



Modélisation Statistique

Rapport du Projet

Julien Debidour
Groupe de TD/TP n°3

À rendre pour le 20 Décembre 2023
M. Jérôme Saracco

Problématiques :

- *Comment expliquer les fluctuations des activations de l'aire de Broca (variable `Prod_G_Frontal_Inf_Tri_1_L`) au cours de la tâche de production à l'aide des autres variables présentes dans notre jeu de données “activation.Rdata” ?*

- *Comment expliquer les interactions entre les différentes régions cérébrales au cours de la production d'une phrase et la notion de réseau qui se cache derrière ?*

- *Quel est le modèle “le plus simple et le meilleur possible” de la variable `Prod_G_Frontal_Inf_Tri_1_L` en fonction des autres variables ?*

Introduction

Tout comme le fonctionnement moteur, où près de 90% de la population utilise préférentiellement la main droite et 10% la main gauche, il existe une latéralisation du cerveau humain pour les activités langagières. D'après de nombreuses études¹, l'hémisphère gauche se révèle dominant pour le langage dans 90% des cas environ. Nous essayerons donc d'estimer cette affirmation grâce à l'étude d'une expérience présentant les résultats d'activation des aires cérébrales au cours d'une tâche de production langagière, sur 124 sujets.

Partie 1: Analyse descriptive préliminaire.

1.1 Description sommaire du jeu de données.

Notre jeu de données “activation.Rdata” contient ainsi des mesures cérébrales pour ces 124 sujets, divisés en groupes homme et femme. De plus, nous disposerons les concernant de renseignements supplémentaires, leur âge, leur sexe, leur index de latérisation et leur volume cérébral.

Pour cette étude, le taux d'activité cérébral est enregistré dans les six régions distinctes suivantes lors de l'exécution d'une tâche de rédaction d'une phrase comprenant un sujet, un verbe et un complément.

- gyrus frontal inférieur triangulaire : `Prod_G_Frontal_Inf_Tri_1`
- sillon supérieur temporal : `Prod_S_Sup_Temporal_4`
- gyrus Occipital Latéral : `Prod_G_Occipital_Lat_1`
- gyrus angulaire : `Prod_G_Angular_2`
- opercule rolandique : `Prod_G_Rolandic_Oper_1`
- hippocampe : `Prod_G_Hippocampus-1`

Ces aires cérébrales sont par ailleurs considérées selon l'hémisphère (droite (R) et gauche (L)). Nous aurons ainsi 12 variables caractérisant les zones du cerveau activées pendant l'exécution de cette action spécifique..

Chaque groupe contient des mesures pour plusieurs variables telles que l'âge, le volume cérébral et plusieurs indicateurs spécifiques pour différentes régions du cerveau.

Nous chercherons donc à élucider les variations des taux d'activation du gyrus frontal inférieur triangulaire gauche (Aire de Broca) pendant la tâche de production, en tenant compte des autres variables. Ceci afin de comprendre les interactions entre les différentes régions cérébrales impliquées dans ce type de tâches.

Commençons par l'analyse descriptive de notre jeu de données "activation.Rdata" pour le groupe général.

1.2 Groupe général. (variable donnees).

Premièrement, nous pouvons observer une synthèse des données afin de repérer celles qui pourraient être pertinentes pour la suite de notre étude. (*grâce à la commande summary*).

Sujet	Sexe	Age	Volume_Cerebral	Index_Lateralisation_Hemispherique
Length:124	F:62	Min. :19.41	Min. :1176	Min. :26.00
Class :character	H:62	1st Qu.:22.69	1st Qu.:1308	1st Qu.:50.00
Mode :character		Median :25.35	Median :1384	Median :61.50
		Mean :27.01	Mean :1397	Mean :59.87
		3rd Qu.:29.98	3rd Qu.:1483	3rd Qu.:69.00
		Max. :46.10	Max. :1743	Max. :86.00
PROD_G_Frontal_Inf_Tri_1_L	PROD_G_Angular_2_L	PROD_G_Occipital_Lat_1_L	PROD_G_Rolandic_Oper_1_L	
Min. :0.0393	Min. :-1.12430	Min. :-0.74180	Min. :-1.30420	
1st Qu.:0.9300	1st Qu.: 0.06968	1st Qu.: -0.07745	1st Qu.: -0.40782	
Median :1.2381	Median : 0.40890	Median : 0.05330	Median : -0.22550	
Mean :1.2456	Mean : 0.39792	Mean : 0.07245	Mean : -0.25195	
3rd Qu.:1.5293	3rd Qu.: 0.68107	3rd Qu.: 0.20215	3rd Qu.: -0.02983	
Max. :2.4050	Max. : 1.60020	Max. : 0.93860	Max. : 0.61790	

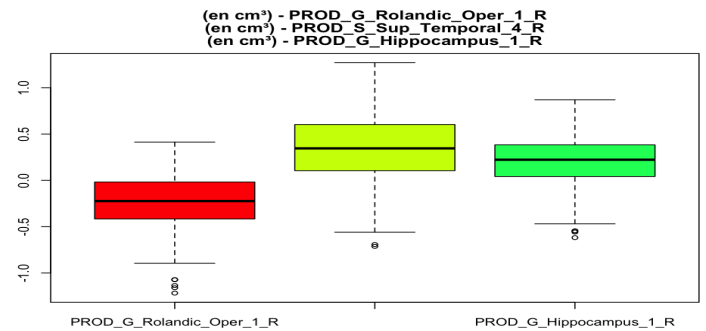
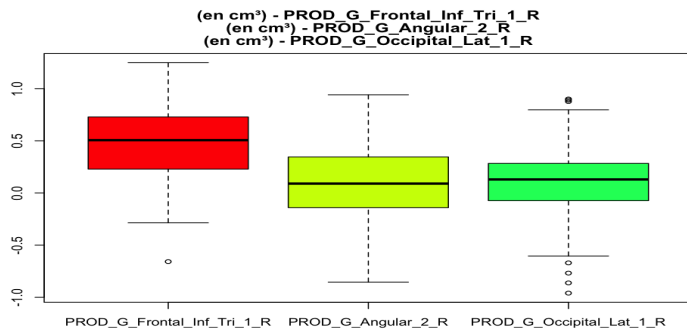
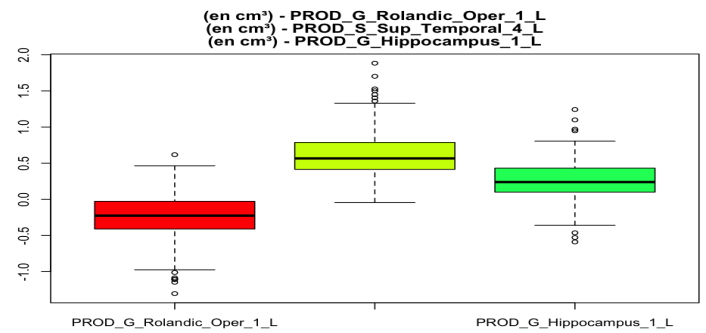
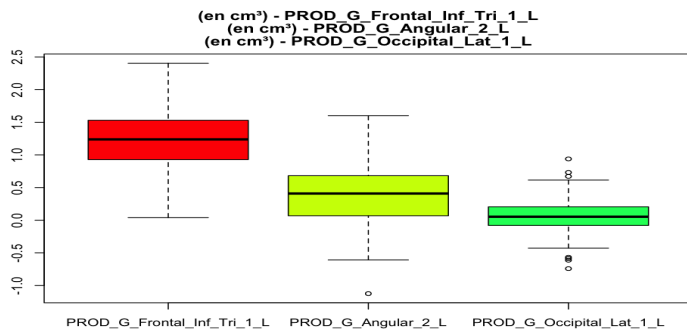
Notre jeu de données se compose de 124 lignes, soit 62 femmes et 62 hommes, parmi lesquels la moyenne d'âge est de 27, 01 ans. De plus, leur volume cérébral moyen est d'environ 1397 cm³, leur index de latéralisation hémisphérique moyen est d'environ 59.87 et leur gyrus frontal inférieur triangulaire moyen est de 1.2456 cm³.

