Statistiques II – Séries

Table des matières

[Série 1 3](#_Toc528314160)

[Exercice 1 3](#_Toc528314161)

[1a : moyenne et écart-type 3](#_Toc528314162)

[1b homogénéité des variances – Test de Levene 3](#_Toc528314163)

[1c) ANOVA 4](#_Toc528314164)

[Série 2 – Droite de régression 4](#_Toc528314165)

[Vue d’ensemble Jamovi 4](#_Toc528314166)

[Exercice 1 4](#_Toc528314167)

[a) Représentation graphique VI🡪VD 4](#_Toc528314168)

[b) Coefficient r de corrélation de Bravais-Pearson 4](#_Toc528314169)

[c) Savoir si le coefficient de corrélation est statistiquement significatif 4](#_Toc528314170)

[d) Equation de la droite de régression 4](#_Toc528314171)

[e) Diagramme de dispersion 5](#_Toc528314172)

[f) calculer l’erreur standard de régression dans Jamovi 6](#_Toc528314173)

[g) limites de confiance imposées à l’ordonnée à l’origine de la droite de régression 6](#_Toc528314174)

[h) L’ordonnée à l’origine est-elle significativement différente de 0 ? 7](#_Toc528314175)

[i) Calculer les limites de confiance imposées à la pente de la droite de régression 7](#_Toc528314176)

[j) La pente de la droite est-elle significativement différente de 0 ? 7](#_Toc528314177)

[k) Prédiction 8](#_Toc528314178)

[Exercice 2 8](#_Toc528314179)

[coefficient de corrélation linéaire r entre X et Y 8](#_Toc528314180)

[Série 3 8](#_Toc528314181)

[Exercice 1 Proposez un modèle linéaire pour décrire Y à l’aide des quatre prédicteurs. 8](#_Toc528314182)

[Exercice 2 – Modèle 9](#_Toc528314183)

[a) Matrice des nuages de points : histogrammes, diagrammes de dispersion de toutes les paires de variables et corrélation 9](#_Toc528314184)

[b) Estimation des paramètres du modèle : équation du modèle 9](#_Toc528314185)

[c) Interprétation • Ordonnée à l’origine, coefficients de régression partiels (Estimate) 9](#_Toc528314186)

[d) Erreur standard de régression + commentaire 9](#_Toc528314187)

[e) Coefficient de corrélation multiple R2 (coefficient de détermination = part de variance expliquée par les prédicteurs. 9](#_Toc528314188)

[f) Coefficient de détermination ajusté 9](#_Toc528314189)

[g) Test global 9](#_Toc528314190)

[Exercice 3 – droite de régression à 3 coefficients 9](#_Toc528314191)

[a) Matrice des corrélations 9](#_Toc528314192)

[b) Equation de la droite de régression et prédiction 9](#_Toc528314193)

[c) Estimation des paramètres du modèle et équation du modèle 9](#_Toc528314194)

[d) Erreur standard et séparation 9](#_Toc528314195)

[e) Coefficient de corrélation multiple au carré (coefficient de détermination) R2 = part de variance du critère expliquée par les prédicteurs. Le modèle décrit-il bien les données ? % 9](#_Toc528314196)

[f) Test global de la régression multiple et hypothèses et conclusion 9](#_Toc528314197)

[g) Tests des coefficients de régression partiel associé à chaque variable (Significativité des paramètres) 9](#_Toc528314198)

[Série 4 – Modèles emboités (linéarité, hétérosc. et normalité) et variables standardisées 9](#_Toc528314199)

[Ex1 – matrice de corrélation, overall model test (global) et specific, modèles emboîtés (blocks) et régression 10](#_Toc528314200)

[f) Test des coefficients de régression partiels 10](#_Toc528314201)

[Exercice 2 (avec données de l’exercice 1) – Linéarité 10](#_Toc528314202)

[Série 4 – Modèles emboités et comparaisons 12](#_Toc528314203)

[Exercice 1 - Estimation des paramètres du modèle de régression, coefficient de détermination, test global de la régression multiple, test des coefficients de régression partiels, comparaison, critères d’emboitement 12](#_Toc528314204)

[Exercice 2 - ajustement, conditions d’application de la régression multiple, homoscédasticité et normalité + graphiques 12](#_Toc528314205)

