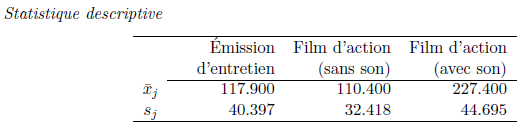
# Série 1

## Exercice 1

**Pop-corn et films d’action.** Pop-corn, sucreries et films d’action font bon ménage. Et pour cause, les amateurs de films d’action consomment 98% plus de friandises que les personnes qui regardent une émission d’entretien. C’est le résultat d’une étude américaine [1](#_bookmark0) réalisée auprès d’une trentaine d’étudiants répartis aléatoirement dans trois groupes à qui l’on proposa de regarder la télévision durant 20 minutes. Dans la première condition, les participants regardèrent un extrait de *The Island*, un film d’action hollywoodien (24.7 plans/min ; 24.5 changements de source sonore/min). Dans la deuxième condition, les participants visionnèrent un extrait de l’émission d’entretien *Charlie Rose* (4.8 plans/min ; 3.2 changements de source sonore/min). Dans la troisième condition, les participants visionnèrent le même extrait de *The Island*, mais sans le son. Durant l’expérience, de la nourriture (M&M’s, biscuits, carottes et raisins) a généreusement été mise à disposition des participants qui pouvaient se servir et consommer à leur guise. Les quantités absorbées ont été mesurées, en voici le relevé :

|  |  |  |
| --- | --- | --- |
| Émission  d’entretien | Film d’action  (sans son) | Film d’action  (avec son) |
| 34 | 66 | 170 |
| 63 | 71 | 186 |
| 119 | 88 | 194 |
| 121 | 103 | 202 |
| 123 | 105 | 203 |
| 129 | 105 | 230 |
| 134 | 110 | 232 |
| 141 | 144 | 277 |
| 141 | 152 | 278 |
| 174 | 160 | 302 |
| 1179 | 1104 | 2274 |

1. Calculez la moyenne et l’écart-type de chacun des trois échantillons.
2. Les échantillons sont-ils issus de distributions ayant même variance ?

*Test d’homogénéité des variances*

F (2, 27) = 0.643, p = 0.534. Comme p > 0.05, on ne rejette pas H0 au seuil de 5%. Les variances sont donc homogènes.

1. La consommation moyenne est-elle la même dans les trois conditions expérimentales ?

*Analyse de variance*

F (2, 27) = 27.493, p < 0.001. Comme p < 0.05, on rejette H0 au seuil de 5%. Les quantités consommées différent selon le type d’´émission regardée.

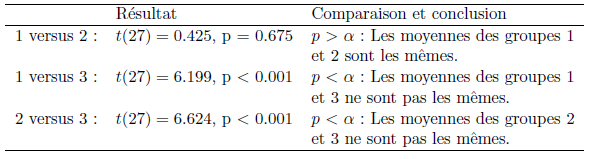
1. Comparez la consommation des trois groupes à l’aide de la méthode de Holm, puis inter- prétez brièvement vos résultats.

*Comparaisons multiples*

1: Emission d’entretien ;

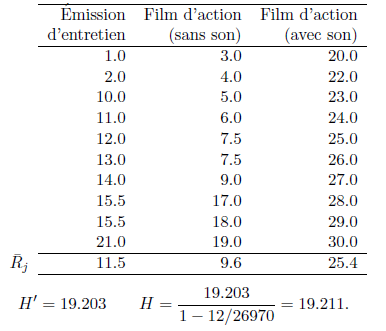
2: Film d’action (sans son) ; 3 : Film d’action (avec son).

3 : Film d'action (avec son).



1. La consommation médiane est-elle la même dans les trois conditions expérimentales ? Veuillez détailler vos calculs.

*Test de Kruskal-Wallis*



*χ*2(2) = 19*.*211, p *<* 0.001. Comme p *<* 0.05, on rejette *H*0 au seuil de 5%. La consom- mation m´ediane est influenc´ee par le type d’´emission regard´ee.

## Exercice 2

**Une thérapie en ligne pour aider les dysmorphophobiques.** Une équipe de chercheurs suédois [2](#_bookmark0) a testé l’efficacité d’une cyberthérapie cognitivo-comportementale (TCC) pour les pa- tients souffrant de dysmorphophobie. Ce trouble, au nom impressionnant, désigne l’altération de la perception de l’image de son propre corps, souvent une préoccupation démesurée pour un défaut physique imaginaire.

Trente patients adultes diagnostiqués dysmorphophobiques y ont participé. Dix-sept ont suivi une thérapie cognitive et comportementale en ligne intitulée BDD-NET (*Body Dysmorphic Disorder* ). Les treize autres ont bénéfi d’une thérapie virtuelle de soutien classique. Le traitement dura

douze semaines. La sévérité des symptômes dysmorphophobiques a été mesurée avant et après le traitement. La différence des sévérités (Après *−* Avant) a été calculée pour chaque sujet et est reportée dans le tableau ci-dessous :

*Différence de sévérité*

|  |  |
| --- | --- |
| Thérapie  de soutien | BDD-NET |
| *−*1 | *−*5 |
| *−*1 | *−*6 |
| *−*1 | *−*9 |
| *−*5 | *−*11 |
| *−*1 | *−*6 |
| *−*7 | *−*4 |
| 5 | *−*3 |
| 1 | *−*12 |
| 1 | *−*3 |
| 3 | *−*16 |
| 6 | *−*14 |
| 0 | *−*3 |
| *−*7 | *−*6 |
| *−*6 |
| *−*6 |
| *−*15 |
| *−*15 |

**Indication :** Dans cette situation, n’utilisez que des tests non-paramétriques.

* 1. Calculez la médiane et l’intervalle inter-quatile de la distribution des différences de sévérité dans chacun des deux groupes.

*Résumé numérique*

* 1. La symptomatologie des patients qui ont bénéficié d’une thérapie de soutien s’est-elle améliorée significativement ?

*Test de Wilcoxon*

T + = 32.5, T− = 45.5, p = 0.316, one-tailed. Comme p > 0.05, on ne rejette pas H0 au seuil de 5%. Les personnes ayant bénéficié d’une thérapie de soutien n’ont pas vu leur symptomatologie s’améliorer significativement.

* 1. Les symptômes des patients ayant bénéficié de la thérapie BDD-NET se sont-ils atténués significativement ?

*Test de Wilcoxon*

T + = 0, T− = 153, p < 0.001, one-tailed. Comme p < 0.05, on rejette H0 au seuil de 5%. Les personnes ayant bénéficié d’une thérapie BDD-NET ont vu leur symptomatologie s’améliorer significativement.

* 1. La thérapie BDD-NET est-elle significativement plus efficace que la thérapie de soutien ?

*Test de Mann-Whitney*

U1 = 196.5, U2 = 24.5, p < 0.001, one-tailed. Comme p < 0.05, on rejette H0 au seuil de 5%. La symptomatologie s’améliore davantage dans le groupe qui a bénéficié d’une thérapie BDD-NET.

## Exercice 3

**Comment la douleur mène à l’hypervigilance.** L’exposition à une souffrance extrême affecte-t-elle durablement la façon dont le cerveau perçoit la douleur ? C’est la question à laquelle répond l’expérience conduite par une équipe de chercheurs israéliens de la Bar-Ilan University [3](#_bookmark0). L’équipe a ainsi analysé les zones du cerveau impliquées dans la perception de la douleur chez un groupe de douze combattants de l’armée israélienne, exposés à la mort d’un compagnon d’armes (ou à des blessures graves chez l’un d’eux). Face à ces vétérans : un groupe contrôle composé de dix militaires non inscrits dans des unités de combat.

Tous les sujets ont été confronté, d’une part, à une image montrant une douleur intense (en l’occurrence, une hache plantée dans le pied), et, d’autre part, à une image ne montrant pas une situation douloureuse – une hache plantée sur une bûche, à côté d’un pied, indemne cette fois. Lors de chaque présentation, l’activité du cortex cingulaire postérieur a été enregistrée (voir tableau ci-dessous).

*Groupe contrôle Groupe des combattants*

1 1

|  |  |
| --- | --- |
| Image  sans douleur | Image  avec douleur |
| 29 | 78 |
| 34 | 57 |
| 30 | 85 |
| 30 | 82 |
| 41 | 85 |
| 44 | 79 |
| 29 | 89 |
| 17 | 83 |
| 38 | 80 |
| 38 | 72 |
| 330 | 790 |

|  |  |
| --- | --- |
| Image  sans douleur | Image  avec douleur |
| 65 | 59 |
| 72 | 52 |
| 56 | 61 |
| 55 | 85 |
| 87 | 62 |
| 53 | 93 |
| 89 | 68 |
| 56 | 83 |
| 75 | 77 |
| 45 | 80 |
| 70 | 82 |
| 77 | 73 |
| 800 | 875 |

2 2

3 3

4 4

5 5

6 6

7 7

8 8

9 9

10 10

11

12

**Indication :** Vous pouvez supposer que les mesures sont toutes issues de distributions normales.

1. En présence de l’image montrant une douleur intense, l’activation du cortex cingulaire postérieur est-elle la même dans les deux groupes ? Pour répondre à cette question, réalisez un test statistique au seuil de 5%, mais préalablement testez l’homogénéité des variances.

*Test de l’´égalité des variances*

F (1, 20) = 2.734, p = 0.114. Comme p > 0.05, on ne rejette pas H0 au seuil de 5%. Les variances sont donc homogènes. Nous allons poursuivre avec les test de Student.

*Test de l’´égalité des moyennes*

t(20) = 1.284, p = 0.214. Comme p > 0.05, on ne rejette pas H0 au seuil de 5%. Lorsque l’image représente une scène douloureuse, l’activité du cortex cingulaire postérieur est la même dans les deux groupes (contrôle et combattants).

1. En présence de l’image ne montrant pas une situation douloureuse, l’activation du cortex cingulaire postérieur est-elle la même dans les deux groupes ? Comme précédemment, ré- pondez à cette question en recourant à un test statistique en tenant compte de l’éventuelle hétérogénéité des variances entre les deux groupes.

*Test de l’´égalité des variances*

*F (1, 20) = 5.009, p = 0.037. Comme p < 0.05, on rejette H0 au seuil de 5%. Les variances ne sont donc pas homogènes. Nous allons poursuivre avec les test de Welch.*

*Test de l’´égalité des moyennes*

*t*(17*.*682) = *−*7*.*127, p *<* 0.001. Comme p *<* 0.05, on rejette *H*0 au seuil de 5%. Lorsque l’image ne représente pas une scène douloureuse, l’activation du cortex cingulaire postérieur n’est pas la même chez les combattants et parmi les individus du groupe contrôle.

1. Au sein du groupe contrôle, l’activation du cortex cingulaire postérieur est-elle plus forte en présence de l’image montrant une douleur intense ?

*Test de Student sur mesures pairées*

t(9) = −11.135, p < 0.001. Comme p < 0.05, on rejette H0 au seuil de 5%. Pour le groupe contrôle, l’activation du cortex cingulaire postérieur est plus forte en présence de l’image montrant une douleur intense.

1. Au sein du groupe des combattants, l’activation du cortex cingulaire postérieur s’atténue- t-elle lors de la présentation d’une image ne montrant pas une situation douloureuse ?

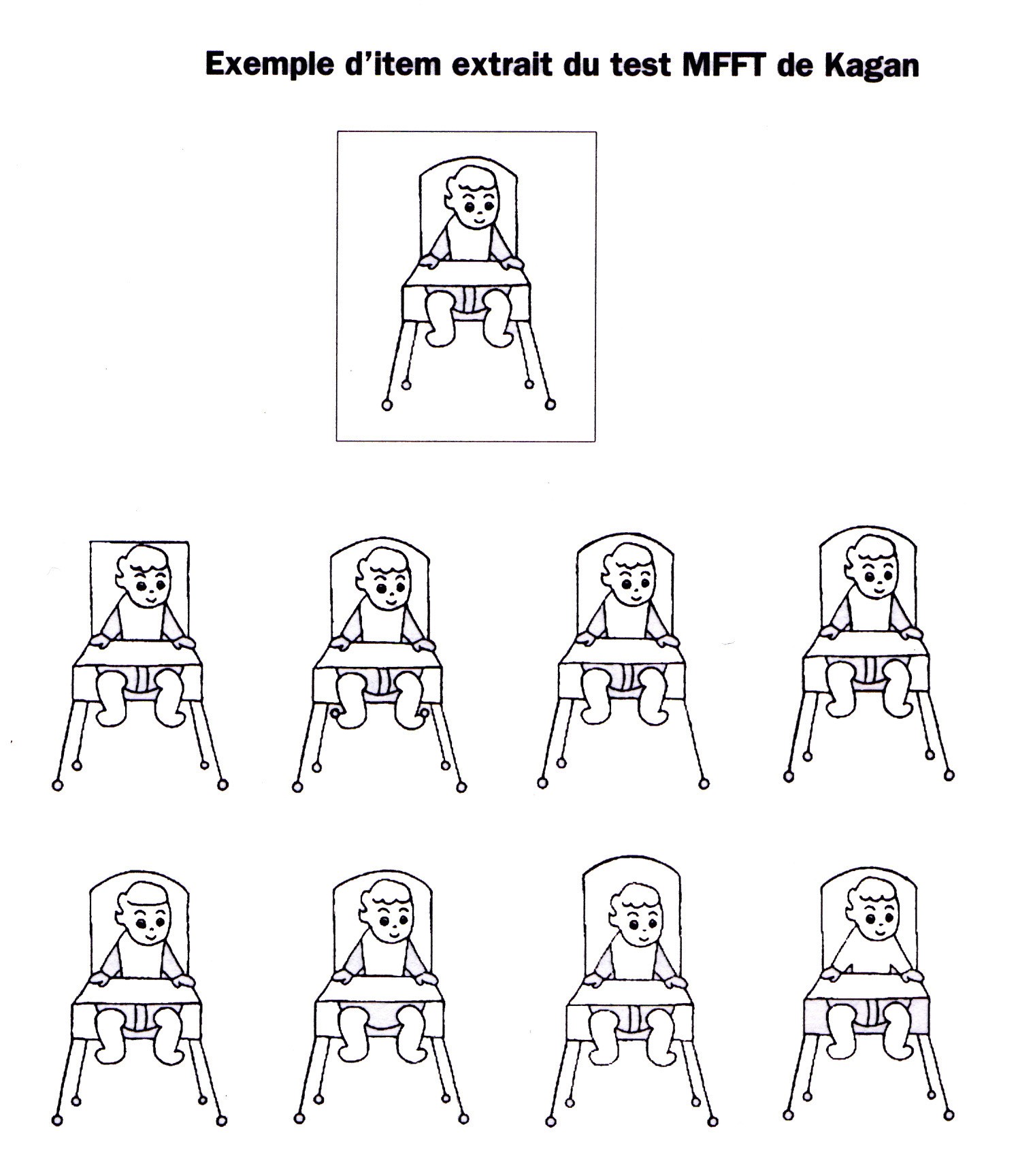
*Test de Student sur mesures pairées*

t(11) = −0.953, p = 0.181. Comme p > 0.05, on ne rejette pas H0 au seuil de 5%. Chez les combattants, l’activation du cortex cingulaire postérieur ne s’atténue pas lors de la présentation d’une image ne montrant pas une situation douloureuse.