PROD_S_Sup_Temporal_4_L	PROD_G_Hippocampus_1_L	PROD_G_Frontal_Inf_Tri_1_R	PROD_G_Angular_2_R
Min. :-0.0430	Min. :-0.5893	Min. :-0.6585	Min. :-0.85570
1st Qu.: 0.4151	1st Qu.: 0.1008	1st Qu.: 0.2314	1st Qu.: -0.13810
Median : 0.5663	Median : 0.2391	Median : 0.5068	Median : 0.09005
Mean : 0.6274	Mean : 0.2580	Mean : 0.4839	Mean : 0.08930
3rd Qu.: 0.7735	3rd Qu.: 0.4315	3rd Qu.: 0.7261	3rd Qu.: 0.34460
Max. : 1.8835	Max. : 1.2426	Max. : 1.2500	Max. : 0.94180
PROD_G_Occipital_Lat_1_R	PROD_G_Rolandic_Oper_1_R	PROD_S_Sup_Temporal_4_R	PROD_G_Hippocampus_1_R
Min. :-0.96080	Min. :-1.21760	Min. :-0.7129	Min. :-0.61840
1st Qu.: -0.07305	1st Qu.: -0.41173	1st Qu.: 0.1122	1st Qu.: 0.04358
Median : 0.12990	Median : -0.22355	Median : 0.3456	Median : 0.22310
Mean : 0.09878	Mean : -0.24737	Mean : 0.3213	Mean : 0.21511
3rd Qu.: 0.28230	3rd Qu.: -0.01932	3rd Qu.: 0.6005	3rd Qu.: 0.38112
Max. : 0.90180	Max. : 0.41310	Max. : 1.2709	Max. : 0.86990

D'autres variables sont également mieux visualisables si besoin avec leurs valeurs minimales, maximales, moyennes et quartiles dans la sortie R reçue à partir de notre code R, en annexe.

Précisons notre étude avec les boxplots des taux d'activation des 9 zones du cerveau étudiées pour le groupe général, afin de distinguer des zones plus stimulées éventuelles et surtout celle de l'aire de Broca :

Nous notons clairement que lors d'une tâche langagière, l'aire de Broca s'active plus que les autres aires (page ci-dessous).



1.3 Étude comparée selon le sexe (homme ou femme).

1.3.1 Groupe des hommes.

Les boîtes à moustaches des principales variables pour le groupe homme sont disponibles au besoin grâce au code R en annexe. (*variable donnees_homme et donnees_numeriques_homme*).

Au sein du groupe des 62 individus homme, l'âge moyen est de 27.75 ans, le volume cérébral moyen est d'environ 1485 cm³, l'index de latéralisation hémisphérique moyen est de 58.45 U.I et le taux d'activation moyen du gyrus frontal inférieur triangulaire est de 1.3528.

1.3.2 Groupe des femmes.

Au sein du groupe des 62 individus femme, l'âge moyen est de 26.26 ans, le volume cérébral moyen est d'environ 1308 cm³, l'index de latéralisation hémisphérique moyen est de 61.29 U.I et l'activation moyenne du gyrus frontal inférieur triangulaire est de 1.1385.

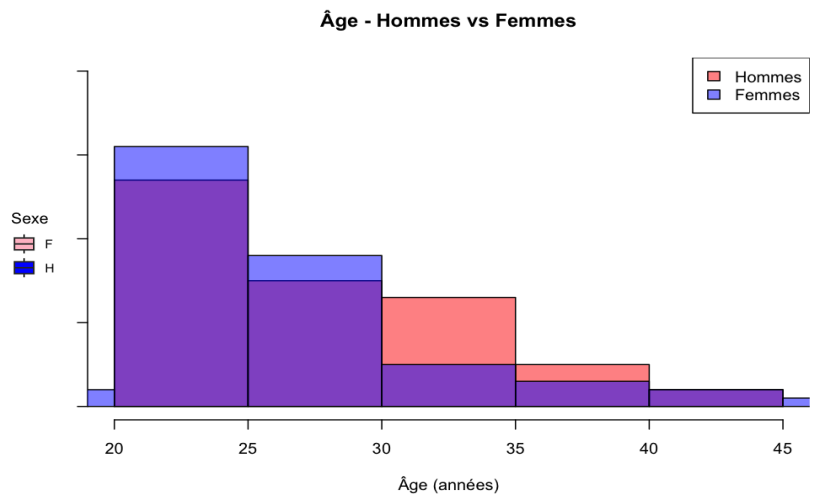
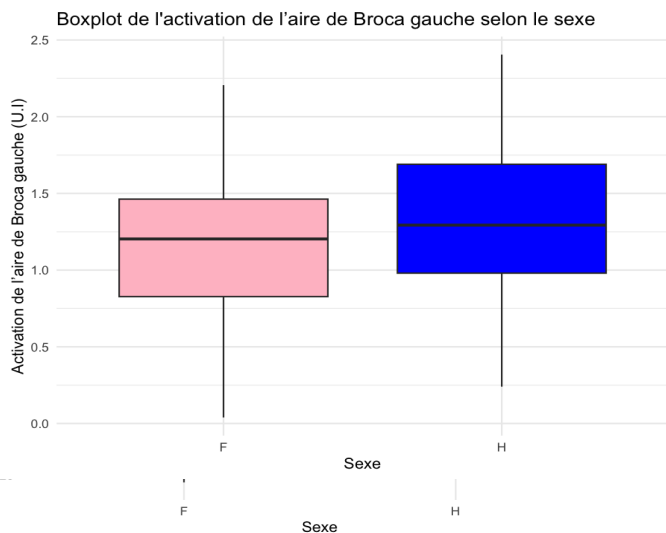
Les boîtes à moustaches des principales variables pour le groupe femme sont disponibles au besoin grâce au code R en annexe. (*variable donnees_femme et donnees_numeriques_femme*).

1.3.3 Comparaison groupe des femmes versus groupe des hommes.

Ensuite, nous avons réalisé pour les 4 variables principales caractérisant chaque individu de l'étude des histogrammes et des boxplots de comparaison (ci-dessous) en fonction du sexe de celui-ci. (*grâce aux commandes hist et boxplot*).