[Exercice 3 – (A faire !) Estimation des paramètres, comparaison des modèles, coefficients de régression standardisés et bruts, simplification de la formule 12](#_Toc528314206)

[Série 5 – Effets d’interaction 14](#_Toc528314207)

[Exercice 1 – Effets de A/B/interaction selon diagramme 14](#_Toc528314208)

[Exercice 2 – ANOVA à 2 facteurs, effet propre + interaction 14](#_Toc528314209)

[Exercice 3 – Diagramme d’interaction avec 1 VI à 2 modalités x 1VI à 3 modalités (et une VD numérique) 14](#_Toc528314210)

[a) Moyenne et écart-type des ss groupes 14](#_Toc528314211)

[b) Représentation graphique (diagramme d’interaction) 14](#_Toc528314212)

[c) Conditions d’application de l’ANOVA 14](#_Toc528314213)

[d) ANOVA (différence intergroupes) et tailles des effets avec de η2 et de 14](#_Toc528314214)

[e) Analyse 14](#_Toc528314215)

[f) Post-hoc 14](#_Toc528314216)

[Série 6 14](#_Toc528314217)

[Exercice 1 14](#_Toc528314218)

[Exercice 2 14](#_Toc528314219)

[Questions 14](#_Toc528314220)

# Série 1

## Exercice 1

### 1a : moyenne et écart-type

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Analyses > Exploration > Descriptive > Variables A > Split by B >Tout désélectionner sauf mean et stand. dev.

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

### 1b homogénéité des variances – Test de Levene

Analyse>Anova>assumption checks>homogeneity tests

Colonne A : Variable dépendante

Colonne B : Split by group

Si p>0.05 on peut faire l’ANOVA (car ça signifie que les variances sont =, car on n’a pas rejeté H0. Si les variances n’étaient pas égales, on ne pourrait pas faire d’ANOVA)

Notation APA :

F (df1, df2)

### 1c) ANOVA

Same

Puis post hoc tests on voit les comparaisons entre chaque groupe on regarde p

1d)

# Série 2 – Droite de régression

## Vue d’ensemble Jamovi

[Imprimer et relier chaque nombre à sa définition]

## Exercice 1

### Représentation graphique VI🡪VD

Exploration>Scatterplot

### Coefficient r de corrélation de Bravais-Pearson

Analyses>Regression>Correlation Matrix>Pearson>Entrer les valeurs X et Y>Préciser l’hypothèse>Cocher correlation matrix>Noter la valeur appelée Pearson’s r

Commenter : Si r = [-1 ;-0.80]U[0.80 ;1] on dit que la corrélation est forte

b) r=-0.80

### Savoir si le coefficient de corrélation est statistiquement significatif

b) puis cocher la case Report significance

sur Jamovi : on trouve r par model fit measure (matrice de corrélation ?) sous R dans les résultats r(10)= -0.50, p= 0.002

