Statistiques II

Table des matières

[Datalab vidéos tutorial 3](#_Toc530659328)

[Modules 3](#_Toc530659329)

[Préparer ses données « Data wrangling » 3](#_Toc530659330)

[Entering data 4](#_Toc530659331)

[Variable types and labels 4](#_Toc530659332)

[Computing variables 4](#_Toc530659333)

[Filtering rows 4](#_Toc530659334)

[Exploration 4](#_Toc530659335)

[Descriptive statistics - graphiques 4](#_Toc530659336)

[Histogrammes 4](#_Toc530659337)

[Density plots 4](#_Toc530659338)

[Box plots et identifier les outliers 4](#_Toc530659339)

[Dot and bar plots 4](#_Toc530659340)

[T-tests 4](#_Toc530659341)

[Independant-samples t-tests 4](#_Toc530659342)

[Test de normalité: La distribution suit-elle une loi normale ? 4](#_Toc530659343)

[Shapiro-Wilk = T-test independent sample 4](#_Toc530659344)

[Paired t-test 4](#_Toc530659345)

[Test d’égalité des variances de Levene 4](#_Toc530659346)

[Inferential tests (?) p.ex t-tests 4](#_Toc530659347)

[One sample t-test 4](#_Toc530659348)

[ANOVA 5](#_Toc530659349)

[ANOVA 5](#_Toc530659350)

[Repeated-measures ANOVA 5](#_Toc530659351)

[ANCOVA 5](#_Toc530659352)

[MANCOVA 5](#_Toc530659353)

[Kruskal-Wallis test 5](#_Toc530659354)

[Friedman test 5](#_Toc530659355)

[Taille d’effet : Cohen’s D 5](#_Toc530659356)

[Régression 5](#_Toc530659357)

[Linear regression 5](#_Toc530659358)

[Binomial logistic regression 6](#_Toc530659359)

[Multinomial logistic regression 6](#_Toc530659360)

[Ordinal logistic regression 6](#_Toc530659361)

[Frequencies 7](#_Toc530659362)

[Binomial test 7](#_Toc530659363)

[Χ2 test of association 7](#_Toc530659364)

[McNemar test 7](#_Toc530659365)

[Log-linear regression 7](#_Toc530659366)

[Factor 7](#_Toc530659367)

[Reliability analysis 7](#_Toc530659368)

[Principal component analysis 7](#_Toc530659369)

[Exploratory factor analysis 7](#_Toc530659370)

[Confirmatory factor analysis 7](#_Toc530659371)

[Next steps (conclusion) 7](#_Toc530659372)

[Choix du test (Datalab et cours) 7](#_Toc530659373)

[Partie I – Méthodes explicatives (Introduction) 10](#_Toc530659374)

[Ajouter un filtre – désélectionner une partie des données 10](#_Toc530659375)

[Importer des données 11](#_Toc530659376)

[QQ Plot 11](#_Toc530659377)

[Cours 2 – Intro 11](#_Toc530659378)

[Variable à expliquer vs variable explicative 11](#_Toc530659379)

[Régression multiple sur Jamovi 11](#_Toc530659380)

[Modèles linéaires généraux 11](#_Toc530659381)

[Analyse confirmatoire 11](#_Toc530659382)

[Chapitre 1: Régression simple 11](#_Toc530659383)

[Modèle 11](#_Toc530659384)

[Estimation des paramètres β0 et β1 (1.2) 12](#_Toc530659385)

[Evaluation de la qualité de l’ajustement 12](#_Toc530659386)

[Créer une droite de régression 13](#_Toc530659387)

[Chapitre 2 – Régression multiple 13](#_Toc530659388)

[Modèle (2.1) 14](#_Toc530659389)

[Estimation (2.2) 14](#_Toc530659390)

[Régression multiple dans Jamovi 16](#_Toc530659391)

[Ajustement et qualité de l’ajustement (2.3) 16](#_Toc530659392)

[Coefficient de détermination de R2 (2.3.2) 17](#_Toc530659393)

[Aspect inférentiel (2.4) 17](#_Toc530659394)

[Test global (2.4.1) 17](#_Toc530659395)