1.3.3.1 Comparaison selon l'âge.

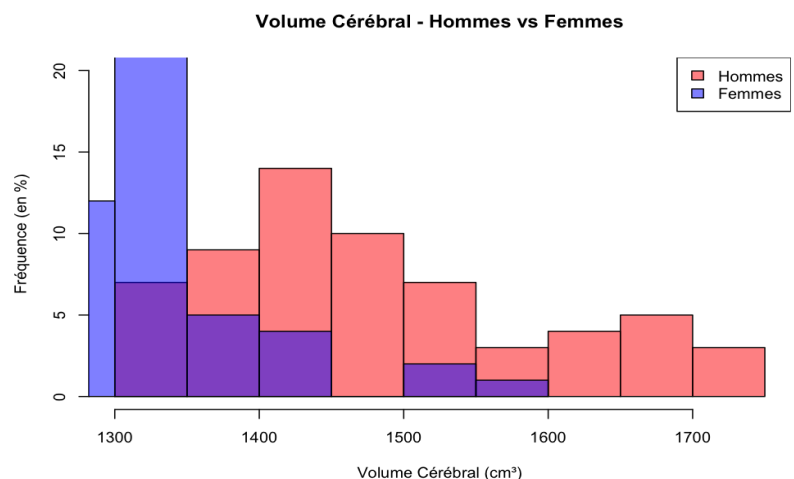
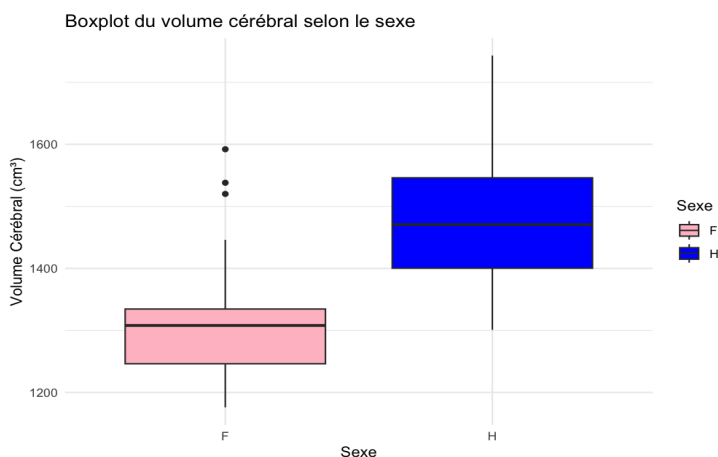
Nous pouvons observer que la population féminine présente généralement un âge légèrement inférieur à celle des hommes (colonnes violettes toujours en dessous de celle des hommes au-delà de 30 ans dans l'histogramme).



De plus, la dispersion des âges au sein des femmes (environ 6 ans) semble être moins importante que chez les hommes (plus proche des 10 années), notons donc la variabilité moindre de l'âge chez les femmes d'après le boxplot.

1.3.3.2 Comparaison selon le volume cérébral.

En ce qui concerne le volume cérébral, nous distinguons clairement qu'il est plus élevé chez les hommes, avec une moyenne considérablement supérieure. L'observation conjointe du boxplot et de l'histogramme indique qu'une partie réduite des femmes présente un volume cérébral comparable à celui des hommes.

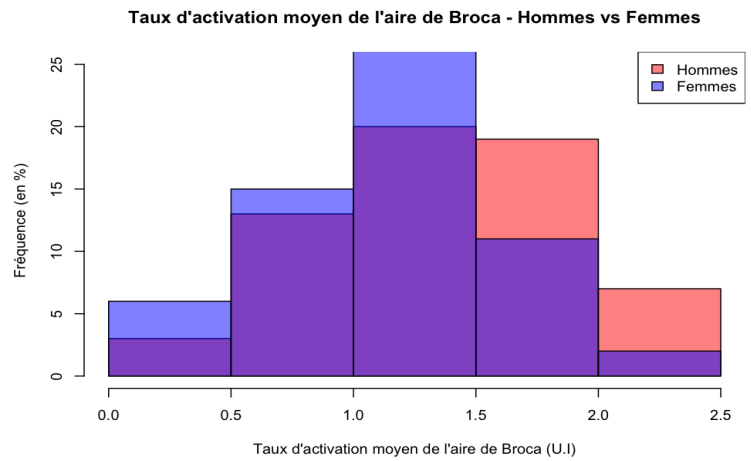
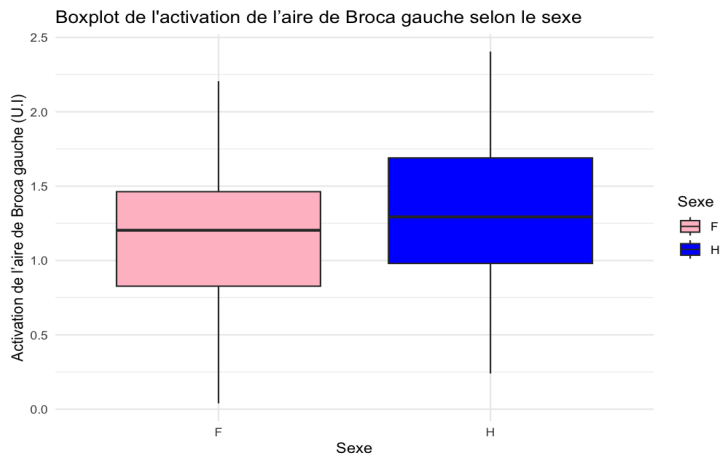


Par ailleurs, la distribution des volumes cérébraux chez les hommes est plus étendue, reflétant ainsi une plus grande variabilité du volume cérébral parmi les hommes (de 1300 à 1750 cm³) par rapport aux femmes (de 1275 à 1600 cm³) d'après le boxplot.

1.3.3.3 Comparaison selon le taux d'activation de l'aire de Broca.

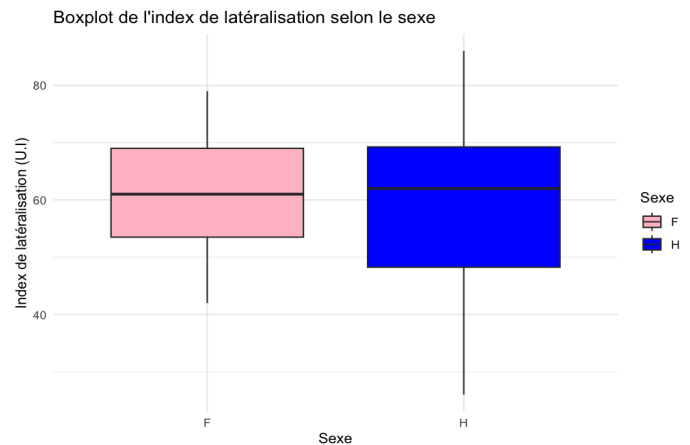
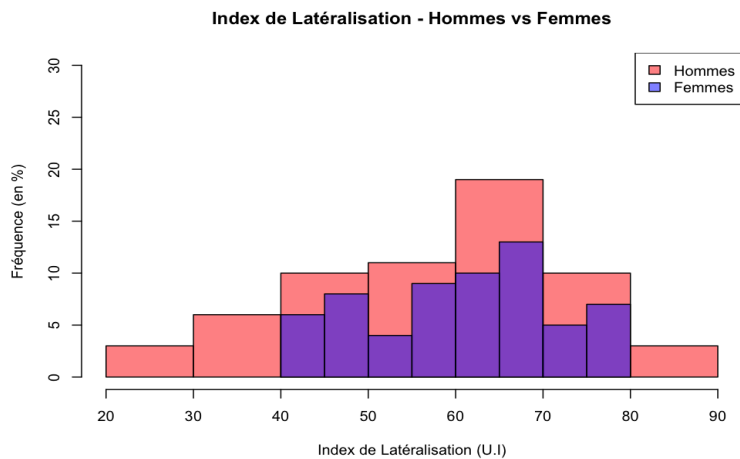
Quant à l'activité de l'aire de Broca gauche, les fluctuations semblent relativement similaires entre hommes et femmes, bien que l'amplitude moyenne puisse légèrement varier chez les hommes, dont nous pouvons noter que le taux d'activation moyen est en moyenne plus élevé.

