

Projet Econométrie linéaire

M2 TIDE

Laurène David - Safa Hamdan - Allan Pendant

24 Mars 2022

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 2 |
| 1.1 | Présentation des données | 2 |
| 1.1.1 | Analyses des données | 3 |
| 2 | Création de modèles et estimations | 6 |
| 2.1 | Analyse de la covariance (ANCOVA) | 6 |
| 2.1.1 | Analyse du modèle Additif | 10 |
| 2.1.2 | ANCOVA supplémentaires | 12 |
| 2.2 | Analyse de la variance (ANOVA): | 17 |
| 2.3 | MANOVA/MANCOVA | 20 |
| 3 | Conclusion | 22 |

1 Introduction

Nous allons travailler sur le jeu de données “Theory of mind in remitted bipolar disorder” étudié par Espinós U, Fernández-Abascal EG et Ovejero M en 2019.

```
bipolar <- read.csv('Theory of mind in remitted bipolar disorder.csv', sep = ";")
```

1.1 Présentation des données

Ce jeu de données comporte 227 Observations et 12 variables

Voici un aperçu des données :

| Group | Type | Age | Right_answers | Audio_prosody | Combined_channel | Face_video | Body_video |
|---------|-------|-----|---------------|---------------|------------------|------------|------------|
| Bipolar | BD I | 47 | 40 | 9 | 11 | 9 | 11 |
| Bipolar | BD I | 49 | 49 | 13 | 13 | 11 | 12 |
| Bipolar | BD I | 45 | 43 | 9 | 11 | 13 | 10 |
| Bipolar | BD I | 53 | 44 | 10 | 10 | 12 | 12 |
| Bipolar | BD II | 50 | 50 | 14 | 13 | 11 | 12 |
| Bipolar | BD I | 31 | 54 | 13 | 14 | 14 | 13 |

| Positive_valence | Negative_valence | Dominant | Submissive |
|------------------|------------------|----------|------------|
| 18 | 22 | 23 | 17 |
| 24 | 25 | 24 | 25 |
| 21 | 22 | 24 | 19 |
| 25 | 19 | 24 | 20 |
| 23 | 27 | 23 | 27 |
| 28 | 26 | 26 | 28 |

1.1.1 Analyses des données

Regardons ce que donne une analyse descriptive de ces données.

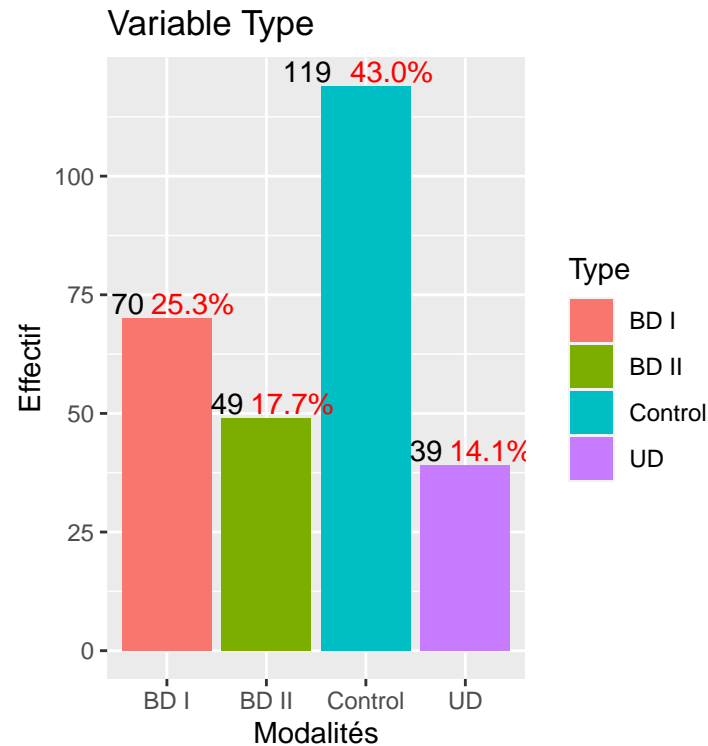
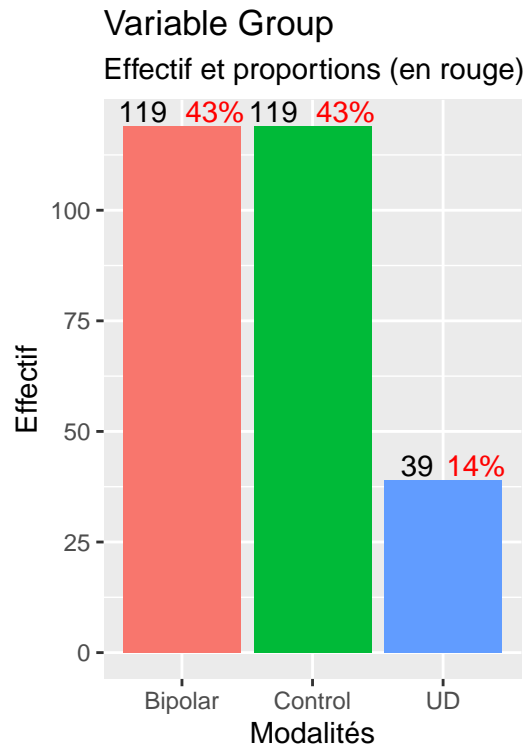
| | Age | Right_answers | Audio_prosody | Combined_channel | Face_video | Body_video |
|--------------|--------|---------------|---------------|------------------|------------|------------|
| nbr.val | 277.00 | 277.00 | 277.00 | 277.00 | 277.00 | 277.00 |
| nbr.null | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| nbr.na | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| min | 21.00 | 28.00 | 3.00 | 6.00 | 8.00 | 5.00 |
| max | 78.00 | 58.00 | 16.00 | 16.00 | 15.00 | 16.00 |
| range | 57.00 | 30.00 | 13.00 | 10.00 | 7.00 | 11.00 |
| median | 50.00 | 47.00 | 12.00 | 12.00 | 12.00 | 12.00 |
| mean | 48.72 | 47.08 | 11.43 | 12.34 | 11.79 | 11.52 |
| SE.mean | 0.75 | 0.31 | 0.13 | 0.12 | 0.10 | 0.11 |
| CI.mean.0.95 | 1.47 | 0.62 | 0.25 | 0.23 | 0.20 | 0.23 |
| var | 155.07 | 27.37 | 4.51 | 3.73 | 2.87 | 3.66 |
| std.dev | 12.45 | 5.23 | 2.12 | 1.93 | 1.69 | 1.91 |

| | Positive_valence | Negative_valence | Dominant | Submissive |
|--------------|------------------|------------------|----------|------------|
| nbr.val | 277.00 | 277.00 | 277.00 | 277.00 |
| nbr.null | 0.00 | 0.00 | 0.00 | 0.00 |
| nbr.na | 0.00 | 0.00 | 0.00 | 0.00 |
| min | 13.00 | 10.00 | 12.00 | 13.00 |
| max | 30.00 | 30.00 | 31.00 | 30.00 |
| range | 17.00 | 20.00 | 19.00 | 17.00 |
| median | 24.00 | 24.00 | 24.00 | 24.00 |
| mean | 23.56 | 23.52 | 23.45 | 23.62 |
| SE.mean | 0.19 | 0.19 | 0.19 | 0.18 |
| CI.mean.0.95 | 0.38 | 0.37 | 0.37 | 0.36 |
| var | 10.34 | 9.82 | 9.77 | 9.20 |
| std.dev | 3.22 | 3.13 | 3.13 | 3.03 |

| Group | Type |
|------------------|------------------|
| Length:277 | Length:277 |
| Class :character | Class :character |
| Mode :character | Mode :character |