[Cours 3 18](#_Toc530659396)

[Tests marginaux (test des coefficients) (2.4.2) 18](#_Toc530659397)

[Hypothèse de linéarité dans jamovi 21](#_Toc530659398)

[2.5.2 Hypothèse d’homoscédasticité = d’homogénéité mais pas pour une analyse de variance= dispersion des erreurs 21](#_Toc530659399)

[2.5.3 Hypothèse de normalité 21](#_Toc530659400)

[3. Analyse de variance à deux facteurs 21](#_Toc530659401)

[Structure des données dans Jamovi 22](#_Toc530659402)

[Le modèle : 22](#_Toc530659403)

[Effet d’interaction 23](#_Toc530659404)

[Test statistique pour savoir si effet de A, B et/ou d’interaction AB 23](#_Toc530659405)

[Avec Jamovi 24](#_Toc530659406)

[Stratégie : 24](#_Toc530659407)

[Puissance des tests (3.5) 24](#_Toc530659408)

[Dans Jamovi 24](#_Toc530659409)

[Chapitre 4 : Plans complexes 25](#_Toc530659410)

[Différents plans d’expérience (4.1) 25](#_Toc530659411)

[Analyse de variance sur mesures répétées (4.2) SxA 26](#_Toc530659412)

[Jamovi et résumé du cours 26](#_Toc530659413)

[Analyse de variance sur plan mixte (4.3) 27](#_Toc530659414)

[Analyse de covariance (5) 27](#_Toc530659415)

[Une variable numérique et une variable norminale (5.1) 27](#_Toc530659416)

[Jamovi – faire les droites et effet du sexe (γ) 28](#_Toc530659417)

[Comparer 2 modèles emboités 28](#_Toc530659418)

[Variable nominale polytomique (5.2) 28](#_Toc530659419)

[Test de l’effet de la variable nominale 29](#_Toc530659420)

[Modélisation des interactions (5.3) 29](#_Toc530659421)

[Représentation graphique des résultats de l’ANCOVA sur Jamovi 30](#_Toc530659422)

[Conclusion 30](#_Toc530659423)

[Questions 30](#_Toc530659424)

[Examen 30](#_Toc530659425)

# Datalab vidéos tutorial

## Modules

## Préparer ses données « Data wrangling »

1 column per variable

1 row per case

Entrer les données dans **Excel**, puis enregistrer le fichier en format .**csv**

**NPO : PREMIERE LIGNE = NOM DE LA VARIABLE (« ID », « min », « score »)**

Dans Jamovi, **Open**>menu déroulant remplacer **“Data files” par « csv »**> ouvrir le doc

NPO de préciser le **type de variable** (nominale, ordinale, numérique)

Aller dans descriptives, choisir les VI/VD et faire un test de normalité si besoin.

### Entering data

### Variable types and labels

### Computing variables

### Filtering rows

## Exploration

### Descriptive statistics - graphiques

NPO préciser le type de variable (nominale, ordinale..) >Descriptives > choisir les variables> tout en bas, choisir le graphique approprié

### Histogrammes

Descriptives>sél. Les variables et split by>tout décocher> Plots >l > histogram

### Density plots

### Box plots et identifier les outliers

Box plots

### Dot and bar plots

## T-tests

### Independant-samples t-tests

One-sample t-test : Q-Q Plot. Same mean as the observed data?

Une image contenant capture d’écran, ordinateur, portable, intérieur

Description générée avec un niveau de confiance très élevé On veut savoir si les points ont l’air de suivre la ligne. On regarde en général ça avant de faire le One-sample t-test, avec la normalité.

### Test de normalité: La distribution suit-elle une loi normale ?

### Shapiro-Wilk = T-test independent sample

Analyses>T-tests>Assumption checks>Normality

Si p<0.05, la distribution n’est **pas** normale.

### Paired t-test

Analyses>T-tests paired t-test>Assumption checks>Normality

### Test d’égalité des variances de Levene

Analyses>T-test>independent samples t-test>Assumption checks>equality of variances

### Inferential tests (?) p.ex t-tests

Dans la vidéo sur les t-tests pour échantillons indépendants, il parle de fonctions utiles pour les tests d’inférences.