# Série 2

## Exercice 1

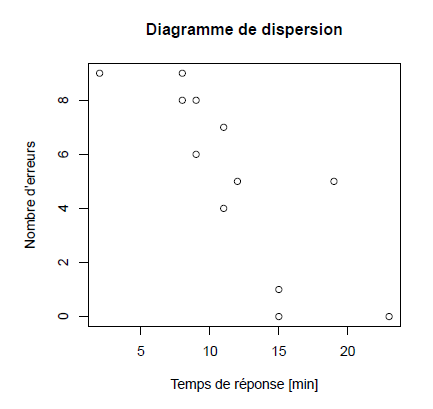
Quand on est face à un problème il y a au moins deux attitudes possibles. Une première attitude consiste à chercher la meilleure solution possible au problème, en prenant tout son temps. C’est l’attitude réfléchie. La seconde attitude consiste à répondre le plus rapidement possible, au risque de donner une mauvaise réponse. C’est l’attitude impulsive. On évalue la réflexion-impulsivité (R-I) au moyen d’une épreuve d’exploration visuelle : le test d’appariement de figures de Kagan.



Dans ce test, le sujet doit choisir parmi huit dessins représentant un objet familier celui qui est rigoureusement identique à un modèle présenté simultanément. L’attitude réfléchie se caractérise par des temps de réponse longs et un nombre d’erreurs faible, alors que l’attitude impulsive correspond à des temps de réponse courts et un nombre élevé d’erreurs.

Voici les résultats de 12 sujets au test d’appariement de figures de Kagan :

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sujets | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Temps  de réponse [min] | 23 | 8 | 15 | 9 | 9 | 11 | 11 | 19 | 12 | 8 | 15 | 2 |
| Nombre d’erreurs | 0 | 8 | 1 | 8 | 6 | 7 | 4 | 5 | 5 | 9 | 0 | 9 |

1. Représentez graphiquement ces données en plaçant en abscisse le *Temps de réponse* et en ordonnée le *Nombre d’erreurs*.

*Représentation graphique - Scatterplot*

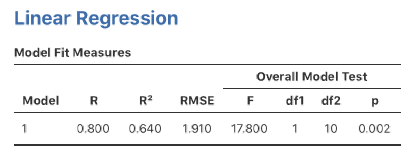
A` partir de la forme du nuage, nous pouvons affirmer que la corrélation entre le *Temps de réponse* et le *Nombre d’erreurs* est n´ergative ; cela signifie que lorsqu’une personne prend plus de temps pour répondre, elle commet moins d’erreurs.

1. Calculez la valeur du coefficient de corrélation de Bravais-Pearson entre *Temps de réponse*

et *Nombre d’erreurs*. Commentez votre résultat.

*Coefficient de corrélation*

Le coefficient de corrélation de Bravais-Pearson vaut rxy = −0.8, ce qui indique une forte corrélation n´négative entre les variables Temps de réponse et Nombre d’erreurs. Nous voyons, en effet, que les points se distribuent grossièrement le long d’une droite de pente n´négative.



1. Effectuez un test afin de savoir si le coefficient de corrélation est statistiquement significatif au seuil de 5%.

*Test de la correlation – Jamovi : Correlation matrix > Pearson > Entrer VD et VI*

Hypothèse : H0 : pxy = 0

H1 : pxy ≠ 0

Conclusion :

Le test du coefficient de corrélation est significatif au seuil de 5% (r(10) = −0.8, p = 0.002). Nous rejetons l’hypothèse nulle et acceptons l’hypothèse alternative qui affirme qu’au sein de la population la corrélation entre le Temps de réponse et le Nombre d’erreurs est différente de zéro.

1. Construisez l’équation de la droite qui permette de prédire le nombre d’erreurs à partir du temps de réponse en appliquant les formules, d’une part, et en utilisant Jamovi, d’autre part.

Formules :

Pente de la droite des moindres carrés vaut :

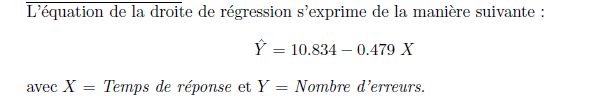


L’ordonnée à l’origine de cette droite vaut :

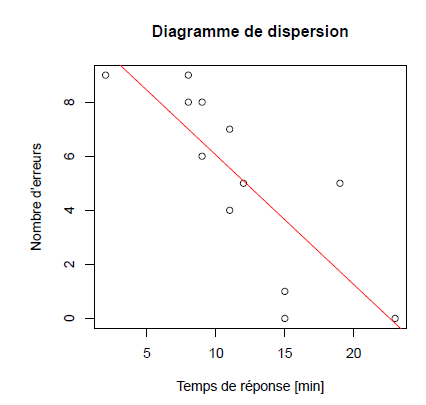
Equation de la droite de régression:

= β0 + β1X

Avec X = VI et Y=VD



1. Placez la droite de régression dans le diagramme de dispersion.

*Diagramme de dispersion et droite de régression – Jamovi : Scatterplot> cocher ‘linear’*

1. Calculez l’erreur standard de régression *σ*ˆ.

*Erreur standard de régression*

L’estimation de l’erreur standard de régression est égale à

RMSE dans Jamovi :

* Calcul :
* Graphique : Scatterplot > cocher linear > cocher standard error

1. Calculez les limites de confiance imposées à l’ordonnée à l’origine de la droite de régression.

*Intervalle de confiance de l’ordonnée `a l’origine*

L’intervalle de confiance à 95% de l’ordonnée à l’origine vaut [7.552; 14.115].

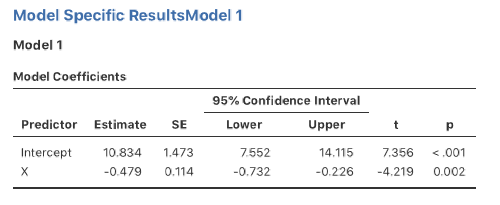
1. L’ordonnée à l’origine est-elle significativement différente de 0 ?

*Test de la valeur de l’ordonnée `a l’origine*

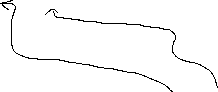


L’ordonnée `a l’origine est significativement différente de 0 au seuil de 5% (t(10) = 7.356, p < .001).

Dans linear regression :



1. Calculez les limites de confiance imposées à la pente de la droite de régression.



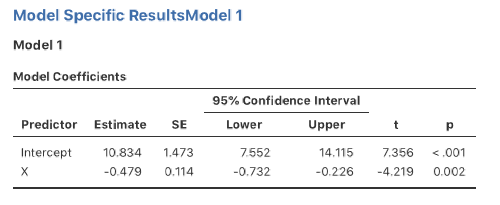
*Intervalle de confiance de la pente*

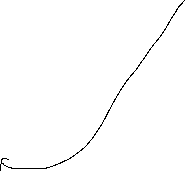
L’intervalle de confiance `a 95% de la pente vaut [−0.732; −0.226].

1. La pente de la droite est-elle significativement différente de 0 ?

*Test de la valeur de la pente*

La pente est significativement différente de 0 au seuil de 5% (t(10) = −4.219, p = 0.002).





1. Prédisez le nombre d’erreurs commises par un individu qui a mis 16 minutes pour passer le test de Kagan.

*Prédiction*

Selon notre modèle linéaire, le nombre d’erreurs commises par un individu ayant mis 16 minutes pour passer le test de Kagan vaut :

## Exercice 2

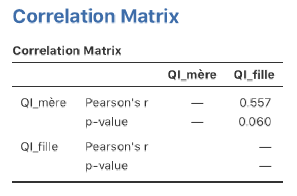
**Telle mère, telle fille ?** On dispose des 12 paires d’observations qui apparaissent dans le tableau ci-dessous. La variable *X* représente le Q.I. des mères et la variable *Y* représente le Q.I. des filles.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| *X* | 123 | 144 | 105 | 110 | 98 | 138 | 131 | 90 | 119 | 109 | 125 | 100 |
| *Y* | 102 | 138 | 136 | 133 | 95 | 146 | 115 | 100 | 142 | 105 | 130 | 120 |

1. Calculez le coefficient de corrélation linéaire *r* entre *X* et *Y* .

*Coefficient de corrélation linéaire*

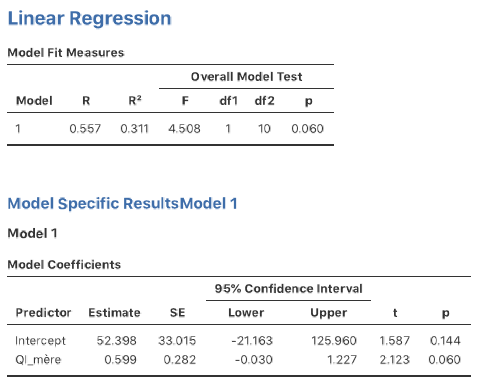
Le coefficient de corrélation rxy vaut 0.557.





1. Calculez la droite de régression qui permette de prédire le Q.I. d’une fille connaissant celui de sa mère.

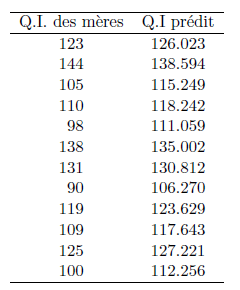
*Droite de régression*

L’équation de la droite de régression qui permet de prédire le Q.I. d’une fille connaissant celui de sa m`ère s’écrit donc :



1. À partir du Q.I. des mères, prédisez le Q.I. de chacune des filles.

*Prédiction*

Pour prédire les Q.I. des filles, on utilise l’équation trouvée au point précédent. Voici l’ensemble des valeurs prédites calculées avec Jamovi :

1. Calculez la variance des Q.I. prédits des filles, c’est-à-dire la variance expliquée *V ar*(*Y*ˆ ).

*Variance expliquée*

On calcule l’´écart-type des Q.I. prédits des filles puis on l’´élève au carré pour obtenir la valeur de la variance. On trouve :

V ar( ) = 10.0392 = 100.779

1. Calculez la variance des Q.I. observés des filles, c’est-à-dire la variance totale *V ar*(*Y* ).

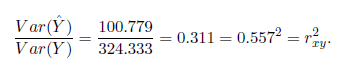
*Variance totale*

En élevant au carré la valeur de l’´écart-type des Q.I. observées des filles, on obtient la variance des Q.I. observés des filles :

V ar(Y ) = 18.0092 = 324.333

1. Quelle est la proportion de la variance des Q.I. des filles expliquée par les Q.I. des mères ?

Indication : déterminez la valeur du rapport *V ar*(*Y*ˆ )*/V ar*(*Y* ).

*Part de variance expliquée*

1. Comparez cette proportion au carré du coefficient de corrélation.

# Série 3

## Exercice 1

Une grosse société veut prédire la satisfaction au travail qu’éprouvent ses employés. Elle a recueilli des données chez 15 employés qui ont chacun donné des informations sur leur satisfaction au travail (*Y* ), le degré de responsabilité (*X*1), le nombre de personnes supervisées (*X*2), l’évaluation de l’environnement de travail (*X*3) et les années de service (*X*4). Les données sont les suivantes :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Satisfaction | 2 | 2 | 3 | 3 | 5 | 5 | 6 | 6 | 6 | 7 | 8 | 8 | 8 | 9 | 9 |
| Responsabilité | 4 | 2 | 3 | 6 | 2 | 8 | 4 | 5 | 8 | 8 | 9 | 6 | 3 | 7 | 9 |
| Personnes supervisées | 5 | 3 | 4 | 7 | 4 | 8 | 6 | 5 | 9 | 8 | 9 | 3 | 6 | 9 | 9 |
| Environnement | 1 | 1 | 7 | 3 | 5 | 8 | 5 | 5 | 6 | 4 | 7 | 2 | 8 | 7 | 9 |
| Années de service | 5 | 7 | 5 | 3 | 3 | 6 | 3 | 2 | 7 | 3 | 5 | 5 | 8 | 8 | 1 |

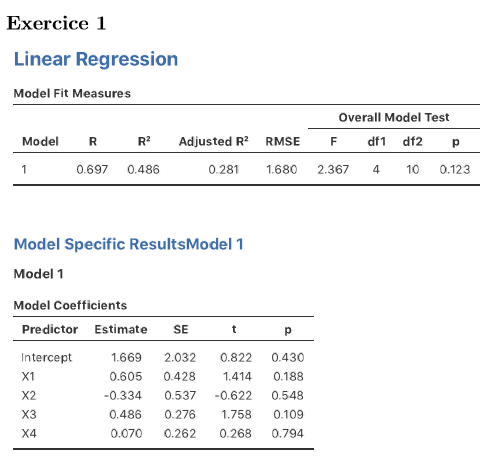
1. Proposez un modèle linéaire pour décrire la satisfaction des travailleurs à l’aide des quatre prédicteurs.

*Modèle*

= β0 + β1 X1 + β2 X2 + β3 X3 + β4 X4

1. Estimez les paramètres de votre modèle.

*Estimation des paramètres du modèle*

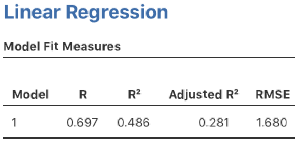


Après estimation, l’´équation du modèle devient :

= 1.669 + 0.605 X1 − 0.334 X2 + 0.486 X3 + 0.07 X4.

Toute chose égale par ailleurs, lorsque le degré de responsabilité augmente la satisfaction au travail augmente aussi. Si le nombre de personnes supervisées augmente, alors la satisfaction au travail diminue. Lorsque la qualité de l’environnement de travail augmente, la satisfaction au travail augmente. Finalement, comme la valeur du coefficient de régression partiel β4 est proche de 0, cela signifie que le nombre d’années de service n’a qu’une influence n´négligeable sur la satisfaction au travail.