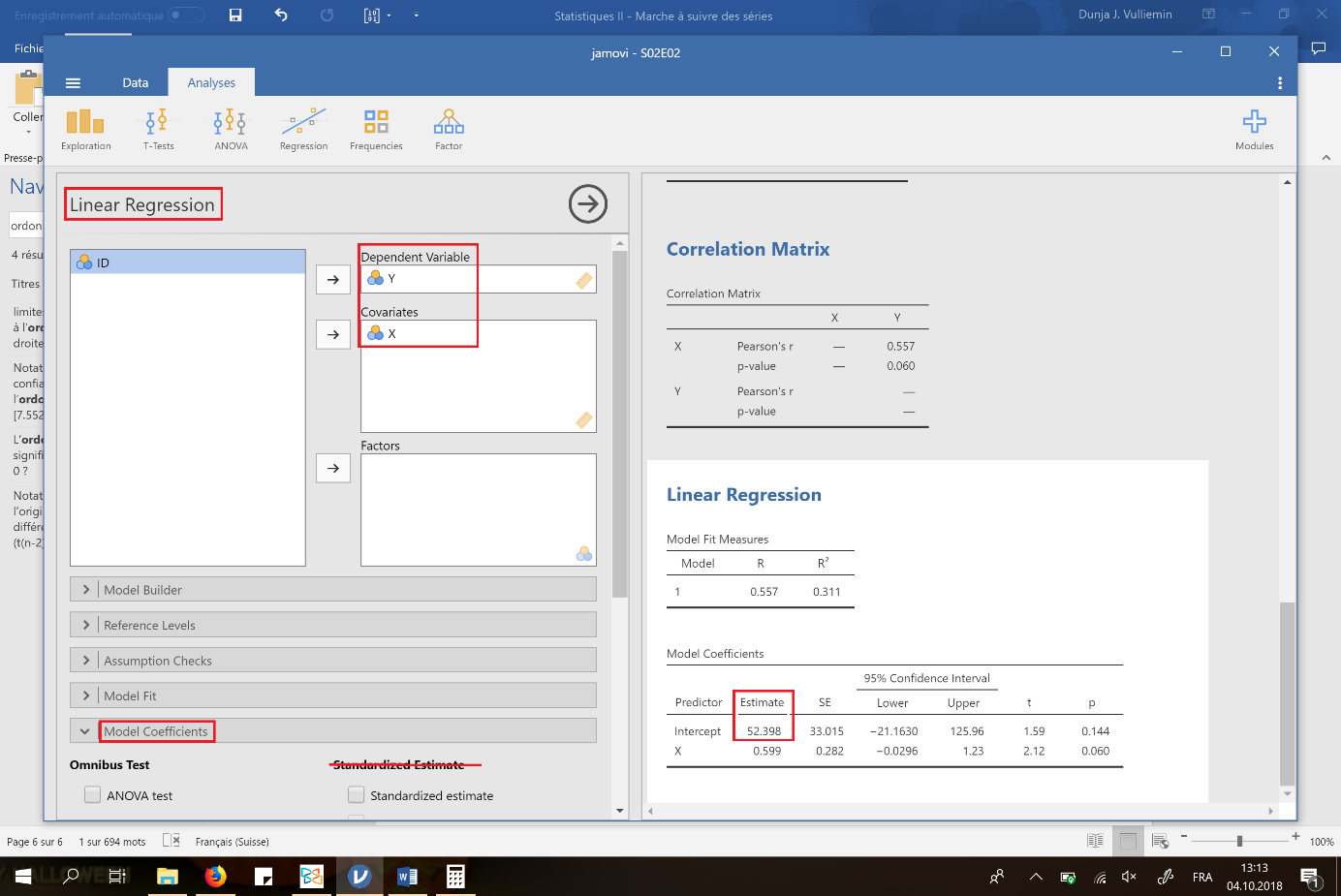
### Equation de la droite de régression

Voir e) pour le diagramme de dispersion

Fonction y=â+b̂ x ou **=0+1X (voir k)**

Utiliser Jamovi pour trouver sy, sx, et (descriptives), rxy (rég>c.matrix>variables)

0: Linear regression>Model coefficient>~~Standardized~~ Estimate> C.I.



A la main :