On Distingue 2 variables qualitatives nominales :

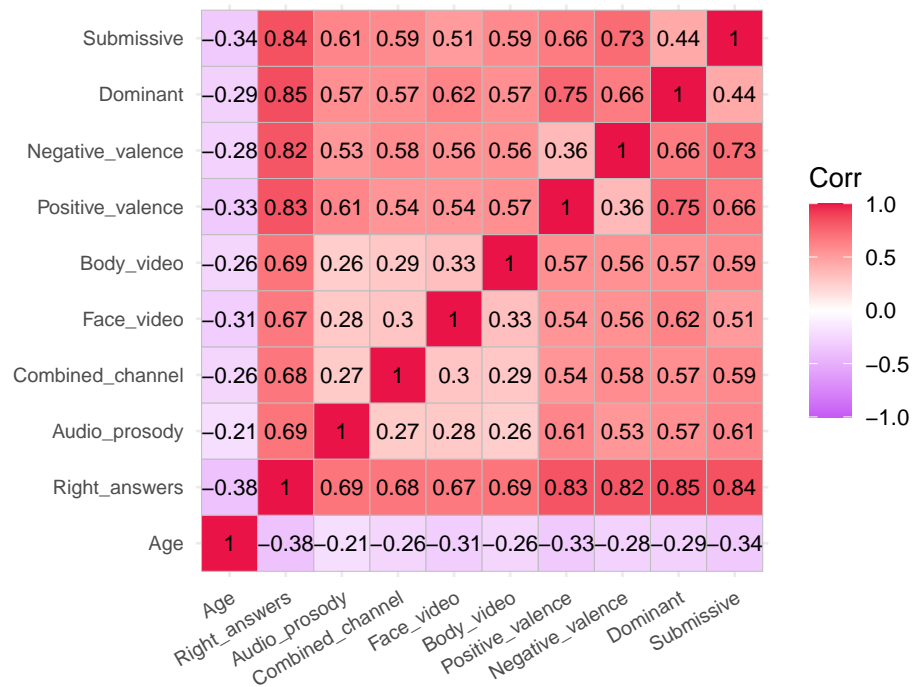
- **Group** ayant pour modalités Bipolar (individus bipolaires) de taille 119, Control (individus sains) aussi de taille 119, UD (individus atteints de dépression unipolaire) de taille 39.
- **Type** avec les modalités BD I (Bipolaire type 1), BD II (bipolaire type 2), control et UD. Nous avons 70 personnes ayant un trouble de la bipolarité de type 1, 49 personnes ayant un trouble de la bipolarité de type 2, 119 personnes “Controlée” et 39 personnes ayant une dépression unipolaire. On observe aussi qu’aucunes des variables n’a de valeurs manquantes

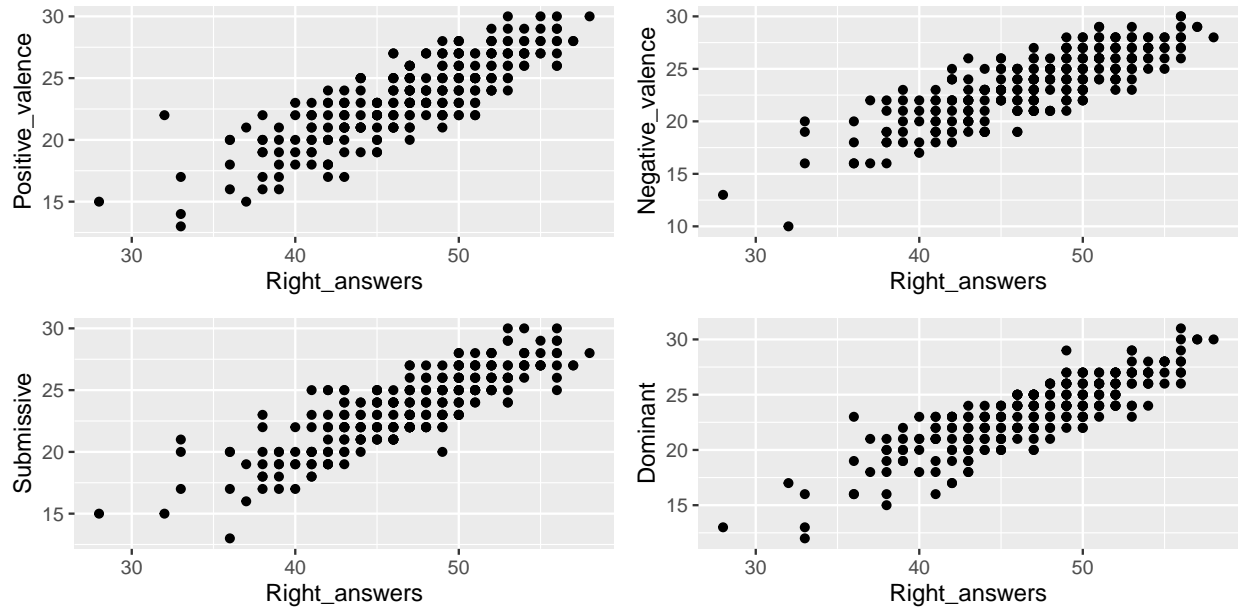


Nous avons aussi 10 variables quantitatives : *Age*, *Right_answers*, *Audio_prosody*

Combined_channel, *Face_video*, *Body_video*, *Positive_valence*, *Dominant* et *Submissive*. Ces variables ne prennent que des valeurs positives et entières.

Étudions la corrélation entre ces variables :



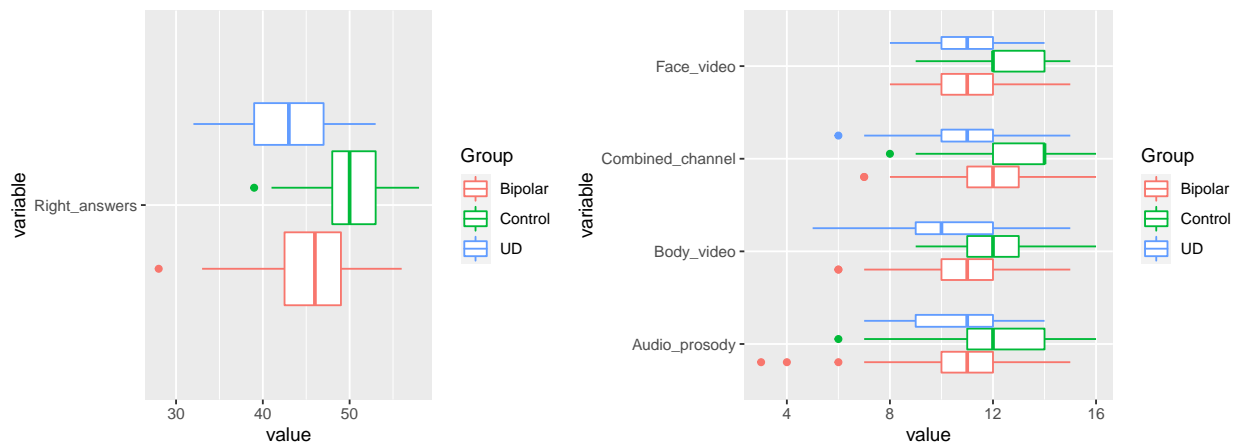


De cette matrice de corrélation nous pouvons voir que si nous posons un seuil à 80%, la variable **Right_answers** est fortement corrélée aux variables **Positive_valence**, **Negative_valence** et **Submissive** ce qui se confirme aussi à travers les nuages de points. Nous pouvons donc garder uniquement **Right_answers**, **Audio_prosody**, **Combined_channel**, **Face_video**, **Body_video**, **Age** et écarter les autres.