1. Calculez l’erreur standard de régression.

*Erreur standard de régression*

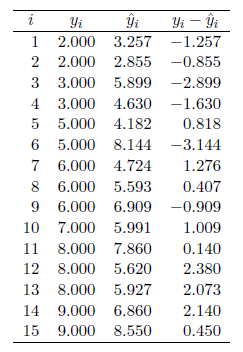
. L’indice de séparation vaut sy/ = 1.179. La séparation est donc importante

1. Calculez la valeur du coefficient de corrélation multiple.

*Coefficient de d´détermination et coefficient de régression multiple*

Le coefficient de d´détermination vaut R2 = 0.486. Ainsi un peu moins de 50% de la variance totale est expliquée par le modèle. Le coefficient de corrélation multiple, quant à lui, est égal à

1. Calculez les résidus *ei* et les scores ajustés *y*ˆ*i*.

*Résidus et valeurs prédites*

ou` yˆ i représente les valeurs prédites et yi − yˆiles résidus.

1. Montrez que la corrélation entre les résidus et les scores ajustés est égale à 0.

*Corrélation entre les résidus et les scores ajustés*

Le coefficient de corrélation entre les résidus et les scores prédits vaut 0.

1. Montrez que les résidus ne sont pas du tout corrélés aux prédicteurs.

*Corrélation entre les résidus et les prédicteurs*

Les coefficients de corrélation entre les résidus et les prédicteurs valent tous 0.

1. Calculez la corrélation entre les scores prédits *y*ˆ*i* et les scores observés *yi*. Comparez ce coefficient de corrélation au coefficientde corrélation multiple.

*Corrélation entre les scores prédits et les scores observés*

Le coefficient de corrélation entre les scores prédits et les scores observés vaut :

*ry*ˆ*y* = 0*.*697*.*

Le coefficient de corrélation multiple est ´égal `a la racine carrée du coefficient de détermination *R*2 :

En comparant *ry*ˆ*y* et *R*, nous constatons que ces deux coefficients sont ´égaux.

s

1. La liaison globale entre *Y* et les variables explicatives *X*1, *X*2, *X*3 et *X*4 est-elle significative au seuil de 5% ?

*Test globale*

Hypothèses :

Conclusion :

Globalement la liaison entre la variable critère Y et les variables prédicatrices *X*1, *X*2, *X*3 et *X*4 n’est pas significative au seuil de 5% (*F* (4*,* 10) = 2*.*367*, p* = 0*.*123). Aucune des quatre variables prédicatrices ne permet d’expliquer la satisfaction au travail. Vu la toute petite taille de l’´échantillon, il est possible que nous commettions ici une erreur de deuxième espèce.

## Exercice 2

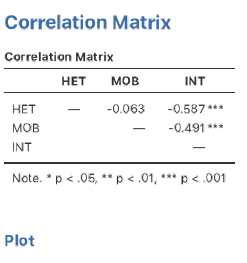
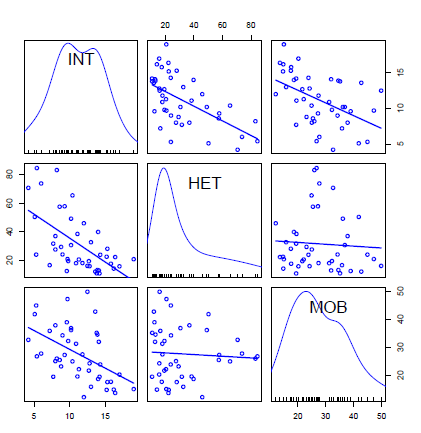
Dans le fichier Angell.csv, 43 villes des États-Unis sont caractérisées selon 3 variables [*1*](#_bookmark1). Ces variables sont l’intégration morale (*i.e.* intégration sociale) (INT), l’hétérogénéité ethnique (HET) et la mobilité géographique (MOB).

L’intégration morale est un indice qui combine des informations sur la criminalité, d’une part, et sur l’assistance sociale, d’autre part. L’hétérogénéité ethnique est un indice calculé à partir de la proportion des habitants de couleur et de celle des habitants blancs nés à l’étranger. La mobilité géographique est une variable construite à partir du nombre d’habitants entrant et sortant quotidiennement de la ville.

1. Représentez toutes les relations binaires entre les variables INT, HET et MOB.

*Matrice des nuages de points*

Dans le graphique ci-dessous sont représentés les histogrammes de chacune des variables ainsi que les diagrammes de dispersion de toutes les paires de variables.

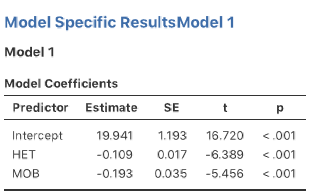


On peut faire les diagrammes sous Corrélation Matrix > Plot

De manière générale, nous voyons que la variable INT d´dépend linéairement des variables HET et MOB. Par contre les variables HET et MOB ne sont pas corrélées l’une avec l’autre.

1. Estimez les paramètres *β*0, *β*1 et *β*2 de l’équation : INT = *β*0 + *β*1 *·* HET + *β*2 *·* MOB.

Jamovi : linear regression

*Estimation des paramètres du modèle*



L’´équation du modèle s’´écrit ainsi :

*INT* = 19*.*941 *−* 0*.*109 *HET −* 0*.*193 *MOB.*

1. Interprétez les résultats de vos calculs.

*Interprétation*

* + Ordonn´ee `a l’origine

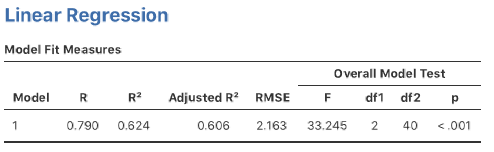
19*.*941 est l’ordonn´ee `a l’origine. C’est la valeur que prend th´eoriquement la variable INT lorsque les variables HET et MOB prennent la valeur de 0.

* + Premier coefficient de r´egression partiel

*−*0*.*109 est le coefficient de r´egression partiel correspondant `a la variable HET (dans le tableau des Coefficients voir la valeur se trouvant dans la colonne Estimate `a la ligne HET). Lorsque la variable HET augmente d’une unit´e et que la variable MOB reste constante, la variable INT diminue de 0*.*109. Le sentiment d’int´egration diminue lorsque l’h´et´erog´en´eit´e augmente.

* + Second coefficient de r´egression partiel

*−*0*.*193 est le coefficient de r´egression partiel correspondant `a la variable MOB. Lorsque la variable MOB augmente d’une unit´e et que la variable HET est maintenue constante, la variable INT diminue de 0*.*193. Le sentiment d’int´egration diminue lorsque la mobilit´e augmente.

1. Les données sont-elles bien décrites par le modèle ? Pour répondre à cette question calculez l’erreur standard de régression.

*Erreur standard de régression*

L’erreur standard de régression vaut : Si les erreurs sedistribuaient normalement, 68% des observations se situeraient dans une bande de *±*2*.*24 autour du plan d´efini par le mod`ele. La s´eparation vaut *sINT /σ* = 1*.*592, elle est forte.

Sint = Descriptives > Std. Deviation

1. Calculez le coefficient de corrélation multiple.

*Le coefficient de d´détermination multiple*

Le coefficient de d´détermination vaut R2 = 0.624. Le coefficient de corrélation multiple au carré, encore appelée coefficient de d´détermination, représente la part de variance du critère expliquée par les prédicteurs. Dans ce modèle les prédicteurs permettent d’expliquer 62.4% de la variance du critère. Le coefficient de corrélation multiple, quant `à lui, vaut :

1. Calculez également le coefficient de détermination ajusté.

*Le coefficient de détermination ajusté*

Le coefficient de détermination ajusté vaut *R2ajusté* = 0.606

1. La liaison globale entre INT et les variables explicatives HET et MOB est-elle significative ? Pour répondre à cette question réalisez un test au seuil de 5%.

*Test globale*

Hypothèses :

Conclusion :

Globalement la liaison entre la variable réponse INT et les variables explicatives HET et MOB est significative au seuil de 5% (*F* (2*,* 40) = 33*.*245, *p <* 0*.*001). Nous rejetons donc l’hypothèse nulle et acceptons l’hypothèse alternative qui affirme qu’au moins l’un des coefficients de régression est non nul. On peut ainsi conclure au seuil de 5% que l’h´et´erog´en´eit´e ethnique ou la mobilité géographique sont des facteurs explicatifs

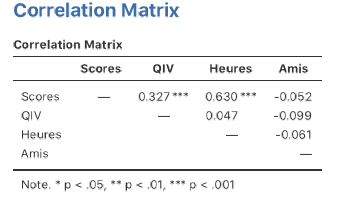
de l’intégration sociale aux E´ tats-Unis.

## Exercice 3

Dans le cadre du projet d’immersion linguistique de l’Université Catholique de Louvain (Bel- gique), des chercheurs se sont intéressés à la performance en français (résultats en français) selon certaines caractéristiques individuelles chez les enfants de 14 ans. Pour mener à bien ce projet, plusieurs informations concernant 360 enfants ont été recueillies, parmi lesquelles, le quotient intellectuel verbal (*QIV* ), le nombre d’amis à 14 ans (*Amis*), le nombre d’heures d’étude moyen par semaine durant la période des examens (*Heures*) et les résultats en français à 14 ans (*Scores*). Ces informations sont contenues dans le fichier Immersion.csv.

* 1. Calculez l’ensemble des corrélations simples de l’ensemble des variables contenus dans le fichier Immersion.csv.

*Matrice des corrélations*



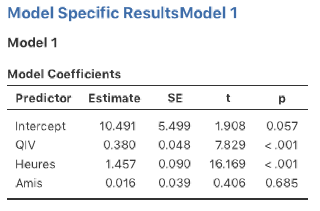
* 1. Etablissez l’équation de régression multiple sous-jacente à la situation du projet puis construisez le modèle complet.

*Equation de la droite de régression*

L’´équation de régression qui va permettre de prédire la performance en français des adolescents de 14 ans a la forme suivante :

= *β*0 + *β*1 QIV + *β*2 Heures + *β*3 Amis.

* 1. Que valent les paramètres de ce modèle ?

*Estimation des paramètres du modèle*

Jamovi : linear regression

Une fois les paramètres estimés, l’équation du modèle devient :

= 10*.*491 + 0*.*380 QIV + 1*.*457 Heures + 0*.*016 Amis.

* + Ordonn´ee `a l’origine

10*.*491 est l’ordonn´ee `a l’origine. C’est la valeur que prend th´eoriquement la variable

Scores lorsque les autres variables prennent la valeur de 0.

* + Premier coefficient de r´egression partiel

0*.*380 est le coefficient de r´egression partiel correspondant `a la variable QIV. Lorsque la variable QIV augmente d’une unit´e et que les autres variables ind´ependantes restent constantes, la variable Scores augmente de 0*.*380. Plus le quotient intellectuel verbal d’un adolescent de 14 ans est ´el´ev´e, meilleure sera sa performance en fran¸cais.

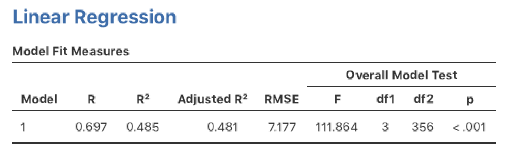
* + Deuxi`eme coefficient de r´egression partiel

1*.*457 est le coefficient de r´egression partiel correspondant `a la variable Heures. Lorsque la variable Heures augmente d’une unit´e et que les autres variables ind´ependantes restent constantes, la variable Scores augmente de 1*.*457. Plus un adolescent ´etudie, meilleure sera sa performance en fran¸cais.

* + Troisi`eme coefficient de r´egression partiel

0*.*016 est le coefficient de r´egression partiel correspondant `a la variable Amis. Lorsque la variable Amis augmente d’une unit´e et que les autres variables ind´ependantes restent constantes, la variable Scores augmente de 0*.*016. Cependant, cette augmentation est toute petite (proche de 0). Le nombre d’amis d’un adolescent a une influence n´egligeable sur sa performance en fran¸cais.

* 1. Que vaut l’erreur standard de régression ?

*Erreur standard*

Si les erreurs se distribuaient normalement, 68%des observations se situeraient dans une bande de ±7.217 autour du plan défini par le modèle. Le rapport entre l’écart-type de la variable Scores et l’erreur standard vaut 10.018/7.217 = 1.388 : la séparation est importante.

* 1. Quelle est la valeur du coefficient de corrélation multiple au carré ?

*Coefficient de corrélation multiple au carré (coefficient de détermination)*

*R*2 = 0*.*485. Le coefficient de corrélation multiple au carré représente la part de variance du critère expliquée par les prédicteurs. Cela nous montre qu’ici le modèle d´ecrit bien les donn´ees. Nous pouvons dire que 48*.*5% de la variance totale de la variable Scores est expliqu´ee par les trois variables QIV, Heures et Amis.

R2*ajusté* = 0.481. La taille de l’échantillon étant relativement grande, ce coefficient de régression ajusté au carré est proche de R2.

* 1. Le modèle de régression multiple est-il globalement significatif au seuil de 5% ?

*Test globale*

Hypothèses :

Conclusion :

Globalement, la liaison entre la variable crit`ere et les variables pr´edictrices QIV, Heures et Amis est significative au seuil de 5% (*F* (3*,* 356) = 111*.*864, *p <* 0*.*001). Il est possible, partiellement au moins, de rendre compte de la performance en fran¸cais des adolescents de 14 ans `a partir de la connaissance des variables QIV, Heures et Amis.

* 1. Quelles sont les variables significatives au seuil de 5% dans le modèle ? Justifiez vos ré- ponses ?

*Tests des coefficients de régression partiel (Significativité des paramètres)*

* + Test du coefficient de r´egression partiel associ´e `a la variable QIV

*t*(356) = 7*.*829, *p <* 0*.*001. Comme *p < α*, on rejette l’hypoth`ese nulle affirmant que *β*1 = 0 au seuil de 5%.

* + Test du coefficient de r´egression partiel associ´e `a la variable Heures

*t*(356) = 16*.*169, *p <* 0*.*001. Comme *p < α*, on rejette l’hypoth`ese nulle affirmant que *β*2 = 0 au seuil de 5%.

* + Test du coefficient de r´egression partiel associ´e `a la variable Amis

*t*(356) = 0*.*406, *p* = 0*.*685. Comme *p > α*, on ne rejette pas l’hypoth`ese nulle affirmant que *β*3 = 0 au seuil de 5%.