1 = r

= -0.8 🡪1 =-0.479

0 = -1

= 5.167-(-0.479∙11.833) = 10.835

Erreur standard

Model fit et cocher RMSE

Notation APA

Bravais-Pearson

r(n-2) = -0.80, p= 0.002

### Diagramme de dispersion

NPO de sélectionner les variables

Une image contenant capture d’écran, ordinateur

Description générée avec un niveau de confiance très élevé

### calculer l’erreur standard de régression dans Jamovi

∙RMSE

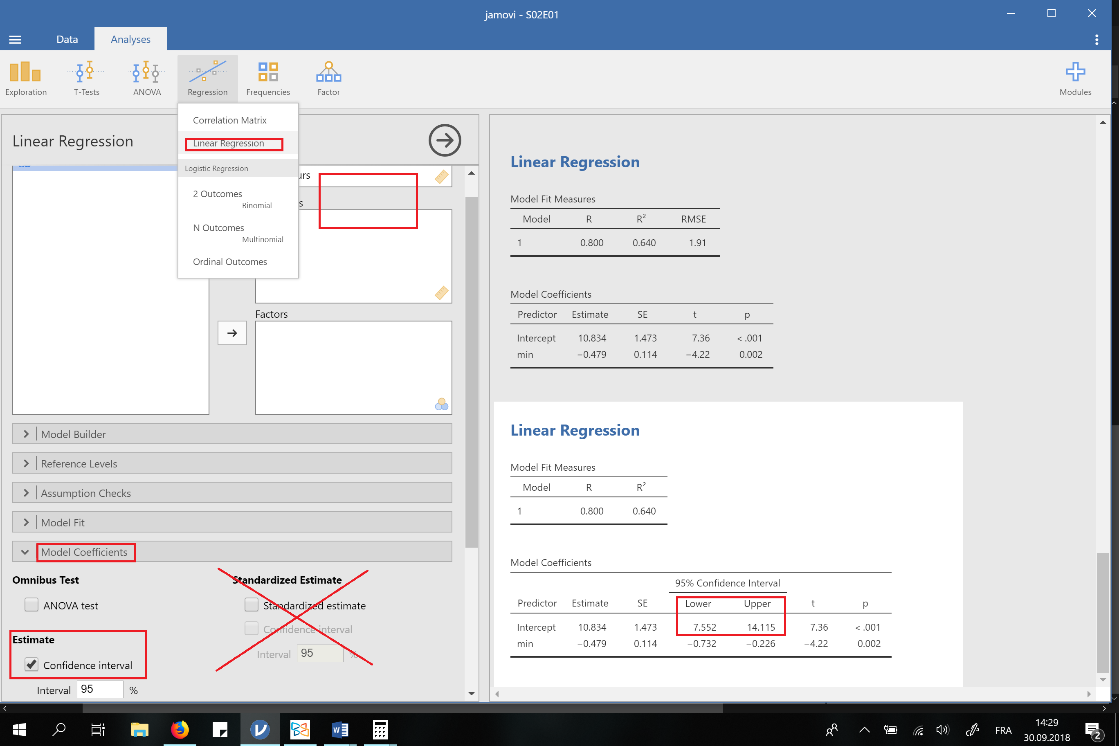
#### RMSE dans Jamovi:

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

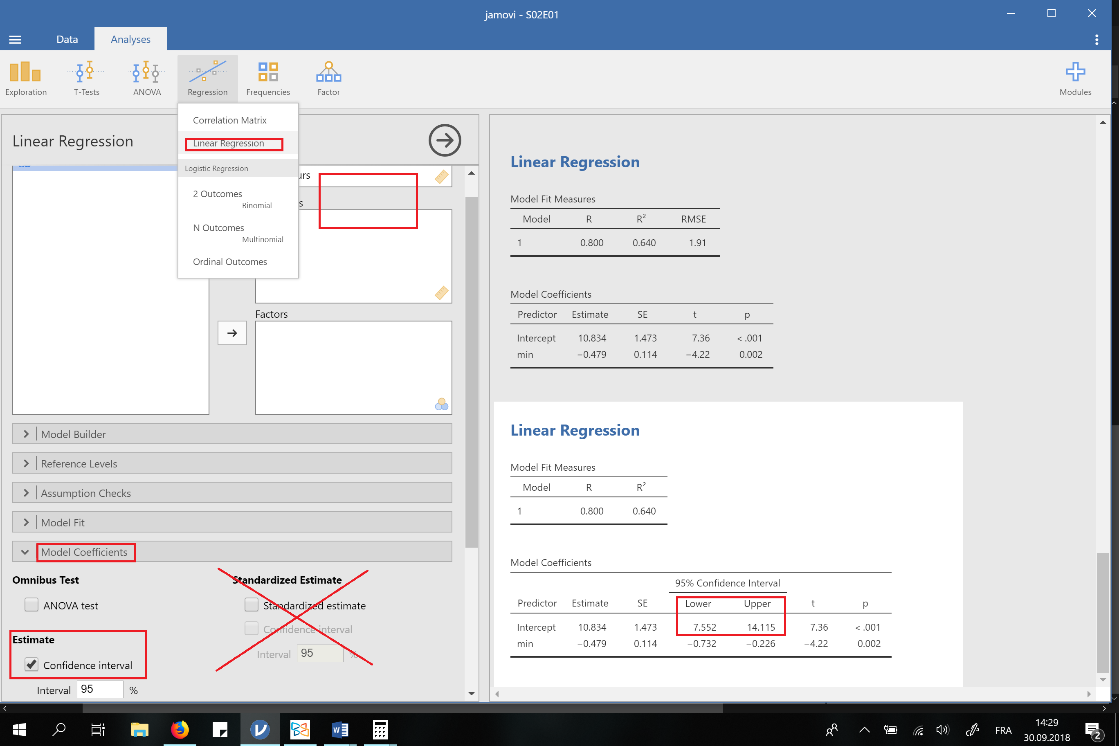
### limites de confiance imposées à l’ordonnée à l’origine de la droite de régression

Linear regression>Model coefficient>~~Standardized~~ Estimate> C.I.