Analyses>independent samples t-tests>additional statistics>descriptives/descriptive plots

* Voir si l’IC overlap la moyenne de l’autre groupe (si oui, pas significatif)

### One sample t-test

Etape 1 : descriptives, QQ plot et normalité

Etape 2 : analyse>one sample t-test>selectionner les variables>hypothesis>test value [entrer la donnée médiane, par exemple 3 si échelle des score est entre 1 et 5]

## ANOVA

### ANOVA

### Repeated-measures ANOVA

### ANCOVA

### MANCOVA

### Kruskal-Wallis test

### Friedman test

### Taille d’effet : Cohen’s D

How many standard deviantions there are between the scores at time 1 and time 2 or the scores on the first VD in the second VD

Really what it is is the mean of the different scores divided by the standard deviation of the different scores

Exemple video: on a un score entre 1 et 5 pour une variable du big 5. Cette variable a une taille d’effet de 1.7 (=1.7 points above the hypothesised mean of 3) comparaison de l’échantillon avec l’échantillon de référence. Il dit que c’est une très grande taille d’effet.

## Régression

= associations.

Scatterplots, correlations, prédiction

Linear:

Binomial logistic regression: outcome =/= quanti mais dichotomique

Multinomial logistic regression : several categories in outcome variable

Ordinal logistic regression: use a collection of variables to predict which ordered categories a case falls into

### Linear regression

Modèles emboîtés (model builder) : Model specific results, predictor\*Estimate+predictor\*estimate…, model comparisons.

Reference Levels

**Assumption checks**

Autocorrelation test

Colinearity (colinéarité): VIF et tolerance 🡪 how associated each variable are with each other

QQ plot of residuals: should be normally distributed

Residual plots:

Cook’s distance: how influential each point is (in the scatterplot)

**Model fit**

R, R2, Adjusted R2, AIC, BIC (bayesian criteria), RMSE, overall model test (F test)

**Model coefficients**

Omnibus test

Standardized estimate: Standardized estimate:

Confidence interval

**Estimated Marginal Means**

Chart. Term 1 and term 2: level of VI per modalité of term 1

### Binomial logistic regression

outcome =/= quanti mais dichotomique

### Multinomial logistic regression

several categories in outcome variable

### Ordinal logistic regression

several categories in outcome variable

## Frequencies

### Binomial test

### Χ2 test of association

### McNemar test

### Log-linear regression

## Factor

### Reliability analysis

### Principal component analysis

### Exploratory factor analysis

### Confirmatory factor analysis

## Next steps (conclusion)

# Choix du test (Datalab et cours)

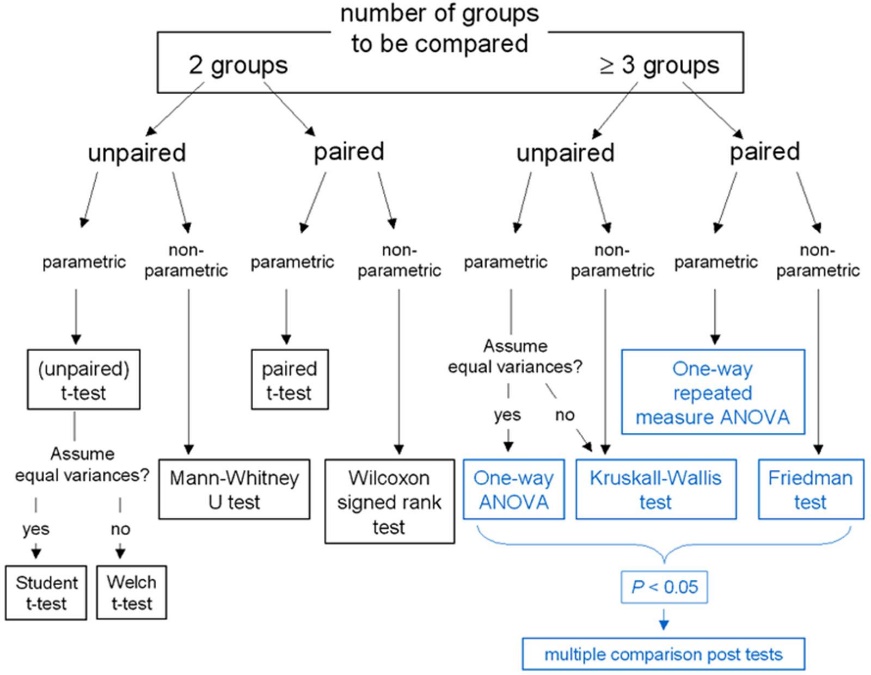
Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevéUne image contenant capture d’écran

