	1.86	1.98	3.92
80	1242	23358	7149	0.65	6578	6497	6666	0.68	0.81	1.35
85	2485	92450	5765	0.32	5397	5363	5432	0.33	0.37	0.90
90-93	3176	33860	5716	0.64	5040	4973	5117	0.78	0.85	1.29
90	225	2258	5410	0.91	4978	4911	5047	0.71	0.70	2.49
91	1125	14174	7059	0.73	6355	6283	6442	0.70	0.68	1.62
92	1066	12475	6034	1.07	5333	5234	5441	1.03	1.05	2.04
93	760	4953	4039	1.50	3593	3473	3721	1.82	1.95	4.56

Légende (voir [1]) :

NOGA2 = les classes NOGA2 ;
n_e = nombre d'entreprises répondantes ; n_s = nombre de salaires fournies ;
Moyenne = salaire moyen ;
CV_{moy} = coefficient de variation (CV) moyen ;
Médiane = salaire médian ;
BI, BS = bornes inférieure et supérieure de l'intervalle de confiance à 95% pour la médiane ;
$CV_{syn95\%}$ = CV synthétique 95% ; CV_{syn} = CV synthétique ; CV_{perc} = CV percentile.

Annexe 3
Intervalles de confiance pour la médiane

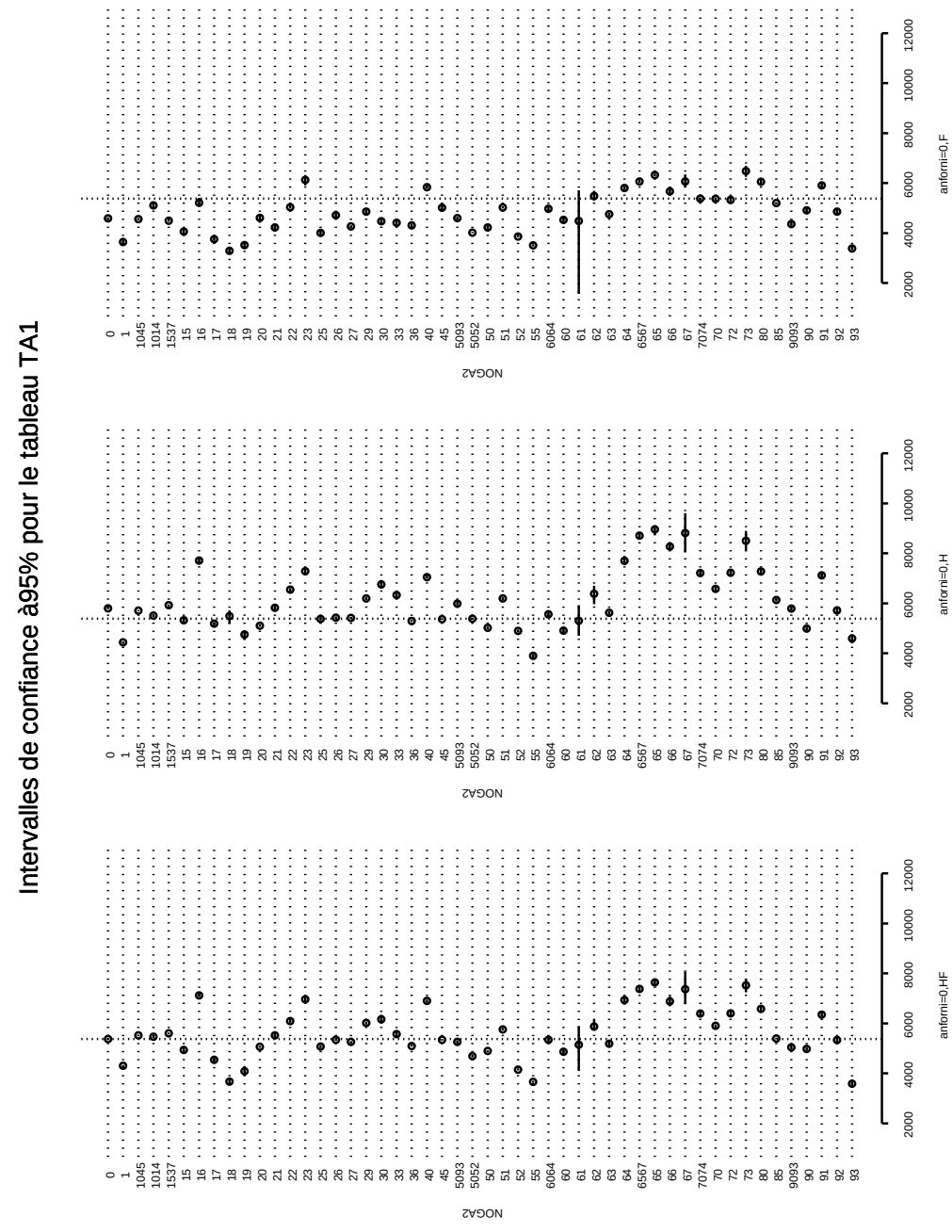


FIG. 11 – Tableau TA1 - niveau de qualification 0, par sexe

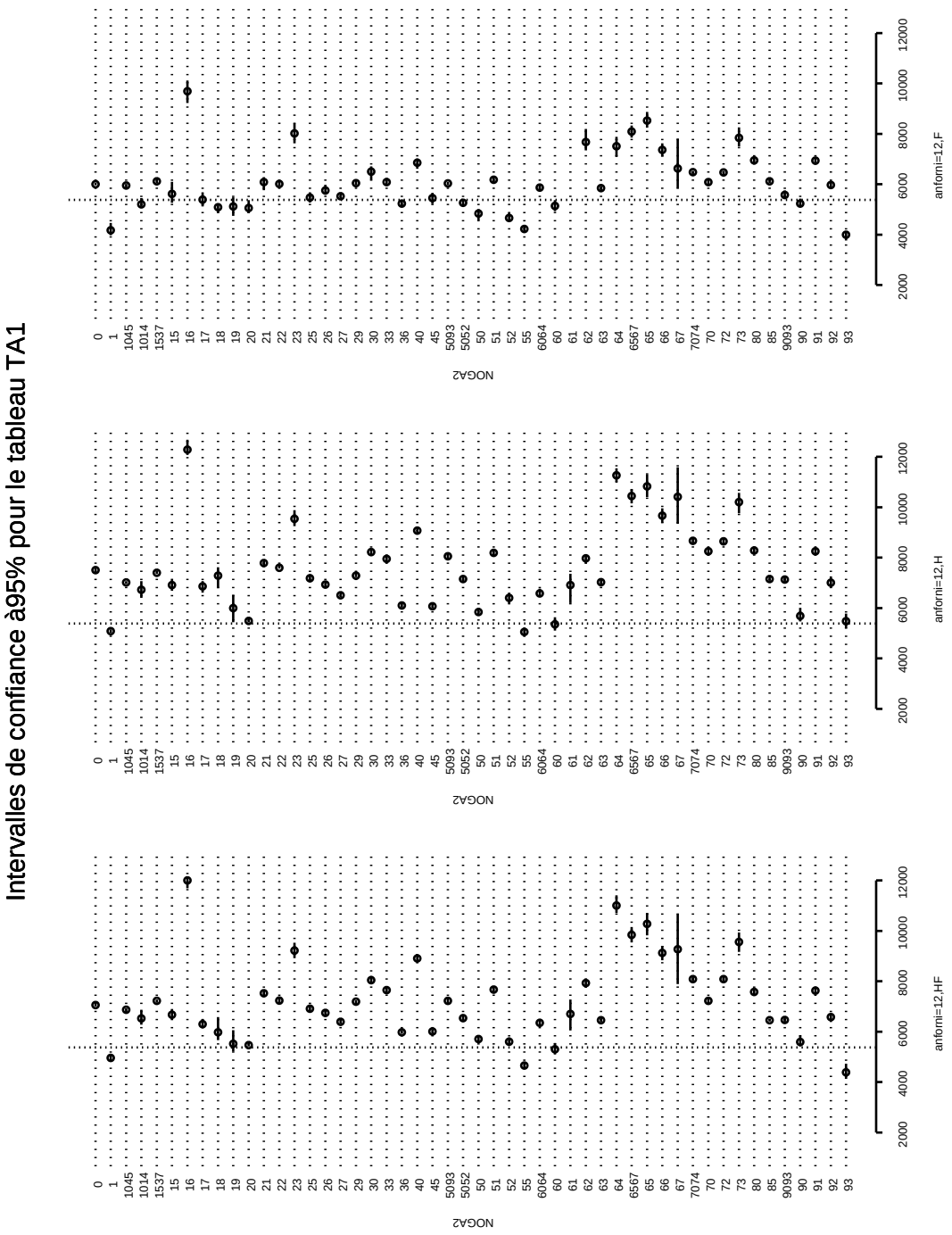


FIG. 12 – Tableau TA1 - niveau de qualification 1+2, par sexe

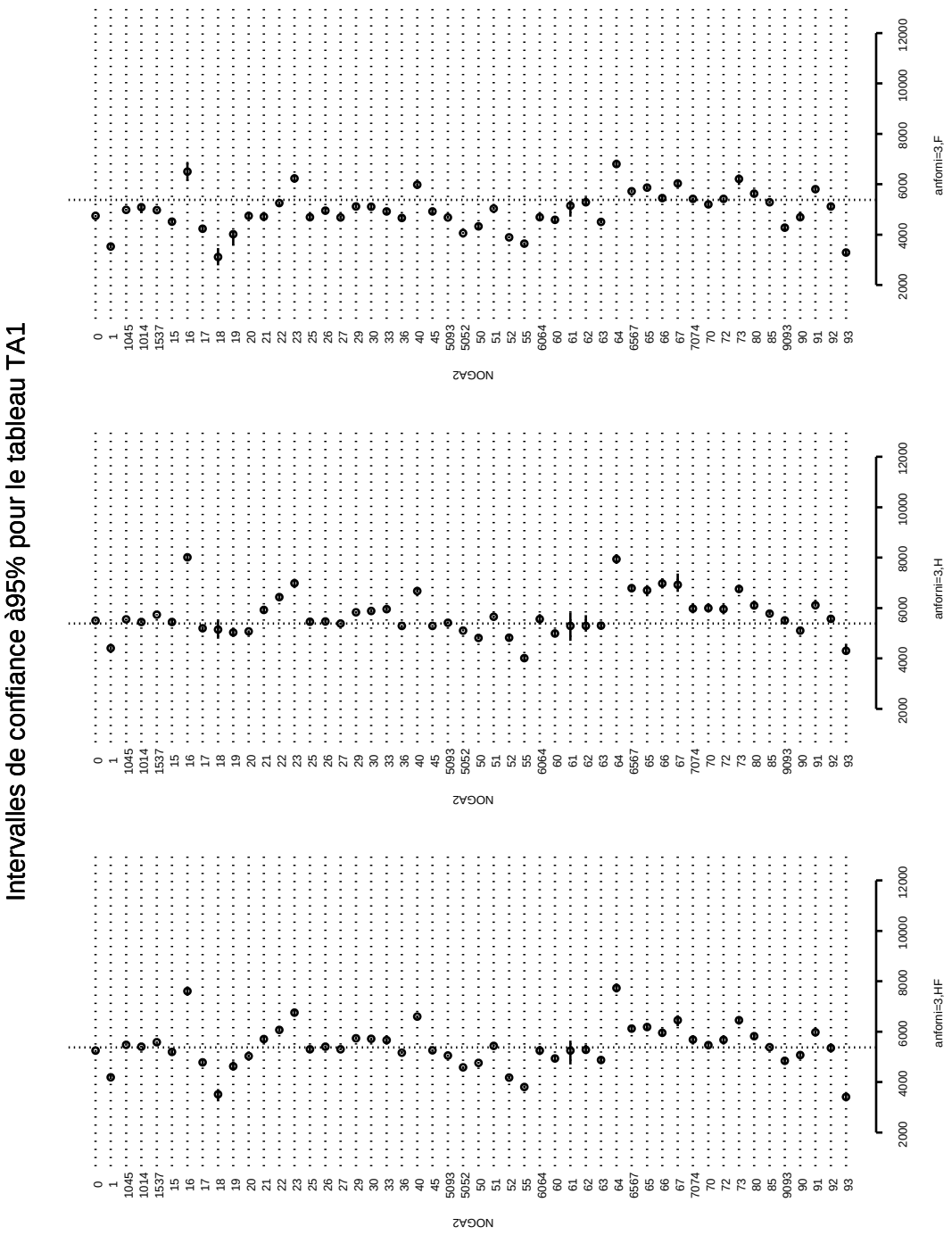


FIG. 13 – Tableau TA1 - niveau de qualification 3, par sexe

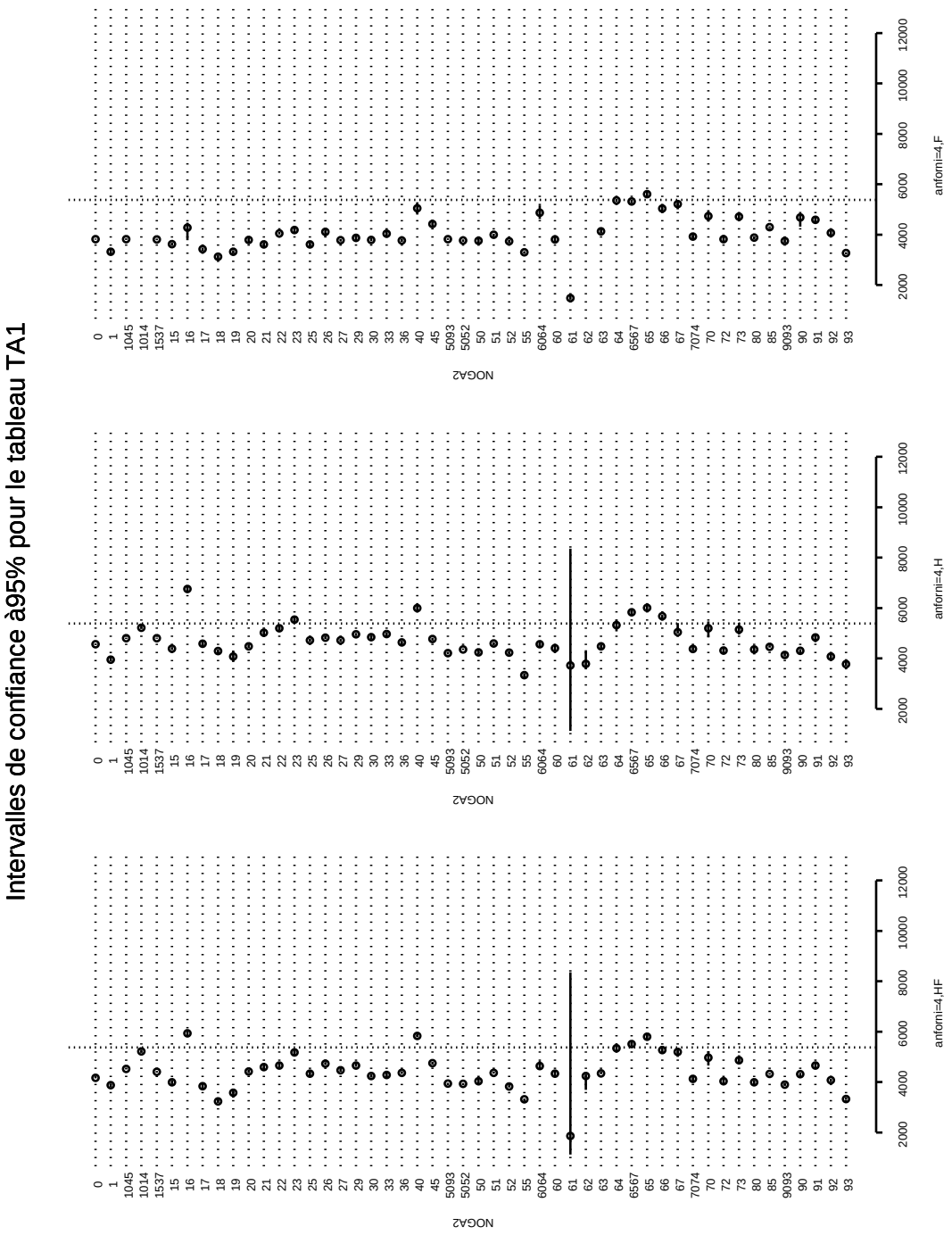


FIG. 14 – Tableau TA1 - niveau de qualification 4, par sexe

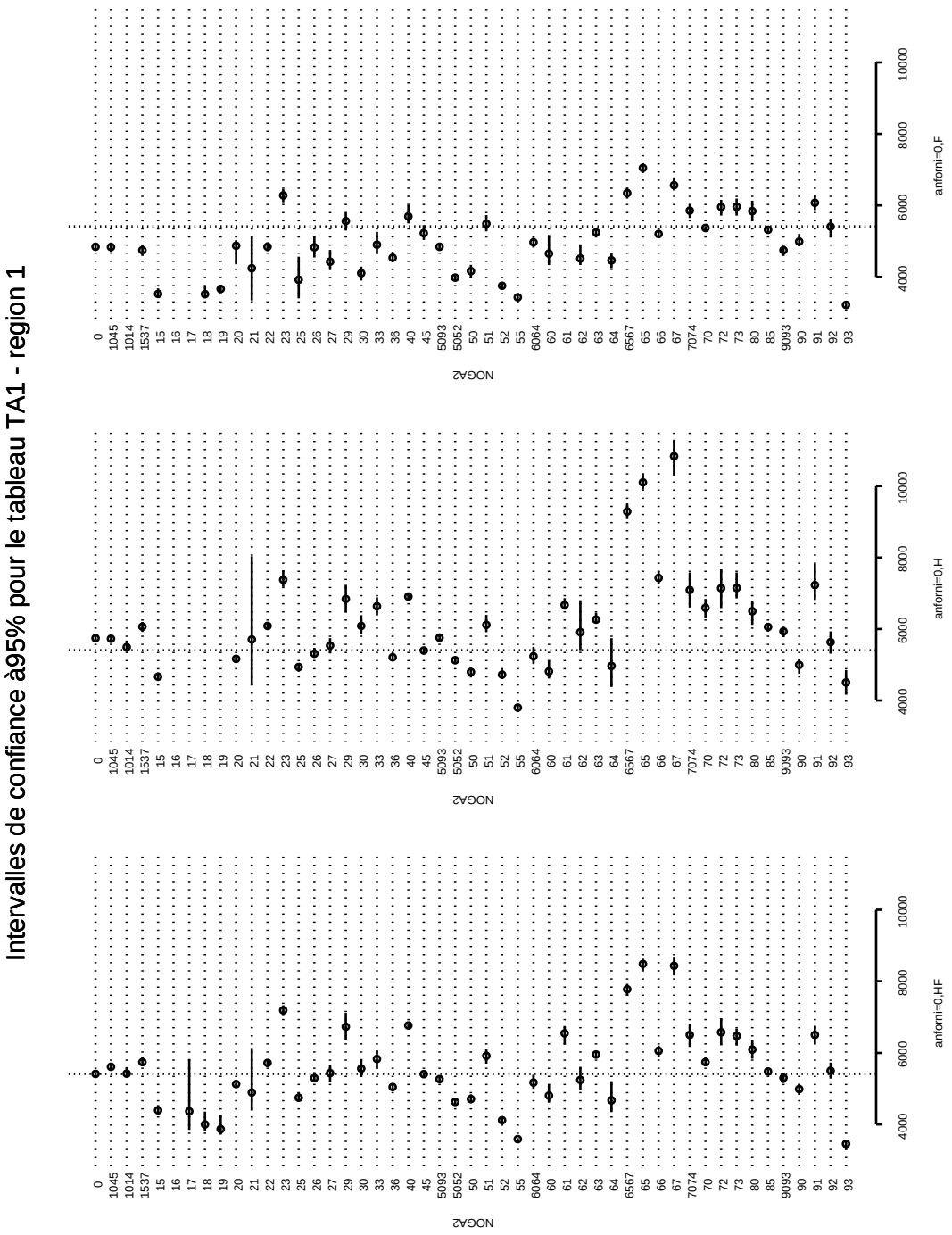


FIG. 15 – Tableau TA1 - région 1, niveau de qualification 0, par sexe

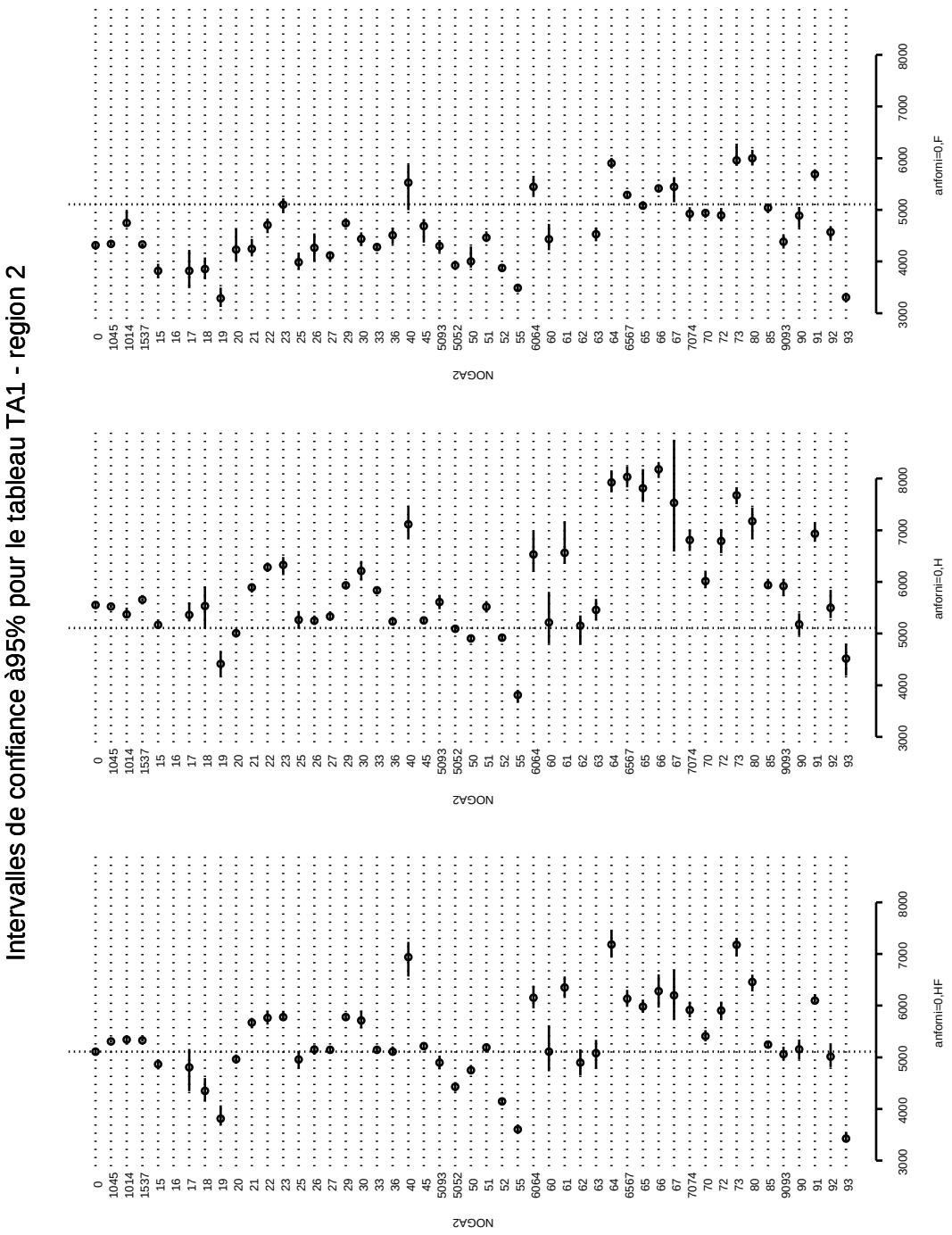


FIG. 16 – Tableau TA1 - région 2, niveau de qualification 0, par sexe