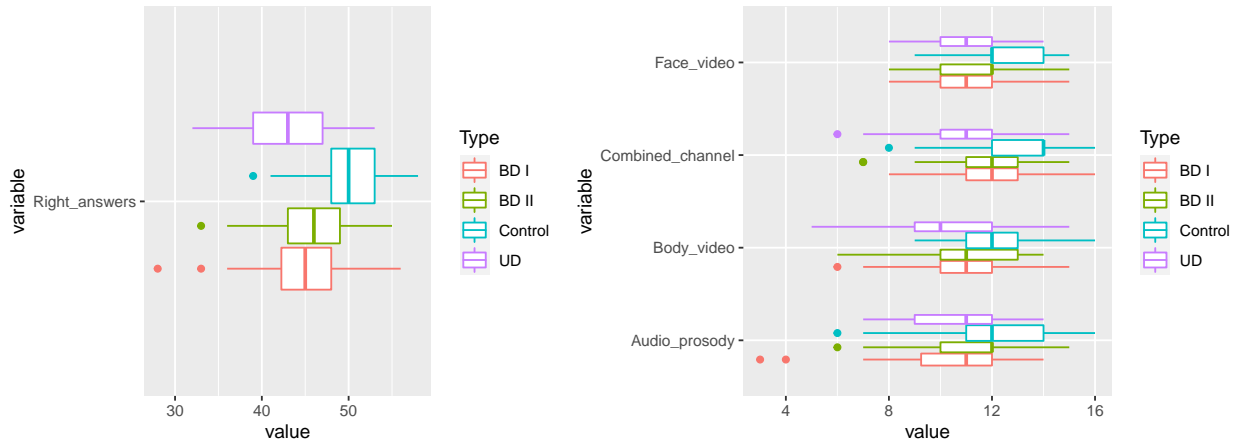
Des pistes d'études seraient de voir si il y a des différences significative ou non entre les différents niveaux des deux variables qualitatives sur les autres variables quantitatives. Ceci nous amènerait à appliquer des modèles d'ANOVA, d'ANCOVA, ou MANOVA.

Pour nous donner un aperçu de l'étude, utilisons des boxplots sur la variable **Group** puis sur la variable **Type**:

Boxplots sur la variable Group



Boxplots sur la variable Type



On observe plusieurs choses intéressantes dans ces boxplots :

- De possibles valeurs aberrantes dans la variable **Audio_prosody**
- Nous remarquons que de manière générale, le groupe **UD** semble avoir moins de bonnes réponses que les autres groupes, tandis que le groupe **Control** a plus de bonnes réponses pour les tests **Combined channel**, **Body video** et **Right answers**.
- Nous voyons également que, les groupes '**BD II**' et '**Control**' ont des moyennes égales pour les tests **Audio prosody** et **Face video**
- Pour les tests **Right answers**, **Body video** et **Combined channel** les groupes **BD I** et **BD II** ont les mêmes moyennes mais pas les mêmes distributions des "bonnes réponses". Cela pourrait notamment venir du déséquilibre des deux populations (70 individus BD I, 49 individus BD II).

Mais une chose intéressante dans ces observations c'est que l'on peut émettre certaines hypothèses sur les possibles similitudes entre les groupes. Faisons une étude plus poussée pour vérifier cela.

2 Création de modèles et estimations

2.1 Analyse de la covariance (ANCOVA)

Nous partons du principe qu'il serait intéressant d'étudier en priorité les niveaux de la variable **Type** avec le nombre de bonnes réponses obtenues (**Right_answers**)

Pour cela nous pouvons effectuer une ANCOVA muni de covariables sous réserve de validation des hypothèses

Ou de faire une ANOVA à un facteur (facteur **Group** ou facteur **Type**).

Ici nous allons nous intéresser à la mise en place d'une ANCOVA si possible.

Voici la liste des hypothèses à vérifier :

- Linéarité du modèle
- La normalité des résidus
- l'homoscédasticité des résidus

Lançons d'abord un modèle additif avec comme variables à expliquer Group et toutes les autres variables quantitatives restantes :

| | Estimate | Std. Error | t value | Pr(> t) |
|------------------|------------|------------|-----------|----------|
| (Intercept) | -1.687e-13 | 6.682e-14 | -2.525 | 0.01216 |
| TypeBD II | 1.164e-14 | 1.607e-14 | 0.7248 | 0.4692 |
| TypeControl | 7.863e-15 | 1.43e-14 | 0.5498 | 0.5829 |
| TypeUD | 1.934e-14 | 1.911e-14 | 1.012 | 0.3122 |
| Audio_prosody | 1 | 2.643e-15 | 3.783e+14 | 0 |
| Combined_channel | 1 | 2.985e-15 | 3.35e+14 | 0 |
| Face_video | 1 | 3.469e-15 | 2.882e+14 | 0 |
| Body_video | 1 | 2.999e-15 | 3.335e+14 | 0 |
| Age | 1.257e-16 | 4.91e-16 | 0.2559 | 0.7982 |

Table 7: Fitting linear model: $\text{Right_answers} \sim \text{Type} + \text{Audio_prosody} + \text{Combined_channel} + \text{Face_video} + \text{Body_video} + \text{Age}$

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|-------|----------------|
| 277 | 8.44e-14 | 1 | 1 |

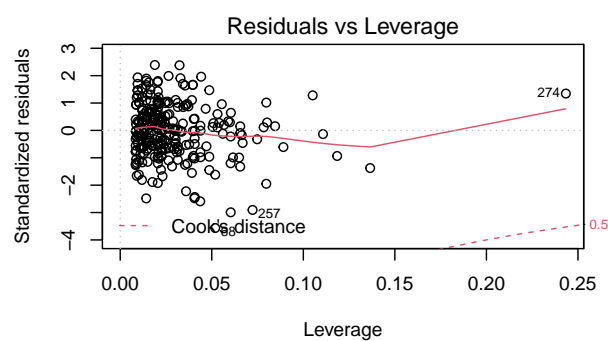
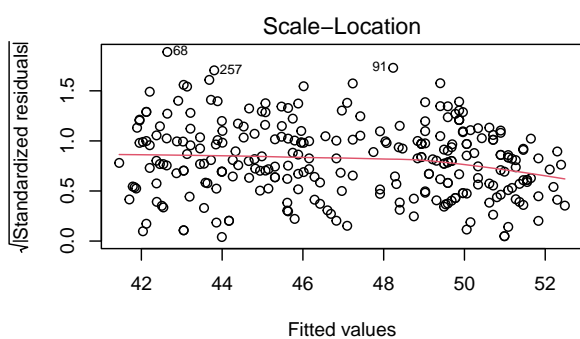
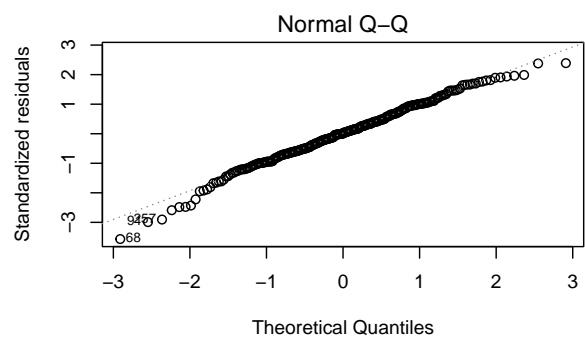
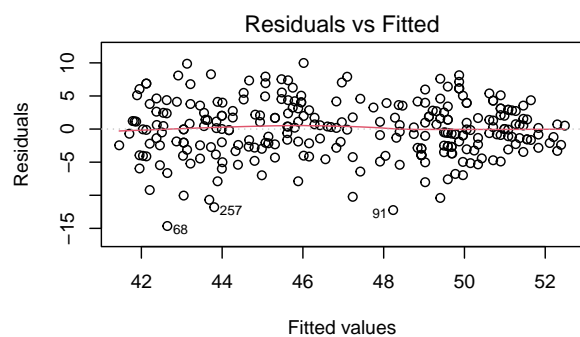
Remarque Importante :