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

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



Model coeff. >C.I. sous t (ici 7.36 sur l’image)

Notation : *L’ordonnée à l’origine est significativement différente de 0 au seuil de 5% (t(n-2) = t, p < .001).*

### Calculer les limites de confiance imposées à la pente de la droite de régression

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Notation : *L’intervalle de confiance à 95% de la pente vaut [−0.732; −0.226].*

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

Même image qu’en dessus, sous t (x min) : *La pente est significativement différente de 0 au seuil de 5% (t(10) = −4.22, p = 0.002).*

### Prédiction

Selon notre modèle linéaire, le nombre d’erreurs commises par un individu ayant mis 16 minutes pour passer le test de Kagan vaut : (juste remplacer x par 16 dans l’équation de rég)

i = 10.834 − 0.479 xi = 10.834 − 0.479 × 16 = 3.171

Appliquer la fct =0+1X ⬄ 3.17

## Exercice 2

Regression>correlation matrix

### coefficient de corrélation linéaire r entre X et Y

R= 0.557

1. = 0 + 1QIm 🡪 =0+1X = 52.398+0.599X
2. Une image contenant capture d’écran

   Description générée avec un niveau de confiance très élevé

Dans la fonction, taper le résultat de la fonction de

1. Exploration>Descriptives>Variance (à tester?)

Notation : Var( = 100.921

1. Var(Y) = 324.333
2. Variance expliquée (ratio, part de variance expliquée)

R2 = pourcentage de variance explique. Si R2 = 0.311, 31.1% de variance est expliqué par la variable X (La VI)

# Série 3

## Exercice 1 Proposez un modèle linéaire pour décrire Y à l’aide des quatre prédicteurs.

1. Yˆ = β0 + β1 X1 + β2 X2 + β3 X3 + β4 X4 (ne pas remplacer les données)

On ne met pas ε car c’est une estimation (avec ^). = y+ε

= β1 + β2resposabilite+β3env+β4annees

Modèle de Karasek ? Si on est + autonome, on est + satisfait

1. 1.669 +0.605- 0.334+0.486+0.07
2. Sigma chap. = (voir formule) = ∙ 1.68 = 2.058
3. R = 0.697

Si un modèle explique 40-48% de la variance, c’est un bon modèle.

1. Entrer l’éq. du pt. b) :

**= 1.669+0.605\*reponsabilite-0.324\*perssup+0.4855\*environnement+0.07\*annees**

puis enter et ça calcule les données

= satisfaction – score.ajuste (attention nom que j’ai donné)

f) – ou + c’est égal.

g)

i) F(4,10) = 2.367, p=0.123

j) = 2f) : calculer le coeff de détermination ajusté

## Exercice 2 – Modèle

### a) Matrice des nuages de points : histogrammes, diagrammes de dispersion de toutes les paires de variables et corrélation

### b) Estimation des paramètres du modèle : équation du modèle

### c) Interprétation • Ordonnée à l’origine, coefficients de régression partiels (Estimate)

### d) Erreur standard de régression + commentaire

### e) Coefficient de corrélation multiple R2 (coefficient de détermination = part de variance expliquée par les prédicteurs.

### f) Coefficient de détermination ajusté

### g) Test global

## Exercice 3 – droite de régression à 3 coefficients

### a) Matrice des corrélations

### b) Equation de la droite de régression et prédiction

### c) Estimation des paramètres du modèle et équation du modèle

### d) Erreur standard et séparation

### e) Coefficient de corrélation multiple au carré (coefficient de détermination) R2 = part de variance du critère expliquée par les prédicteurs. Le modèle décrit-il bien les données ? %

### f) Test global de la régression multiple et hypothèses et conclusion

### g) Tests des coefficients de régression partiel associé à chaque variable (Significativité des paramètres)

# Série 4 – Modèles emboités (linéarité, hétérosc. et normalité) et variables standardisées

Ex1 : modèles Ex 2 conditions d’application Ex 3 : variables standardisées

Il va nous montrer tous les exos sur les données du premier exercice

On a des variables qui ont chacun une sous échelles et qui sont rentrées dans 3 groupes = 3 concepts