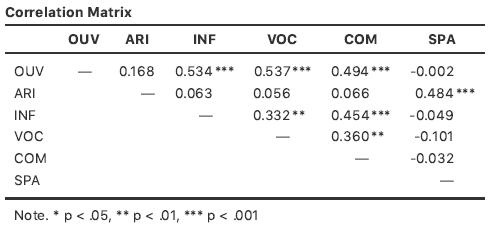
Le mod`ele lin´eaire est significatif mais seuls le quotient intellectuel verbal et le nombre d’heures d’´etude moyen par semaine durant la p´eriode des examens affectent v´eritable- ment la performance en fran¸cais chez les adolescents de 14 ans.

# Série 4

## Exercice 1

Les psychométriciens ont souvent voulu mettre en lien la personnalité et les aptitudes. C’est ainsi, par exemple, qu’un concept comme l’intelligence émotionnelle est apparu. Certains auteurs pos- tulent que le facteur *Ouverture* du BIG-5 est lié à l’intelligence. Cette hypothèse a été corroborée par Ashton, Lee et Vernon (2000). Ces derniers montrent plus précisément, dans leur étude, que l’*Ouverture* est liée à l’intelligence cristallisée (à l’exception de l’arithmétique), mais pas à l’intelligence fluide. Nous avons observé quelques variables sur un échantillon de 75 personnes. Les données enregistrées dans le fichier Ouverture.csv (les séparateurs de champ sont des “,”) contiennent six variables :

|  |  |  |
| --- | --- | --- |
| Variable | Sous-échelle | Concept mesuré |
| OUV | Ouverture | Trait de personnalité |
| ARI | Arithmétique | Facette de l’intelligence cristallisée |
| INF | Information | Facette de l’intelligence cristallisée |
| VOC | Vocabulaire | Facette de l’intelligence cristallisée |
| COM | Compréhension | Facette de l’intelligence cristallisée |
| SPA | Spatial | Facette de l’intelligence fluide |

1. Calculez l’ensemble des corrélations simples entre les six variables du fichier Ouvertu- re.csv.

*Matrice des corrélations*

1. Dérivez l’équation de régression qui prédit l’*Ouverture* sur la base des cinq autres variables d’intelligence.

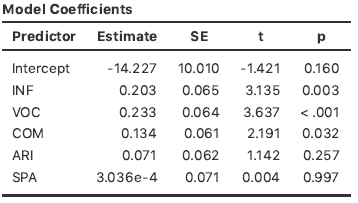
*Equation de la droite de régression*

L’équation de régression de droite qui va permettre de prédire l’*Ouverture* a la forme suivante :

= *β*0 + *β*1 ARI + *β*2 INF + *β*3 VOC + *β*4 COM + *β*5 SPA.

1. Que valent les paramètres du modèle ?

*Estimation des paramètres du modèle de régression*



Une fois les paramètres estimés l’équation du modèle devient :

= −14*.*227 + 0*.*071 ARI + 0*.*203 INF + 0*.*233 VOC + 0*.*134 COM + 0*.*0003 SPA.

* + −14*.*227 est l’ordonné `a l’origine. C’est la valeur que prend théoriquement la variable *Ouverture* lorsque les autres variables prennent la valeur de 0.
  + 0*.*071 est le coefficient de régression partiel correspondant `à la variable *Arithmétique*. Lorsque la variable *Arithmétique* augmente d’une unité et que les autres variables in- dépendantes restent constantes, la variable *Ouverture* augmente de 0*.*071. L’ouverture augmente lorsque les résultats au test d’arithmétique augmentent.
  + 0*.*203 est le coefficient de r´egression partiel correspondant `a la variable *Information*. Lorsque la variable *Information* augmente d’une unit´e et que les autres variables in- d´ependantes restent constantes, la variable *Ouverture* augmente de 0*.*203. L’ouverture augmente lorsque les r´esultats au test d’information augmentent.
  + 0*.*233 est le coefficient de r´egression partiel correspondant `a la variable *Vocabulaire*. Lorsque la variable *Vocabulaire* augmente d’une unit´e et que les autres variables ind´e- pendantes restent constantes, la variable *Ouverture* augmente de 0*.*233. L’ouverture augmente lorsque les r´esultats au test de vocabulaire augmentent.
  + 0*.*134 est le coefficient de r´egression partiel correspondant `a la variable *Comprehension*. Lorsque la variable *Comprehension* augmente d’une unit´e et que les autres variables ind´ependantes restent constantes, la variable *Ouverture* augmente de 0*.*134. L’ouver- ture augmente lorsque les r´esultats au test de comprehension augmentent.
  + 0*.*0003 est le coefficient de r´egression partiel correspondant `a la variable *Spatial*. Lorsque la variable *Spatial* augmente d’une unit´e et que les autres variables ind´ependantes restent constantes, la variable *Ouverture* augmente de 0*.*0003. L’ouverture n’augmente pas lorsque les r´esultats au test d’aptitudes spatiales augmentent.
  + *Erreur standard*

*σ*ˆ = 7*.*724. Si les erreurs se distribuaient normalement, 68% des observations se situe- raient dans une bande de ±7*.*724 autour du plan d´efini par le mod`ele. Le nuage des observations est relativement proche du plan d´efini par le mod`ele.

1. Quelle est la valeur du coefficient de corrélation multiple au carré (coefficient de détermination) ?

*Coefficient de détermination*

*R*2 = 0*.*481. Le coefficient de corr´elation multiple au carr´e repr´esente la part de variance du crit`ere expliqu´ee par les pr´edicteurs. Ici, 48*.*1% de la variance totale de la variable *Ouverture* est expliqu´ee par les cinq variables explicatives.

R2ajusté= 0*.*444. Le coefficient de r´egression ajust´e au carr´e est toujours plus petit ou egal à *R*2. Sa valeur est influenc´ee par le nombre de variables explicatives introduites dans le modèle. Plus le nombre de prédicteurs est grand, plus R2ajusté sera petit par rapport à R2.

1. La régression multiple est-elle significative au seuil de 5% ?

*Test globale*

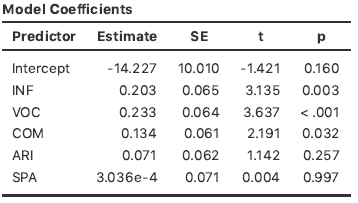
Hypothèses :

Conclusion :

Globalement, la liaison entre la variable crit`ere *Ouverture* et les variables explicatives *Arithm´etique*, *Information*, *Vocabulaire*, *Compr´ehension* et *Spatial* est significative au seuil de 5% (*F* (5*,* 69) = 12*.*806, *p <* 0*.*001). Il est possible, partiellement au moins, de rendre compte de l’*Ouverture* `a partir de la connaissance des variables *Arithm´etique*, *Information*, *Vocabulaire*, *Compr´ehension* et *Spatial*.

1. Quelles sont les variables qui jouent un rôle significatif au seuil de 5% dans notre modèle ?

*Test des coefficients de régression partiels*



1. Test du coefficient de r´egression partiel associ´e `a la variable *Arithm´etique*
   * R´esultat

*t*(69) = 1*.*142, *p* = 0*.*257.

* + Conclusion

Comme *p > α*, on ne rejette pas l’hypoth`ese nulle affirmant que *β*1 = 0.

1. Test du coefficient de r´egression partiel associ´e `a la variable *Information*
   * R´esultat

*t*(69) = 3*.*135, *p* = 0*.*003.

* + Conclusion

Comme *p < α*, on rejette l’hypoth`ese nulle affirmant que *β*2 = 0.

1. Test du coefficient de r´egression partiel associ´e `a la variable *Vocabulaire*
   * R´esultat

*t*(69) = 3*.*637, *p* = 0*.*001.

* + Conclusion

Comme *p < α*, on rejette l’hypoth`ese nulle affirmant que *β*3 = 0

1. Test du coefficient de r´egression partiel associ´e `a la variable *Compr´ehension*
   * R´esultat

*t*(69) = 2*.*191, *p* = 0*.*032.

* + Conclusion

Comme *p < α*, on rejette l’hypoth`ese nulle affirmant que *β*4 = 0.

1. Test du coefficient de r´egression partiel associ´e `a la variable *Spatial*
   * R´esultat

*t*(69) = 0*.*004, *p* = 0*.*997.

* + Conclusion

Comme *p > α*, on ne rejette pas l’hypoth`ese nulle affirmant que *β*5 = 0.

**Interpr´etation** : Le mod`ele lin´eaire est significatif mais il semble que seules les aptitudes d’information, de vocabulaire et de compr´ehension ont un effet sur l’ouverture du BIG-5. Cela corrobore les trouvailles de Ashton, Lee et Vernon (2000).

1. Comparez le modèle complet estimé au point c) à un deuxième modèle où vous supprimez la variable explicative ARI et à un troisième modèle où vous supprimez INF. Pour chaque test (il y en a deux), prenez comme seuil de signification *α* = 5%.

*Comparaison du modèle complet aux deux modèles réduits d’un prédicteur*

1. Suppression de la variable *Arithm´etique*

Hypothèses :

Conclusion :

Le test de Fisher permettant de comparer deux mod`eles emboˆıt´es n’est pas signi- ficatif au seuil de 5% (*F* (1*,* 69) = 1*.*304, *p* = 0*.*257). Le mod`ele r´eduit – sans la variable *Arithm´etique* – ou` *β*1 = 0 est meilleur que le mod`ele complet : le mo- d`ele r´eduit est plus parcimonieux et d´ecrit les donn´ees presque aussi bien que le mod`ele complet. Dans le mod`ele complet, le test d’arithm´etique n’influence pas l’ouverture.

1. Suppression de la variable *Information*

Hypothèses :

Conclusion :

Le test de Fisher permettant de comparer deux mod`eles emboˆıt´es est significatif au seuil de 5% (*F* (1*,* 69) = 9*.*828, *p* = 0*.*003). Le mod`ele complet est meilleur que le mod`ele sans le test d’information ou` *β*2 = 0. Cela signifie que dans le mod`ele complet de r´egression, le test d’information influence de mani`ere non n´egligeable l’ouverture.

1. Comparez les résultats obtenus en f) à ceux obtenus en g).

*Comparaison des tests marginaux et des tests de modèles emboités*

Le test portant sur le coefficient de r´egression partiel associ´e `a la variable *Xj* est ´equivalent au test qui compare le mod`ele complet au mod`ele obtenu par la suppression de l’unique variable ind´ependante *Xj* . Il existe en effet une correspondance entre les statistiques *t*(*ν*) et *F* (1*, ν*) :

(*t*(*ν*))2 = *F* (1*, ν*).

Par exemple, pour la variable *Information*, on a :

(*t*(69))2 = 3*.*1352 = 9*.*828 = *F* (1*,* 69).

1. Dans leur étude, Ashton, Lee et Vernon nous disent que l’*Ouverture* corrèle uniquement avec les facteurs d’intelligence cristallisée (à l’exception de l’arithmétique). Construisez donc le modèle proposé par les auteurs :

OUV ∼ INF + VOC + COM

1. Est-ce que ce nouveau modèle est emboîté dans le modèle estimé en c) ?

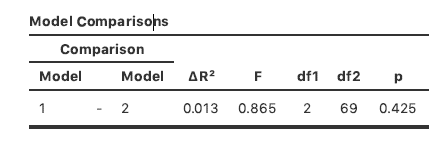
*Critère d’emboîtement*

Oui ils sont emboˆıt´es ! Toutes les variables du mod`ele r´eduit sont pr´esentes dans le mod`ele complet :

{INF*,* VOC*,* COM} ⊆ {ARI*,* INF*,* VOC*,* COM*,* SPA}

1. Comparez les deux modèles à l’aide d’un test, que concluez-vous ?

*Comparaison du modèle complet `a un modèle parcimonieux*



Modèle réduit

Hypothèses : 

Conclusion :

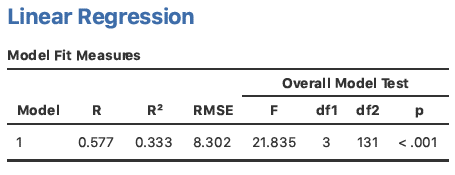
Le test de Fisher permettant de comparer deux mod`eles emboˆıt´es n’est pas significatif au seuil de 5% (*F* (2*,* 69) = 0*.*865, *p* = 0*.*425). Le mod`ele r´eduit est donc pr´ef´erable au mod`ele complet, il d´ecrit les donn´ees aussi bien que le mod`ele complet. En effet, la part de variance expliqu´ee par chacun des mod`eles est presque la mˆeme alors que le mod`ele r´eduit est plus simple !

L’*Ouverture* ne d´epend donc ni des performances arithm´etiques ni des performances spatiales. Les r´esultats de nos analyses confirment les r´esultats de Ashton, Lee et Vernon (2000) : l’ouverture est li´ee `a l’intelligence cristallis´ee (`a l’exception de l’arith- m´etique) mais pas `a l’intelligence fluide.

## Exercice 2

Utilisez l’ensemble de données Mireault.csv pour examiner la relation entre le niveau actuel de dépression et les autres variables. Un modèle plausible pourrait suggérer que la dépression (Depres) est fonction des éléments suivants : (1) la perception actuelle par l’individu de sa propre vulnérabilité vis-à-vis de toute perte additionnelle (VulPer), (2) le niveau de soutien social dont bénéficie la personne (SupTot) et (3) l’âge auquel la personne a perdu un de ses parents durant l’enfance (AgePer).

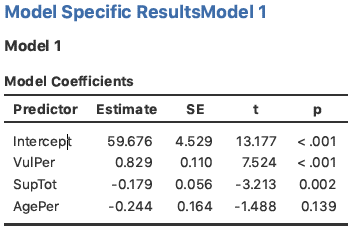
1. Évaluez le modèle esquissé ci-dessus.

*Ajustement du modèle*

Test globale de la r´egression

H0 : R2 = 0

H1 : R2 ≠ 0

*F* (3*,* 131) = 21*.*84, *p <* 0*.*001. La degr´e de d´epression d’une personne d´epend significativement au seuil de 5% de sa vuln´erabilit´e vis-`a-vis d’une perte additionnelle, du soutien social dont elle b´en´eficie ainsi que de l’ˆage auquel elle a perdu l’un de ses parents durant l’enfance. La part de variance expliqu´ee vaut 33.3%. Nous voyons cependant que l’erreur standard est assez importante (*σ*ˆ = 8*.*427).