Description générée avec un niveau de confiance très élevéUne image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Rappel Stat I :

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Mean differences between 2 groups 🡪 (t-tests)

ANOVA: mean + std mais sur + de groupes

Standard factorial anova: **1** variable splits nos personnes dans différents groupes, et on veut voir comment les modalités de cette VI influencent notre (nos) **VD QUANTI. On a donc vraiment une variable qui détermine plusieurs cas de figures, plusieurs conditions expérimentales qu’on veut comparer**. Il dit aussi qu’on peut avoir plusieurs facteurs ?

Exemple : Orange Juice vs Vitamin C x 500mg vs 1000mg vs 2000mg +VD =long. dents

Density avec VD comme var (pas de split by).

Box plot

QQ plot. Density avec VD, (avec split by OJ/VC et mg)

ANOVA : One factor anova ou interaction: fixed factors (on met toutes les VI)

On regarde p par ligne.

η2 si on a pas d’interaction (taille d’effet)

partial η2 si on a une interaction pour avoir + de details (taille d’effet)

* + - * η2p = proportion de variance expliquée par le facteur. Valeur entre 0 et 1 (ds l’exemple, un peu pour le supp et bcp pour la dose)

Model : model terms = prédicteurs

Repeated measures :

Ancova : analyse de covariance :

Mancova = multiple ancova plusieurs VD + quanti

Non-parametric : on ne fait pas des tests sur des moyennes et des écarts type, mais sur des rangs

# Partie I – Méthodes explicatives (Introduction)

## Ajouter un filtre – désélectionner une partie des données

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

De manière générale : On veut faire CE TEST sur CES DONNEES qu’on sépare par CES GROUPES

### Importer des données

Format CSV (document texte brut)

Virgule = colonne suivante

1,2,2,4,5,

### QQ Plot

Test d’hypothèse : Est-ce que la différence entre les 2 groupes est significative ?

Comparer la loi normale en N(µ, σ2) avec les points de notre distribution pour savoir s’il y a, par exemple, 2 modes ou si notre échantillon suit bien cette loi normale

## Cours 2 – Intro

Note finale : 1x (note 1) + 2x (note 2) divisé par 3

Contrôle continu : réaliser des analyses (Jamovi) ensuite, extraire des paramètres à partie des sorties informatiques. Puis interpréter, mettre dans Word

### Variable à expliquer vs variable explicative

Variables explicatives servent à déterminer les variables à expliquer

Voir tableau

Anova à 1 facteur de classification : sert à rendre numérique un truc nominal

Slide 23

### Régression multiple sur Jamovi

Analyse > régression > régression multiple

### Modèles linéaires généraux

### Analyse confirmatoire

Notre modèle peut-il être décrit comme un autre modèle préétabli ?

Exemple : est-ce que le Big 5 se reflète dans mes données ? (=ressort-il une structure en 5 facteurs ?)

Sur Jamovi: Factor Reliability, exploratory factor analysis et confirmatory…

## Chapitre 1: Régression simple

### Modèle

Exemple du Phubbing (télésnobbage): Y a-t-il un lien entre le phubbing et la satisfaction conjugale ?