Cependant, la dispersion des données est approximativement similaire (environ 0,6 U.I.) entre les deux groupes selon le boxplot (page ci-dessous).



1.3.3.4 Comparaison selon l'index de latéralisation.

Concernant l'index de latéralisation en fonction du sexe, les valeurs moyennes sont comparables entre hommes et femmes. Néanmoins, l'écart des valeurs chez les hommes est plus large, indiquant une plus grande variabilité. Les données chez les femmes montrent une certaine concentration autour de valeurs moyennes, suggérant une dispersion plus restreinte des valeurs d'index de latéralisation chez les femmes par rapport aux hommes.



1.3.3.4 Conclusion partielle de la comparaison selon l'analyse descriptive des 4 variables .

Globalement, dans notre échantillon, que ce soit chez les individus masculin ou féminin, l'hémisphère gauche semble être dominant, comme l'indiquent les valeurs positives de l'index de latéralisation pour l'ensemble des participants. Cependant, cette dominance semble être plus variable chez les hommes, avec une plus grande variabilité des valeurs d'index de latéralisation et la présence de valeurs extrêmes dans chacune des variables.

1.3.4 Un premier bilan.

L'analyse initiale des données révèle des similitudes marquées entre les données des hommes et des femmes, à l'exception notable du volume cérébral. Il apparaît clairement que la distinction la plus significative entre les sexes réside dans la supériorité moyenne du volume cérébral masculin .

Par conséquent, dans les analyses ultérieures, nous considérerons conjointement les données des deux sexes, hormis dans le cas d'analyses spécifiques centrées sur le volume cérébral.

Partie 2: Analyse en composantes principales (ACP).

L'analyse en composantes principales (ACP) est un outil puissant pour comprendre les relations et les liens entre les variables.

2.1. Détermination des valeurs propres pertinentes.

Lors de cette analyse, nous commencerons par explorer les caractéristiques des différentes dimensions afin de choisir celles sur lesquelles nous pourrions nous concentrer par la suite.

Dimensions	Eigenvalue (<i>valeurs propres</i>)	Proportion	Cumulative
dim 1	4.42	29.48	29.48
dim 2	02.07	13.82	43.30
dim 3	1.50	10.00	53.30
dim 4	1.36	09.04	62.34
dim 5	01.02	6.79	69.13
dim 6	0.98	6.56	75.70
dim 7	0.89	5.93	81.62
dim 8	0.77	5.12	86.74
dim 9	0.50	3.31	90.04
dim 10	0.41	2.74	92.79
dim 11	0.30	1.98	94.77
dim 12	0.24	1.61	96.38
dim 13	0.22	1.50	97.87
dim 14	0.18	1.17	99.05
dim 15	0.14	0.95	100.00

Classement des dimensions selon leur valeur propre et leur proportion de représentation (ci-contre).

(Nous obtenons les valeurs propres pour chaque dimension et son pourcentage de représentation grâce à la commande `round(resACP$eig, digits = 3)`.

Selon le critère de Kaiser, il est recommandé de retenir uniquement les dimensions dont les valeurs propres (*Eigenvalues*) sont supérieures à 1.

Dans notre cas, nous avons identifié 5 (voire 6) dimensions répondant à ce critère. Cependant, afin de simplifier notre analyse, nous nous concentrerons sur les 4 premières dimensions (*en rouge ci-contre*) et nous laisserons de côté les dimensions 5 et 6 (*en orange ci-contre*).

2.2. Étude des proportions de représentations pour chaque variable et du cercle de corrélations.

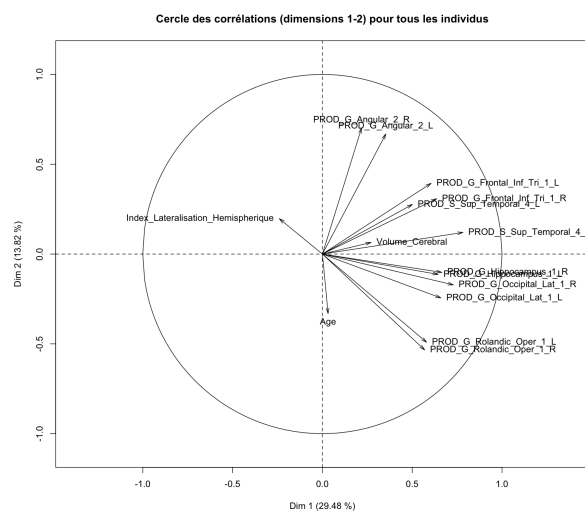
Ainsi, nous pouvons passer à une étude plus poussée en examinant les proportions de représentation pour chaque dimension. (grâce à la commande `round(resACPs$quanti$cos2, digits = 3)`).

Variable / Pourcentage de représentation par dimension	dim 1	dim 2	dim 3	dim 4	dim 5
Age	0.001	0.110	0.025	0.080	0.148
Volume_Cerebral	0.073	0.004	0.201	0.072	0.001
Index_Lateralisation_Hemispherique	0.058	0.039	0.282	0.004	0.285
PROD_G_Frontal_Inf_Tri_1_L	0.366	0.155	0.182	0.029	0.012
PROD_G_Angular_2_L	0.125	0.447	0.080	0.023	0.120
PROD_G_Occipital_Lat_1_L	0.435	0.060	0.074	0.101	0.013
PROD_G_Rolandic_Oper_1_L	0.336	0.241	0.033	0.052	0.028
PROD_S_Sup_Temporal_4_L	0.251	0.076	0.221	0.087	0.052
PROD_G_Hippocampus_1_L	0.415	0.013	0.032	0.330	0.071
PROD_G_Frontal_Inf_Tri_1_R	0.404	0.095	0.073	0.117	0.049
PROD_G_Angular_2_R	0.048	0.493	0.194	0.051	0.037
PROD_G_Occipital_Lat_1_R	0.531	0.029	0.002	0.048	0.007
PROD_G_Rolandic_Oper_1_R	0.324	0.285	0.000	0.039	0.055
PROD_S_Sup_Temporal_4_R	0.612	0.014	0.002	0.018	0.103
PROD_G_Hippocampus_1_R	0.442	0.010	0.098	0.305	0.038

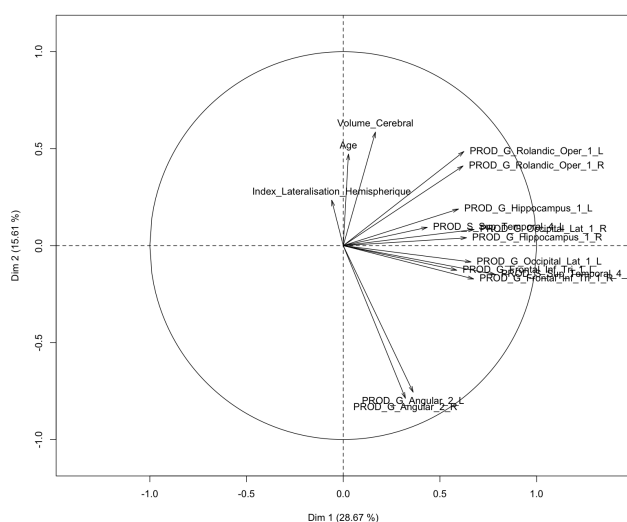
Tout d'abord, commençons par se concentrer sur les dimensions 1 et 2 (en bleu ci-dessus), nous constatons plusieurs points :

Étude pour tous les individus :

- L'activation de l'hippocampe est liée à celle de la partie occipitale.
- L'activation de la partie frontale semble être corrélée à celle de la partie temporale gauche.
- Pour chaque zone cérébrale, une activation dans la partie gauche est souvent associée à une activation similaire dans la partie droite (hormis légèrement les parties temporales R et L).
- L'âge, le volume cérébral et l'index de latéralisation hémisphérique ont une faible représentation dans ces dimensions.