Correlation matrix

Modèle emboité : si on veut savoir si un modèle aurait vraiment besoin de la variable INF

Quand on veut voir l’effet d’interaction des 2 variables (l’effet de 2 variables ensembles ) : pas moyen de le tester 🡪 on fait un modèle emboité où on enlève les 2 variables (et on laisse les autres)

## Ex1 – matrice de corrélation, overall model test (global) et specific, modèles emboîtés (blocks) et régression

1. Corr. Matrix

Equation de régression

1. = β0 + β1ARI+ β2INF+ β3VOC+ β4COM+ β5SPA
2. Estimation des paramètres : Linear regression>Model fit+model coefficient

Estimate : si on augmente d’une unité, VOC augmente de 0.233

VDép OUV

COV : autres

RMSE pour connaitre l’estimation standard

n’est pas bcp plus bas que R2 car n est beaucoup plus grand que le nombre de variables

à la main = ∙RMSE

1. Valeur du

48= « 48% de la variable est expliquée par le modèle »

1. Overall model test > F test 🡪 tester si notre regression est significative

F(5, 69) = 12 806, p<0.001

H0 (β1 = 0) ∩ (β2 = 0) …

H1 (β1 ≠ 0) U …

### Test des coefficients de régression partiels

Regarder dans le tableau model coefficient les variables qui ont p<0.005 INF : t(69) = 3 , 135, p=.03

1. Modèles emboités

Enlever une variable (ici INF) et on compare les modèles

Linear regression > Model builder > Blocks > on **glisse** la variable dans le block 2

* Le block 2 est complet (Block 1 + block 2)

Tableau (Model specific results)

* Si on voit qu’il y a le même pouvoir avec un modèle plus simple (=avec moins de variables)🡪 la variable du block 2 est « inutile » 🡪 on préférera le modèle plus simple

COMPARAISON : Si p<0.05 = Modèle 1 ≠ Modèle 2 🡪 on rejette H0 (β2 = 0) et on accepte H1 (β2 ≠0)

F(1, 69) = 9.828, p=0.003

Valeur au carré 9.28

1. Même test que test marginal t car on n’enlève qu’une seule variable (INF) dans le block 2

Same avec ARI

1. H0 (β1 = 0) ∩ ( β2 = 0) …

H1 (β1 ≠ 0) U …

1. Oui
2. F(2.69) = 0.865, p=0.425 (Linear regression (même si multiple) > model builder > Blocks)

## Exercice 2 (avec données de l’exercice 1) – Linéarité

Linéarité : vérifiez

1. Calculer le score prédit et résidus //série 3

Comp var

Score prédit : -7.889 + 0.205\*INF+0.235\*VOC+0.136\*COM

β0 + β1INF+ β2VOC

calculer les résidus = OUV – scores prédits

fx = OUV-scorespred

Illustration : avec scatr : image de points catterplot X =scorepred Y=residus

Exemple où on met la droite de régression en abscisse :

Une image contenant texte, tableau blanc

Description générée avec un niveau de confiance très élevé

En cas de linéarité, (//Ex 2b) 🡪 smooth

On voit que la droite est assez horizontale et pas en U

Ex 2b : oui, semble linéaire (vaut 0 pour presque tous les scores ajustés)

Valeur des erreurs est constante ? Hétéroscédasticité

Une image contenant texte

Description générée avec un niveau de confiance très élevé

Scatterplot : X : scorespred Y : racineresidusabs • Linear (car on teste la linéarité)

Une image contenant texte

Description générée avec un niveau de confiance très élevé

* Petite hétéroscédasticité
* La variance n’est pas constante

Les résidus sont-ils distribués normalement ?

Manière 1 : QQ plot (graphique). On voit que ça suit la ligne 🡪 se distribue normalement

Manière 2 : One Sample t-test > Shapiro Wilk W=0. … 85 p=.518 🡪 H0 🡪 les données sont distribuées normalement

c) standardiser les variables. Créer une fx « stdOUV »=SCALE(OUV) = fc qui standardise

qc/σ (=z ?)