Modèle conceptuel (Fig. 1 slide 30) avec des variables intermédiaires entre le phubbing et la satisfaction conjugale

X = variable explicative (Télésnobbage)

Y = variable à expliquer (Satisfaction conjugale)

Y = β0 + β1X

Avec les paramètres β0 = ordonnée à l’origine (intercept) et β1 = pente

Une image contenant texte, tableau blanc

Description générée avec un niveau de confiance très élevé

yi = β0 + β1Xi +

= +

### Estimation des paramètres β0 et β1 (1.2)

On obtient une estimation des paramètres β0 et β1 en minimisant

= rxy ∙

=

Avec =

sx =

### Evaluation de la qualité de l’ajustement

#### Erreur standard

=

= (fraction multipliée par n)

= ∙ MSRE (moyenne scolastique résiduelle)

MSRE =

Sachant que εi suit une loi normale N(0, )

#### Corrélation rxy

SCT = SCreg + SCR

SCT = +

=

rxy = =

### Créer une droite de régression

#### avec les formules

Bravais-Pearson : on utilisera plutôt la droite des moindres carrés :

Fonction qui permet de prédire Y en fonction de X

Fonction y=â+b̂ x

Voir dans OneNote comment trouver â et b̂

Attention : on ne peut pas juste inverser la fonction pour trouver X

#### dans Jamovi

On a une colonne Phubbing et une colonne Satisfaction

Onglet régression>lin

Variables dépendantes : Satisfaction

Covar : Phubbing

>l Model fit (ajustement) RMSE

>l Coefficient

b) r=-0.80

c)sur Jamovi : on trouve r par model fit measure sous R dans les r

Résultats r(10)= -0.50, p= 0.002

d) 1 = r = -0.8 = 1 =-0.479

0 = -1 = 5.167-(-0.479∙11.833) = 10.835

Coefficient de corrélation significatif

Erreur standard

Model fit et cocher RMSE (moyenne scolastique résiduelle)

Notation APA

Bravais-Pearson

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

Résultats : R = R associé

Intervalle de confiance IC : si on n’a aucune valeur en 0, 🡪 significatif

## Chapitre 2 – Régression multiple

(cours 2)

Risque = RSQ estime de soi = EST âge= AGE

RSQ = β0+ β1EST + β2AGE + ε

On estime notre loi normale ainsi : ε ~ N(0, )

### Modèle (2.1)

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

2 variables explicatives

yi = β0 + β1 x1i + β2 x2i + εi

ε ~ N(0, )

β1 et 2 : coefficients de régression partielle qu’on va apprendre à estimer (comme la pente ?). Permettent de pondérer (notre équation ?)

β0: ordonnée à l’origine

β1:pente pour trouver β1, on parcourt x2 d’une unité et on voit que Y change d’une valeur de β2 (ou 1 ?)

β2: pente de l’autre droite (3D)

Représentation géométrique :

Variable explicative par une droite

À expliquer : graduations

3Dimension pour modéliser + de variables (cf. image dans fragments)

On a une observation i

traitillé = sous la courbe

trait droit = au-dessus (erreur sur le graphique)

