

Partie A : Questions de réflexion

1. Chacune des trois méthodes, l'analyse de la variance, la régression linéaire multiple et l'analyse en composantes principales, met en jeu un principe de décomposition de variance. Pouvez-vous l'expliquer pour chacune d'elles.
2. On considère une régression linéaire simple (une seule variable explicative). Pourquoi les deux tests de Fisher-Snedecor (Globalité du modèle) et le test de Student sur l'unique coefficient donnent le même résultat ? Justifier votre réponse.
3. Dans une régression, pour tester la nullité d'un coefficient a_i , la valeur de décision est

$$t = \frac{a_i}{\sigma_i}$$

Pouvez-vous expliquer pourquoi on divise a_i par σ_i .

Partie B : Analyse de variance

On cherche à vérifier l'influence de la teneur en molybdène présent sur le dosage d'aluminium. Pour cela, dans une solution d'aluminium, on ajoute du molybdène et on mesure le nouveau dosage d'aluminium pour voir s'il a changé. Pour faire l'expérience, on dispose de 3 laboratoires à qui on envoie une solution d'aluminium.

Chaque laboratoire fractionne en deux parties la solution d'aluminium. Dans la 1^{ère} partie, on verse 10% de molybdène et dans la 2^{ème} partie, on verse 20% de molybdène. Pour faire plusieurs mesures de chaque solution, on prend une partie de la solution à 10%, on la dose en aluminium et une deuxième partie de la solution à 10% et on la dose aussi. On fait de même avec la solution à 20%.

En résumé pour chaque laboratoire et pour chacune des proportions (10% et 20%), on a fait deux mesures de dosage d'aluminium.

Afin de pouvoir faire l'analyse de la variance, on suppose que la variance des dosages d'aluminium est la même pour les 3 laboratoires et les 2 teneurs en molybdène.

On veut contrôler (dans cet ordre) les influences des laboratoires et de la teneur en molybdène sur le dosage en aluminium.

1. Préciser le facteur A ainsi que son nombre p de modalités et le facteur B ainsi que son nombre q de modalités.
2. Représenter de façon schématique, le fichier de données ayant permis de faire cette analyse de la variance. Indiquer clairement la position sur le tableau du facteur A et du facteur B.
3. Préciser l'hypothèse nulle et l'hypothèse alternative dans le test d'analyse de la variance.

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

On vous donne ci-dessous, les éléments pour faire l'analyse de la variance des dosages d'aluminium par rapport aux facteurs A et B

Variance	Sommes des carrés	ddl	Carrés moyens	F	ddl x ddl	p-value
Facteur A	182	2	91	5,6875	2;6	0,04117933
Facteur B	47	1	47	2,9375	1;6	0,13737465
Interaction facteurs	16	2	8	0,5	2;6	0,0202855
Résiduelle	96	6	16			
Totale	341	11				

4. Avec un risque de 5%, vérifier si l'hypothèse nulle peut être retenue.
5. Avec le même risque, vérifier si le facteur A influe sur les résultats.
6. Même question pour le facteur B.
7. Faire une conclusion générale.

Partie C : RLM, ACP et AFC

On voudrait examiner si, pour un pays, la consommation d'énergie électrique, les importations des biens et le PIB sont des facteurs qui influent sur le taux des maisons reliées à Internet. On dispose des données concernant l'année 2010, pour 22 pays européens.

Pays	Internet (Tx)	Énergie Electrique	Importations	PIB1000
Autriche	63,7	3,94	140	36,131
Belgique	70	5,23	390,1	33,217
Tchequie	53,6	4,04	110	24,124
Danemark	80,1	3,56	84,5	32,611
Estonie	64,5	4,09	12	18,126
Finlande	75,8	6,65	68,2	32,025
France	66,8	4,09	599,2	29,938
Allemagne	75,2	4,06	1066,8	34,581
Grèce	41,2	2,39	63,3	22,287
Hongrie	52,2	2,55	87,4	17,301
Irlande	57,5	3,29	60,5	37,21
Italie	48,9	2,84	486,6	27,081
Pays bas	81	5,01	360	37,123
Norvège	82,6	6,39	77,3	46,734
Pologne	56,8	2,67	160	17,933
Portugal	50,3	2,21	75,2	21,414
Slovaquie	49,4	3,19	65,9	20,767
Slovenie	62	3,44	26,4	25,133
Espagne	57,4	2,77	280	26,981
Suède	82,6	5,41	148,4	35,151
Suisse	77	3,39	176,3	39,6
Royaume Uni	75	3,28	558,6	32,814

TABLE 1. Données Internet

GS-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

1. On cherche à établir s'il y a une relation linéaire entre le taux de liaisons Internet d'une part et les autres variables, d'autre part. On effectue toutes les régressions. Les résultats sont donnés à l'annexe 1.
 - (a) Considérons la régression complète, c'est-à-dire de la variable internet avec toutes les autres variables. En examinant le graphique des résidus, présenté sur la figure 1, quelles conclusions pouvez-vous établir concernant la validité de cette régression ? Justifier votre réponse.
 - (b) Établir la meilleure, selon vous, régression expliquant le taux de liaisons à internet. Justifier votre réponse.
 - (c) Donner la forme mathématique du modèle linéaire retenu à la question précédente.
 - (d) A votre avis, cette régression est-elle fiable ? Justifier votre réponse.
 - (e) Pour l'année 2011 la France a eu les valeurs suivantes pour les variables précitées :

Internet : 68.2	Energie électrique : 4.16
Importations : 610	PIB_1000 : 29.938

Calculer l'estimation selon le modèle que vous avez retenu et conclure.
2. On procède maintenant à une ACP. Les résultats sont donnés en annexe 2 et les graphiques concernant les deux premiers axes sont donnés par les figures 2 et 3.
 - (a) Au vu des résultats, combien d'axes faut-il retenir pour l'analyse ? Justifier votre réponse.
 - (b) Examiner le graphique des variables. Quels sont les renseignements que vous pouvez en tirer, en particulier concernant la formation des axes et les relations entre les variables ? Justifier vos réponses.
 - (c) Examiner le graphique des observations ainsi que les résultats concernant les observations (pays). Quels sont les renseignements que vous pouvez en tirer, en particulier concernant la formation des axes et la contribution des observations à la formation de ces axes ?
3. On décide maintenant de lire le tableau initial autrement. On considère que c'est un tableau d'effectifs (peu importe si ces effectifs sont des nombres décimaux). On peut dans ce cas faire une AFC avec ce tableau. Les résultats sont donnés en annexe 3 et les graphiques concernant les deux premiers axes sont donnés par les figures 4, 5 et 6. On étudiera les profils lignes donc les pays.
 - (a) Examiner le graphique des colonnes (Internet, énergie, importation et PIB), ainsi que les résultats concernant ces dernières. Quels sont les renseignements que vous pouvez en tirer, en particulier concernant la formation des axes et les relations entre ces colonnes ? Justifier vos réponses.