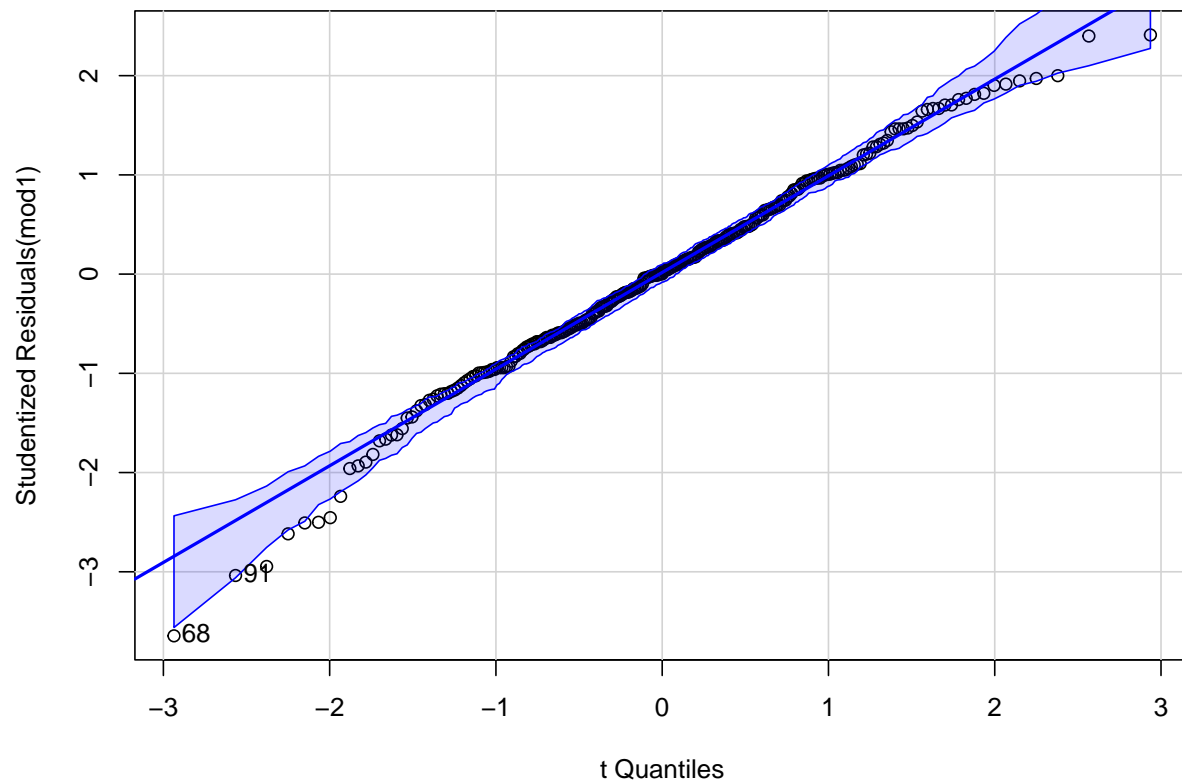
L'encodage du facteur Type se fait automatiquement, et ici la modalité de référence est BD I . c'est pourquoi elle n'apparaît pas explicitement.

Nous observons que l'on obtient un R^2 à 1 ainsi qu'un niveau d'erreur à 0 ce qui veut dire que 100% de la variance de **Right_answers** est expliquée par ce modèle ce qui est cohérent puisque effectivement tout les test réalisés pour le MinPONS expliquent la variable **Right_answers**. Il est donc inutile d'utiliser ce modèle.

Gardons **Right_answers** comme variable à expliquer, et comme facteur nous utilisons toujours **Type** et comme covariable **Age** : Il serait intéressant maintenant de voir si il y une différence significative entre les différentes modalités de **Type** sur le nombre de réponses obtenues en fonction de l'Age. Mettons en place le modèle complet (avec interactions) et vérifions en premier lieu les hypothèses citées plus haut:

```
mod1 <- lm(Right_answers~Type*Age, data=bipolar)
par(mfrow=c(2,2))
plot(mod1)
```





```
## [1] 68 91
```

Residuals Vs Fitted :

Nous pouvons confirmer que l'on ne retrouve aucune structure dans les résidus, le caractère linéaire des données est vérifié

Scale-Location (Homoscédasticité) :

Ici nous ne remarquons pas de forme particulière si ce n'est une très légère tendance décroissante au niveau des valeurs hautes, mais cela reste tout de même satisfaisant

Normal QQplot (Normalité des résidus) :

Les points épousent suffisamment bien la première bissectrice, nous pouvons valider la normalité.

L'ensemble des hypothèses ont été validées, maintenant regardons quels termes sont significatifs.

On étudie ainsi l'effet du facteur **Type** et de la covariable **Age** sur la variable réponse **Right_answers**:

```
pander(anova(mod1))
```

Table 8: Analysis of Variance Table

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------------|-----|--------|---------|---------|-----------|
| Type | 3 | 2289 | 763.1 | 42.89 | 1.106e-22 |
| Age | 1 | 451 | 451 | 25.35 | 8.78e-07 |
| Type:Age | 3 | 27.59 | 9.197 | 0.5169 | 0.671 |
| Residuals | 269 | 4786 | 17.79 | NA | NA |

Ici nous voyons que l'interaction de **Type** avec **Age** (Type:Age) est non significative pour $\alpha=0.05$ (p-value = 0.6513), ce qui peut nous amener à penser que le modèle additif avec simplement les effets principaux serait plus adapté.

vérifions cela :

```
mod1_ad <- lm(Right_answers~Type + Age, data=bipolar)
pander(BIC(mod1,mod1_ad))
```

| | df | BIC |
|----------------|----|------|
| mod1 | 9 | 1626 |
| mod1_ad | 6 | 1611 |

```
pander(anova(mod1, mod1_ad))
```

Table 10: Analysis of Variance Table

| Res.Df | RSS | Df | Sum of Sq | F | Pr(>F) |
|--------|------|----|-----------|--------|--------|
| 269 | 4786 | NA | NA | NA | NA |
| 272 | 4814 | -3 | -27.59 | 0.5169 | 0.671 |

En utilisant le critère BIC, nous observons une plus faible valeur pour le modèle additif sans interactions (BIC mod1_ad : 1611) que pour le modèle complet avec interactions (BIC mod1 : 1626), ce qui nous conforte dans l'idée d'utiliser le modèle sans interaction. De plus, on peut voir que la somme des carrés résiduels (RSS) pour le modèle complet est légèrement plus petite que pour le modèle sans interaction ($4786 < 4814$) ce pendant avec une p-value = 0.671 ce qui n'est pas significatif pour préférer le modèle complet.