Cercle des corrélations (dimensions 1-2) pour les femmes



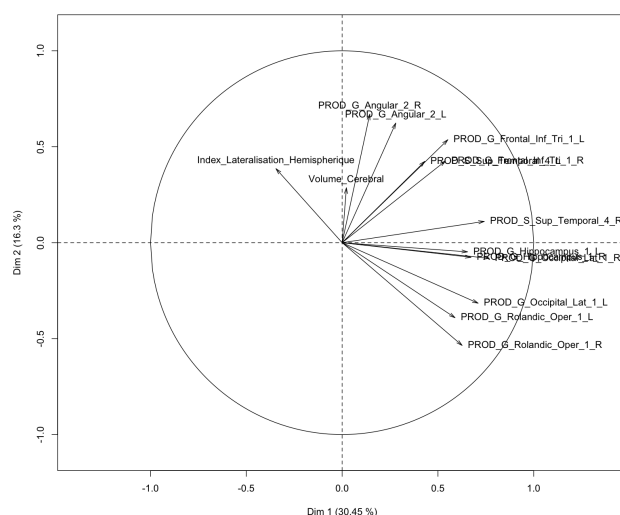
Étude pour tous les individus féminins :

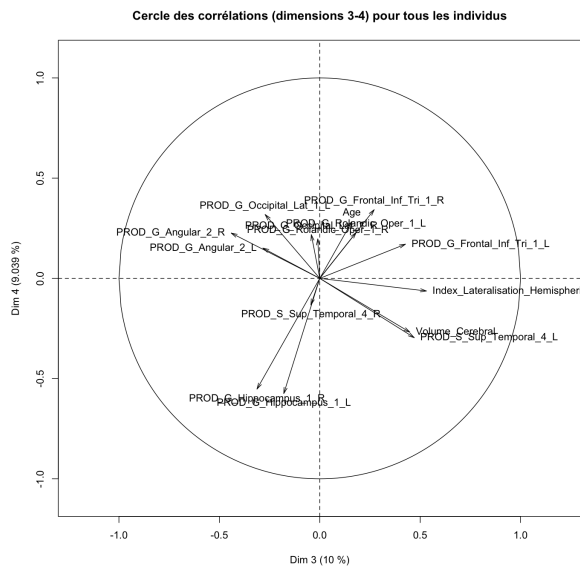
- L'activation de l'hippocampe est liée à celle de la partie occipitale R et temporale L.
- L'activation de la partie frontale semble être corrélée à celle de la partie temporale R et occipitale L.
- Pour les zones cérébrales de l'hippocampe, frontale, du gyrus angulaire de l'opercule rolandique, une activation dans la partie gauche est associée à une activation similaire dans la partie droite.
- L'âge, le volume cérébral et l'index de latéralisation hémisphérique ont toujours une faible représentation dans ces dimensions.

Étude pour tous les individus masculins :

- L'activation de l'hippocampe est liée à celle de la partie occipitale R et temporale L.
- L'activation de la partie frontale semble être corrélée à celle de la partie temporale R mais pas occipitale L contrairement aux femmes.
- Pour les zones cérébrales de l'hippocampe, frontale, du gyrus angulaire de l'opercule rolandique, une activation dans la partie gauche est toujours associée à une activation similaire dans la partie droite.
- L'âge, le volume cérébral et l'index de latéralisation hémisphérique ont toujours une faible représentation dans ces dimensions.

Cercle des corrélations (dimensions 1-2) pour les hommes

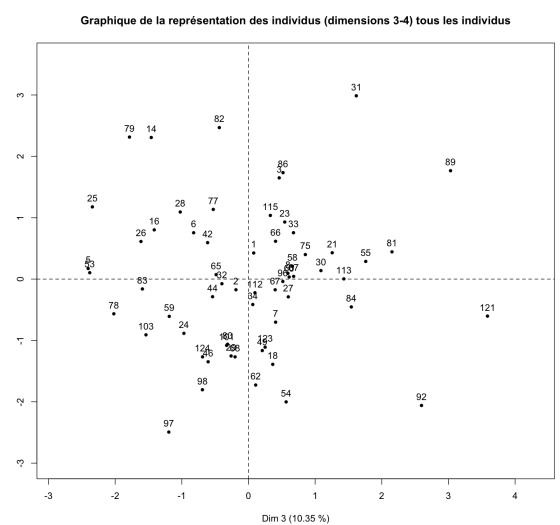
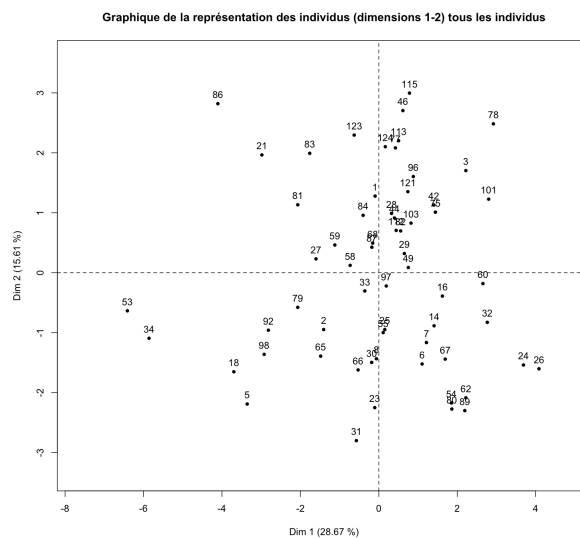




Étude pour tous les individus (dimensions 3-4):

- En revanche, pour les dimensions 3 et 4, nous observons une représentation très limitée des zones cérébrales (sensiblement inférieure à 0,2 voire très inférieure), à l'exception notable de l'hippocampe.
- Ainsi, ces deux dimensions ne semblent pas apporter d'informations pertinentes pour notre analyse centrée sur l'aire de Broca selon le cercle de corrélations des dimensions 3-4.

2.3. Étude de la représentation des individus (ci-contre)



De plus, lorsque nous examinons la représentation des individus dans ces 4 dimensions, nous constatons une répartition déséquilibrée et désorganisée (aucun groupe distinct), nous ne pouvons donc tirer plus d'informations de l'Analyse en Composantes Principales de fait.

Partie 3: Analyse de modélisation statistique

L'objectif de notre étude de modélisation statistique est de réaliser une régression linéaire multiple capable d'expliquer la variable `Prod_G_Frontal_Inf_Tri_1_L` en fonction des autres variables, soit une régression linéaire multiple du gyrus frontal inférieur triangulaire en fonction des autres variables.