GS-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

- (b) Examiner le graphique des observations (pays) ainsi que les résultats concernant les observations (pays). Quels sont les renseignements que vous pouvez en tirer, en particulier concernant la formation des axes et la contribution des observations à la formation de ces axes ? Justifier vos réponses.
- (c) Quelles sont les informations que nous pouvons obtenir avec le graphique de l'AFC où figurent ensemble les colonnes et les profils lignes (pays) ? Justifier vos réponses.

ANNEXE 1

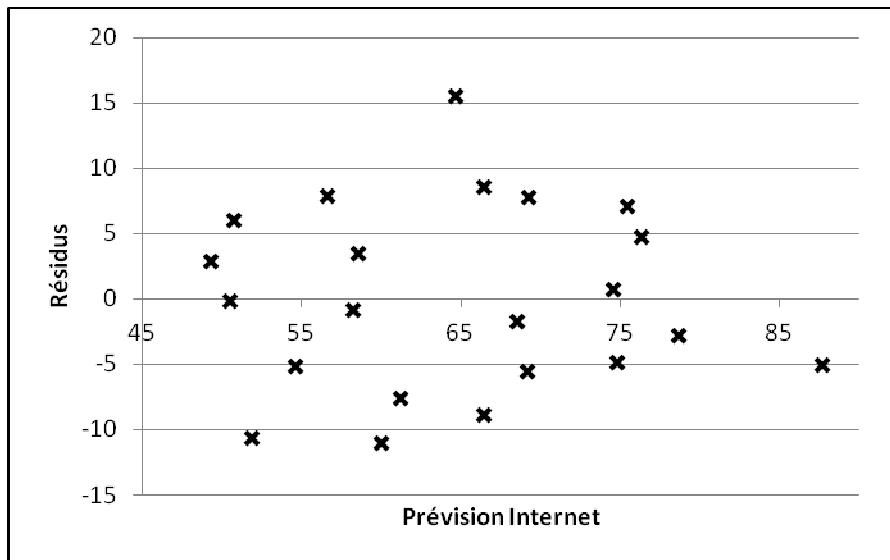


Figure 1. Résidus de la régression complète en fonction des prévisions

	Internet (Tx)	Énergie Elec	Importations	PIB1000
Internet (Tx)	1			
Énergie Elec	0,73	1		
Importations	0,25	0,03	1	
PIB1000	0,75	0,61	0,24	1

Tableau des corrélations

Régression n° 1

Variable expliquée : Internet

Variable explicative : Energie électrique

Statistiques de la régression

Coefficient de détermination multiple	0,73
Coefficient de détermination R ²	0,53
Coefficient de détermination R ²	0,50
Erreur-type	8,91
Observations	22

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1	1777,07	1777,07	22,37	1,28E-04
Résidus	20	1588,60	79,43		
Total	21	3365,68			

	Coefficients	Erreur-type	Statistique t	Probabilité	Inf IDC 95%	Sup IDC 95%
Constante	36,00	6,36	5,66	1,53E-05	22,74	49,27

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

Énergie Elec	7,47	1,58	4,73	1,28E-04	4,18	10,77
--------------	------	------	------	----------	------	-------

Régression n° 2

Variable expliquée : Internet

Variable explicative : Importations

Statistiques de la régression

Coefficient de détermination multiple	0,25
Coefficient de détermination R ²	0,06
Coefficient de détermination R ²	0,01
Erreur-type	12,57
Observations	22

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1	205,34	205,34	1,30	0,27
Résidus	20	3160,34	158,02		
Total	21	3365,68			

	Coefficients	Erreur-type	Statistique t	Probabilité	Inf IDC 95%	Sup IDC 95%
Constante	61,88	3,65	16,95	2,48E-13	54,27	69,50
Importations	0,01	0,01	1,14	0,27	-0,01	0,03

Régression n° 3

Variable expliquée : Internet

Variable explicative : PIB1000

Statistiques de la régression

Coefficient de détermination multiple	0,75
Coefficient de détermination R ²	0,56
Coefficient de détermination R ²	0,53
Erreur-type	8,65
Observations	22

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1	1869,13	1869,13	24,98	6,91E-05
Résidus	20	1496,55	74,83		
Total	21	3365,68			

	Coefficients	Erreur-type	Statistique t	Probabilité	Inf IDC 95%	Sup IDC 95%
Constante	29,71	7,24	4,10	5,53E-04	14,61	44,82
PIB1000	1,19	0,24	5,00	6,91E-05	0,69	1,68

Régression n° 4

Variable expliquée : Internet

Variables explicatives : Energie électrique, Importations

Statistiques de la régression

Coefficient de détermination multiple	0,76
Coefficient de détermination R ²	0,58
Coefficient de détermination R ²	0,53
Erreur-type	8,64
Observations	22

ANALYSE DE VARIANCE

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	2	1947,16	973,58	13,04	2,72E-04
Résidus	19	1418,52	74,66		
Total	21	3365,68			

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	<i>Inf IDC 95%</i>	<i>Sup IDC 95%</i>
Constante	33,70	6,35	5,31	4,03E-05	20,41	46,99
Énergie Elec	7,40	1,53	4,83	1,16E-04	4,20	10,61
Importations	0,01	0,01	1,51	0,15	0,00	0,03

Régression n° 5

Variable expliquée : Internet

Variables explicatives : Energie électrique, PIB1000

Statistiques de la régression

Coefficient de détermination multiple	0,82
Coefficient de détermination R ²	0,67
Coefficient de détermination R ²	0,64
Erreur-type	7,62
Observations	22