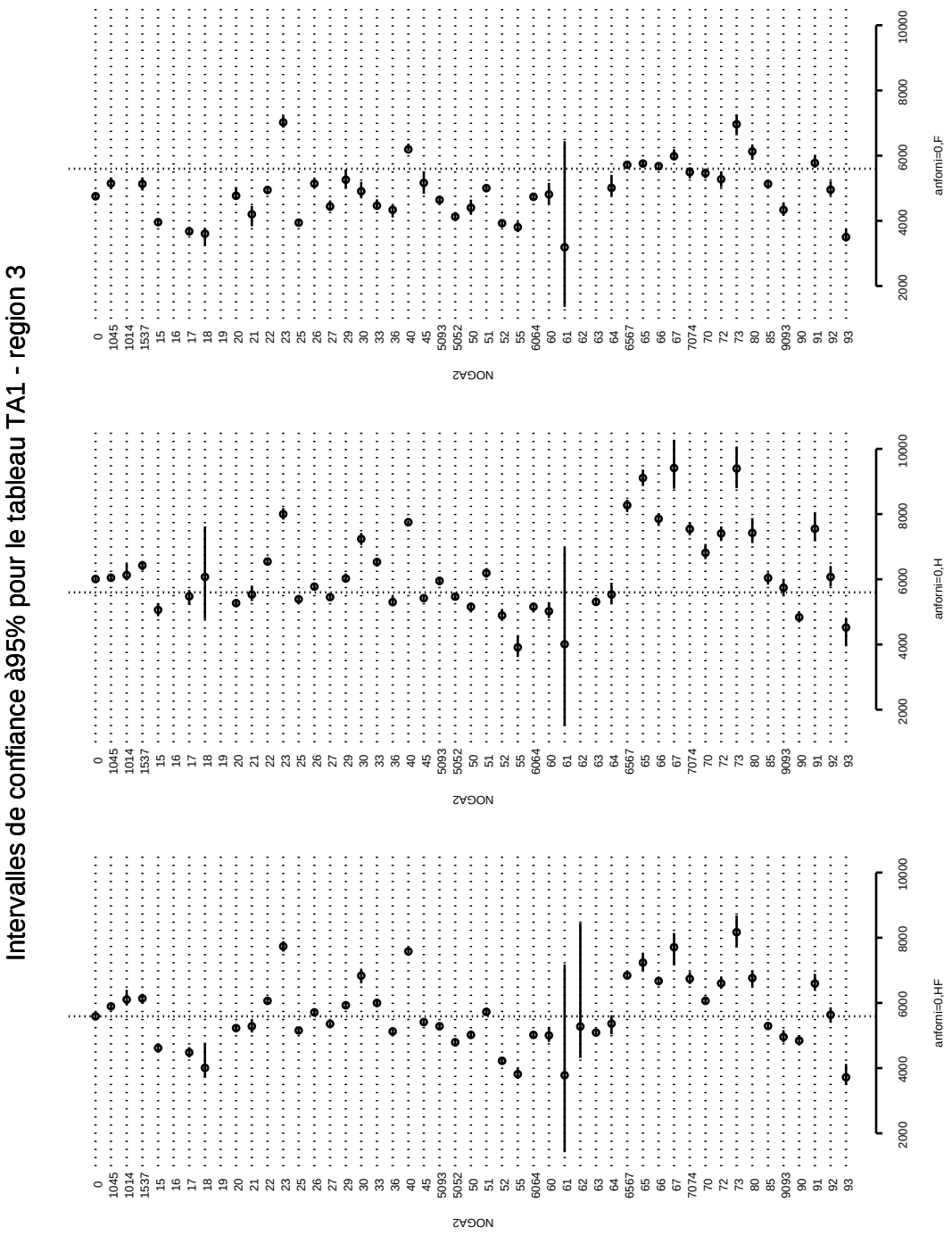


FIG. 17 – Tableau TA1 - région 3, niveau de qualification 0, par sexe

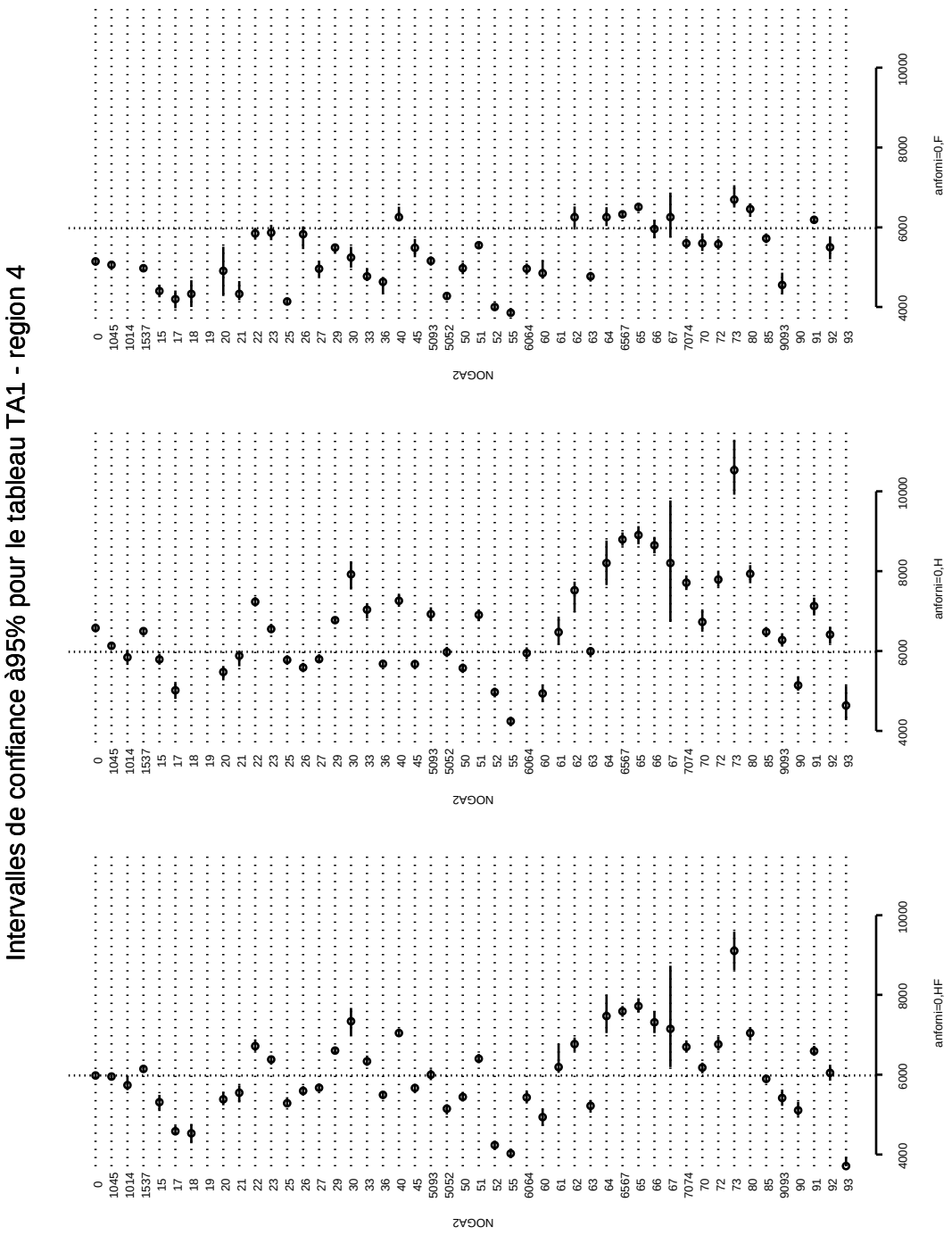


FIG. 18 – Tableau TA1 - région 4, niveau de qualification 0, par sexe

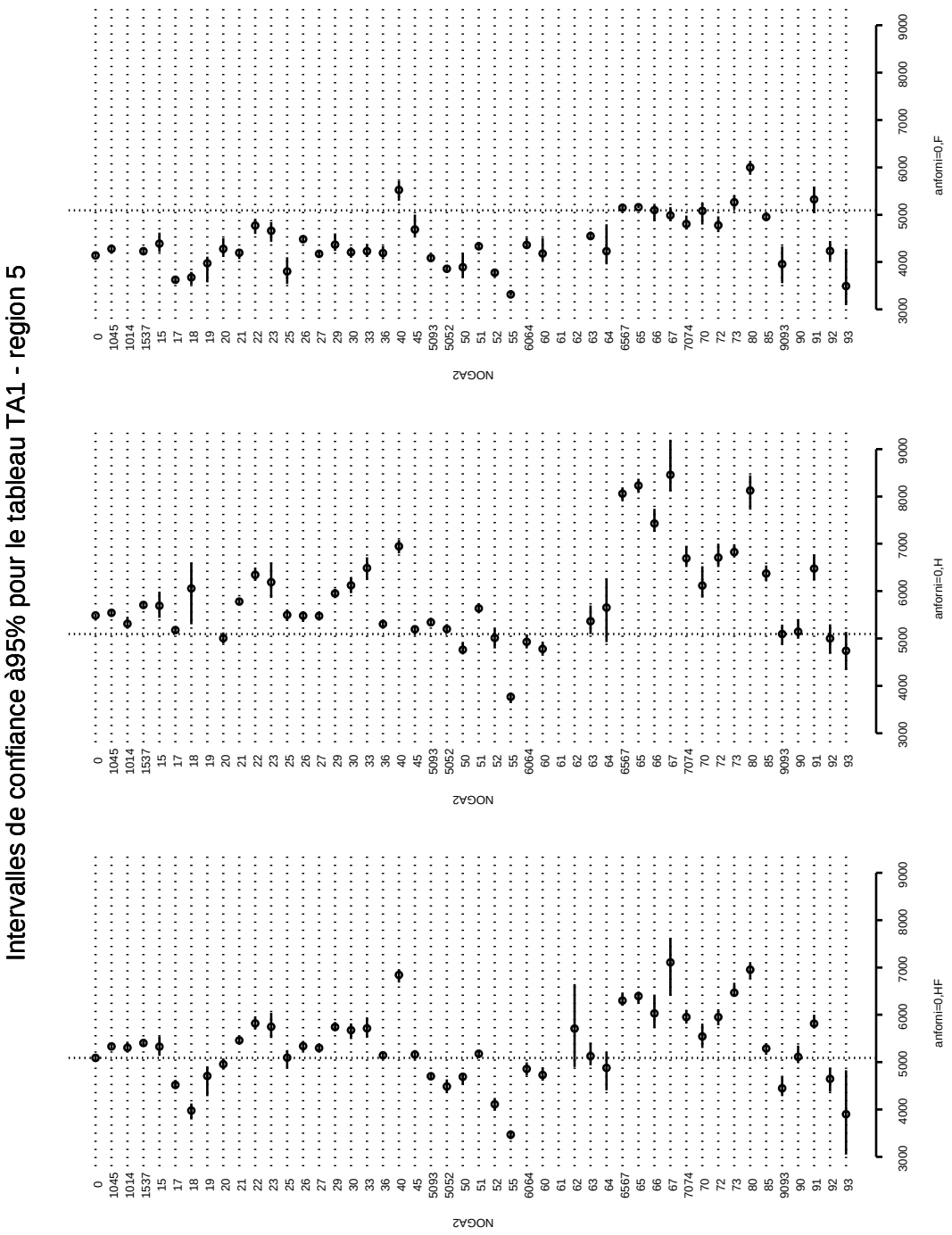


FIG. 19 – Tableau TA1 - région 5, niveau de qualification 0, par sexe

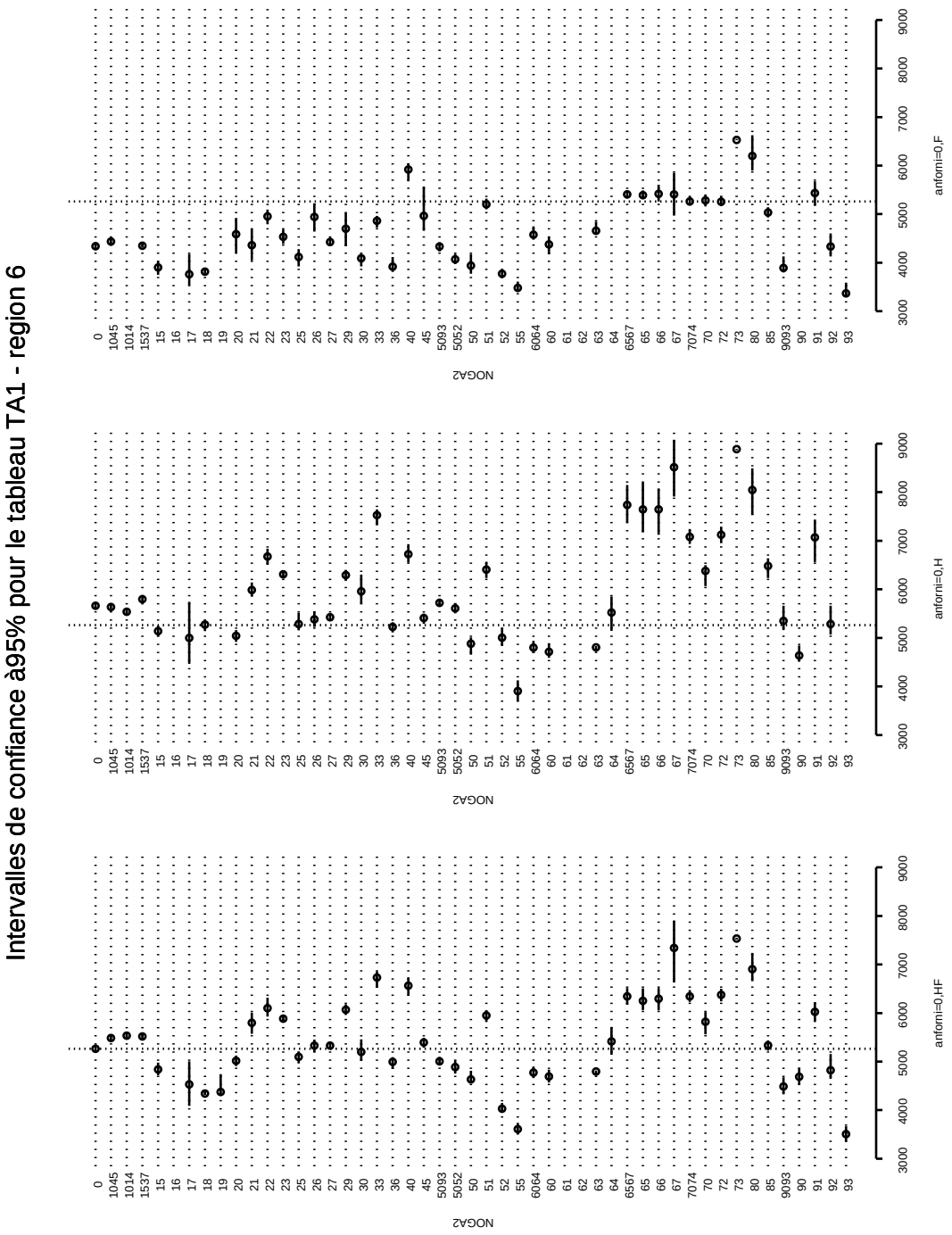


FIG. 20 – Tableau TA1 - région 6, niveau de qualification 0, par sexe

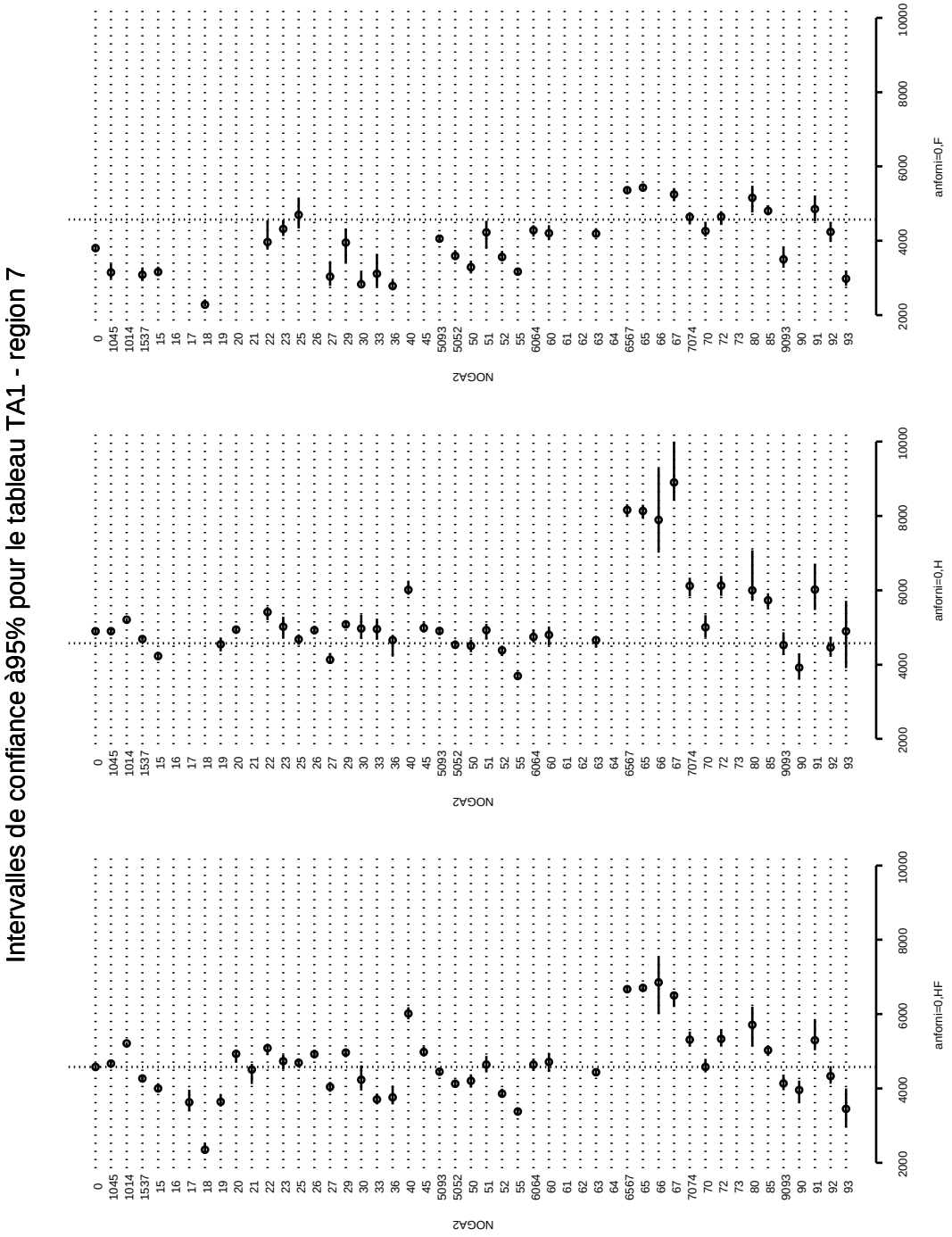


FIG. 21 – Tableau TA1 - région 7, niveau de qualification 0, par sexe

FIG. 22 – Tableau TA2

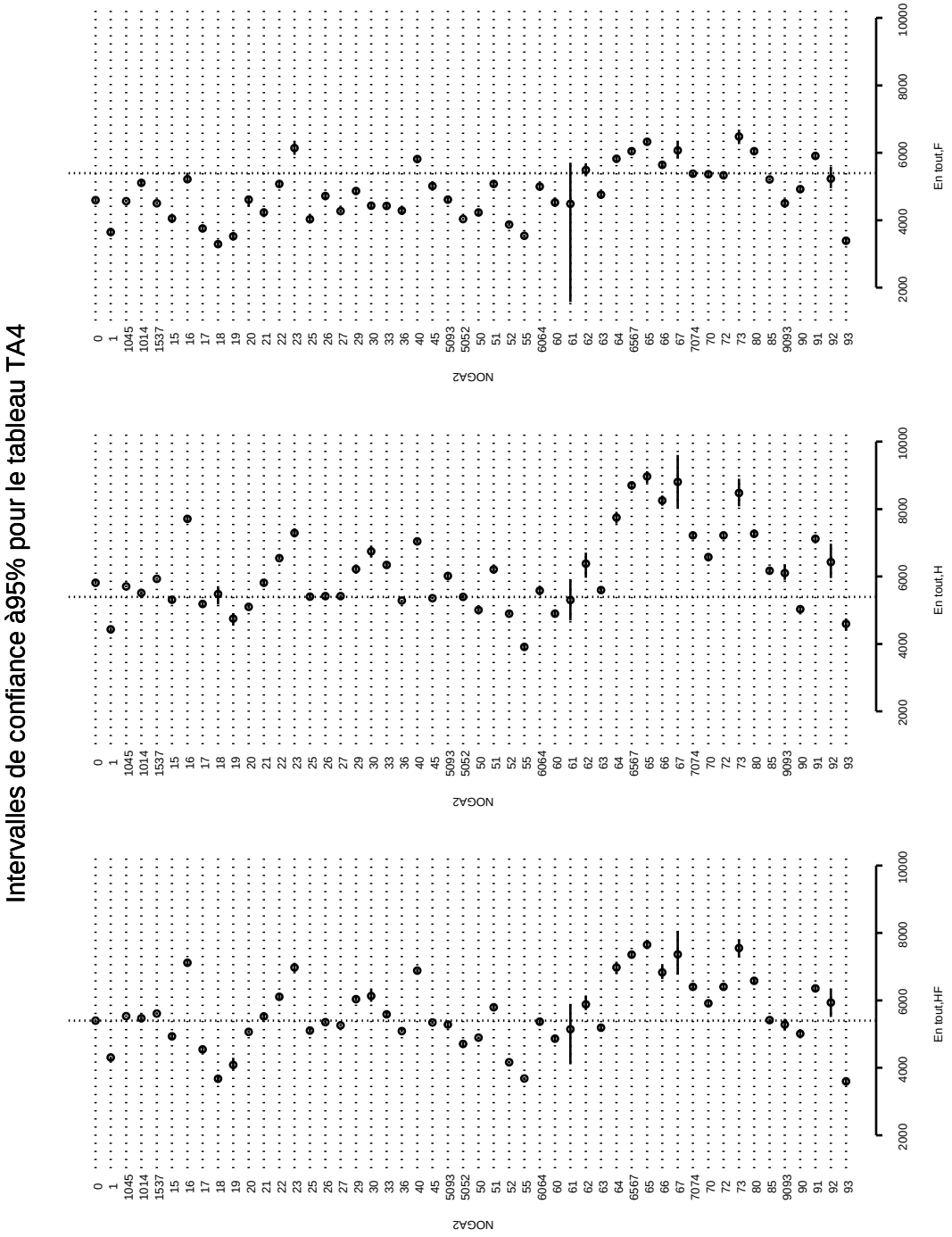


FIG. 23 – Tableau TA4 - en tout, par sexe

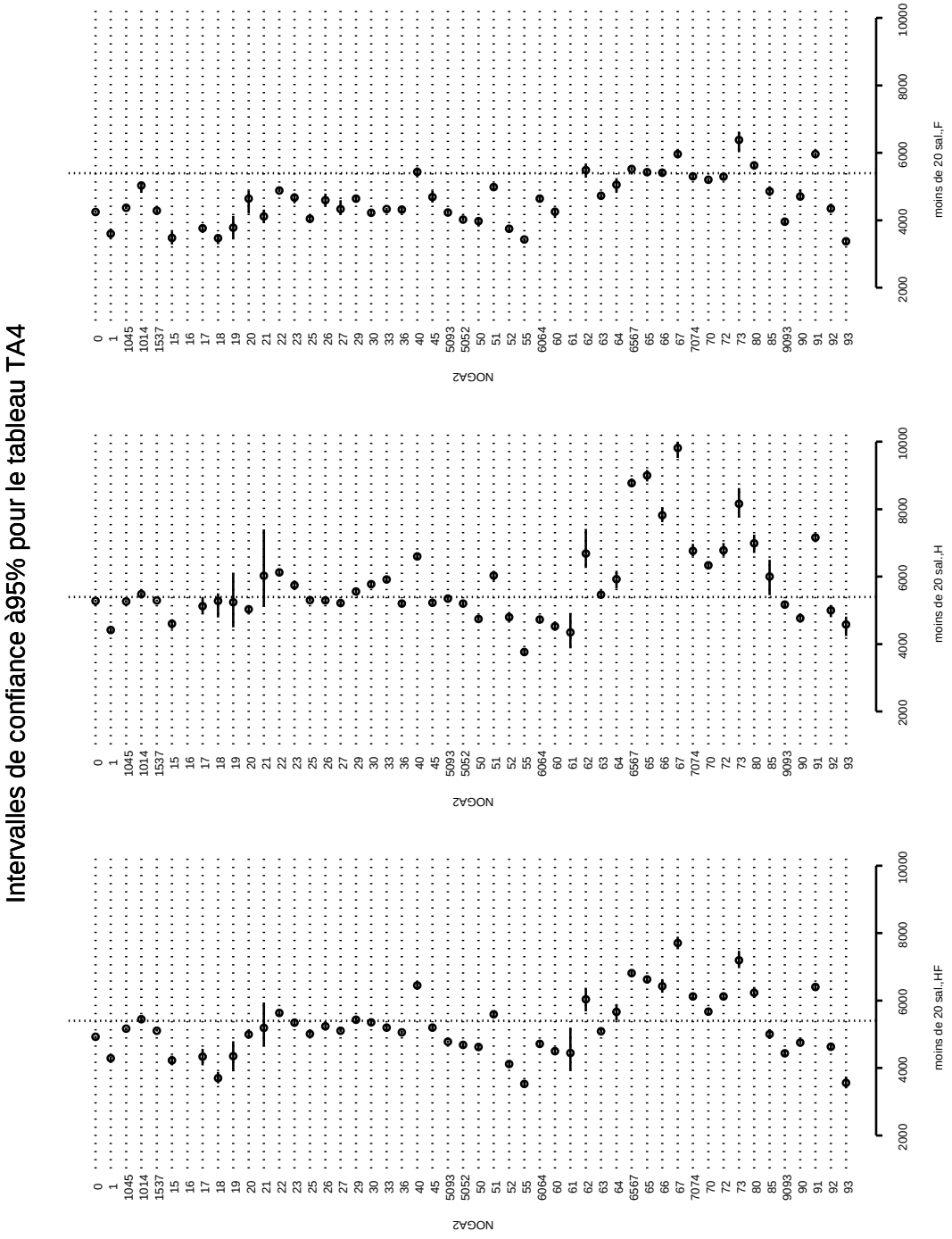