3.1. Régression linéaire multiple de la variable `Prod_G_Frontal_Inf_Tri_1_L`.

- Les hypothèses ici testées par le test de Fisher (ou test de significativité du modèle) sont :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_{16} = 0$ ("le modèle est inutile") et $H_1 : \text{non } H_0$. ($\beta + \text{numéro de la variable}$).

```
Call:
lm(formula = donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Sexe +
    donnees$Age + donnees$Volume_Cerebral + donnees$Index_Lateralisation_Hemispherique +
    donnees$SPROD_G_Angular_2_L + donnees$SPROD_G_Occipital_Lat_1_L +
    donnees$SPROD_G_Rolandic_Oper_1_L + donnees$SPROD_S_Sup_Temporal_4_L +
    donnees$SPROD_G_Hippocampus_1_L + donnees$SPROD_G_Frontal_Inf_Tri_1_R +
    donnees$SPROD_G_Angular_2_R + donnees$SPROD_G_Occipital_Lat_1_R +
    donnees$SPROD_G_Rolandic_Oper_1_R + donnees$SPROD_S_Sup_Temporal_4_R +
    donnees$SPROD_G_Hippocampus_1_R)

Residuals:
    Min       1Q   Median       3Q      Max
-0.76962 -0.19228 -0.03095  0.18119  1.03232

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.4412311    0.4525361    0.975   0.3317
donnees$SexeH    0.0223463    0.0876527    0.255   0.7993
donnees$Age      0.0010733    0.0052474    0.205   0.8383
donnees$Volume_Cerebral -0.0001456    0.0003154   -0.462   0.6451
donnees$Index_Lateralisation_Hemispherique 0.0052752    0.0027181    1.941   0.0549
donnees$SPROD_G_Angular_2_L 0.2454403    0.0946899    2.592   0.0109
donnees$SPROD_G_Occipital_Lat_1_L -0.0403664    0.1660007   -0.243   0.8083
donnees$SPROD_G_Rolandic_Oper_1_L 0.0595780    0.1306876    0.456   0.6494
donnees$SPROD_S_Sup_Temporal_4_L 0.4467297    0.1094909    4.080  8.65e-05 ***
donnees$SPROD_G_Hippocampus_1_L 0.5152240    0.1671351    3.083   0.0026 ***
donnees$SPROD_G_Frontal_Inf_Tri_1_R 0.6079279    0.1123012    5.413  3.77e-07 ***
donnees$SPROD_G_Angular_2_R 0.0129407    0.1238976    0.104   0.9170
donnees$SPROD_G_Occipital_Lat_1_R 0.3058756    0.1435050    2.131   0.0353 *
donnees$SPROD_G_Rolandic_Oper_1_R 0.0096152    0.1447620    0.066   0.9472
donnees$SPROD_S_Sup_Temporal_4_R -0.2260151    0.1308668   -1.727   0.0870
donnees$SPROD_G_Hippocampus_1_R -0.4259909    0.1809285   -2.354   0.0204 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3274 on 108 degrees of freedom
Multiple R-squared:  0.6063,    Adjusted R-squared:  0.5516
F-statistic: 11.09 on 15 and 108 DF,  p-value: 8.636e-16
```

➤ Tout d’abord, la p-value du test est très proche de zéro ($p\text{-value} = 8.636e-16$, dernière ligne). Elle est donc très inférieure au risque de première espèce $\alpha = 5\%$, nous pouvons donc rejeter fortement l’hypothèse H_0 . Ainsi, le modèle est donc utile pour expliquer la variable Prod_G_Frontal_Inf_Tri_1_L.

➤ En outre, le R-squared nous permet de dire que notre modèle explique 60% ($Multiple\ R\text{-squared} = 0,6063$) des valeurs de Prod_G_Frontal_Inf_Tri_1_L. Cela veut dire qu’on a 40% de variabilité. D’autre part, l’erreur résiduelle montre que l’écart-type est acceptable, soit de 0.3274.

➤ Les étoiles disposées à droite des p-values autorisent à conclure que certaines variables décrivent mieux le modèle que d’autres comme Prod_S_Sup_Temporal_4_L (3 étoiles) et Prod_G_Hippocampus_1_L (2 étoiles).

Au contraire, des variables comme Prod_G_Angular_2_R et Prod_S_Sup_Temporal_4_R ne contribuent pas du tout au modèle. Nous pouvons constater que ce modèle est cohérent avec notre étude d’ACP, où nous avons vu que l’Hippocampe gauche et droit et l’aire de Wernicke étaient aussi associés avec l’aire de Broca sur notre cercle des corrélations (1-2).

3.2. Ajustement du modèle. (commande step).

Maintenant, nous pouvons passer à l’étape de réduction du modèle en supprimant les variables qui ne sont pas utiles pour notre modèle.

Ainsi, nous pouvons supprimer étape par étape une variable inutile en suivant le critère AIC de façon automatique (*grâce à la commande step*). L’objectif de ce *step* est de déterminer le meilleur modèle en accord entre un modèle qui correspond le mieux aux données (ayant une somme des carrés des résidus faibles) et un modèle qui soit le plus

```
> res_step <- step(res_donnees_numeriques)
Start: AIC=-262.01
donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Sexe + donnees$Age +
  donnees$Volume_Cerebral + donnees$Index_Lateralisation_Hemispherique +
  donnees$SPROD_G_Angular_2_L + donnees$SPROD_G_Occipital_Lat_1_L +
  donnees$SPROD_G_Rolandic_Oper_1_L + donnees$SPROD_S_Sup_Temporal_4_L +
  donnees$SPROD_G_Frontal_Inf_Tri_1_R +
  donnees$SPROD_G_Angular_2_R + donnees$SPROD_G_Occipital_Lat_1_R +
  donnees$SPROD_G_Rolandic_Oper_1_R + donnees$SPROD_S_Sup_Temporal_4_R +
  donnees$SPROD_G_Hippocampus_1_R

Step: AIC=-264
donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Sexe + donnees$Age +
  donnees$Volume_Cerebral + donnees$Index_Lateralisation_Hemispherique +
  donnees$SPROD_G_Angular_2_L + donnees$SPROD_G_Occipital_Lat_1_L +
  donnees$SPROD_G_Rolandic_Oper_1_L + donnees$SPROD_S_Sup_Temporal_4_L +
  donnees$SPROD_G_Frontal_Inf_Tri_1_R +
  donnees$SPROD_G_Angular_2_R + donnees$SPROD_G_Occipital_Lat_1_R +
  donnees$SPROD_S_Sup_Temporal_4_R + donnees$SPROD_G_Hippocampus_1_R

Step: AIC=-265.99
donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Sexe + donnees$Age +
  donnees$Volume_Cerebral + donnees$Index_Lateralisation_Hemispherique +
  donnees$SPROD_G_Angular_2_L + donnees$SPROD_G_Occipital_Lat_1_L +
  donnees$SPROD_G_Rolandic_Oper_1_L + donnees$SPROD_S_Sup_Temporal_4_L +
  donnees$SPROD_G_Hippocampus_1_L + donnees$SPROD_G_Frontal_Inf_Tri_1_R +
  donnees$SPROD_G_Occipital_Lat_1_R + donnees$SPROD_S_Sup_Temporal_4_R +
  donnees$SPROD_G_Hippocampus_1_R

- donnees$PROD_G_Rolandic_Oper_1_R      1  0.00047 11.580 -264.00
- donnees$PROD_G_Angular_2_R             1  0.00452 11.585 -265.95
- donnees$Age                           1  0.00617 11.586 -265.94
- donnees$Sexe                           1  0.00673 11.587 -265.93
- donnees$Volume_Cerebral                1  0.02291 11.603 -265.76
- donnees$PROD_G_Rolandic_Oper_1_L       1  0.04908 11.629 -265.48
<none>                                  11.580 -264.00
- donnees$PROD_S_Sup_Temporal_4_R        1  0.32071 11.901 -262.62
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf_Tri_1_R     1  3.15671 14.737 -236.11

- donnees$Age                           1  0.00431 11.984 -261.75
- donnees$Index_Lateralisation_Hemispherique 1  0.40431 11.984 -261.75
- donnees$PROD_G_Occipital_Lat_1_R       1  0.48945 12.069 -260.87
- donnees$PROD_G_Hippocampus_1_R         1  0.59679 12.177 -259.77
- donnees$PROD_G_Angular_2_L             1  0.72761 12.308 -258.45
- donnees$PROD_G_Hippocampus_1_L         1  1.02515 12.605 -255.49
- donnees$PROD_S_Sup_Temporal_4_L        1  1.80666 13.387 -248.03
- donnees$PROD_G_Frontal_Inf
```