ANALYSE DE VARIANCE

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>	<i>Valeur critique de F</i>
Régression	2	2261,24	1130,62	19,45	2,53E-05
Résidus	19	1104,44	58,13		
Total	21	3365,68			

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>	<i>Inf IDC 95%</i>	<i>Sup IDC 95%</i>
Constante	25,09	6,63	3,79	1,25E-03	11,22	38,95
Énergie Elec	4,45	1,71	2,60	1,77E-02	0,86	8,03

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

PIB1000	0,77	0,27	2,89	9,46E-03	0,21	1,32
---------	------	------	------	----------	------	------

Régression n° 6

Variable expliquée : Internet

Variables explicatives : Importations, PIB1000

Statistiques de la régression

Coefficient de détermination multiple	0,75
Coefficient de détermination R ²	0,56
Coefficient de détermination R ²	0,51
Erreur-type	8,83
Observations	22

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	2	1884,65	942,32	12,09	4,10E-04
Résidus	19	1481,03	77,95		
Total	21	3365,68			

	Coefficients	Erreur-type	Statistique t	Probabilité	Inf IDC 95%	Sup IDC 95%
Constante	29,71	7,39	4,02	7,32E-04	14,24	45,18
Importations	3,46E-03	0,01	0,45	0,66	-0,01	0,02
PIB1000	1,16	0,25	4,64	1,78E-04	0,64	1,68

Régression n° 7

Variable expliquée : Internet

Variables explicatives : Energie électrique, Importations, PIB1000

Statistiques de la régression

Coefficient de détermination multiple	0,83
Coefficient de détermination R ²	0,69
Coefficient de détermination R ²	0,63
Erreur-type	7,65
Observations	22

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	3	2311,42	770,47	13,15	8,67E-05
Résidus	18	1054,26	58,57		
Total	21	3365,68			

	Coefficients	Erreur-type	Statistique t	Probabilité	Inf IDC 95%	Sup IDC 95%
Constante	24,82	6,66	3,73	1,54E-03	10,84	38,81
Énergie Elec	4,69	1,74	2,70	0,01	1,04	8,35
Importations	0,01	0,01	0,93	0,37	-0,01	0,02

ANNEXE 2 Résultats de l'ACP

Valeurs propres

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.4555794	61.389486	61.38949
comp 2	0.9813840	24.534599	85.92408
comp 3	0.3506088	8.765221	94.68931
comp 4	0.2124278	5.310695	100.00000

Résultats sur les axes

\$coord

	Dim.1	Dim.2	Dim.3	Dim.4
Internet	0.9277614	-0.037794466	0.02094464	-0.37066375
Elec	0.8423367	-0.324913630	0.37930851	0.20254636
Import	0.3227930	0.935053652	0.13627834	0.05392175
PIB1000	0.8838052	0.007832214	-0.43327060	0.17636208

\$cor

	Dim.1	Dim.2	Dim.3	Dim.4
Internet	0.9277614	-0.037794466	0.02094464	-0.37066375
Elec	0.8423367	-0.324913630	0.37930851	0.20254636
Import	0.3227930	0.935053652	0.13627834	0.05392175
PIB1000	0.8838052	0.007832214	-0.43327060	0.17636208

\$cos2

	Dim.1	Dim.2	Dim.3	Dim.4
Internet	0.8607413	1.428422e-03	0.0004386781	0.137391617
Elec	0.7095312	1.055689e-01	0.1438749444	0.041025029
Import	0.1041953	8.743253e-01	0.0185717873	0.002907555
PIB1000	0.7811117	6.134358e-05	0.1877234119	0.031103585

\$contrib

	Dim.1	Dim.2	Dim.3	Dim.4
Internet	35.052472	0.145551759	0.1251189	64.676858
Elec	28.894653	10.757142038	41.0357457	19.312459
Import	4.243207	89.091055481	5.2970108	1.368726
PIB1000	31.809668	0.006250722	53.5421245	14.641957