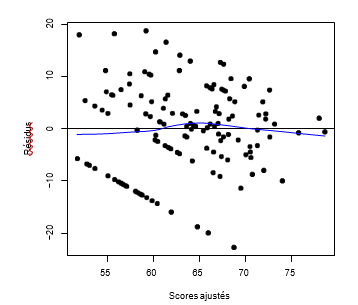
Evaluation des coefficients de régression partiels

Les coefficients de r´egression *β*1 correspondant `a VulPer (*β*1 = 0*.*829, *t*(131) = 7*.*524*, p <* 0*.*001) et *β*2 correspondant `a SupTot (*β*2 = −0*.*179*, t*(131) = −3*.*213*, p* = 0*.*002) sont significativement diff´erents de 0 au seuil de 5%.

Par contre nous constatons que le mod`ele r´eduit excluant la variable AgePer est aussi bon que le mod`ele complet (*β*3 = −0*.*244*, t*(131) = −1*.*488*, p* = 0*.*139).

1. Les données utilisées pour construire ce modèle satisfont-elles les conditions nécessaires à la mise en application d’une régression multiple ?
   * Vérifiez que la moyenne des résidus est égale à 0 conditionnellement à .

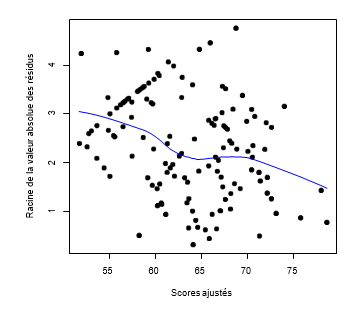
Linéarité

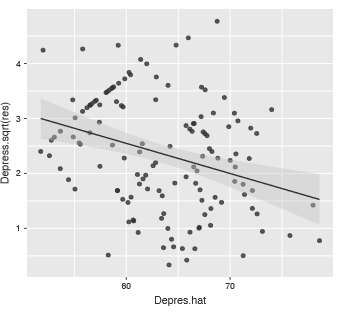


La moyenne des r´esidus vaut 0 pour grosso modo tous les scores ajust´es (la r´egression polynomiale locale suit relativement bien l’horizontale passant par 0).

* + La variance des erreurs est-elle constante ?

Homoscédasticité





La dispersion des résidus semble dépendre de la valeur des scores ajustés . En effet, lorsque la valeur des scores ajustés augmente, la variance résiduelle diminue.

Notons qu’en cas d’hétéroscédasticité les estimations des paramètres sont toujours fiables par contre celles des erreurs standards peuvent être fausses.

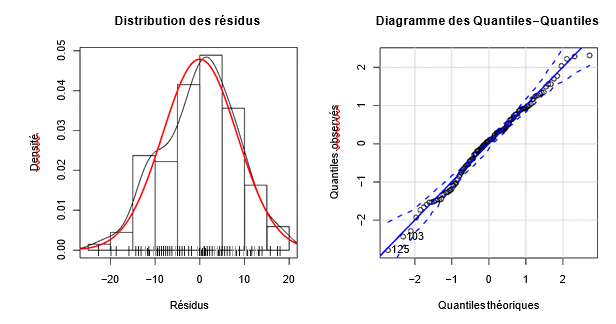
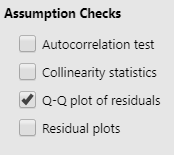
* + Les résidus se distribuent-ils normalement ? Pour répondre à cette question, utilisez deux approches différentes.

Normalité

1. Diagnostic graphique

1. Graphique: Linear regression > Assumption Checks > Q-Q plot of residuals





Nous voyons sur ces graphiques que les r´esidus suivent bien une distribution de Laplace-Gauss. Cela corrobore l’hypoth`ese selon laquelle la distribution des *ε* est normale.

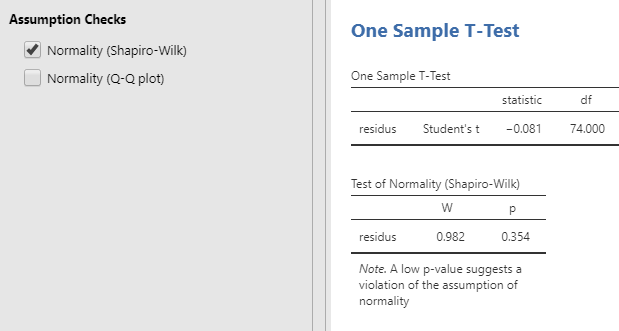
2. Test de normalité de Shapiro Wilk

1. Shapiro-Wilk: T-tests > One sample T-test > Residus in *Dependant Variables* > Contrôler hypotheses > Assumption Checks > Normality (Shapiro-Wilk)



*H*0 : La distribution des r´esidus suit une loi normale.

*H*1 : La distribution des r´esidus ne suit pas une loi normale.

La valeur empirique de la variable de d´ecision du test de Shapiro-Wilk vaut *W* = 0*.*992. *p* = 0*.*611. Comme *p > α*, nous acceptons *H*0. Les r´esidus se distribuent bien selon une loi de Laplace-Gauss. En cons´equence, les erreurs *ε* suivent une loi normale.



## Exercice 3

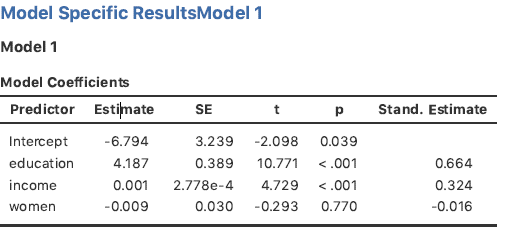
Lorsque les prédicteurs sont incommensurables, il est courant de les standardiser avant de construire un modèle de régression. La comparaison de l’impact des prédicteurs sur la variable dépendante est alors plus aisée.

1. À partir des données du fichier PrestigeCA.csv, construisez le modèle qui permette de prédire le prestige d’une profession à partir du niveau d’instruction des personnes qui l’exercent, de leur revenu moyen et de la proportion de femmes qui pratiquent le métier.

*Estimation des paramètres du modèle (Variables non-standardisées)*

A` partir de l’estimation des param`etres, le mod`ele s’´ecrit :

prestige = −6*.*794 + 4*.*187 education + 0*.*001 income − 0*.*009 women.



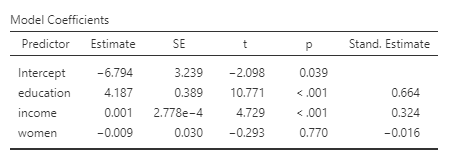
1. Construisez un modèle similaire au précédent mais en standardisant au préalable toutes les variables (prédicteurs et critère).

JAMOVI :

Ajouter colonne «stdY» (Y = Variable indépendant/ à expliquer)

avec :

Fx = SCALE(Y) (=fonction qui standardise)

Pour construire le modèle des βj\*: Linear regression > Model coefficients > **Stand. Estimate** : Standardized Estimate



*Estimation des paramètres du modèle (Variables standardisées)*

Apr`es avoir standardis´e toutes les variables, le mod`ele devient :

prestige∗ = 0*.*664 education∗ + 0*.*324 income∗ − 0*.*016 women∗.

* + Lorsque les variables income et women sont maintenues constantes, une hausse d’un ´ecart-type de la variable education se traduit sur la variable prestige, en moyenne, par une hausse de 0*.*664 ´ecart-type.
  + Lorsque les variables education et women sont maintenues constantes, une hausse d’un ´ecart-type de la variable income se traduit sur la variable prestige, en moyenne, par une hausse de 0.324 ´ecart-type.
  + Lorsque les variables education et income sont maintenues constantes, une hausse d’un ´ecart-type de la variable women se traduit sur la variable prestige, en moyenne, par une baisse de 0.016 ´ecart-type.

L’impact de l’´education sur le prestige d’une profession est deux fois plus important que le revenu et pr`es de 40 fois plus important que la “f´eminit´e” de la profession.

1. Comparez les deux modèles.

*Comparaison des modèles*

En utilisant des variables standardis´ees, on met toutes les variables sur un mˆeme pied d’´egalit´e. En effet toutes les variables standardis´ees poss`edent des moyennes nulles et des variances unitaires. Il est ainsi plus facile de comparer l’effet de chaque variable explicative sur la variable r´eponse.

Par ailleurs le test global et les tests portant sur la significativit´e des coefficients par- tiels de r´egression conduisent aux mˆemes conclusions que les variables explicatives soient standardis´ees ou non.

1. Le modèle qui permet d’exprimer la variable *Y* en fonction des variables non standardisées

*X*1, *X*2 et *X*3 s’écrit de la manière suivante :

*Y* = *B*0 + *B*1*X*1 + *B*2*X*2 + *B*3*X*3 + *ε*

Celui qui permet d’exprimer la même réalité mais à l’aide des variables standardisées s’écrit :

Zy = β0 + β1Z1 + β2Z2 + β3Z3 + ε

Avec Zy = et Zj = pour j = 1, 2, 3

Comment calculer les coefficients *β*0, *β*1, *β*2 et *β*3, à partir des coefficients *B*0, *B*1, *B*2 et *B*3 ? Vérifiez l’exactitude de vos formules sur les modèles construits aux deux premiers points de cet exercice.

*Coefficients de r´egression standardis´es et coefficients de r´egression bruts*

*Y*ˆ = *B*0 + *B*1 *X*1 + *B*2 *X*2 + *B*3 *X*3 (1)

Comme *B*0 = *Y*¯ − *B*1 *X*¯1 − *B*2 *X*¯2 − *B*3 *X*¯3, l’´equation (1) devient :

*Y*ˆ =

*Y*¯ − *B*1 *X*¯1 − *B*2 *X*¯2 − *B*3 *X*¯3 + *B*1 *X*1 + *B*2 *X*2 + *B*3 *X*3 (2)

Apr`es d´eplacements et mises en ´evidence, l’´equation (2) s’´ecrit :

*Y*ˆ − *Y*¯

= *B*1 (*X*1 − *X*¯1) + *B*2 (*X*2 − *X*¯2) + *B*3 (*X*3 − *X*¯3) (3)

En multipliant les termes du membre de droite de l’´equation (3) par tivement, on obtient l’expression :

*s*1 , *s*2 *s*1 *s*2

et *s*3

*s*3

respec-

*Y*ˆ − *Y*¯

= *B*1 *s*1

(*X*1 − *X*¯1)

*s*1

+ *B*2 *s*2

(*X*2 − *X*¯2)

*s*2

+ *B*3 *s*3

(*X*3 − *X*¯3)

*s*3

(4)

En divisant chaque membre de l’´equation (4) par l’´ecart-type de la variable *Y* , l’´equa- tion (4) devient :

*Y*ˆ − *Y*¯

*sy*

*s*1

= *B*1

*s*

*y*

(*X*1 − *X*¯1) + *B s*2

*s*1 2 *sy*

(*X*2 − *X*¯2) + *B s*3

*s*2 3 *sy*

(*X*3 − *X*¯3)

*s*3

(5)

En recourant `a la définition d’une variable standardisée, il est possible de simplifier l’écriture de l’´équation (5) :

*s*1

ˆ

*Zy* = *B*1

*s*

*y*

*s*2

*Z*1 + *B*2

*s*

*y*

*s*3

*Z*2 + *B*3 *Z*3

*s*

*y*

Par ailleurs comme *Z*ˆ*y* = *β*0 + *β*1 *Z*1 + *β*2 *Z*2 + *β*3 *Z*3, on d´eduit par correspondance que :

*β*0 = 0*, β*1 = *B*1

*s*1

*, β*2 = *B*2

*s*

*y*

*s*2 *s*3

*, β*3 = *B*3 *.*

*s*

*y sy*

Utilisons ces formules pour calculer les coefficients de régression standardisés `a partir des coefficients de régression bruts :

* Premier coefficient de régression standardisé *β*1

*s*1

*β*1 = *B*1

*sy*

2*.*728

= 4*.*187

= 0*.*664

17*.*204

* Deuxième coefficient de régression standardisé *β*2

*s*2

*β*2 = *B*2

*sy*

4245*.*922

= 0*.*001

= 0*.*324

17*.*204

* Troisième coefficient de régression standardisé *β*3

*s*3

*β*3 = *B*3

*s*

*y*

31*.*725

= −0*.*009 17*.*204

= −0*.*016

Les r´esultats de ces calculs sont les mˆemes que les r´esultats des estimations faites au point b) de cet exercice.

# Série 5

## Exercice 1

À partir de l’examen des diagrammes d’interaction représentés ci-dessous, dites s’il y a un effet principal de A, un effet principal de B, ou un effet d’interaction.

1. Afin de savoir s’il y a un effet principal de A, il faut d´eterminer la position de *µ*1*.* ainsi que celle de *µ*2*.* :

La valeur de µ1. correspond `a l’ordonn´ee du milieu du segment d´eﬁni par les points (a1,b1) et (a1,b2). La valeur de µ2. correspond `a l’ordonn´ee du milieu du segment d´eﬁni par les points (a2,b1) et (a2,b2). Si µ1. = µ2., alors il n’y a pas d’eﬀet principal de A. Si au contraire µ1. ≠ µ2., alors il y a un eﬀet principal de A.

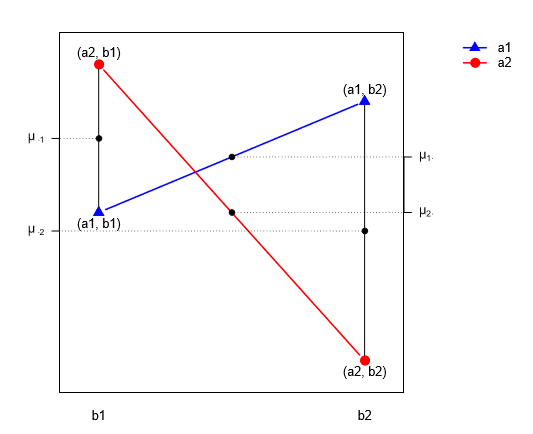
1. Concernant l’eﬀet principal de B, il faut déterminer la position de µ.1 et de µ.2 :

La valeur de *µ.*1 correspond `a l’ordonn´ee du milieu du segment d´efini par les points (*a*1*, b*1) et (*a*2*, b*1). La valeur de *µ*.2 correspond `a l’ordonn´ee du milieu du segment d´efini par les points (*a*1*, b*2) et (*a*2*, b*2).

Si *µ*.1 = *µ*.2, alors il n’y a pas d’effet principal de B. Si au contraire *µ*.1 ≠ *µ*.2, alors il y a un effet principal de B.

1. L’effet d’interaction quant `a lui s’observe si les deux droites d´efinies par (*a*1*, b*1) et (*a*1*, b*2), d’une part, et par (*a*2*, b*1) et (*a*2*, b*2), d’autre part, ne sont ni confondues ni parall`eles. Les droites n’ont pas besoin de se croiser pour qu’il y ait un effet d’interaction, il suffit qu’elles ne soient ni parall`eles ni confondues.