k variables explicatives

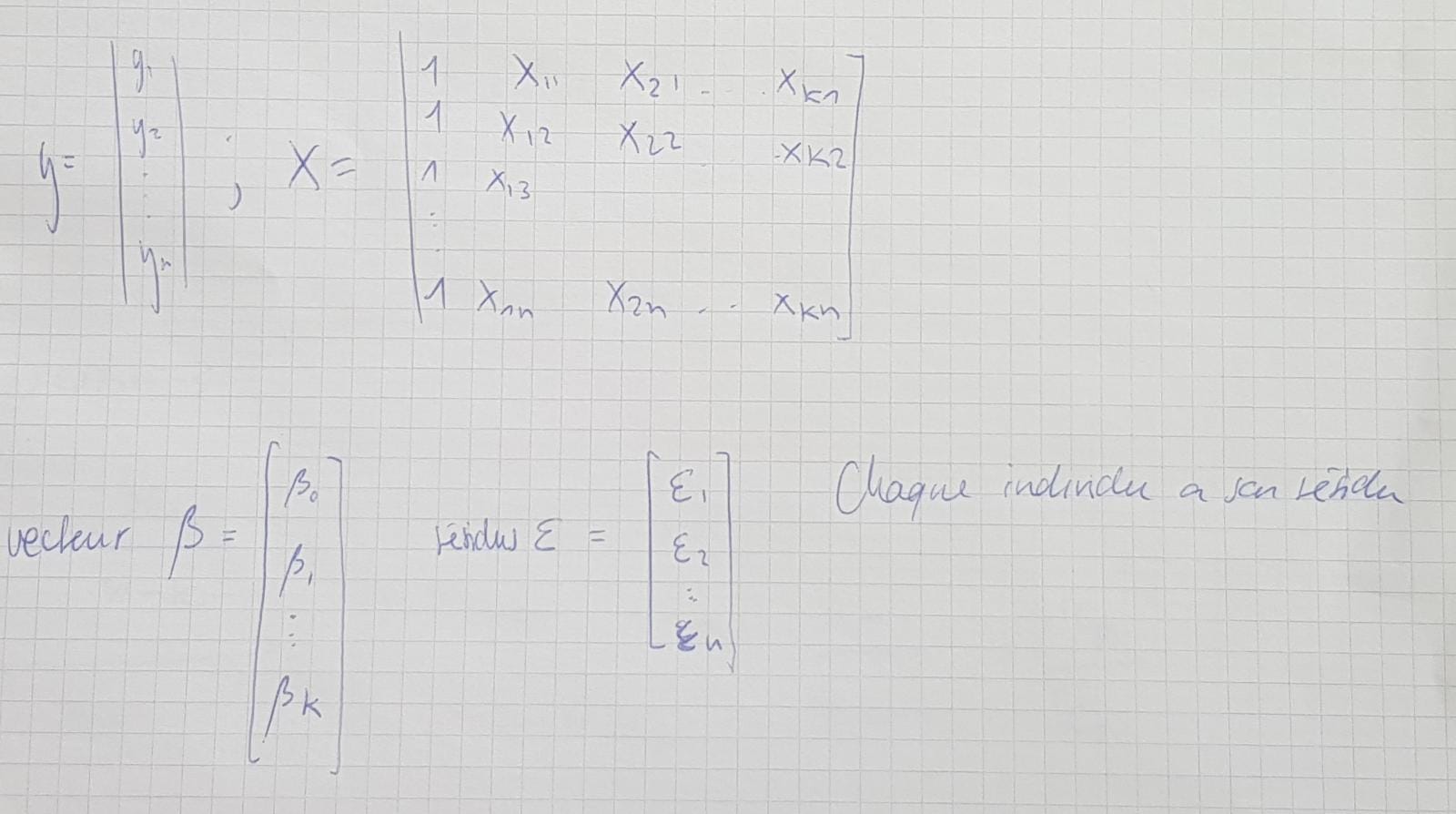
individu **yi = β0 + β1 x1i + β2 x2i + … + βjxji + … + βh xki + εi**

ε ~ N(0, )

### Estimation (2.2)

On obtient l’estimation des paramètres du modèle en minimisant

On va minimiser la somme des erreurs



n= taille de l’échantillon 🡪 autant d’individus que de résidus

k = nb de paramètres, de variables

en général, n>k

Le modèle sous sa forme matricielle :

y = xβ + ε

y a n lignes et 1 colonne 🡪 n x 1

X : n x (k+1)

β : (k+1) x 1

ε : n x 1

L’estimation des paramètres vaut (On permute les lignes et les colonnes):

= ( XtX)-1Xty

On va devoir interpréter les matrices

Interprétation des coefficients de régression partiels 1, 2, … , k

j est une mesure de l’effet de la variable explicative Xj sur le critère Y (lorsqu’on maintient tous les autres critères). j est le changement que l’on observe au niveau de Y lorsque Xj change d’une unité et que toutes les autres variables explicatives restent constantes.

|  |  |
| --- | --- |
| Situation 1 | Situation 2 |
| x1, x2, …, xj, …, xk | x1, x2, …, xj+1, …, xk |
| (1) = 0+1x1+2x2+…+jxj+…+kxk | (2) = 0+1x1+2x2+…+j(xj+1)+…+kxk |
|  | j(xj+1) = jxj+j 🡪 annule sit 1 sauf j |

|  |
| --- |
| DONC : Δ= (2) - (1) = j |

### Régression multiple dans Jamovi

Régression : analyse regression linear regression

Sél. VD et Vis (variables explicatives, prédictives)