Représentation des variables

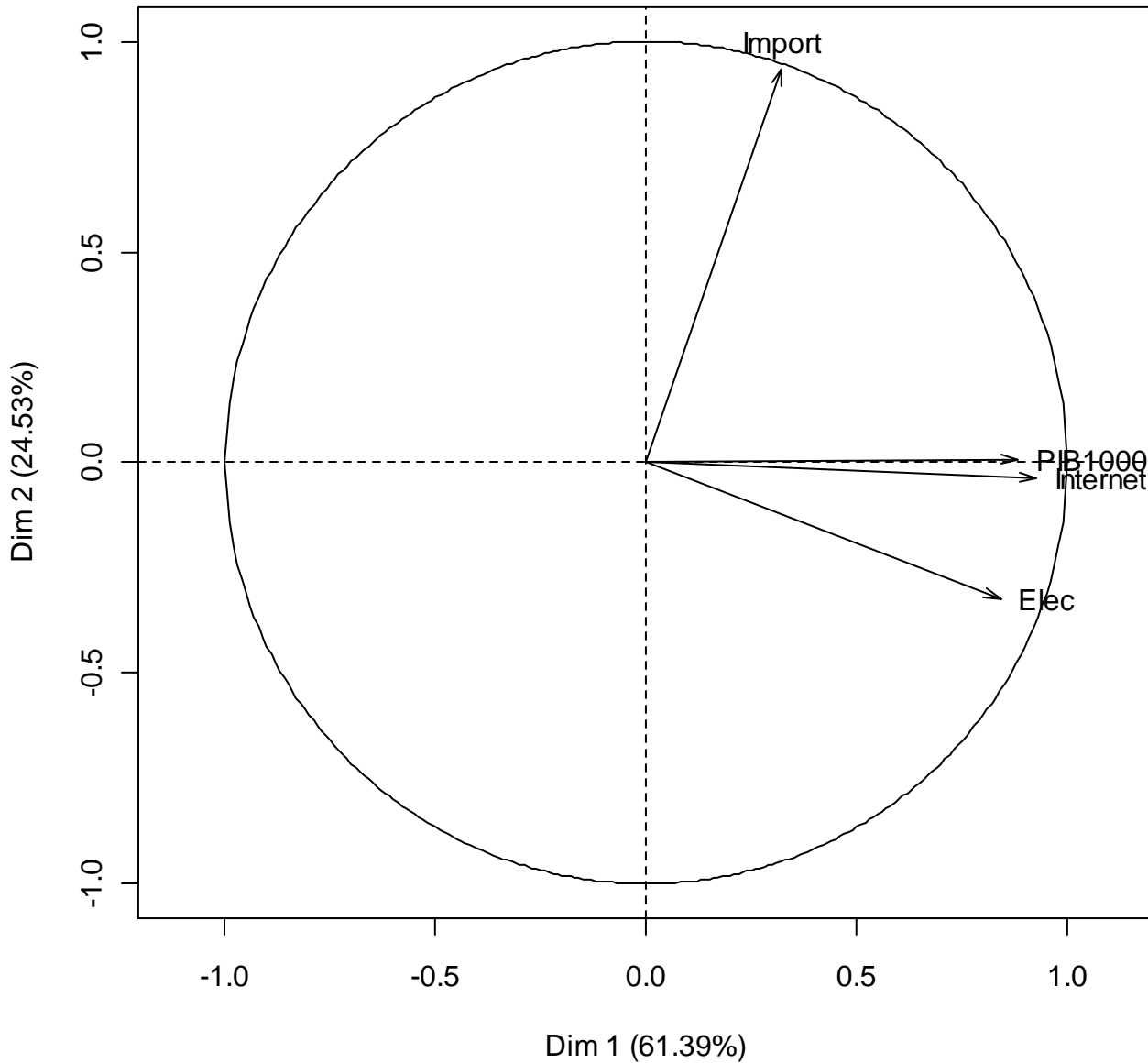


Figure 2. Résultat de l'ACP sur les variables

Résultats sur les individus

Score

	Dim.1	Dim.2	Dim.3	Dim.4
Autriche	0.4050503	-0.36276276	-0.662357911	0.38771197
Belgique	1.2772913	0.20560865	0.547535468	0.42274560
Tchequie	-0.9309328	-0.48419293	0.466506749	0.47496212
Danemark	0.7187259	-0.52246425	-0.537048179	-1.01698892
Estonie	-0.9033223	-0.90693797	0.999753197	-0.55699248
Finlande	1.8383486	-1.41412812	1.137209595	0.35547924
France	0.5481702	1.31123445	0.432340143	0.15016353
Allemagne	1.6588822	3.06068375	0.432376729	0.04039723
Grèce	-2.4340549	-0.17374629	-0.317763661	0.56575562

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

Hongrie	-2.1785018	-0.16556384	0.291177316	-0.32555844
Irlande	-0.1693009	-0.46487799	-1.201189581	0.56932297
Itamie	-1.1677231	1.28010856	-0.118945669	0.66369129
Pays bas	1.9645638	0.12221924	0.065853608	-0.19436968
Norvège	3.1240051	-1.31489648	-0.360294791	0.54777143
Pologne	-1.7990200	0.06179652	0.375402841	-0.51571537
Portugal	-2.1325925	-0.10875125	-0.294394823	-0.12920215
Slovaquie	-1.7922130	-0.40901264	0.277607707	0.25125954
Slovenie	-0.7925348	-0.66050556	-0.001101194	-0.27981623
Espagne	-0.9693596	0.49418065	-0.312379593	-0.01601616
Suède	1.9026162	-0.79138541	0.274993426	-0.34828514
Suisse	1.0777450	-0.11340362	-1.211053034	-0.49008949
Royaume Uni	0.7541572	1.35679729	-0.284228344	-0.55622648

§cos2

	Dim.1	Dim.2	Dim.3	Dim.4
Autriche	0.18544763	0.148747205	4.958940e-01	0.1699111343
Belgique	0.75802896	0.019642134	1.392934e-01	0.0830355502
Tchequie	0.56118499	0.151812062	1.409241e-01	0.1460788530
Danemark	0.24456077	0.129232990	1.365484e-01	0.4896577998
Estonie	0.27676903	0.278989093	3.390140e-01	0.1052278538
Finlande	0.49706983	0.294129898	1.902141e-01	0.0185861800
France	0.13479186	0.771246945	8.384630e-02	0.0101149029
Allemagne	0.22358081	0.761097636	1.518896e-02	0.0001325887
Grèce	0.92922670	0.004734695	1.583687e-02	0.0502017310
Hongrie	0.95604723	0.005521975	1.707963e-02	0.0213511561
Irlande	0.01424763	0.107424151	7.172112e-01	0.1611170033
Itamie	0.39445209	0.474032510	4.092718e-03	0.1274226860
Pays bas	0.98543269	0.003813940	1.107271e-03	0.0096460991
Norvège	0.81886410	0.145067958	1.089192e-02	0.0251760226
Pologne	0.88739018	0.001047058	3.864006e-02	0.0729227086
Portugal	0.97529811	0.002536239	1.858583e-02	0.0035798195
Slovaquie	0.91263324	0.047532483	2.189677e-02	0.0179375098
Slovenie	0.54968397	0.381794210	1.061217e-06	0.0685207628
Espagne	0.73312837	0.190538038	7.613345e-02	0.0002001367
Suède	0.81472326	0.140956080	1.701972e-02	0.0273009415
Suisse	0.40313807	0.004463501	5.090356e-01	0.0833628552
Royaume Uni	0.20313873	0.657504854	2.885385e-02	0.1105025612

§contrib

	Dim.1	Dim.2	Dim.3	Dim.4
Autriche	0.30369748	0.60951409	5.687743e+00	3.216506326
Belgique	3.01997429	0.19580381	3.886682e+00	3.824055425
Tchequie	1.60420543	1.08586353	2.821437e+00	4.827073924
Danemark	0.95620256	1.26430403	3.739220e+00	22.130867606
Estonie	1.51045841	3.80972414	1.295806e+01	6.638418875
Finlande	6.25574569	9.26223682	1.676623e+01	2.703924005
France	0.55622978	7.96340978	2.423291e+00	0.482497379