Nous utilisons donc le modèle additif sans interactions

2.1.1 Analyse du modèle Additif

```
pander(summary(mod1_ad))
```

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|----------|------------|---------|------------|
| (Intercept) | 50.36 | 1.15 | 43.81 | 2.996e-125 |
| TypeBD II | 1.143 | 0.7936 | 1.44 | 0.151 |
| TypeControl | 5.279 | 0.6348 | 8.316 | 4.4e-15 |

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------|----------|------------|---------|-----------|
| TypeUD | -0.2622 | 0.9435 | -0.2779 | 0.7813 |
| Age | -0.1173 | 0.02324 | -5.048 | 8.181e-07 |

Table 12: Fitting linear model: Right_answers ~ Type + Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|--------|----------------|
| 277 | 4.207 | 0.3627 | 0.3534 |

Avec toujours la modalité BD I comme référence, nous remarquons que **36% de la variance de Right_answers est expliquée par ce modèle**, le test global de fisher est significatif avec une p-value de 2.2e-16.

Interprétations des coefficients :

- De manière significative (p-value<2e-125), nous pouvons dire qu'un individu de la modalité BD I aura en moyenne 50 bonnes réponses (Intercept).
- La différence entre le type BD I et BD II n'est pas significative (p-value=0.151)
- Un individu faisant partie de la modalité Control aura quant à elle en moyenne 5 bonnes réponses en plus qu'un individu de type BD I.
- Nous ne pouvons pas dire de manière significative qu'un individu dépressif (UD) ait en moyenne une mauvaise réponse en moins qu'un individu de type BD I.
- De manière significative, l'Age des individus ne semble pas influencer sur le nombre de bonnes réponses obtenues (coefficient à -0.11).

Pour avoir plus d'information nous pouvons changer la référence pour voir ce que donnerait ces résultats mais du point de vue du Type BD II :

```
bipolar <- bipolar %>% mutate(Type2 = relevel(factor(Type), ref="BD II"))
pander(summary(lm(Right_answers~Type2 + Age, data=bipolar)))
```

| | Estimate | Std. Error | t value | Pr(> t) |
|---------------------|----------|------------|---------|------------|
| (Intercept) | 51.5 | 1.306 | 39.45 | 1.521e-114 |
| Type2BD I | -1.143 | 0.7936 | -1.44 | 0.151 |
| Type2Control | 4.136 | 0.7195 | 5.749 | 2.412e-08 |
| Type2UD | -1.405 | 0.9523 | -1.475 | 0.1413 |
| Age | -0.1173 | 0.02324 | -5.048 | 8.181e-07 |

Table 14: Fitting linear model: Right_answers ~ Type2 + Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|--------|----------------|
| 277 | 4.207 | 0.3627 | 0.3534 |

- En moyenne les type BD II ont 51 bonnes réponses
- Les individus de controle ont significativement 4 bonnes réponses en plus que les types BD II en moyenne.

2.1.2 ANCOVA supplémentaires

Effectuons les mêmes démarches vu précédemment mais en changeant la variable réponse ***Right_answers***

Faisons une ANCOVA sur toutes nos variables de testes :

Face_video

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|----------|------------|---------|-----------|
| (Intercept) | 12.95 | 0.4151 | 31.19 | 8.072e-92 |
| TypeBD II | 0.3557 | 0.2865 | 1.241 | 0.2156 |
| TypeControl | 1.333 | 0.2292 | 5.814 | 1.704e-08 |
| TypeUD | 0.4426 | 0.3407 | 1.299 | 0.195 |
| Age | -0.03795 | 0.00839 | -4.524 | 9.091e-06 |

Table 16: Fitting linear model: Face_video ~ Type + Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|--------|----------------|
| 277 | 1.519 | 0.2068 | 0.1951 |

Audio_prosody

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|----------|------------|---------|-----------|
| (Intercept) | 11.97 | 0.541 | 22.12 | 9.783e-63 |
| TypeBD II | 0.6753 | 0.3735 | 1.808 | 0.07171 |
| TypeControl | 1.513 | 0.2988 | 5.064 | 7.573e-07 |
| TypeUD | 0.1036 | 0.4441 | 0.2332 | 0.8158 |
| Age | -0.02721 | 0.01094 | -2.488 | 0.01344 |

Table 18: Fitting linear model: Audio_prosody ~ Type + Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|--------|----------------|
| 277 | 1.98 | 0.1425 | 0.1299 |

Combined_channel

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|----------|------------|---------|-----------|
| (Intercept) | 13.09 | 0.4796 | 27.3 | 6.921e-80 |
| TypeBD II | -0.1197 | 0.3311 | -0.3617 | 0.7179 |
| TypeControl | 1.171 | 0.2648 | 4.422 | 1.413e-05 |
| TypeUD | -0.6079 | 0.3936 | -1.544 | 0.1237 |
| Age | -0.02359 | 0.009694 | -2.434 | 0.0156 |

Table 20: Fitting linear model: Combined_channel ~ Type + Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|--------|----------------|
| 277 | 1.755 | 0.1866 | 0.1746 |