Model coefficient stadardized estimate

Une image contenant tableau blanc, texte

Description générée avec un niveau de confiance très élevé

Estimate =

Standard estimate = = 0.75

Exerci

# Série 4 – Modèles emboités et comparaisons

## Exercice 1 - Estimation des paramètres du modèle de régression, coefficient de détermination, test global de la régression multiple, test des coefficients de régression partiels, comparaison, critères d’emboitement

## Exercice 2 - ajustement, conditions d’application de la régression multiple, homoscédasticité et normalité + graphiques

## Exercice 3 – (A faire !) Estimation des paramètres, comparaison des modèles, coefficients de régression standardisés et bruts, simplification de la formule

Etape 1) Définir les variables (3 : notes, genre, mode) 🡪 en colonne

Mettre le mode décimal pour pouvoir mettre genre 4**.1**

CC : régressions multiples et Anova

Faire bien l’exercice 3 !

Diagramme d’interaction : d’abord faire une Anova et cocher la case

One way Anova > vd note > facteurs genre et mode > estimate of the marginal means > glisser les 2 facteurs à droite

Si on veut inverser celui en X et celui à droite, on enlève et on reglisse les trucs dans le sens inverse.

Interprétation : dans les conditions anonyme et standard, les H ont des meilleures notes que les F

Dans la condition informatique, H et F ont des notes similaires

Scores – la moyenne = les résidus. ON peut les mettre dans une nouvelle colonne

Dans la logique, on fait l’assumption check d’homogeneity

On note en c) F(5,36) = 0.121, p = 0.987

Jamovi ne permet pas de faire le test des résidus directement mais on peut faire le QQ plot directement. 🡪 si les points suivent la ligne, on considère que c’est normal

Ici, le qq plot semble respecter la normalité et on ferait une ANOVA

d) F c’est pour normes APA, etc. mais ce qui nous intéresse le plus c’est le p : on voit qu’il y a un effet du genre, et la colonne des noms pour les différencier. Ensuite effet d’interaction (lequel ?)

F(1,36) = 11.019, p=0.002, n2 = 0.176 (etc. avec chaque mode et genre ⬄ mode)

Dans le corrigé le prof a noté que la moyenne globale des hommes est la même moyenne globale que les femmes. (H0)

Et pour le mode : le truc avec mu.j. = mu.j’. moyenne femme standar mu 11 femmes … mu12

Avec le graphique: effet principal du genre et pas du mode

Si le barres sont parallèles, les écarts sont toujours les mêmes

Gamlj effet d’interaction ou effet simple

Effet d’interaction si oui effets simples

Si pas d’effet d’interaction 🡪 post hoc

Effet simple par exemple effet du genre sur le facteur info

General linear model dependeant notes > genre et mode en facteur > simple effect

Omnibus test là : on peut dire que les différentes moyennes à chaque mode sont significativement différentes. Par exemple : en mode standard, la moyenne f et la moyenne h sont significativement différentes

L’utilité du post hoc est nulle s’il y a une

Interaction

# Série 5 – Effets d’interaction

## Exercice 1 – Effets de A/B/interaction selon diagramme

## Exercice 2 – ANOVA à 2 facteurs, effet propre + interaction

## Exercice 3 – Diagramme d’interaction avec 1 VI à 2 modalités x 1VI à 3 modalités (et une VD numérique)

### Moyenne et écart-type des ss groupes

### Représentation graphique (diagramme d’interaction)

### Conditions d’application de l’ANOVA

### ANOVA (différence intergroupes) et tailles des effets avec de η2 et de

### Analyse

### Post-hoc

# Série 6

## Exercice 1

## Exercice 2

Dans post-hoc comparison Age : t(df de la ligne)= t du tableau, ptukey=0.017

Exemple 3 comparaisons (car 3 groupes d’ages) non-fumeuses + 3 comparaisons fumeuses

Anova 2 fact 🡪 si int 🡪

Si pas sign 🡪 anova

Anova mixte avec intra 🡪 post hoc

Si on fait une analyse factorielle

Non mixte

# Série 7

# Questions

Série 1 Exercice 1b

Comment se fait la notation APA ? Que désignent 2 et 27 ? Pourquoi n’ai-je pas les mêmes résultats ?

Série 2 Exercice 2b