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

Allemagne	5.09394684	43.38856396	2.423701e+00	0.034919586
Grèce	10.96690477	0.13982005	1.309070e+00	6.848946077
Hongrie	8.78494802	0.12696071	1.099182e+00	2.267900028
Irlande	0.05305693	1.00095911	1.870586e+01	6.935590040
Itamie	2.52407973	7.58982858	1.834221e-01	9.425366142
Pays bas	7.14423288	0.06918589	5.622295e-02	0.808393887
Norvège	18.06536729	8.00795252	1.682947e+00	6.420439311
Pologne	5.99094471	0.01768750	1.827046e+00	5.690968065
Portugal	8.41858480	0.05477809	1.123608e+00	0.357195111
Slovaquie	5.94569422	0.77483961	9.991197e-01	1.350862371
Slovenie	1.16267950	2.02065102	1.572108e-05	1.675374110
Espagne	1.73937469	1.13112301	1.265085e+00	0.005488865
Suède	6.70078555	2.90077768	9.803905e-01	2.595588792
Suisse	2.15008357	0.05956514	1.901432e+01	5.139451464
Royaume Uni	1.05280284	8.52645094	1.047344e+00	6.620172613

Représentation des individus

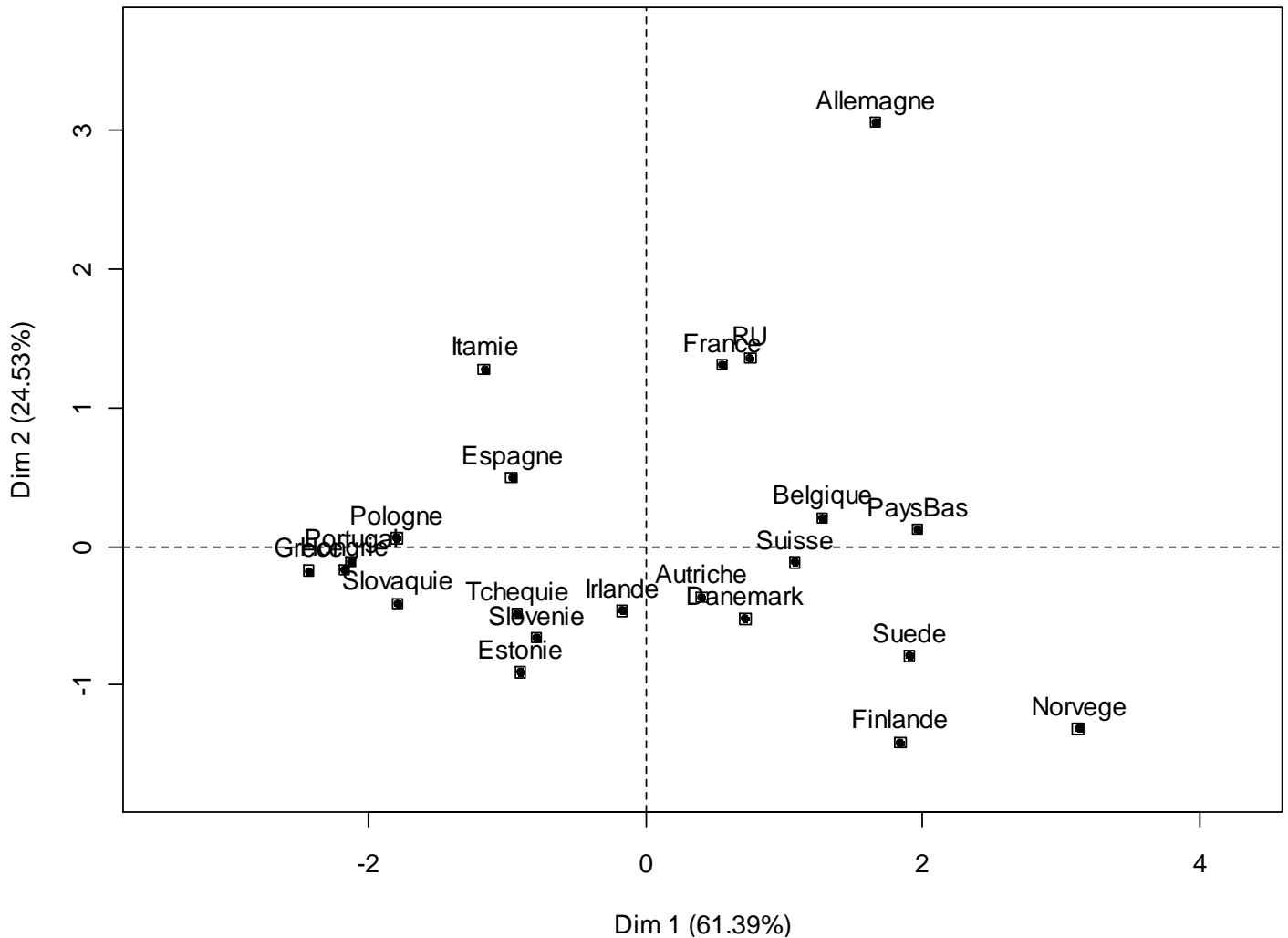


Figure 3. Résultats de l'ACP sur les individus

ANNEXE 3

Résultats de l'AFC

Valeurs propres

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	1.796835e-01	9.754411e+01	97.5441
dim 2	3.807534e-03	2.066982e+00	99.6111
dim 3	7.164007e-04	3.889098e-01	100.0000
dim 4	1.127835e-32	6.122637e-30	100.0000

Résultats sur les colonnes

§coord

	Dim 1	Dim 2	Dim 3
Internet	0.6595952	-0.0737073334	-0.0132953786
Elec	0.7158337	-0.0442231302	0.2416035412
Import	-0.2756277	-0.0008596336	0.0002611668
PIB1000	0.6252083	0.1743800793	-0.0043451323

§contrib

	Dim 1	Dim 2	Dim 3
Internet	47.523975	28.00551437	4.842965563
Elec	3.321998	0.59832678	94.914789320
Import	29.710136	0.01363798	0.006690321
PIB1000	19.443891	71.38252086	0.235554796

§cos2

	Dim 1	Dim 2	Dim 3
Internet	0.9872706	1.232828e-02	4.011277e-04
Elec	0.8946689	3.414578e-03	1.019165e-01
Import	0.9999894	9.726943e-06	8.978122e-07
PIB1000	0.9277798	7.217540e-02	4.481274e-05

§inertia

	Internet	Elec	Import	PIB1000
	0.086493760	0.006671835	0.053384783	0.037657067

Résultats sur les lignes

§coord

	Dim1	Dim2	Dim3
Aut	0.27908133	0.087141776	-0.002364600
Belg	-0.17432184	0.002153230	0.021767102
Tch	0.28306791	-0.001452667	0.036501789
Dnk	0.61665081	-0.036107496	-0.060382523
Est	1.27845331	-0.292958816	0.020808689
Fin	0.72296275	-0.031512120	0.097666555