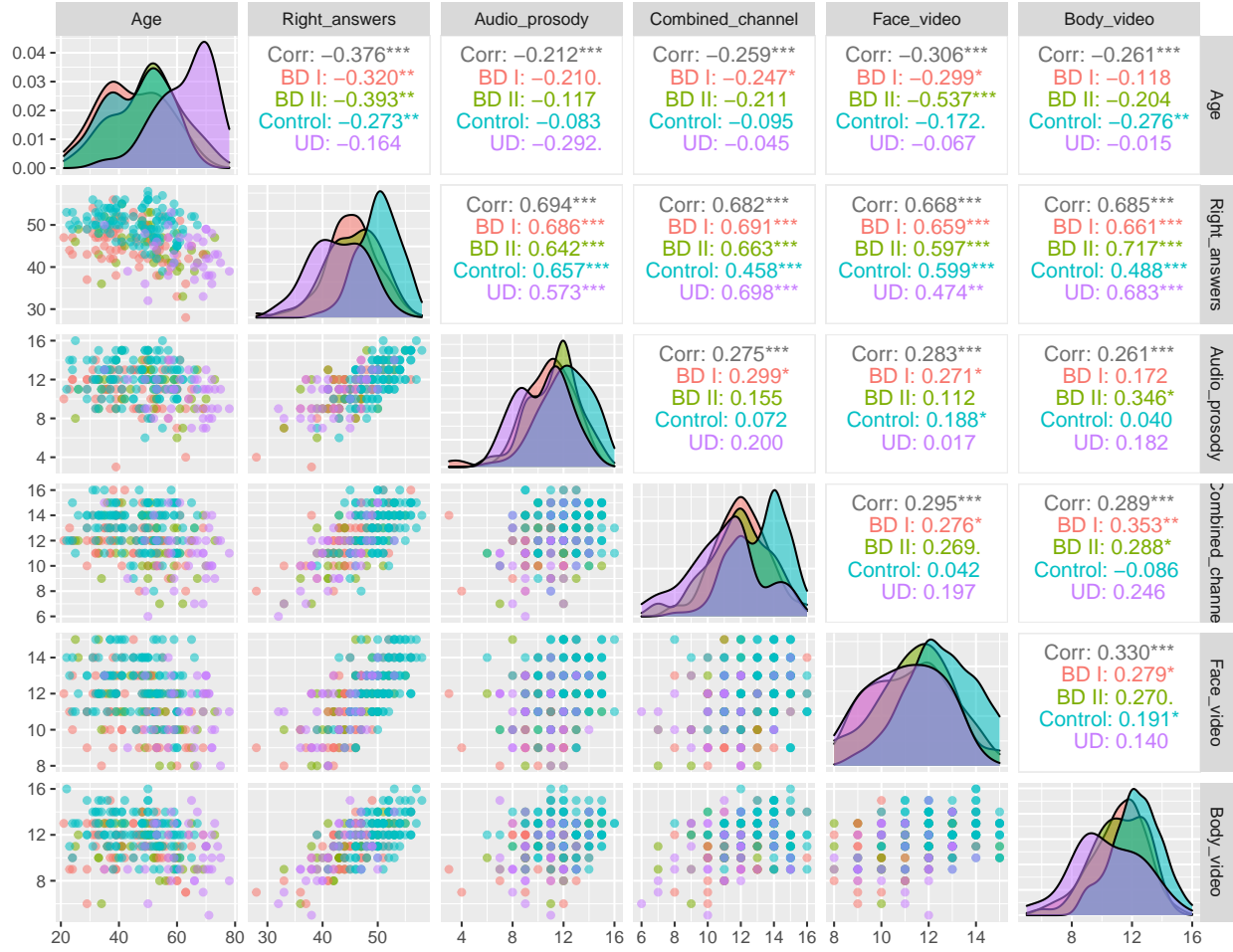
Body_video

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------|----------|------------|---------|-----------|
| (Intercept) | 12.36 | 0.4795 | 25.77 | 5.823e-75 |
| TypeBD II | 0.2315 | 0.331 | 0.6992 | 0.485 |
| TypeControl | 1.262 | 0.2648 | 4.767 | 3.048e-06 |
| TypeUD | -0.2005 | 0.3935 | -0.5094 | 0.6109 |
| Age | -0.02854 | 0.009692 | -2.945 | 0.00351 |

Table 22: Fitting linear model: Body_video ~ Type + Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|-------|----------------|
| 277 | 1.755 | 0.17 | 0.1578 |

Cependant, une chose à vérifier pour pouvoir réellement valider toutes ces Analyses de covariances, c'est la linéarité de la covariable avec la variable réponse. Traçons des nuages de points :



On observe que l'interaction entre 'Age' et 'Right_answers' et toutes les autres variables ne semblent pas linéaires. Si nous appliquons une régression linéaire simple sur ces deux variables, nous obtenons ceci :

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|------------|
| (Intercept) | 54.78 | 1.18 | 46.41 | 4.268e-132 |
| Age | -0.158 | 0.02348 | -6.729 | 9.895e-11 |

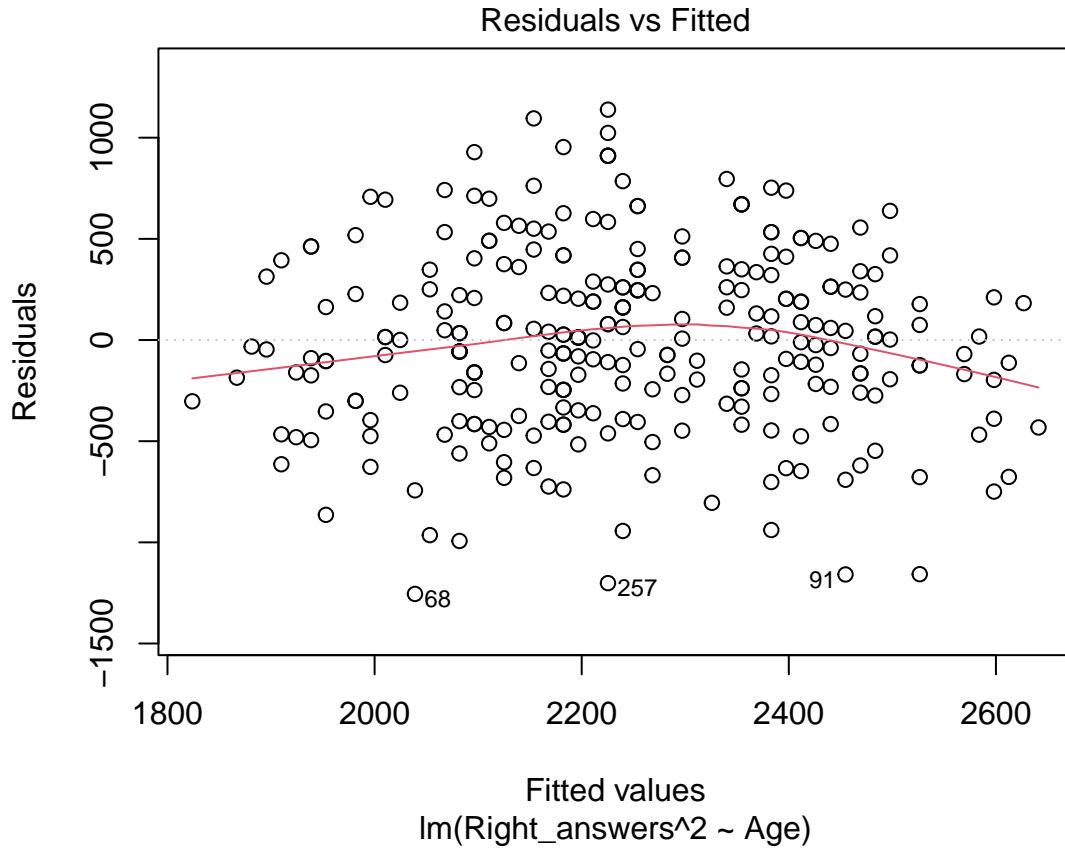
Table 24: Fitting linear model: Right_answers ~ Age

| Observations | Residual Std. Error | R^2 | Adjusted R^2 |
|--------------|---------------------|--------|----------------|
| 277 | 4.857 | 0.1414 | 0.1383 |



Nous remarquons une forme de “U” renversé, le caractère linéaire ne semble pas vérifié. Il convient alors de transformer `Right_answers` :

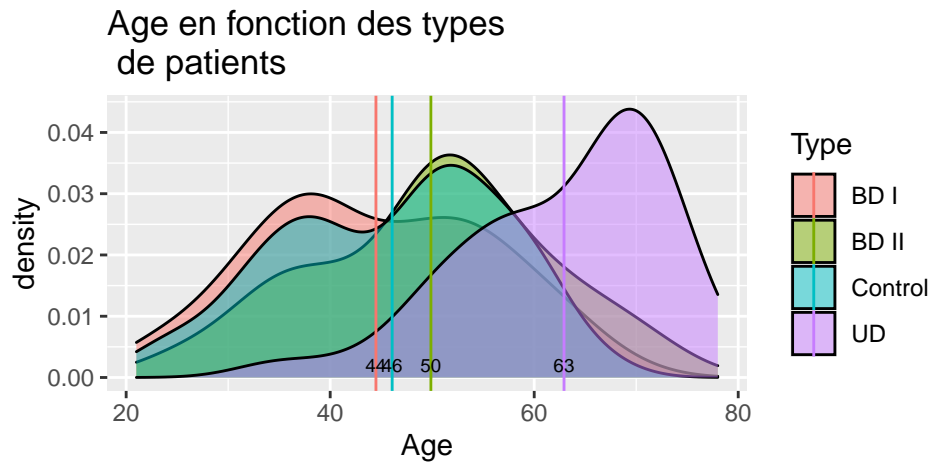
```
mod_transform<-lm(Right_answers~Age,Bipolar)
plot(mod_transform, which=1)
```



Cette transformation ne change rien au problème de linéarité. (Nous pourrions appliquer une transformation de boxcox mais nous n'allons pas l'utiliser)

Cependant, nous remarquons que '*Right_answers*' semble avoir une relation linéaire avec toutes les variables restantes, néanmoins ces interactions semblent plus dispersées que celles de '*Rights_answers*' et '*Negative/Positive_valence*', '*Submissive*' et '*Dominante*'.