FIG. 24 – Tableau TA4 - moins de 20 salariés, par sexe

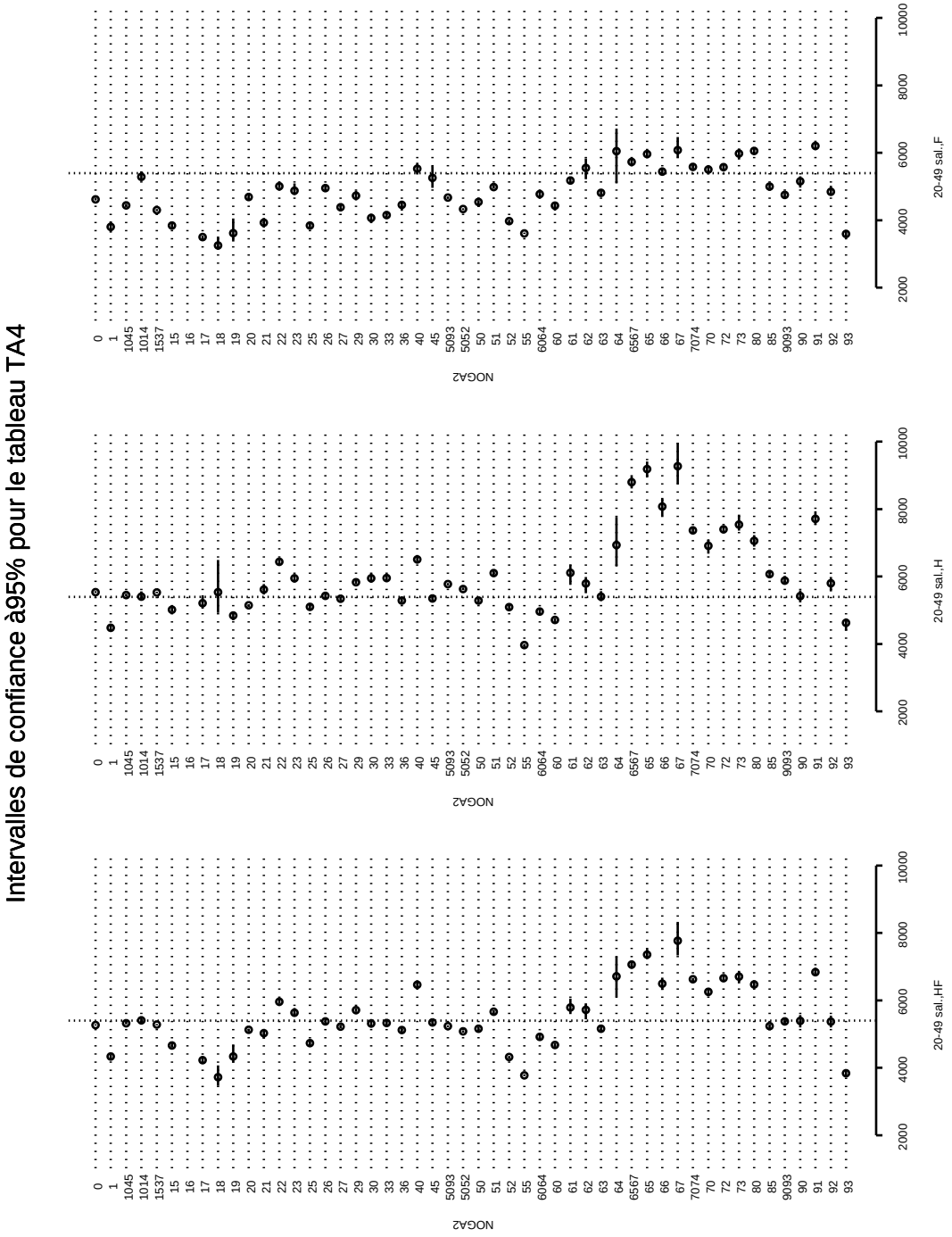


FIG. 25 – Tableau TA4 - entre 20-49 salariés, par sexe

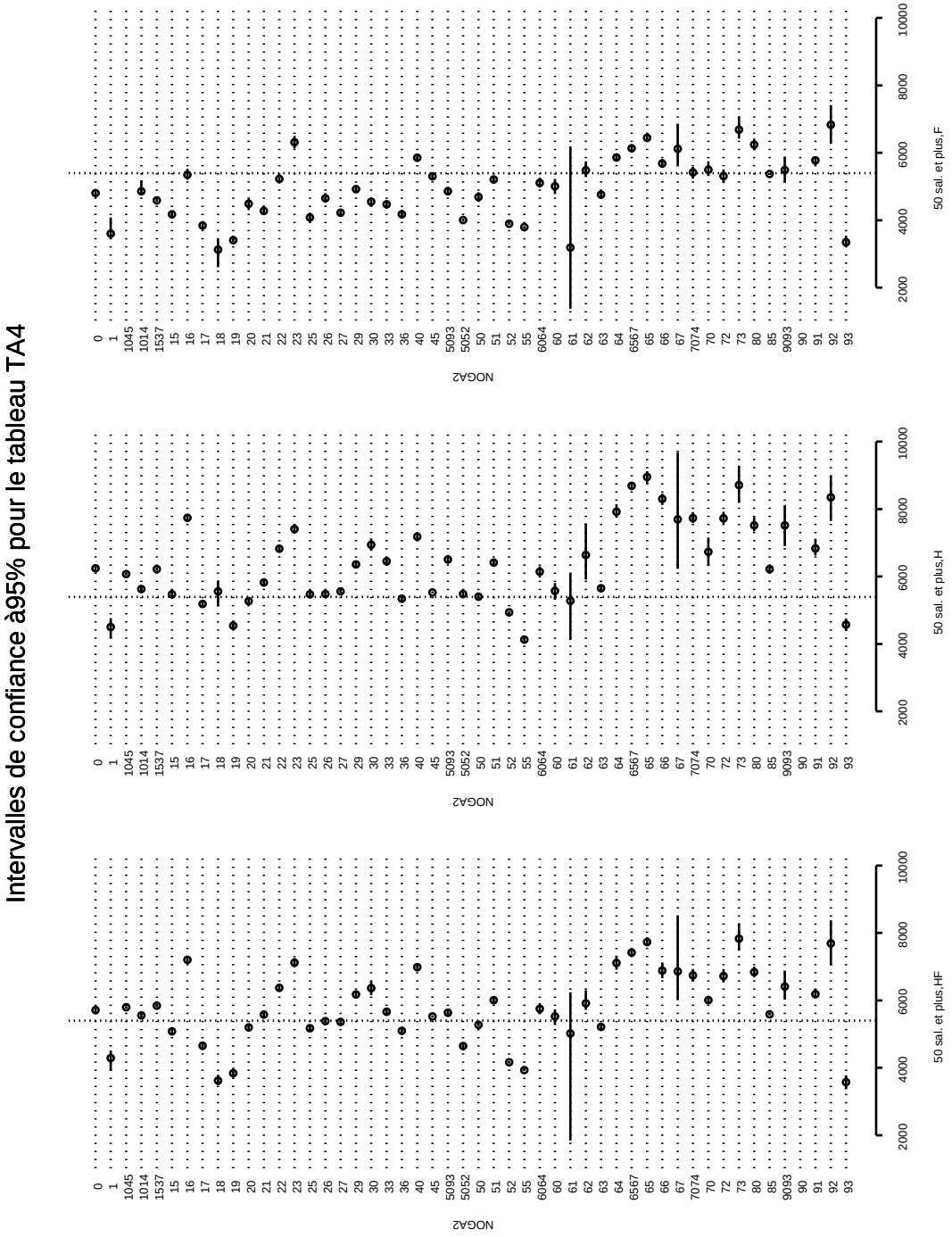


FIG. 26 – Tableau TA4 - 50 salariés et plus, par sexe

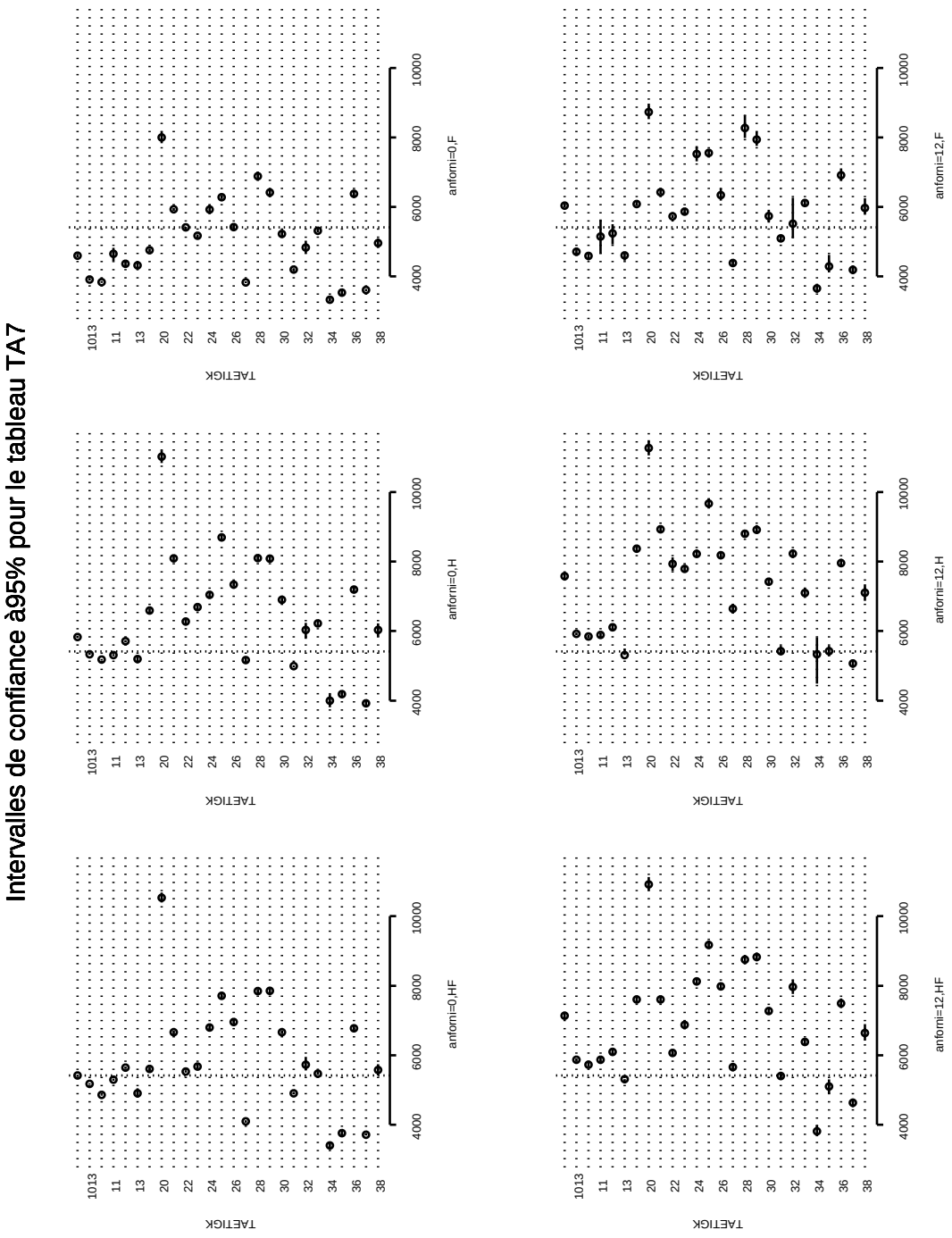


FIG. 27 – Tableau TA7 - niveaux de qualification 0 et 1+2, par sexe

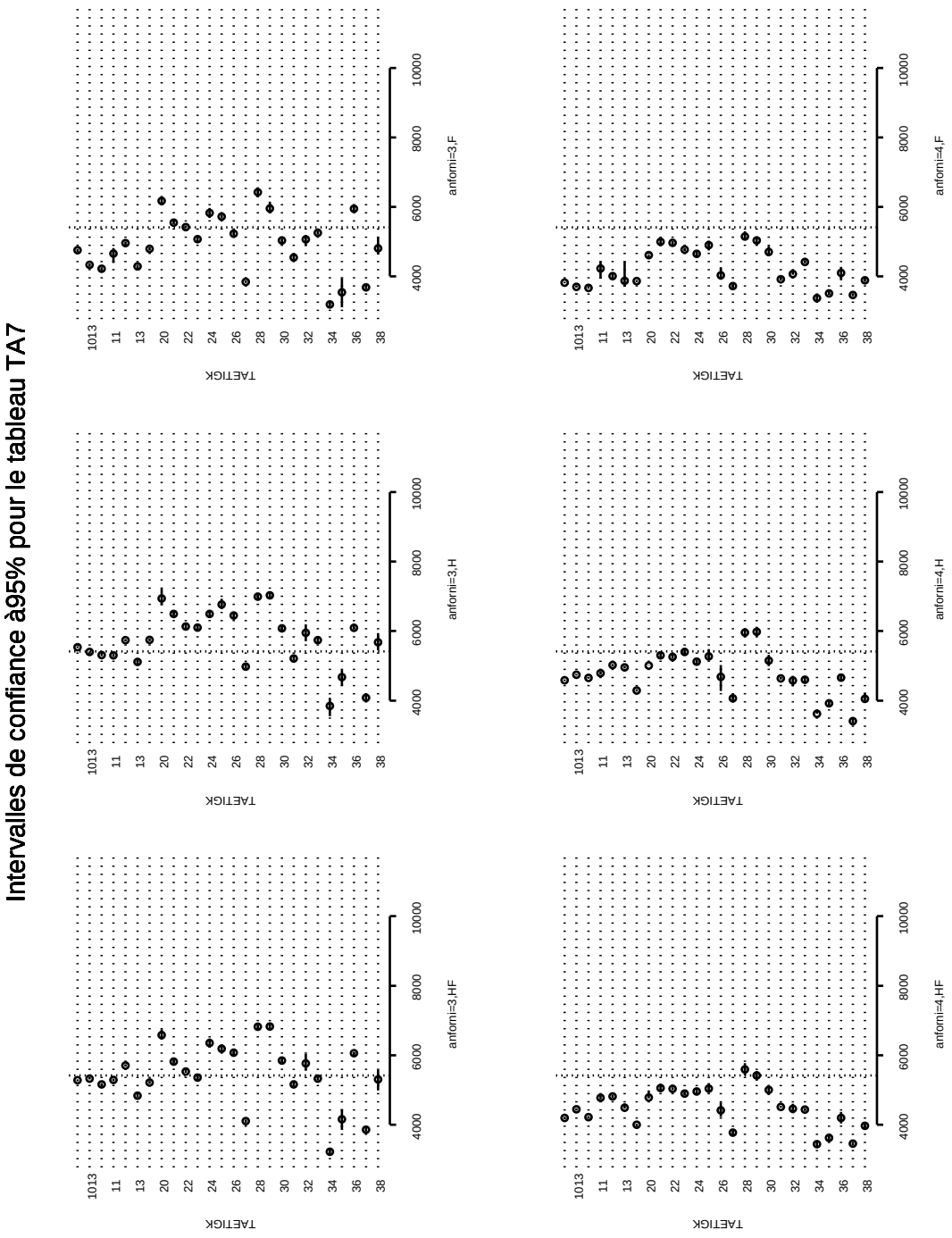


FIG. 28 – Tableau TA7 - niveaux de qualification 3 et 4, par sexe

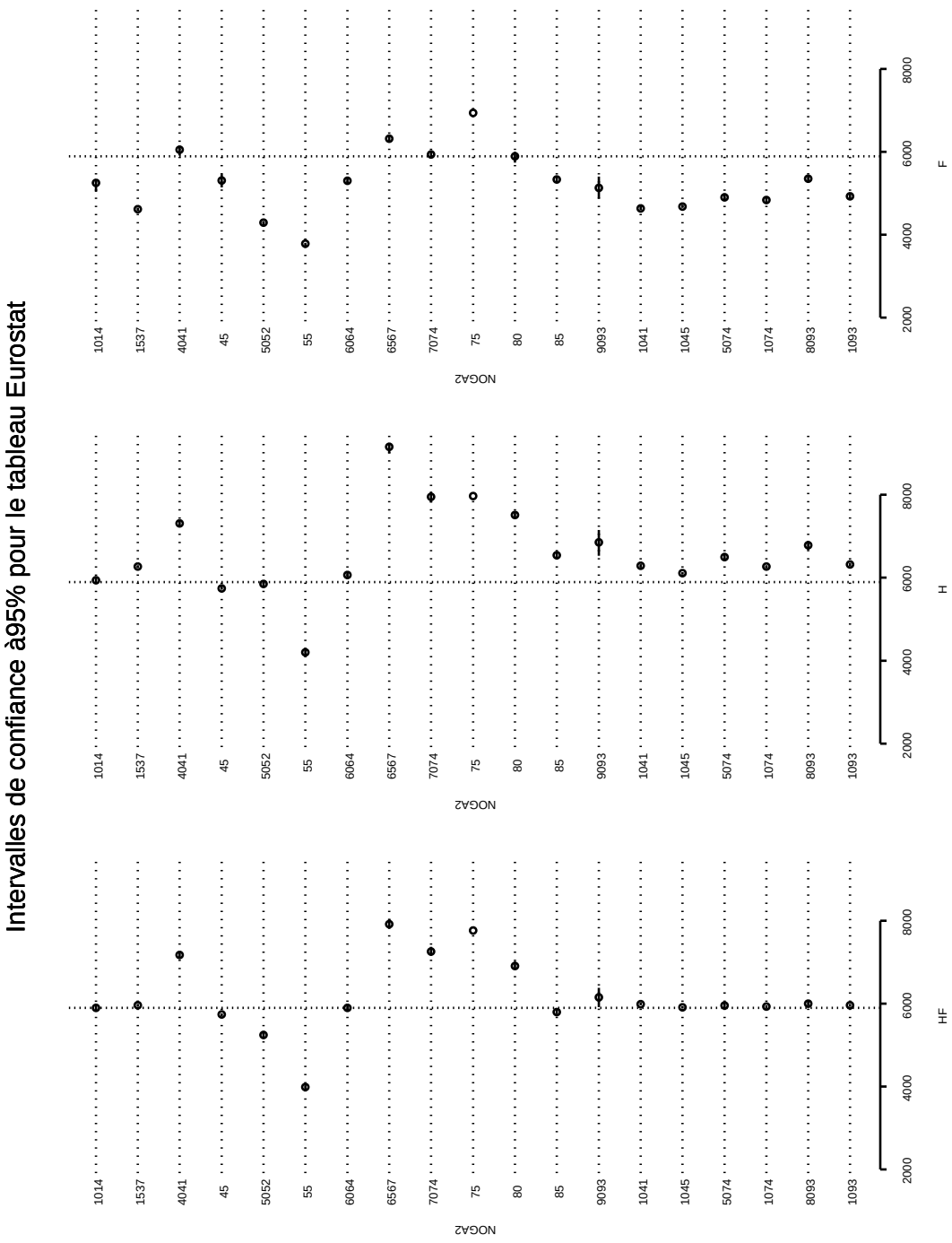
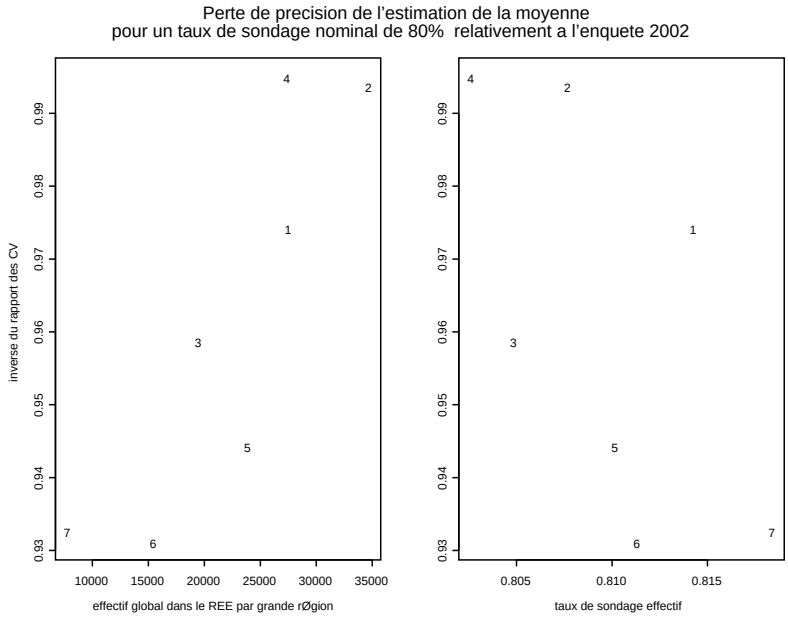
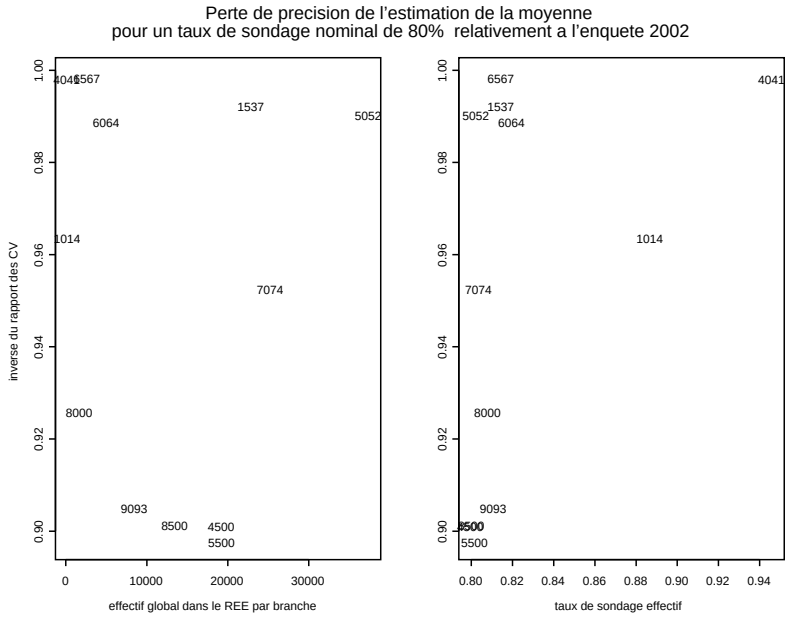


FIG. 29 – Tableau Eurostat

Annexe 4
Perte de précision de l'estimation





Références

[1] Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. *Rapport de méthode*, OFS.