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

Fr	-0.33514669	-0.009237656	0.006747916
Al	-0.43941606	-0.008361601	0.003463589
Gr	0.46337395	0.086509554	-0.014647902
Ho	0.34003801	-0.103514404	-0.030525921
Ir	0.69761326	0.209913177	-0.027223164
It	-0.34589080	0.016457906	0.003001432
PBas	-0.09278923	-0.000932794	0.005121170
Nor	0.74163886	0.130259215	0.046086223
Pol	0.06446901	-0.089769628	-0.023013046
Por	0.43378635	-0.014745208	-0.052167282
Slova	0.50292154	-0.025313131	0.010972838
Slove	1.04457840	-0.050150985	-0.030505579
Esp	-0.13148579	0.004897455	-0.014045148
Suede	0.34252454	-0.019419558	0.013056086
Suisse	0.23393186	0.050786793	-0.041704295
RU	-0.28756266	-0.010434355	-0.011235112

§contrib

	Dim1	Dim2	Dim3
Aut	1.45684513	6.702996984	0.02623127
Belg	1.16246398	0.008369919	4.54599871
Tch	1.17901182	0.001465323	4.91719425
Dnk	5.85800017	0.947828816	14.08787466
Est	12.38018362	30.678552777	0.82261942
Fin	7.32623345	0.656853505	33.53457758
Fr	6.03330211	0.216308306	0.61344648
Al	17.49200160	0.298903445	0.27257904
Gr	2.12822827	3.500640492	0.53340613
Ho	1.41465966	6.186756063	2.85947822
Ir	5.91872977	25.289651522	2.26062458
It	5.19062617	0.554568586	0.09802818
PBas	0.31917701	0.001522202	0.24385246
Nor	8.99048839	13.088190998	8.70748553
Pol	0.07571084	6.927542953	2.41967141
Por	2.15312429	0.117404168	7.81026829
Slova	2.70263434	0.323104452	0.32268366
Slove	9.79349048	1.065316695	2.09491377
Esp	0.48704841	0.031887456	1.39386063
Suede	2.44467180	0.370834614	0.89087217
Suisse	1.24412936	2.767277717	9.91746225
RU	4.24923932	0.264023005	1.62687131

§cos2

	Dim1	Dim2	Dim3
Aut	0.9111046	8.882997e-02	6.540665e-05
Belg	0.9844996	1.502079e-04	1.535017e-02
Tch	0.9836182	2.590463e-05	1.635587e-02
Dnk	0.9871503	3.384538e-03	9.465143e-03

GSI-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

Est	0.9498705	4.987786e-02	2.516425e-04
Fin	0.9802482	1.862342e-03	1.788941e-02
Fr	0.9988362	7.588362e-04	4.049149e-04
Al	0.9995760	3.619452e-04	6.210357e-05
Gr	0.9653868	3.364854e-02	9.646926e-04
Ho	0.9084875	8.419098e-02	7.321525e-03
Ir	0.9156966	8.290897e-02	1.394438e-03
It	0.9976662	2.258689e-03	7.512150e-05
PBas	0.9968627	1.007423e-04	3.036538e-03
Nor	0.9664545	2.981353e-02	3.731975e-03
Pol	0.3261229	6.323217e-01	4.155540e-02
Por	0.9846222	1.137679e-03	1.424014e-02
Slova	0.9969997	2.525726e-03	4.746052e-04
Slove	0.9968521	2.297775e-03	8.501727e-04
Esp	0.9873641	1.369811e-03	1.126607e-02
Suede	0.9953544	3.199433e-03	1.446173e-03
Suisse	0.9268572	4.368533e-02	2.945748e-02
RU	0.9971649	1.312905e-03	1.522148e-03

\$inertia

Aut	Belg	Tch	Dnk	Est	Fin	0.0028731173	0.0021216423	0.0021537724	0.0106628748	0.0234191380	0.0134292852
Fr	Al	Gr	Ho	Ir	It	0.0108534797	0.0314435758	0.0039611846	0.0027979583	0.0116140886	0.0093485169
PBas	Nor	Pol	Por	Slova	Slove	0.0005753134	0.0167151429	0.0004171430	0.0039292323	0.0048708022	0.0176528576
Esp	Suede	Suisse	RU			0.0008863454	0.0044131739	0.0024119091	0.0076568900		