Une autre remarque que nous pouvons faire, de manière générale, il semble que le type '*UD*' contribue très peu à la corrélation entre les variables. Cela peut être dû au fait que la population de personnes atteintes de dépression unipolaire soit assez petite comparé à celle des autres types. Il se peut aussi que nous observons cette différence car cette population semble significativement plus vieille que pour les autres type.

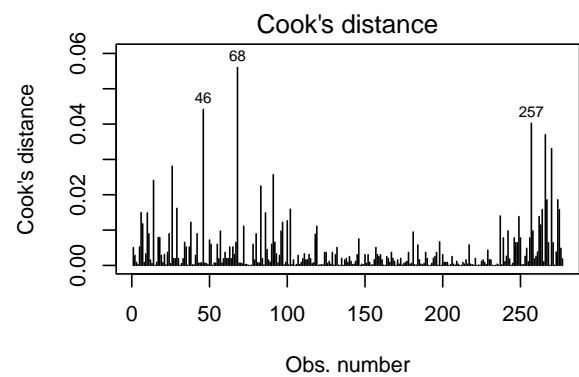
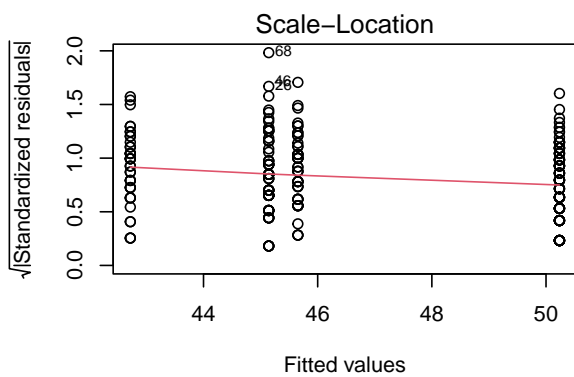
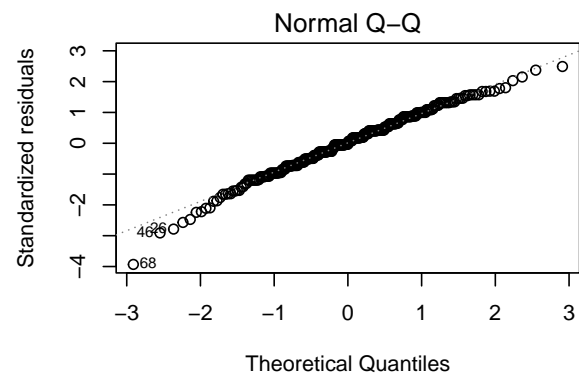
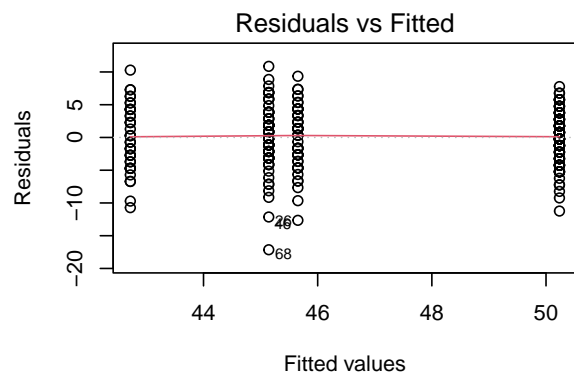


2.2 Analyse de la variance (ANOVA):

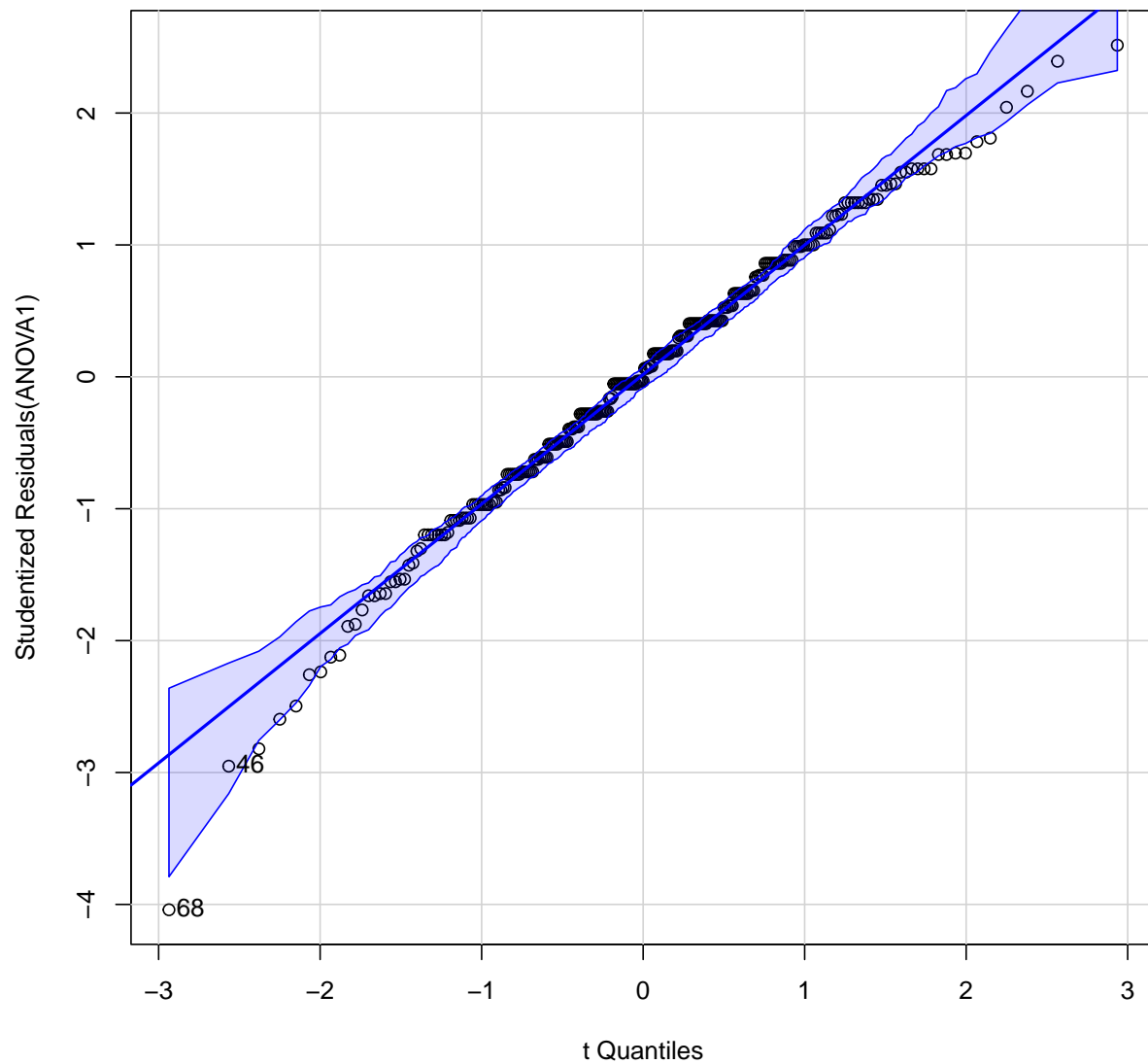
Nous pouvons alors tout simplement faire une Analyse de la variance (ANOVA) dans un premier temps sur `Right_answers` pour voir si les moyennes des modalités de Type et/ou Group sont significativement différentes.