simple (contenant le moins possible de variables explicatives). Il ne faudra ni trop supprimer ni trop conserver de variables explicatives. C'est le principe de sélection des variables pour un modèle avec la méthode basée sur le critère d'information d'Akaike (soit *AIC*, en effectuant une sélection de modèles pas à pas par élimination successive des variables les moins significatives pour minimiser l'*AIC*).

Lors de l'utilisation de la commande *step*, nous obtenons ainsi comme meilleur modèle de compromis pour expliquer la variable *Prod_G_Frontal_Inf_Tri_1_L* le modèle : (*image ci-dessous*)

Prod_G_Frontal_Inf_Tri_1_L ~ Index_Lateralisation_Hemispherique + PROD_G_Angular_2_L + PROD_S_Sup_Temporal_4_L + PROD_G_Hippocampus_1_L + PROD_G_Frontal_Inf_Tri_1_R + PROD_G_Occipital_Lat_1_R + PROD_S_Sup_Temporal_4_R + PROD_G_Hippocampus_1_R).

Il ne reste plus qu'à analyser ce modèle.

➤ Les hypothèses testées par le test de Fisher (soit test de significativité du modèle) sont :

$H_0 : \beta_4 = \beta_6 = \dots = \beta_{16} = 0$ ("le modèle est inutile") et H_1 : non H_0

À nouveau, le test de Fisher révèle une *p-value* très proche de zéro (*p-value* < $2.2e-16$), et elle est donc encore très inférieure au risque de première espèce $\alpha = 5\%$, nous pouvons donc rejeter fortement l'hypothèse H_0 ci-dessus. Par conséquent, le modèle est donc utile pour expliquer la variable *Prod_G_Frontal_Inf_Tri_1_L*.

```
Step: AIC=-273.66
donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Index_Lateralisation_Hemispherique +
donnees$PROD_G_Angular_2_L + donnees$PROD_G_Rolandic_Oper_1_L +
donnees$PROD_S_Sup_Temporal_4_L + donnees$PROD_G_Hippocampus_1_L +
donnees$PROD_G_Frontal_Inf_Tri_1_R + donnees$PROD_G_Occipital_Lat_1_R +
donnees$PROD_S_Sup_Temporal_4_R + donnees$PROD_G_Hippocampus_1_R

Step: AIC=-275.09
donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Index_Lateralisation_Hemispherique +
donnees$PROD_G_Angular_2_L + donnees$PROD_S_Sup_Temporal_4_L +
donnees$PROD_G_Hippocampus_1_L + donnees$PROD_G_Frontal_Inf_Tri_1_R +
donnees$PROD_G_Occipital_Lat_1_R + donnees$PROD_S_Sup_Temporal_4_R +
donnees$PROD_G_Hippocampus_1_R
```

	Df	Sum of Sq	RSS	AIC
- donnees\$PROD_G_Rolandic_Oper_1_L	1	0.0529	11.665	-275.09
<none>			11.613	-273.66
- donnees\$PROD_S_Sup_Temporal_4_R	1	0.3085	11.921	-272.40
- donnees\$Index_Lateralisation_Hemispherique	1	0.4378	12.050	-271.07
- donnees\$PROD_G_Hippocampus_1_R	1	0.6761	12.289	-268.64
- donnees\$PROD_G_Occipital_Lat_1_R	1	0.7506	12.363	-267.89
- donnees\$PROD_G_Hippocampus_1_L	1	1.0736	12.686	-264.69
- donnees\$PROD_G_Angular_2_L	1	1.3208	12.933	-262.30
- donnees\$PROD_S_Sup_Temporal_4_L	1	2.0368	13.649	-255.62
- donnees\$PROD_G_Frontal_Inf_Tri_1_R	1	3.3567	14.969	-244.17

```
> summary(res_step)

Call:
lm(formula = donnees$PROD_G_Frontal_Inf_Tri_1_L ~ donnees$Index_Lateralisation_Hemispherique +
  donnees$PROD_G_Angular_2_L + donnees$PROD_S_Sup_Temporal_4_L +
  donnees$PROD_G_Hippocampus_1_L + donnees$PROD_G_Frontal_Inf_Tri_1_R +
  donnees$PROD_G_Occipital_Lat_1_R + donnees$PROD_S_Sup_Temporal_4_R +
  donnees$PROD_G_Hippocampus_1_R)

Residuals:
    Min       1Q   Median       3Q      Max
-0.76348 -0.17608 -0.02683  0.19010  1.07483

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.256851   0.157736   1.628 0.106184
donnees$Index_Lateralisation_Hemispherique  0.005241   0.002518   2.081 0.039653 *
donnees$PROD_G_Angular_2_L                0.244979   0.069273   3.536 0.000586 ***
donnees$PROD_S_Sup_Temporal_4_L            0.448018   0.100054   4.478 1.79e-05 ***
donnees$PROD_G_Hippocampus_1_L             0.525300   0.157495   3.335 0.001147 **
donnees$PROD_G_Frontal_Inf_Tri_1_R         0.606265   0.102441   5.918 3.43e-08 ***
donnees$PROD_G_Occipital_Lat_1_R           0.297926   0.102682   2.901 0.004453 **
donnees$PROD_S_Sup_Temporal_4_R            -0.211052   0.125749  -1.678 0.095991 .
donnees$PROD_G_Hippocampus_1_R             -0.437309   0.169326  -2.583 0.011059 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3185 on 115 degrees of freedom
Multiple R-squared:  0.6034,    Adjusted R-squared:  0.5758
F-statistic: 21.87 on 8 and 115 DF, p-value: < 2.2e-16
```

Cependant, il persiste toujours la variabilité considérable sur ce modèle (*40 % environ*), nous devons la noter.

Au bilan de l'utilisation de la commande *step*, la variabilité de notre modèle est encore très proche de celle de notre modèle initial avec 15 variables, alors que celui-ci en a 8. De façon plus concrète, l'utilisation de la simplification par le critère AIC nous permet de déterminer un *meilleur compromis* de modèle entre simplification et ajustement, conservant cependant un pourcentage d'explications de la variable *Prod_G_Frontal_Inf_Tri_1_L* semblable au précédent.

De même, l'erreur résiduelle nous indique l'écart-type qui est de 0.3185, un écart-type faible et donc une preuve pour notre modèle final qu'il semble cohérent. Notons qu'il est très légèrement inférieur à celui du modèle initial, ce qui est plus favorable car cela caractérise le fait que le modèle final est légèrement moins bruité que le modèle initial.