Graphique AFC

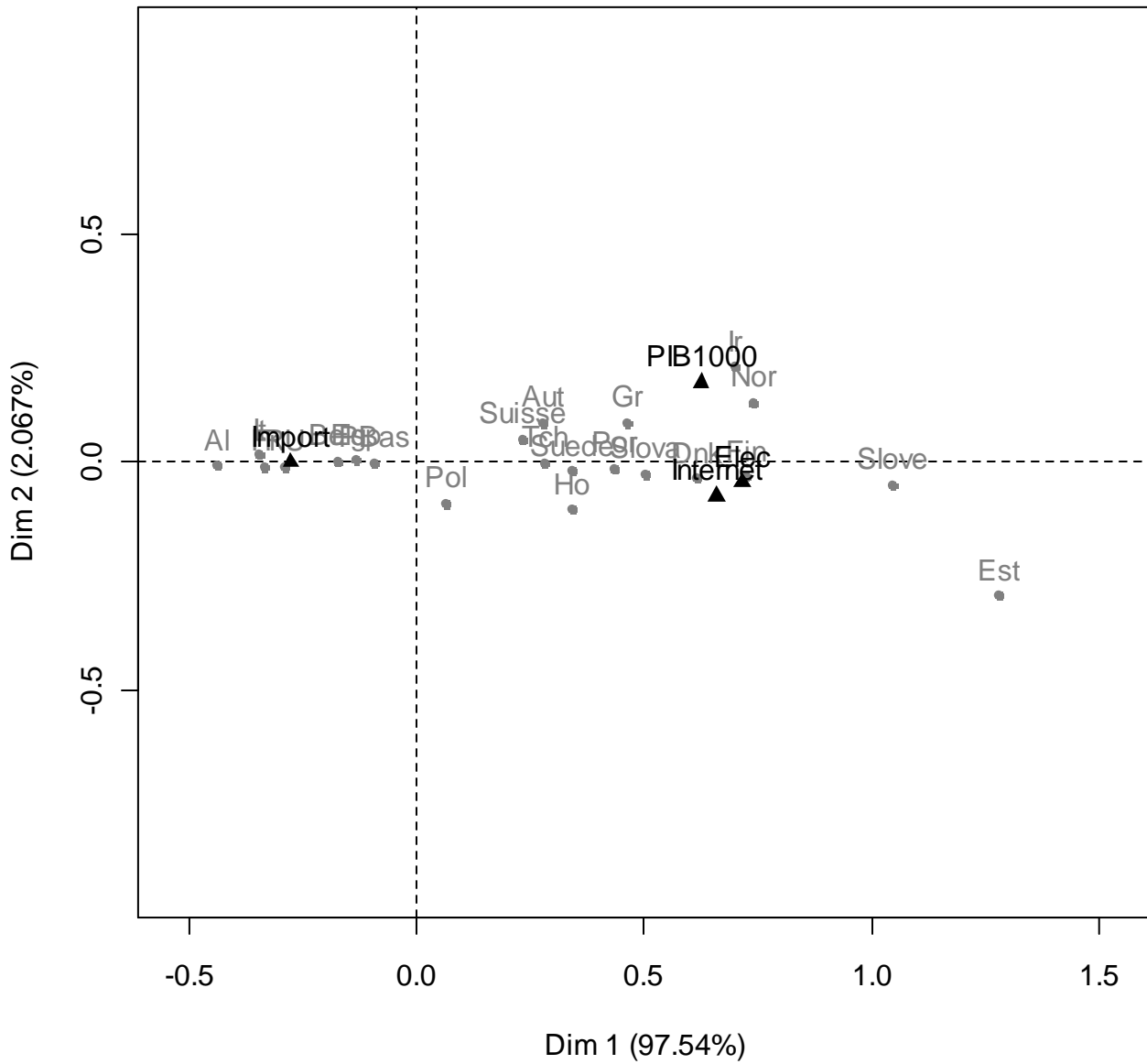


Figure 4. Résultats de l'AFC

BAREME

Partie A

Question	1.	2.	3.
Points	2	1	1

Partie B

Question	1.	2.	3.	4.	5.	6.	7.
Points	0.5	0.5	0.5	1	0.5	0.5	0.5

Partie C – RLM

Question	(a)	(b)	(c)	(d)	(e)
Points	1	1	0.5	0.5	1

Partie C – ACP

Question	(a)	(b)	(c)
Points	1	2	2

Partie C – AFC

Question	(a)	(b)	(c)
Points	1	1	1

CORRECTION

Partie A

1. Pour chaque méthode la variance se décompose de la façon suivante

Variance totale = variance expliquée + variance résiduelle

- Dans le cas de l'ANOVA, la variance totale est celle de la variable quantitative quelles que soient les modalités du ou des variables qualitatives, la variance expliquée est la variance inter modalités
- Dans le cas de la régression linéaire la variance expliquée correspond à la variance de Y expliquée par le modèle (fonction des X_i)
- Dans l'ACP, la variance totale est l'inertie et la variance expliquée est l'inertie prise en compte lorsqu'on projette le nuage sur un axe.

Plus la variance expliquée très importante par rapport à la variance résiduelle, plus on retire de l'information pertinente de l'analyse effectuée.

2. Notons $a_i, i=1, \dots, p$ les coefficients associés aux variables explicatives X_i dans la régression.

Les hypothèses du test de Fisher sont

$$H_0 : a_1 = a_2 = \dots = a_p = 0$$

$$H_1 : \exists i \in \{1, \dots, p\}, a_i \neq 0$$

Si on a une seule variable explicative alors les hypothèses deviennent

$$H_0 : a_1 = 0$$

$$H_1 : a_1 \neq 0$$

On retrouve les mêmes hypothèses que pour le test de Student sur a_1 . A noter qu'il est possible de montrer mathématiquement que les deux tests sont alors les mêmes car la loi de Fisher à 1 (car une variable explicative) et k degrés de liberté correspond au carré d'une loi de Student à k ddl.

3. Diviser le coefficient par son écart-type permet de « normaliser » la valeur du coefficient afin de prendre en compte les ordres de grandeurs différents des variables.

Partie B

- 1) Facteur A = Laboratoires / nb modalités $p=3$

Facteur B = teneur en molybdène / nb modalités $q=2$

- 2) Faire dessin d'un tableau croisé dynamique

- 3) $H_0 : \mu_{(i,j)} = \mu \Leftrightarrow$ égalité des valeurs moyennes de Y à l'intérieur de chaque croisement de modalités de A et B

$H_1 : \exists (i,j), \exists (i',j') \mu_{(i,j)} \neq \mu_{(i',j')} \Leftrightarrow$ Il existe au moins deux valeurs moyennes de croisement de traitement différentes

- 4) La p-valeur associée à l'interaction est $p=0.02 < 5\%$ donc on rejette H_0 , autrement dit, on accepte que l'interaction des deux facteurs (laboratoire et teneur) a une influence sur le dosage (sans savoir pour quelle(s) croisement de modalités la différence de moyennes est significative)
- 5) La p-valeur associé au facteur A est $p=0.05 < 5\%$ donc là encore on considère que le laboratoire a une influence significative sur le dosage (sans savoir de quel(s) laboratoires provient la différence de moyennes)
- 6) La p-valeur associée au facteur B est $p=0.14 > 5\%$ donc on accepte H_0 , c'est-à-dire que la teneur en molybdène n'a pas d'influence sur le dosage.
- 7) Les conclusions ci-dessus sont valables sous condition des hypothèses du test de l'ANOVA.

Partie C – RLM

- (a) Sur la figure 1, il semble que les résidus soient centrés car équidistribués autour de 0, de variance constante car amplitude du nuage constante, et non corrélés car pas de forme particulière dans le nuage.

- (b) Procédure descendante :

Régression 7 :

GS-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

- $R^2=0.69$ et p-valeur du test de Fisher $=8.67 \times 10^{-5}$ donc modèle justifié
- p-valeur du test de Student pour les importations = 0.37 donc cette variable n'a pas d'influence dans le modèle

On supprime importation et on passe à la régression 5

- $R^2=0.67$ et p-valeur du test de Fisher $=2.53 \times 10^{-5}$ donc modèle justifié
- Les p-valeurs du test de Student pour les deux variables Énergie Elec et PIB1000 sont inférieures à 1% donc les deux variables ont une influence dans ce modèle.