Voici un exemple :



On remarque que µ1. ≠ µ2. : il y a donc un eﬀet de A. De même, µ.1 ≠ µ.2 : il y a donc un eﬀet de B. Finalement, comme les droites ne sont ni parallèles ni confondues, il y a un eﬀet d’interaction.

a)

●

●

a1



~~●~~

a2 Effet de A

Effet de B

Effet d’interaction

b1 b2

b)

●

●

a1



~~●~~

a2 Effet de A

Effet de B



Effet d’interaction



b1 b2

c)

●

●

a1

~~●~~

a2 Effet de A

Effet de B

Effet d’interaction



b1 b2

d)

●

●

a1

~~●~~

a2 Effet de A

Effet de B



Effet d’interaction

b1 b2

e)

●

●

a1

~~●~~

a2 Effet de A

Effet de B



Effet d’interaction



b1 b2

f)

●

●

a1



~~●~~

a2 Effet de A

Effet de B



Effet d’interaction



b1 b2

g)

●

●

a1

~~●~~

a2 Effet de A

Effet de B



Effet d’interaction



b1 b2

h)

●

●

a1

~~●~~

a2 Effet de A

Effet de B



Effet d’interaction

b1 b2

i)

●

●

a1



~~●~~

a2 Effet de A

Effet de B

Effet d’interaction



b1 b2

j)

●

●

a1



~~●~~

a2 Effet de A

Effet de B



Effet d’interaction

b1 b2

k)

●

●

a1

~~●~~

a2 Effet de A

Effet de B



Effet d’interaction



b1 b2

l)

●

●

a1



~~●~~

a2 Effet de A

Effet de B

Effet d’interaction



b1 b2

## Exercice 2

Des chercheurs veulent étudier le stress perçu au sein d’une grande entreprise de restauration. Pour cela, les chercheurs repèrent les employés suivant leur âge (moins de 50 ans ou plus de 50 ans) et leur statut professionnel (équipier ou manager), deux facteurs qui sont susceptibles d’influencer le niveau de stress perçu. On demande à 32 employés (sélectionnés de manière aléatoire dans chaque catégorie) de passer le PSS*−*10 (*Perceived Stress Scale*), qui est un des tests d’évaluation du stress perçu le plus utilisé. Le score total à ce test se situe entre 0 et 40 (un score de 40 traduit un stress perçu très élevé).

*Score total au PSS−10*

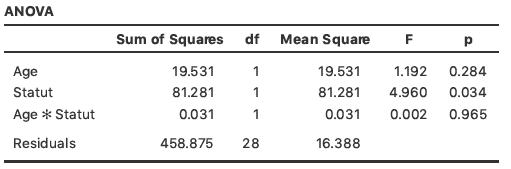
Statut

|  |  |  |
| --- | --- | --- |
| Âge | Équipier | Manager |
|  | 25 | 27 |
|  | 27 | 23 |
|  | 23 | 32 |
| Moins de 50 ans | 30 | 34 |
|  | 21 | 35 |
|  | 26 | 24 |
|  | 31 | 26 |
|  | 22 | 29 |
|  | 22 | 23 |
|  | 25 | 27 |
|  | 29 | 30 |
| Plus de 50 ans | 21 | 32 |
|  | 19 | 28 |
|  | 26 | 21 |
|  | 30 | 26 |
|  | 20 | 31 |

Faites une analyse de variance à deux facteurs. Testez l’effet propre de chaque facteur, à savoir l’âge et le statut, ainsi que leur interaction sur le niveau de stress perçu. Autrement dit, est-ce que l’âge, le statut ou l’interaction entre les deux, a un effet significatif au seuil 5% sur le niveau de stress perçu ? Afin de répondre à ces questions, construisez au préalable la table de l’ANOVA.

*Construction de la table de l’ANOVA*

a = 2 b = 2 n= 8



Effet du facteur A (Aˆ ge)

1. **Hypothèses :** H0 : µ1·· = µ2··

H1 : µ1·· ≠ µ2··

1. **Seuil :** α = 0.05
2. **Valeur empirique :** Femp = 1.192
3. **Probabilité critique :** p = Prob(F (a − 1, ab(n − 1)) ≥ Femp) = Prob(F (1, 28) ≥ 1.192) = .284
4. **Conclusion :**

On ne rejette pas l’hypothèse H0 car p > α. Le niveau de stress perçu n’est pas influencé par l’ˆâge. Que les employés aient moins de 50 ans ou plus de 50 ans, leur niveau de stress est le même.

Effet du facteur B (Statut)

1. **Hypothèses :** H0 : (∀*j*)(∀*j’*) *µ·j·* = *µ·j’·*

H1 : (∃*j*)(∃*j’*) *µ·j·* ≠ *µ·j’·*

1. **Seuil :** α = 0.05
2. **Valeur empirique :** Femp = 4.960
3. **Probabilité critique :** p = Prob(F (b − 1, ab(n − 1)) ≥ Femp) = Prob(F (1, 28) ≥ 1.192) = .284
4. **Conclusion :**

On rejette l’hypoth`ese *H*0 car *p < α*. Le stress per¸cu est influenc´e par le statut professionnel. Le niveau de stress des managers est plus ´elev´e que celui des ´equipiers.

Effet de l’interaction entre A et B (Age et Statut)

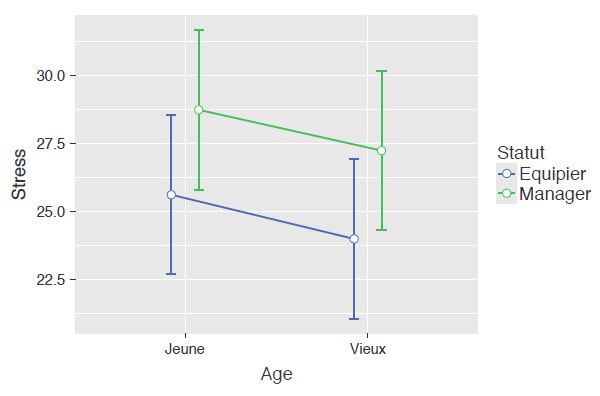
1. **Hypothèses :** H0 : (∀*i*)(∀*i’*)(∀*j*)(∀*j’*) *µij·* − *µij’·* = *µi’j·* − *µitj’·*

H1 : (∃*i*)(∃*i’*)(∃*j*)(∃*j’*) *µij·* − *µij’·* ≠ *µi’j·* − *µitj’·*

1. **Seuil :** α = 0.05
2. **Valeur empirique :** Femp = 0.002
3. **Probabilité critique :** p = Prob(F ((a − 1)(b − 1)), ab(n − 1) ≥ Femp) = Prob(F (1, 28) ≥ 0.002) = .965
4. **Conclusion :**

On ne rejette pas l’hypoth`ese *H*0 car *p > α*. Il n’y a pas d’effet d’interaction.

*Diagramme d’interaction*



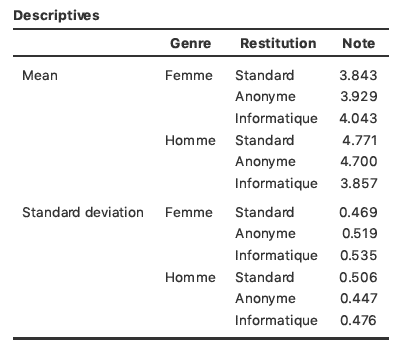
## Exercice 3

Des chercheurs s’intéressent à l’impact du genre sur les notes obtenues par des élèves de gymnase à un test de mathématiques. L’hypothèse est que la présence du stéréotype selon lequel les femmes seraient moins douées que les hommes en mathématiques pourraient pousser l’enseignant à donner de meilleures notes aux hommes dans cette branche, que ce soit de manière consciente ou non. Les chercheurs ont donc récolté les notes de 42 élèves à une épreuve de mathématiques, et ont formé des groupes sur la base de deux critères : le genre et le mode de restitution de l’épreuve. Le mode de restitution pouvait être “standard” : la copie est manuscrite et le nom figure en haut de la feuille. La restitution pouvait se faire sur le mode “anonyme” : dans ce cas la copie est également manuscrite, mais c’est un code à la place du nom qui permet d’identifier l’élève. Finalement, avec le mode “informatique”, la copie est tapée à l’ordinateur, et le nom de l’élève est également codé. Les chercheurs supposent que ces différents modes de restitution permettent d’empêcher, avec une effi respectivement croissante, d’identifier le genre de l’élève lors de la correction de l’épreuve. Les chercheurs supposent en eff t que bien qu’anonyme, les copies manuscrites pourraient laisser aux correcteurs la possibilité d’identifier le genre à travers l’écriture. Les notes obtenues se trouvent dans le tableau suivant :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Genre | Mode de restitution | Notes obtenues | | | | |  | |
| Femme | Standard | 3.5 | 4.1 | 4.2 | 3.0 | 4.1 | 3.7 | 4.3 |
| Femme | Anonyme | 3.3 | 4.2 | 3.7 | 4.1 | 4.2 | 4.7 | 3.3 |
| Femme | Informatique | 4.8 | 4.0 | 3.4 | 3.6 | 4.4 | 3.6 | 4.5 |
| Homme | Standard | 5.2 | 4.6 | 4.9 | 5.3 | 5.0 | 3.8 | 4.6 |
| Homme | Anonyme | 4.5 | 5.0 | 5.2 | 4.0 | 5.1 | 4.8 | 4.3 |
| Homme | Informatique | 3.4 | 3.2 | 4.1 | 3.9 | 4.4 | 3.6 | 4.4 |

1. Calculez la moyenne et l’écart-type de la distribution des notes de chacun des sous-groupes.

*Résumé numérique*

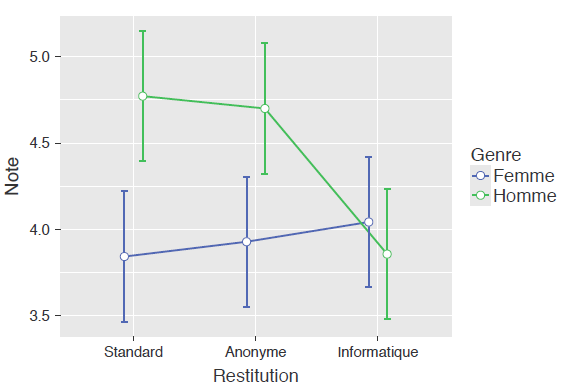


= écart-type

= moyenne

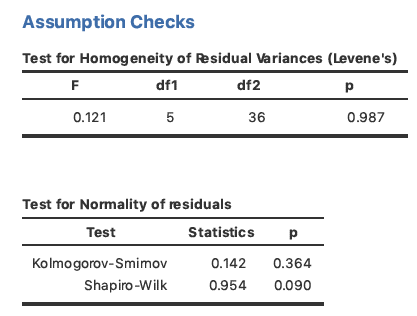
1. Représentez les notes obtenues par les élèves par un diagramme d’interaction.

*Diagramme d’interaction*



1. Expliquez la procédure à suivre si l’on voulait vérifier les conditions d’application d’une analyse de variance sur ces données.

*Conditions d’application*



Levene’s = Sous Anova > Assumptions Checks

Le prof fait aussi un QQPLot pour verifier la loi normale !! (Aussi assumptions checks)

Pour appliquer une analyse de variance `a deux facteurs, deux conditions doivent ˆetre respect´ees. Il faut tout d’abord que les distributions suivent une loi normale. Ceci peut ˆetre v´erifi´e en appliquant un test de Shapiro-Wilk sur la distribution des r´esidus. Il faut également que les variances soient homog`enes. Nous pourrions utiliser par exemple un test de Levene afin de v´erifier ce point. Si ces deux conditions sont respect´ees, nous pouvons alors r´ealiser l’analyse de variance. Afin de se concentrer sur la r´ealisation de l’analyse de variance, nous supposerons que ces conditions sont satisfaites.

1. Effectuez une analyse de variance au seuil de 5% afin de vérifier s’il y a bien des différences entre les groupes. Déterminez la taille de chacun des effets à l’aide de *η*2 et de *η*2.

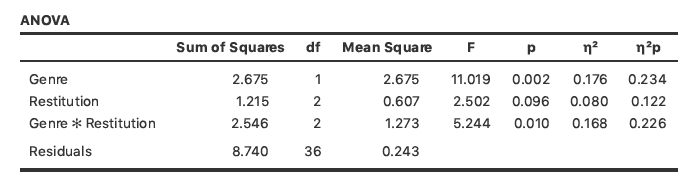
*p*

*Analyse de variance `a deux facteurs*

Nous nous demandons si le genre et le mode de restitution de l’´epreuve peuvent influencer la note obtenue. Calculons les valeurs interm´ediaires avec *a* = 2, *b* = 3, et *n* = 7 :

Nous avons les sommes des carr´es suivantes :

Cela nous donne la table de l’ANOVA :



Taille d’effet : *η*2

Ex : Genre : 17.6% de la variance expliquée est expliquée par l’effet de genre.

Effet du facteur A (Genre)

1. **Hypothèses :** H0 : µ1·· = µ2··

H1 : µ1·· ≠ µ2··

1. **Seuil :** α = 0.05
2. **Valeur empirique :** Femp = 11.019
3. **Probabilité critique :** p = Prob(F (a − 1, ab(n − 1)) ≥ Femp) = Prob(F (1, 36) ≥ 11.019) = 0.002
4. **Conclusion :**

On rejette H0 car p < α. Le genre influence la note obtenue à l’épreuve.

Effet du facteur B (Statut)

1. **Hypothèses :** H0 : (∀*j*)(∀*j’*) *µ·j·* = *µ·j’·*

H1 : (∃*j*)(∃*j’*) *µ·j·* ≠ *µ·j’·*

1. **Seuil :** α = 0.05
2. **Valeur empirique :** Femp = 2.502
3. **Probabilité critique :** p = Prob(F (b − 1, ab(n − 1)) ≥ Femp) = Prob(F (2, 36) ≥ 2.502) = .284
4. **Conclusion :**

On rejette l’hypoth`ese *H*0 car *p < α*. Le stress per¸cu est influenc´e par le statut professionnel. Le niveau de stress des managers est plus ´elev´e que celui des ´equipiers.