Enfin, il est nécessaire de tester la normalité du modèle final pour confirmation, pour cela on utilise le test de Shapiro-Wilk (ou sa variante, le test de Bartlett d'homoscédasticité) :

```
> shapiro.test(res_donnees_numeriques$residuals) # test de normalité des résidus pour res_donnees_numeriques.
```

Shapiro-Wilk normality test

```
data: res_donnees_numeriques$residuals  
W = 0.98677, p-value = 0.2729
```

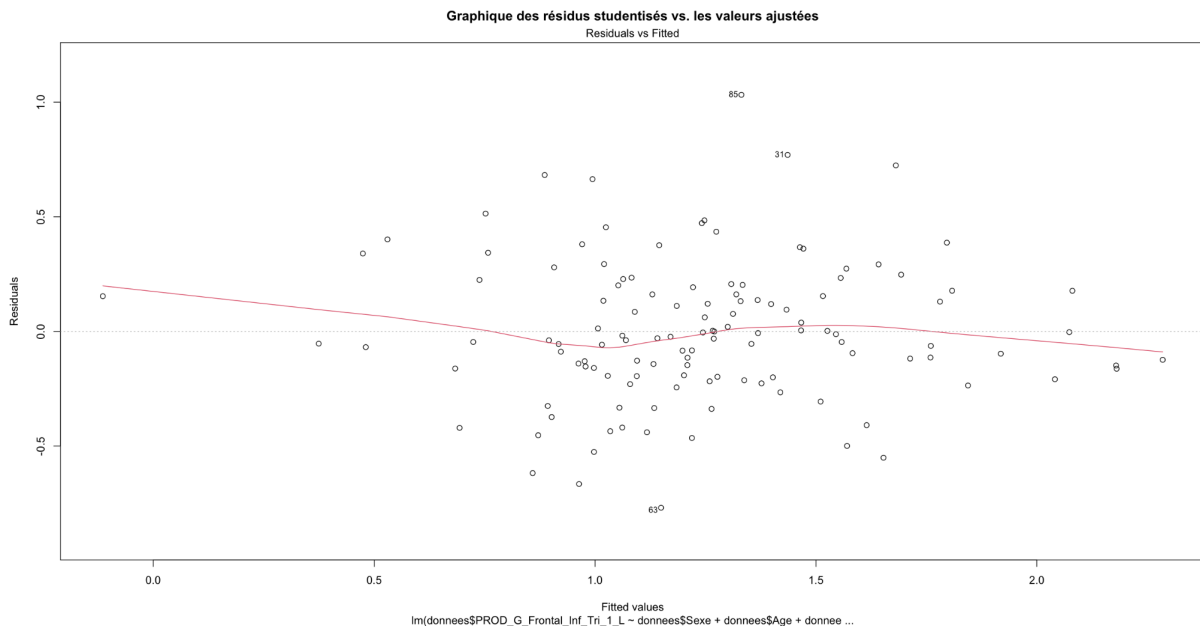
```
> shapiro.test(res_step$residuals) # test de normalité des résidus pour res_step.
```

Shapiro-Wilk normality test

```
data: res_step$residuals  
W = 0.98472, p-value = 0.1767
```

D'après le test de normalité des résidus (ou test de Shapiro-Wilk), nous pouvons valider l'hypothèse H_0 de normalité des résidus avec respectivement une p-valeur de 17,67% (27,29%) pour le modèle final (*initial*), largement supérieure à $\alpha = 5\%$.

3.3. Étude terminale du modèle déterminé. (*test de normalité et étude des résidus*).



De façon globale, le graphique des résidus ne relève aucune structure particulière. Nous observons bien le comportement aléatoire attendu (les résidus se traduisent comme les réalisations du terme d'erreur aléatoire ε du modèle).

En conclusion, nous pouvons dire qu'il n'y a donc pas d'élément supplémentaire dans les résidus qui aurait dû être pris en compte dans notre modèle final.

Conclusion

D'après notre étude de statistique inférentielle, nous pouvons conclure que la variabilité des activations de l'aire de Broca sont dans la plupart du temps pour des individus hommes et femmes liées aux 8 variables suivantes : L'index de latéralisation hémisphérique, le gyrus angulaire droit, le sillon supérieur temporal gauche, l'hippocampe gauche, le gyrus frontal inférieure triangulaire droit, le gyrus occipital latéral droit, le sillon supérieur temporal droit et l'hippocampe droit.

Lors d'une tâche de production, que ce soit l'hippocampe droit ou gauche et le sillon supérieur temporal, ils sont liés à l'activation de l'aire de Broca. En revanche, il subsiste toujours une variabilité non expliquée à notre modèle final : 40% des valeurs restent inexpliquées, nous pourrions supposer qu'ajouter dans notre modèle d'autres variables, notamment explicatives, serait une solution afin d'expliquer davantage la variable.

Finalement, dans la tâche de production d'une phrase en jeu dans cette étude, sont prises en compte les 8 variables évoquées précédemment. Le réseau neuronal est ainsi très complexe car même dans une tâche plutôt habituelle à première vue (production langagière), de nombreuses régions cérébrales interviennent.

Table des matières :

Introduction	2
Partie 1: Analyse descriptive préliminaire.	2
1.1 Description sommaire du jeu de données.	2
1.2 Groupe général. (variable donnees).	3
1.3 Étude comparée selon le sexe (homme ou femme).	4
1.3.1 Groupe des hommes.	4
1.3.2 Groupe des femmes.	4
1.3.3 Comparaison groupe des femmes versus groupe des hommes.	4
1.3.3.1 Comparaison selon l'âge.	4
1.3.4 Un premier bilan.	6
Partie 2: Analyse en composantes principales (ACP).	7
2.1. Détermination des valeurs propres pertinentes.	7
2.2. Étude des proportions de représentations pour chaque variable et du cercle de corrélations.	7
2.3. Étude de la représentation des individus (ci-contre)	9
Partie 3: Analyse de modélisation statistique	9
3.1. Régression linéaire multiple de la variable Prod_G_Frontal_Inf_Tri_1_L.	9
3.2. Ajustement du modèle. (commande step).	10
3.3. Étude terminale du modèle déterminé. (test de normalité et étude des résidus).	12
Conclusion	12
Annexe	13

Annexe

1. <https://www.cea.fr/presse/Pages/actualites-communiques/sante-sciences-du-vivant/preference-manuelle-langage-existe-t-il-hemisphere-dominant.aspx>¹, Mazoyer, B., zago, L., Jobard, G., Crivello, F., Joliot, M., Perchey, G., Mellet, E., Petit, L., et Tzourio-Mazoyer, N. (2014). Mazoyer, B., Zago, L., Jobard, G., Crivello, F., Joliot, M., Perchey, G., Mellet, E., Petit, L., & Tzourio-Mazoyer, N. (2014). Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PloS one*, 9(6), e101165. [Gaussian Mixture Modeling of Hemispheric Lateralization for Language in a Large Sample of Healthy Individuals Balanced for Handedness | PLOS ONE, https://doi.org/10.1371/journal.pone.0101165](https://doi.org/10.1371/journal.pone.0101165)¹
2. Lien du github public où se situe le code R utilisé : <https://github.com/juldebidour/DEBIDOUR-projet-stat-2A-2023-2024.code>