[2] Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. *Rapport de méthode*, OFS.

[3] (2003). Enquête suisse sur la structure des salaires 2002 (ESS 2002). Actualité OFS.

[4] (2004). Enquête suisse sur la structure des salaires 2002. Grandes régions. Statistique de la Suisse, OFS.

[5] (2004). Enquête suisse sur la structure des salaires 2002. Résultats commentés et tableaux. Statistique de la Suisse, OFS.

[6] Bureau of Labor Statistics, U.S. Department of Labor, Labor Force Data Derived from the Current Population Survey, Chapter 1, http ://www.bls.gov/opub/hom/homch1/k.htm.

[7] Starsinic, M. et Town, M. (1999). Analysis of generalized variance estimation for the Census 2000 dress rehearsal. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodological reports of the Statistical Methods Unit of SFSO

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

Diffusionsmittel	*Kontakt* *N° à composer*	*Moyen de diffusion*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch	Service de renseignements individuels
Das BFS im Internet	www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch>>News>>Neuerscheinungen.

La Liste des publications mise à jour régulièrement, donne davantage de détails sur les divers moyens de diffusion. Elle se trouve sur Internet à l'adresse www.statistique.admin.ch >>Actualités>>Nouvelles publications.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodological reports of the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte, werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site internet de l'OFS.

Cette étude est une étude systématique de la précision des salaires mensuels médians publiés dans le cadre de l'enquête suisse sur la structure des salaires 2002. Selon la méthode développée pour l'enquête 2000, des intervalles de confiance non paramétriques à 95% pour la médiane ont été calculés, pour les secteurs privé et public, par branche économique, le niveau de qualification, le sexe, la taille de l'entreprise et la région. Les intervalles de confiance sont représentés ici sous forme graphique. Les intervalles de confiance calculés n'étant pas symétriques, on définit dans ce cas une mesure analogue celle du coefficient de variation et dénommée coefficient de variation synthétique 95%. Pour résumer la masse d'informations produites sur la précision, nous avons établi un modèle de prévision de ce coefficient de variation.

Pour le secteur privé, deux modèles linéaires ont été choisis pour estimer le coefficient de variation synthétique 95%: un modèle pour les données agrégées au niveau Suisse (où on a pris en compte la branche économique, le niveau de qualification et le sexe) et un modèle pour les données complètes (où on a pris en compte la branche économique, le sexe et la région). Les deux modèles sont utilisables pour une estimation globale de la précision.

N° de commande: 338-0031
Commandes: 032 713 60 60
Fax: 032 713 60 61
E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 3-303-00317-3