Effet de l’interaction entre A et B (Age et Statut)

1. **Hypothèses :** H0 : (∀*i*)(∀*i’*)(∀*j*)(∀*j’*) *µij·* − *µij’·* = *µi’j·* − *µitj’·*

H1 : (∃*i*)(∃*i’*)(∃*j*)(∃*j’*) *µij·* − *µij’·* ≠ *µi’j·* − *µi’j’·*

1. **Seuil :** α = 0.05
2. **Valeur empirique :** Femp = 5.244
3. **Probabilité critique :** p = Prob(F ((a − 1)(b − 1)), ab(n − 1) ≥ Femp) = Prob(F (2, 36) ≥ 5.224) = .010
4. **Conclusion :**

On rejette H0 car p < α. Il y a un eﬀet de l’interaction entre le genre et le mode de restitution sur les notes. Cela signiﬁe que les ﬁlles et les gar¸cons ne sont pas not´es de la mˆeme mani`ere selon le mode de restitution.

1. Discutez les résultats de vos analyses, notamment en les comparant au graphique obtenu en b).

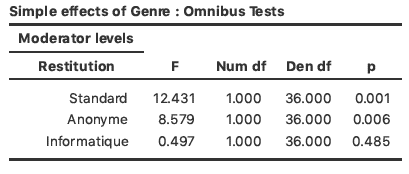
L’eﬀet principal du genre est signiﬁcatif (F(1,36) = 11.019, p = 0.002, η2 p = 0.234). Globalement, les femmes et les hommes ne reçoivent pas les mêmes notes.

En moyenne, il n’y a pas d’inﬂuence du mode de restitution (F(2,36) = 2.502, p = 0.096, η2 p = 0.122). Cela peut s’observer sur le graphique ou` l’on remarque que les moyennes entre hommes et femmes pour les trois modes de restitution se situent à peu près sur une même horizontale.

L’eﬀet de l’interaction est signiﬁcatif (F(2,36) = 5.244, p = 0.010, η2 p = 0.226). La diﬀérence de performance entre les femmes et les hommes s’atténue lorsque le mode de restitution devient de plus en plus anonyme. Lorsque l’anonymat augmente, les notes des femmes s’améliorent alors que celles des hommes se péjorent. Cela pourrait vouloir dire que les femmes sont pénalisées et que les hommes sont favorisés.

1. Quels tests serait-il intéressant de réaliser afin d’affiner cette recherche ?

*Pour aller plus loin*



L’eﬀet de l’interaction étant signiﬁcatif, il est important de tester les eﬀets simples. On s’aperçoit alors que la diﬀérence entre les femmes et les hommes est signiﬁcative dans la condition ”Standard” (F(1, 36) = 12.431, p = .001), d’une part, et dans la condition ”Anonyme” (F(1, 36) = 8.579, p = .006), d’autre part. Par contre, dans la condition ”Informatique”, il n’y a plus de diﬀérence (F(1, 36) = 0.497, p = .485). Les moins bons résultats en mathématiques qu’obtiennent les femmes seraient donc dus `a un biais de correction. Lorsque l’évaluateur ne connaît pas le genre du répondant, il n’y a plus de diﬀérence entre les femmes et les hommes!

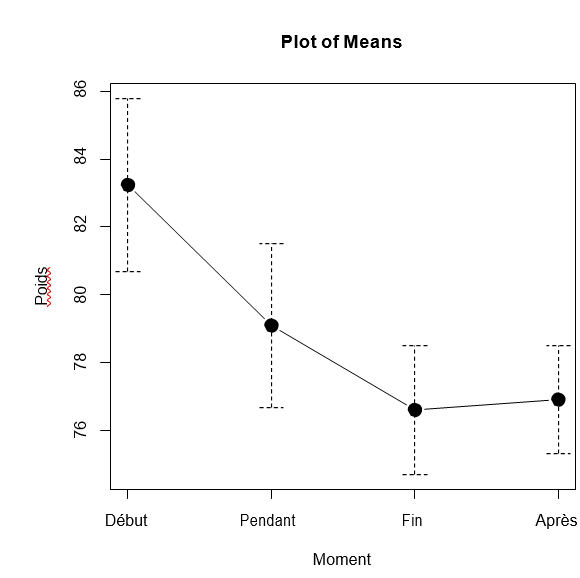
# Série 6

## Exercice 1

La montée fulgurante du nombre de personnes en surpoids aux États-Unis a amené une asso- ciation des troubles du comportement alimentaire de l’Arizona à s’intéresser au traitement de la boulimie. Des psychologues prirent en charge 15 sujets. Pour évaluer l’efficacité de leur interven- tion, ils enregistrèrent le poids des sujets avant le traitement (Poids0), au milieu du traitement (Poids1), à la fin du traitement (Poids2), puis 6 mois après la fin de la prise en charge (Poids3). Ces données sont présentées dans le tableau ci-dessous, elles sont également enregistrées dans le fichier Boulimie.csv.

Il n’y a qu’un facteur avec 4 modalités dont chaque individu fait partie des 4 modalités !

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sujet | Poids0 | Poids1 | Poids2 | Poids3 |
| 1 | 83.6 | 76.1 | 76.4 | 79.8 |
| 2 | 83.5 | 81.4 | 78.4 | 74.4 |
| 3 | 81.0 | 79.2 | 78.2 | 77.5 |
| 4 | 86.1 | 76.9 | 73.8 | 76.9 |
| 5 | 79.2 | 77.7 | 79.0 | 76.0 |
| 6 | 86.1 | 83.0 | 78.7 | 77.0 |
| 7 | 85.5 | 77.7 | 75.1 | 77.0 |
| 8 | 83.8 | 76.2 | 77.3 | 77.3 |
| 9 | 81.6 | 76.8 | 78.5 | 76.6 |
| 10 | 83.3 | 81.5 | 76.0 | 76.0 |
| 11 | 87.4 | 77.2 | 75.8 | 73.4 |
| 12 | 82.3 | 80.1 | 74.2 | 79.2 |
| 13 | 81.8 | 80.1 | 78.2 | 77.5 |
| 14 | 85.0 | 83.5 | 76.1 | 77.2 |
| 15 | 78.5 | 79.0 | 73.3 | 77.8 |

1. Représentez graphiquement l’évolution du poids moyen des 15 sujets ayant été suivis.

*Evolution du poids moyen des sujets*

1. Quel est le plan d’expérience utilisé ? Donnez l’équation du modèle.

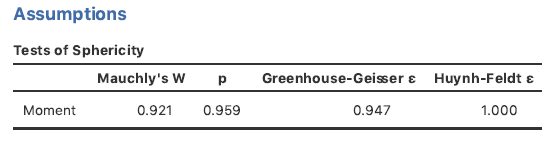
*Evolution du poids moyen des sujets*

Vu que l’´etude mesure le poids des mˆemes sujets `a 4 moments diff´erentes, le plan d’exp´erience est un plan `a mesures r´ep´et´ees (*S \* A*). L`’analyse a appliquer est donc une analyse de variance `a un facteur intra-sujet.

avec

1. L’hypothèse de sphéricité est-elle acceptable ?

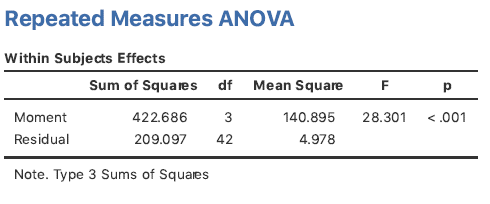
*Test de la sphéricité*



*W* = 0*.*921, *p* = 0*.*959. Comme *p > α*, on ne rejette pas *H*0 au seuil *α* = 5%. L’hypoth`ese de sph´ericit´e est corrobor´ee. Nous n’aurons pas besoin de corriger les r´esultats de l’analyse de variance.

1. Le poids des sujets a-t-il changé de manière significative durant la prise en charge ?

*Efficacité de la prise en charge*



*F* (3*,* 42) = 28*.*301, *p <* 0*.*001. Comme *p < α*, on rejette donc *H*0 au seuil *α* = 5%. Le poids moyen des personnes participant `a l’exp´erience n’est pas toujours le mˆeme, il ´evolue.

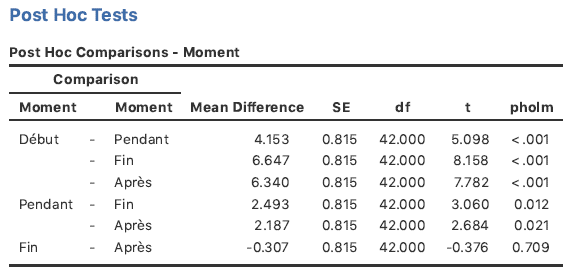
1. Que tirez-vous des comparaisons multiples réalisées selon la méthode de Holm ?

* Il faut cocher Holm dans Post Hoc !!! (Normalement c’est Tukey !)

*Comparaisons multiples*

Comme le facteur intra-sujet poss`ede 4 modalit´es, nous allons r´ealiser la comparaison de 6 paires de variables.

* + Tableau des r´esultats



* + Interpr´etation

Toutes les diff´erences sont statistiquement significatives au seuil de 5% `a l’exception d’une seule : celle qui compare le poids mesur´e en fin de traitement `a celui mesur´e six mois plus tard (*p* = 0*.*709). Du d´ebut `a la fin de l’intervention, les sujets perdent donc r´eguli`erement du poids : la perte de poids est significative entre le d´ebut et le milieu du traitement (*p < .*001) ; elle l’est aussi entre le milieu et la fin du traitement (*p* = *.*012). Par contre, une fois le traitement termin´e, leur poids ne change plus, il reste stable (*p* = 0*.*709).

## Exercice 2

**Mère fumeuse, enfant hyperactif...** Quand la mère fume pendant la grossesse, il est démontré que l’enfant a plus de chances d’être hyperactif. Des chercheurs néerlandais [1](#_bookmark2) font un pas de plus en vérifi t si cet eff se situe dans le cadre du concept de la *programmation fœtale*, selon lequel certains comportements de la femme enceinte s’inscrivent dans le programme de développement du fœtus, et résisteront par la suite à toute modification favorable de l’environnement.

Le *Good Behavior Game* (Jeu de bonne conduite) est utilisé dans les écoles primaires des Pays- Bas comme outil pour réduire l’hyperactivité, les conduites asociales et le recours précoce à la cigarette. Or cet outil ne s’avère efficace que sur les enfants de non-fumeuses : il réduit le nombre d’hyperactifs et, surtout, divise par deux le nombre de fumeurs précoces (dès 10-11 ans). En revanche, chez les enfants dont les mères étaient fumeuses pendant la grossesse, le *Good Behavior Game* n’a aucun eff ni sur l’hyperactivité ni sur l’âge de la première cigarette. C’est donc à juste titre, estiment les auteurs, que l’on essaie de dissuader les futures mères de fumer... Nous reportons ci-dessous les scores à une échelle d’hyperactivité de 24 enfants évalués successive- ment à 7, 8 et 9 ans. Tous bénéfi rent du programme mis au point pour réduire l’hyperactivité,

des conduites asociales et le recours précoce à la cigarette. Les mères de douze de ces enfants fumaient lors de leur grossesse, les autres pas.

*Exposé à la fumée durant la grossesse*

*Pas exposé à la fumée durant la grossesse*

1 13

|  |  |  |
| --- | --- | --- |
| 7 ans | 8 ans | 9 ans |
| 2.4 | 1.9 | 2.4 |
| 3.3 | 3.3 | 3.6 |
| 4.5 | 4.2 | 3.8 |
| 1.4 | 1.6 | 1.6 |
| 3.8 | 4.3 | 3.5 |
| 3.6 | 3.6 | 3.6 |
| 3.4 | 2.8 | 3.0 |
| 2.9 | 3.0 | 3.3 |
| 2.9 | 2.8 | 3.6 |
| 2.7 | 2.5 | 2.6 |
| 2.7 | 2.8 | 3.6 |
| 2.4 | 2.4 | 2.4 |

|  |  |  |
| --- | --- | --- |
| 7 ans | 8 ans | 9 ans |
| 2.1 | 1.5 | 1.7 |
| 3.3 | 2.4 | 2.3 |
| 3.2 | 3.5 | 2.9 |
| 2.2 | 2.1 | 1.8 |
| 2.1 | 1.4 | 1.6 |
| 2.3 | 1.3 | 1.0 |
| 1.5 | 1.6 | 1.4 |
| 4.7 | 3.9 | 3.9 |
| 3.0 | 2.2 | 2.6 |
| 2.2 | 1.8 | 1.9 |
| 2.2 | 1.5 | 1.5 |
| 2.7 | 2.5 | 2.5 |

2 14

3 15

4 16

5 17

6 18

7 19

8 20

9 21

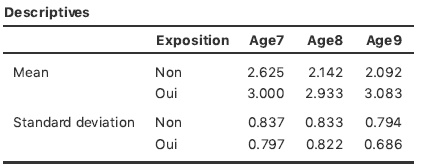
10 22

11 23

12 24

* 1. Résumez numériquement ces observations.

*Résumé numérique*

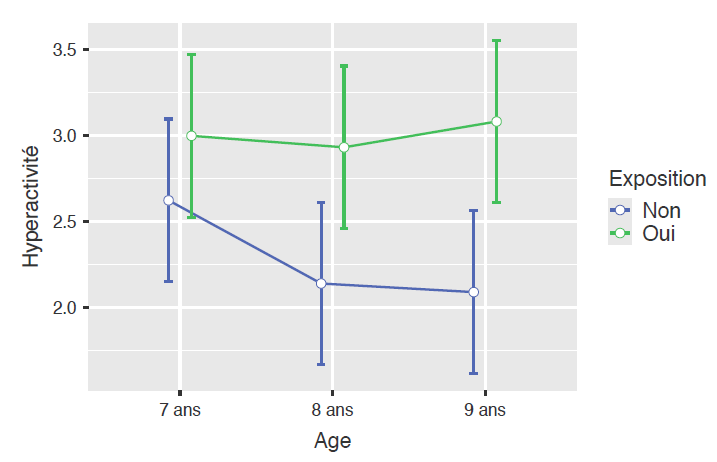


* 1. Représentez ces données à l’aide d’un diagramme d’interaction.

*Diagramme d’interaction*

2.8

3.0



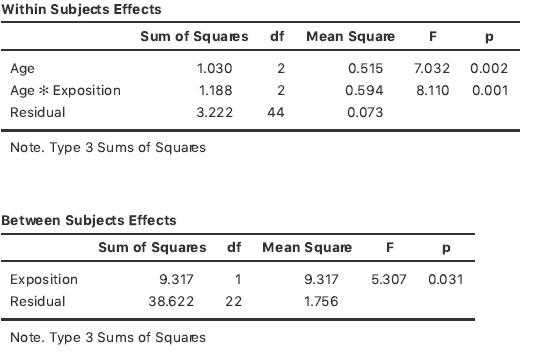
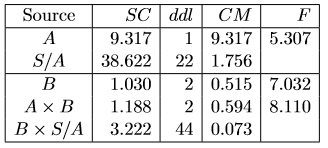
* 1. Faites une analyse de variance de ces données.