Mettons en place le modèle:

```
ANOVA1 <- aov(Right_answers~Type, data=bipolar)
```



Tout semble correct pour la linéarité des résidus, l'homoscédasticité et la distance de cook qui ne montre pas de valeurs aberrantes mais étudions la normalité des résidus de plus près.



```
## [1] 46 68
```

Nous observons ce qui pourrait être un problème : des points créant une forme d'escalier, on pouvait s'attendre car la variable `Right_answers` est à valeur entières allant de 28 à 58. Considérons que cela ne soit pas significatif et analysons les résultats du modèle :

Table 25: Analysis of Variance Model

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------------|-----|--------|---------|---------|-----------|
| Type | 3 | 2289 | 763.1 | 39.57 | 2.905e-21 |
| Residuals | 273 | 5265 | 19.29 | NA | NA |

L'hypothèse null H_0 signifie que toutes les moyennes des différents niveaux du facteur sont égales. Ici nous pouvons écarté H_0 avec une forte significativité ($p\text{-value} < 2e-16$)

Il serait maintenant intéressant de voir quels niveaux se différencient le plus pour la variable réponse `Right_answers` :

- **Type:**

| | diff | lwr | upr | p adj |
|----------------------|--------|--------|---------|-----------|
| BD II-BD I | 0.5102 | -1.604 | 2.625 | 0.9244 |
| Control-BD I | 5.092 | 3.383 | 6.802 | 1.777e-12 |
| UD-BD I | -2.425 | -4.693 | -0.1566 | 0.0309 |
| Control-BD II | 4.582 | 2.655 | 6.509 | 1.67e-08 |
| UD-BD II | -2.935 | -5.371 | -0.4991 | 0.01091 |
| UD-Control | -7.517 | -9.612 | -5.423 | 2.608e-13 |

Ceci nous conforte dans l'idée que les niveaux du facteur **Type** diffèrent significativement concordent avec ce que l'on pouvait observer sur le boxplot et qui rejoint aussi les analyses faites par ANCOVA:

- Les individus Control semblent avoir en moyenne 5 bonnes réponses en plus que les individus Bipolar de type BD I et 4 pour les individus de type BD II
- Les individus atteint de dépression (UD) semblent avoir en moyenne 3 bonnes réponses en moins que les Bipolar et 7 de moins que les individus Control.
- Avec ce tableau, il est difficile d'analyser ce qui différencie le type BD I du type BD II à travers la variable ***Right_answers***.

Si on considère l'hypothèse de normalité des résidus non-satisfaisant, alors il convient de faire des test de Mann-whitney Wilcoxon qui sont non paramétriques.

2.3 MANOVA/MANCOVA

Pour plus de facilité nous allons réencoder le facteur **Type** sous forme de 4 nouvelles variables qualitatives `BD_I`, `BD_II`, `Control`, et `UD`:

MANCOVA avec les variables réponses `Audio_prosody`/`Face_video` par exemple.

```
MANCO <- aov(cbind(Audio_prosody,Face_video)~(BD_I+BD_II+Control+UD)*Age,bipolar)
pander(anova(MANCO))
```

Table 27: Analysis of Variance Table

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|--------------------|----|----------|----------|--------|--------|------------|
| (Intercept) | 1 | 0.9884 | 11438 | 2 | 268 | 3.446e-260 |
| BD_I | 1 | 0.06808 | 9.789 | 2 | 268 | 7.888e-05 |
| BD_II | 1 | 0.03651 | 5.077 | 2 | 268 | 0.006852 |
| Control | 1 | 0.1543 | 24.45 | 2 | 268 | 1.769e-10 |
| Age | 1 | 0.08228 | 12.01 | 2 | 268 | 1.007e-05 |
| BD_I:Age | 1 | 0.001786 | 0.2398 | 2 | 268 | 0.787 |
| BD_II:Age | 1 | 0.0237 | 3.253 | 2 | 268 | 0.04017 |

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|--------------------|-----|----------|----------|--------|--------|--------|
| Control:Age | 1 | 0.006619 | 0.8929 | 2 | 268 | 0.4107 |
| Residuals | 269 | NA | NA | NA | NA | NA |

```
MANCO_1 <- aov(cbind(Audio_prosody,Face_video)~BD_I+BD_II+Control+UD+ Age ,bipolar)
pander(anova(MANCO_1))
```

Table 28: Analysis of Variance Table

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|--------------------|-----|---------|----------|--------|--------|------------|
| (Intercept) | 1 | 0.9882 | 11395 | 2 | 271 | 3.151e-262 |
| BD_I | 1 | 0.06731 | 9.779 | 2 | 271 | 7.931e-05 |
| BD_II | 1 | 0.03584 | 5.038 | 2 | 271 | 0.007111 |
| Control | 1 | 0.1527 | 24.42 | 2 | 271 | 1.778e-10 |
| Age | 1 | 0.08094 | 11.93 | 2 | 271 | 1.08e-05 |
| Residuals | 272 | NA | NA | NA | NA | NA |

On observe avec cette MANCOVA que la combinaison des tests ***Right_answers*** et ***Face_video*** arrive à reconnaître la population touché par des troubles de bipolarité, ainsi que les individu sains et dépressif. On observe que l'interaction **BD_II:Age** est légèrement significative, ce qui signifie que la combinaison des deux tests arrivent à distinguer (plus ou moins) l'age d'un individu ayant un trouble de la bipolarité type 2.

3 Conclusion

Il est difficile d’avoir un réel recul sur notre étude car nous ne pouvons pas affirmer avec certitude nos résultats basés sur ces tests. Nous pouvons néanmoins tirer certaines hypothèses des analyses de variances et covariances effectuées.

Nous avons fait plusieurs ANCOVA qui nous ont montré plusieurs choses :

- La première est que dans la majorité des modèles (donc pour chaque variable de test étant considéré comme variables à expliquer) les interactions entre le facteur *Type* et la covariable *Age* ne sont pas significatives.
- Nous observons que le R^2 est resté assez bas pour tous les modèles construits, donc les modèles ne semblent pas assez fiables.

Après avoir observé que l’hypothèse de linéarité entre la covariable Age et les autres variables réponses reste à discuter, nous avons décidé de réaliser une ANOVA.

L’ANOVA sur la variable *Type* nous a montré que les moyennes obtenues sur la variable *Right_answers* étaient significativement différentes pour chaque groupe (Bipolar, UD, Control). Par contre, cela nous a montré qu’il était compliqué de différencier BD_I et BD_II en fonction de leurs moyennes.

Et enfin nous avons fait une MANCOVA pour voir si la combinaison de deux tests arrive à mieux déceler les différents niveaux du facteur Type. Nous avons basé notre MANCOVA sur les variables *Face_video* et *Audio_prosody*. Nous avons choisi ces deux variables aléatoirement, en faisant attention à ne pas prendre *Right_answers* pour éviter de fausser le modèle. Nous observons que les deux variables réunies, nous voyons que toutes les modalités de *Type* sont significatives ainsi que *Age* et la variable d’interaction *BD_II:Age*.

C’est étonnant car avec l’ANCOVA, aucune interaction n’a été repérée par aucun des modèles. Cela nous dit que la MANCOVA peut fournir des informations sur l’observation faite sur un groupe de variables. Ici, les variables *Type*, *Age* et *BD_II:Age* sont significativement importantes dans la distinction de différents troubles ainsi que l’âge.