La régression 5 est donc retenue comme modèle.

(c) Avec la régression 5, on a

$$\text{Internet} = 25.09 + 4.45 \times \text{Energie} + 0.77 \times \text{PIB100} + \varepsilon$$

où $\varepsilon \sim N(0, \sigma^2)$ avec σ estimée par 7.62 %.

(d) Dans la régression 5, on note que :

- Le $R^2=0.69$ n'est pas très élevé
- les deux variables semblent corrélées (0.61), autrement dit l'hypothèse de variables indépendantes n'est pas nécessairement vérifiée.
- Pas de vérification possible sur les résidus

En l'absence d'informations supplémentaires, cette régression est à utiliser avec précaution.

(e) Avec le modèle de la question (c), on obtient

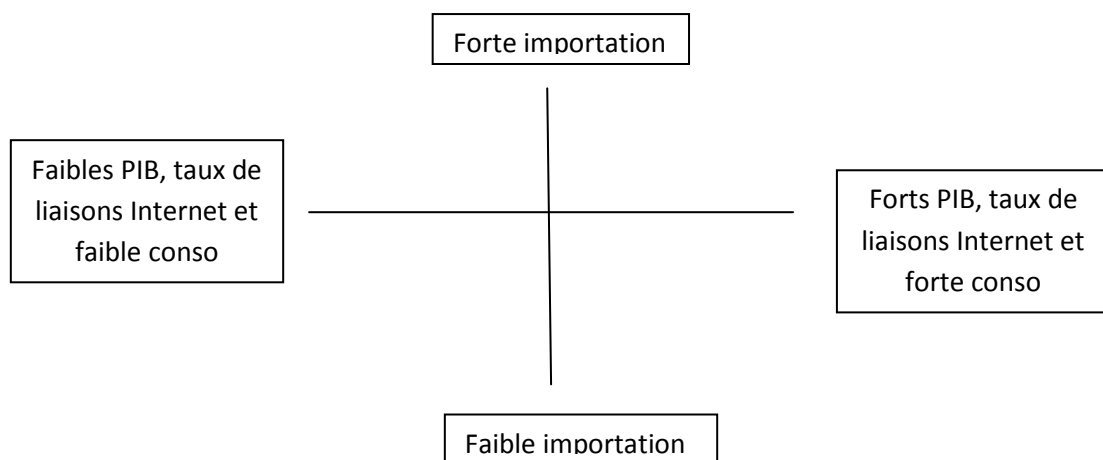
$$\hat{\text{Internet}} = 25.09 + 4.45 \times 4.16 + 0.77 \times 29.94 = 66.65$$

L'estimation du taux de liaisons Internet en France en 2011 est 66.65%. On remarque que c'est une estimation assez proche de la valeur réelle 68.2%.

Partie C – ACP

- (a) Le premier axe explique 61% de l'inertie du nuage et si on ajoute l'axe 2, on monte à 86%, c-a-d un très bon pourcentage d'inertie expliquée.
- (b) Axe 1 : Les variables Internet, Energie Electrique et PIB contribuent à part quasiment égale à la formation de l'axe 1. Elles sont toutes corrélées positivement avec l'axe 1 et fortement corrélées entre elles (confirmé par la matrice de corrélation). C'est-à-dire qu'un pays ayant un PIB important a aussi une forte consommation électrique et un fort taux de liaison Internet.
- Axe 2 : L'axe 2 est formé à 89% par la variables Importations qui se trouve non corrélée aux autres variables (angle 90%).

Nous pouvons donc proposer la cartographie suivante



- (c) Il y a 22 pays donc une contribution uniforme de chaque pays serait de $1/22=4.5\%$. Concernant l'axe 1, on note que la contribution des pays est bien répartie. En revanche sur l'axe 2, on note que l'Allemagne contribue à 43% ce qui est très important. Il faudrait peut-être refaire une ACP en

GSi-MI-SIE : EXAMEN D'ANALYSE MULTIVARIEE

supprimant l'Allemagne et en l'ajoutant en tant qu'individus supplémentaire afin de noter s'il y a des différences.

Concernant l'axe 2, seuls 3 pays sont bien représentés ($\cos^2 \sim 1$), la France, l'Allemagne et le royaume Uni. L'interprétation des autres pays sur l'axe 2 est plus hasardeuse. Ces trois pays se trouvent dans la zone des pays ayant une forte importation des biens.

Concernant l'axe 1, la plus part des pays sont bien représentés. On note les pays les plus à droite sur le graphique, Pays Bas, Suède, Finlande et Norvège, c-à-d les pays ayant un PIB important, un taux de liaison Internet élevé et une forte consommation électrique (impossible de discerner entre les trois caractéristiques). Un bémol en ce qui concerne la Finlande car moyennement représentée sur cet axe ($\cos^2=0.49$). A l'autre extrémité, on trouve La Grèce, le Portugal, la Pologne et la Slovaquie. Les 4 sont bien représentés sur l'axe 1, on peut donc en conclure que ces pays ont un PIB faible, un taux de liaisons Internet peu élevé et une faible consommation d'électricité.

Partie C – AFC

Pour l'AFC, on s'intéresse essentiellement à l'axe 1 qui explique 97% de la distance du chi-deux.

(a) Concernant les colonnes (la variable *économie* prenant les modalités : PIB, Elec, Internet et Import), les modalités Internet, Elec et PIB contribuent fortement à l'axe 1. On retrouve les mêmes résultats qu'avec l'ACP mais pas la même répartition car ici il y a une prédominance de Internet.

L'axe 1 oppose les pays ayant une forte importation aux pays ayant une consommation électrique et un taux de liaisons Internet importants.

(b) et (c) Tous les pays sont bien représentés sur l'Axe1. On note une contribution légèrement plus importante de l'Allemagne et l'Estonie. On retrouve le rapprochement entre l'Allemagne, la France, le RU et les importations. En revanche, nous avons une information complémentaire sur la Norvège et la Finlande. On peut maintenant dire que la Norvège a un fort PIB et la Finlande une forte consommation et taux de liaison. Ces trois caractéristiques étaient indiscernables avec l'ACP.