*Analyse de variance appliquée à un plan mixte*

Rappelons que *a* correspond au nombre de modalit´es du facteur inter-sujets et que *b* cor- respond au nombre de modalit´es du facteur intra-sujet (*i.e.* nombre de mesures r´ep´et´ees). Ici c’est le facteur *Age* qui est le facteur intra-sujet. On a donc :

*a* = 2 *b* = 3 *n* = 12

La table de l’ANOVA est pr´esent´ee ci-dessous :



Effet du facteur inter-sujets A (Exposition `a la fum´ee)

*F* (1*,* 22) = 5*.*307, *p* = *.*031. On rejette l’hypoth`ese *H*0 car *p < α*. Le niveau d’hyperacti- vit´e n’est pas le mˆeme dans les groupes *Expos´e `a la fum´ee* et *Non expos´e `a la fum´ee*.

Effet du facteur intra-sujet B (Age)

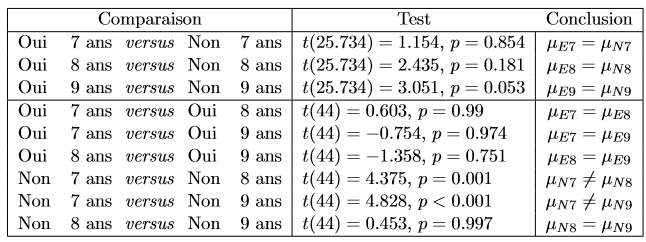
*F* (2*,* 44) = 7*.*032, *p* = *.*002. On rejette l’hypoth`ese *H*0 car *p < α*. Le niveau d’hyperacti- vit´e n’est pas le mˆeme `a *7 ans*, *8 ans* ou *9 ans*.

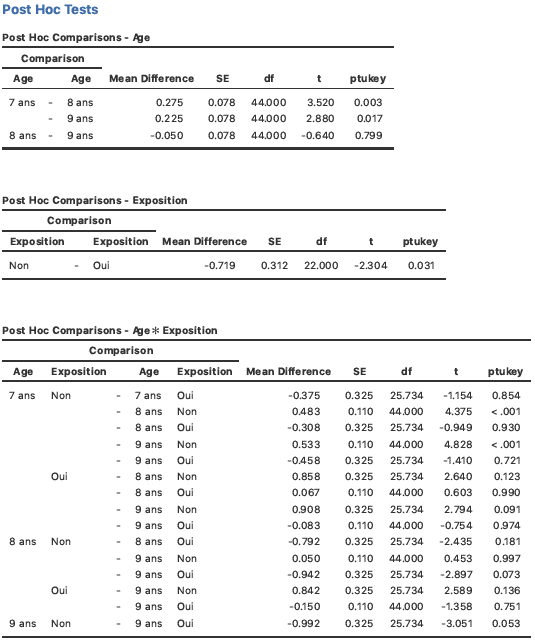
Effet de l’interaction entre A et B

*F* (2*,* 44) = 8*.*110, *p* = *.*001. On rejette l’hypoth`ese *H*0 car *p < α*. Il y a un effet significatif de l’interaction. Il est donc n´ecessaire d’analyser les effets simples pour pouvoir interpr´eter correctement les r´esultats de l’exp´erience.

* 1. Utilisez la procédure de Tukey pour identifier les différences franchement significatives.

*Analyse des effets simples selon la méthode de Tukey*





* 1. Commentez les résultats de vos analyses.

*Conclusion*

L’analyse de variance met en ´evidence un effet principal de *A*, un effet principal de *B*, et un effet d’interaction. Nous pouvons donc conclure que :

* + le niveau d’hyperactivit´e des enfants dont la m`ere fumait durant la grossesse est plus

´elev´e que celui des enfants dont la m`ere ne fumait pas ;

* + globalement, le niveau d’hyperactivit´e diminue entre 7 et 8 ans, puis reste stable entre 8 ans et 9 ans ;
  + l’´evolution des enfants n’est pas la mˆeme dans les deux groupes.

Grˆace `a l’analyse des effets simples, nous pouvons affirmer que les enfants ayant une m`ere fumeuse sont insensible au programme mis en place pour r´eduire l’hyperactivit´e, les conduites asociales et le recours pr´ecoce `a la cigarette. Durant toute l’´etude, leur niveau d’hyperactivit´e est rest´e stable. Par contre, les enfants dont la m`ere ne fumait pas ont tir´e profit du *Jeu de bonne conduite*. En deux ans, leur niveau d’hyperactivit´e a diminu´e de mani`ere significative.

## Exercice 3

Dans le cadre d’une étude sur la consommation de cigarettes, un expérimentateur a décidé de comparer trois procédures différentes visant à arrêter de fumer (diminution progressive, arrêt immédiat et thérapie de l’aversion). Chaque groupe se composait de cinq sujets. L’expérimenta- teur leur a demandé d’évaluer (sur une échelle en 10 points) leur désir de fumer “en ce moment même” dans deux environnements différents (foyer ou travail) avant et après l’arrêt de toute consommation de cigarettes. Nous avons donc une variable inter-sujets (le groupe de traitement) et deux variables intra-sujets (l’environnement et avant/après).

Avant Après

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Foyer | Travail |  | Foyer | Travail |
| **Diminution** | 7 | 6 |  | 6 | 4 |
|  | 5 | 4 |  | 5 | 2 |
|  | 8 | 7 |  | 7 | 4 |
|  | 8 | 8 |  | 6 | 5 |
|  | 9 | 5 |  | 5 | 3 |
| **Arrêt immédiat** | 8 | 7 |  | 7 | 6 |
|  | 5 | 5 |  | 5 | 4 |
|  | 7 | 6 |  | 6 | 5 |
|  | 8 | 7 |  | 6 | 5 |
|  | 7 | 6 |  | 5 | 4 |
| **Aversion** | 9 | 8 |  | 5 | 4 |
|  | 4 | 4 |  | 3 | 2 |
|  | 7 | 7 |  | 5 | 3 |
|  | 7 | 5 |  | 5 | 0 |
|  | 8 | 7 |  | 6 | 3 |

1. Effectuez l’analyse de variance appropriée.
2. Interprétez les résultats.

[Howell, exercice 14.14]

## Exercice 4

1. Dans le cadre d’études neuropsychologiques, une équipe de chercheurs tente de prouver les effets bénéfi ues du sport sur la mémoire des personnes âgées. Un échantillon de 400 séniors en bonne santé a été rassemblé, et les participants vont suivre 10 mois de cours d’aérobic sous la supervision d’un coach. Les chercheurs effectuent des tests psychologiques visant à évaluer la mémoire des personnes âgées avant le début de l’entraînement, puis chaque 2 mois durant l’entraînement et à la fin de celui-ci. Ils cherchent à montrer que les résultats aux tests évoluent eff ement au fil de l’entraînement physique. Quelle analyse statistique utilisent-ils ?

Il faut faire une *analyse de variance sur mesures r´ep´et´ees*, car le mˆeme groupe de parti- cipants est mesur´e aux mˆemes tests plusieurs fois dans le temps.

1. Ayant reçu une donation importante d’un riche mécène vieillissant inquiet de ses récentes pertes de mémoires, l’équipe de chercheurs mentionnée au point précédant se lance dans 10 nouveaux mois d’études avec le même échantillon de 400 séniors. Cette fois-ci, les scientifiques tentent de savoir si des exercices d’aérobic, des exercices de force, des exercices d’endurance et des exercices de pilates provoquent les mêmes eff chez les séniors. Ils répartissent pour ce faire équitablement les participants dans les 4 groupes et mesurent leurs performances cognitives à la fin des 10 mois. Quelle analyse statistique utilisent-ils pour comparer les 4 groupes ?

Il faut faire une analyse de variance a` un facteur ayant 4 modalités, car il y a une seule mesure faite sur 4 ´échantillons indépendants.

1. Les 400 séniors mentionnés aux deux points précédents étant très motivés par les résultats des deux études, ils se portent volontaires à nouveau pour une troisième étude de 10 mois et se cotisent afin d’assumer les frais des tests. Les scientifiques les répartissent en 4 groupes de 100 sujets (cf. point précédant : (1) exercices d’aérobic, (2) de force, (3) d’endurance et (4) de pilates) et décident de faire des mesures à trois moments de l’expérience : au tout début (mois 0) puis chaque 2 mois durant l’entraînement et finalement après la fin de celui-ci (mois 10). Quelle analyse statistique utilisent-ils pour comparer les résultats des trois groupes dans le temps ?

Il faut faire une analyse de variance sur plan mixte, car on a 3 modalités `a la variable “entraînement” et 6 modalités / mesures pour la variable “temps”.

1. Deux psychothérapeutes adeptes d’équitation souhaitent évaluer l’efficacité d’une thérapie contre la dépression qu’elles ont inventé et qui propose des séances de psychothérapie à cheval. Après avoir obtenu le consentement de certains de leurs clients, elles les répartissent en deux groupes (un premier qui suit des psychothérapies classiques, *n*1 = 50 ; un second qui pratique la psychothérapie à cheval, *n*2 = 50) et mesurent leur degré de dépression avant le début de la thérapie, après 3 mois et après 6 mois de thérapie. Quelle analyse statistique vont-elles utiliser afin de comparer les résultats de leurs clients ?

Il faut faire une analyse de variance sur plan mixte, car il y a deux groupes de sujets et chacun des groupes subit 3 mesures dans le temps.

1. Un scientifique passionné de papillons désire vérifi s’il est possible d’apprivoiser ses in- sectes préférés. Pour ce faire, il réunit 300 spécimens de *Parnassius apollo* (le Grand Apollon) dans une grande serre. Afin de voir s’il peut apprivoiser ses papillons, le scien- tifique compte combien de ses insectes viennent spontanément se poser sur lui en 1 heure de présence. Chaque jour pendant 3 mois, il vient passer 3 heures de son temps dans la serre aux papillons. Il eff une mesure pré-test, une mesure après 1 mois, une après 2 mois et une mesure finale après les 3 mois de l’expérience. Quelle analyse statistique doit-il utiliser pour comparer les « degrés d’apprivoisement » de ses papillons au fil du temps ?

Il faut effectuer une analyse de variance sur mesures répétées, car il y a un groupe de papillons et 4 mesures répétées sur ce même échantillon.

1. Encouragé par ses résultats (cf. point précédant), le scientifique amoureux des papillons se lance dans une expérience plus complète sur son échantillon de 300 papillons. Il les sépare en 3 groupes de 100 papillons, chaque groupe étant placé dans des serres séparées et identifiées par les numéros 1, 2 et 3. Chaque jour pendant 3 mois, le scientifique reste 2 heures dans la serre 1 en compagnie des papillons, et 5 heures dans la serre 2. Il ne se rend pas dans la serre 3.

Il répète le test lui permettant de juger du « degré d’apprivoisement » avant l’expérience, après 1 mois, après 2 mois et après 3 mois afin de clore son expérience. Quelle analyse statistique doit-il utiliser pour comparer les « degrés d’apprivoisement » de ses trois groupes de papillons au fil du temps ?

Il faut effectuer une analyse de variance sur plan mixte, car il y a 3 groupes de papillons et 4 mesures effectuées sur chacun des groupes de papillons.

1. Dans le cadre d’un projet d’éco-quartier, une enquête municipale désire remoduler un quartier existant et veut connaître le niveau satisfaction des habitants qui y logent. Les habitants remplissent un questionnaire de satisfaction et un score leur est attribué entre 1 (insatisfaction complète) et 100 (satisfaction intégrale) ; puis ils sont répartis selon le nombre de personnes logeant avec eux (*célibataire* ou *famille*) et le type de logement qu’ils occupent (*appartement* ou *villa*). Quelle analyse statistique doit utiliser la municipalité afin de savoir s’il y a des différences de satisfaction entre les habitants ?

Il faut faire une analyse de variance à deux facteurs inter-sujets, car on l’échantillon est reparti selon deux variables (nombre de personnes dans le foyer et type de logement) et on ne fait qu’une mesure dans le temps.

1. Une équipe de psychologues s’intéresse au profil du « patron parfait ». Au sein de leurs études, ils désirent connaître le niveau moyen d’appréciation des patrons par leurs em- ployés. Pour ce faire, ils rassemblent 300 travailleurs intéressés à remplir un questionnaire d’appréciation globale de leur chef. Leurs réponses sont triées selon leur place dans la hiérarchie (*basse* ou *haute*) et leur nationalité (*suisse* ou *étrangère*). Les chercheurs sou- haitent savoir si ces caractéristiques influencent d’une quelconque manière l’appréciation globale du patron. Quelle analyse statistique doivent-ils effectuent?

Il faut effectuer une analyse de variance a` deux facteurs inter-sujets.

1. Intéressé par l’étude sur le profil du « patron parfait » (cf. point précédent), un des patrons dont les employés ont participé à l’enquête souhaite améliorer la manière qu’il a de diriger son équipe. Il fait appel pour cela à un coach à raison de 5 heures par semaine pour une période de 2 mois et demande à ses employés de remplir le questionnaire d’appréciation légèrement modifié : la catégorie « nationalité » a été enlevée, ne laissant que la variable

« place dans la hiérarchie ». Les employés répondent au questionnaire 3 fois : avant les 2 mois, après un mois et à la fin des 2 mois. Quelle analyse statistique doit utiliser le patron afin de savoir si ses performances de leader ont été modifiées ?

Il faut effectuer une *analyse de variance sur plan mixte* car il y a un facteur inter-sujets

`a 2 modalités et un facteur intra-sujets `a 3 modalités. La normalité des distributions a déjà été vérifié au point précédent.

1. Des psychologues travaillant dans le secteur marketing d’une grande entreprise de cosmé- tiques veulent prouver l’influence de leur nouvelle campagne publicitaire. Pour ce faire, ils relèvent en Suisse les ventes par canton du produit venté par leur slogan accrocheur. Ils font en tout 5 relevés sur une période de 12 mois. Quelle analyse statistique doivent-ils effectuent pour prouver que leur campagne a eu un effet sur les ventes du produit ?

Il faut effectuer une analyse de variance sur mesures répétées, car on a un même échantillon de canton dans lequel on fait plusieurs fois des relevés de vente dans le temps.

courbe de lissage: scatterplot et cocher smooth pour faire apparaître la courbe de lissage