Faire ceci fait apparaitre :

Les coeff. Dans la colonne « estimate »

Intercept x estimate : 0 = un indiv. D’âge 0 et qui aurait une estime de soi de 0

Age x estimate : 1 = si le signe de notre coeff est positif, 🡪 âge augmente 🡪 prise de risque augmente (correl. Pos)

EST x estimate : 2 = si le signe de notre coeff est négatif comme dans notre exemple, corrélation négative entre cette VI (estime de soi) et la VD (Prise de risque)

### Ajustement et qualité de l’ajustement (2.3)

Bon ajustement entre nos données et notre modèle ?

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

#### Erreur standard de régression (résiduelle) (2.3.1)

Ecart-type des résidus , qu’on estime ainsi :

ε = = RMSE ∙

RMSE=Erreur quadratique moyenne

(Il nous donne un exemple en nous disant que c’est un mauvais exemple DiDnT wRiTe lOl),

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevéIl dit que les différences ne transparaissent pas vraiment : Entre 0 et 0.8, la séparation n’apparait pas bcp, puis dès 0.8 bcp et très rapidement. Il y a donc bcp de séparation entre 1.010, 1.072 et 1.162

Autre tableau (Cohen). On se concentre sur la dernière ligne (f2) part de variance expliquée vs part de variance... (totale ? voir slides)

Autre tableau : on voit que les différences numériques sont faibles, mais fortes en matière de corrélation :

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

### Coefficient de détermination de R2 (2.3.2)

SCReg (Somme des carrés dûs à la régression) = 2

SCR = 2

SCT = 2 = SCReg + SCR

R2 = = 1 -

R = + = Coefficient de régression multiple

Les statisticiens proposent un indice ajusté :

R2 ajusté =  **1** - = **1** – = **1** - ∙

##### Dans Jamovi

Model fit> R, R2, Adjusted R2, et RMSE

### Aspect inférentiel (2.4)

### Test global (2.4.1)

Test qu’on réalise en premier, global, sur l’ensemble des coefficients. Si tous les coeff ne sont pas égaux à 0, on doit faire + de tests.

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

Tests marginaux = sur les coefficients

H0 : (β1 = 0) ∩ (β2 = 0) ∩ … ∩ (βk = 0) tous nos coefficients sont = à 0

H1 : (β1 ≠ 0) U (β2 ≠ 0) U … U (βk ≠ 0)

##### Table de l’ANOVA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | SC | ddl | CM | F |
| Régression | SCReg | k | CMReg= | F = |
| Résidu | SCR | n-k-1 | CMR= |  |

p = Prob (F (k, n-k-1) > Femp) (Concrètement?)

si p<a 🡪 on rejette H0

Dans Jamovi

Sous F = Femp

Sous df1 = k

Sous df2 = n- k-1

On voit que la p valeur est + petite 🡪 on rejette H0

## Cours 3

3 types de tests : globaux, marginaux (coefficients) et 2 modèles emboités

### Tests marginaux (test des coefficients) (2.4.2)

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

L’estimation est

E(j) = βj

Var (j) = ∙ on s’attend à ce que rj2 soit grand si … ε est grand (à cause de la division (what) )

est le coefficient de détermination du modèle

Xj ~ X1 + X2 + … + Xj-1 + Xj+ … + XR

Le facteur se nomme « facteur d’inflation de la variance » (VIF)

Erreur standard = racine carrée de la variance

H0 : βj = 0

H1 : βj ≠ 0

Sous H0, t = suit une loi de Student à n-k-1 degrés de liberté avec

(Photo notes Gaetan)

(slide tests marginaux (model coefficients) où β1 = AGE et β2= EST)Lorsqu’on fait une rég mult dans jamovi, on interprète un tableau des coefficients. (intercept age est)

Estimate = SE=SE( t = temp= /SE() p = 2x P( T>ltemp l l H0)

Comparaison de deux modèles emboîtés

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

H0 : (β1 = 0) ∩ (β2 = 0) ∩ … ∩ (βq = 0)

H1 : (βj ≠ 0) U (β2 ≠ 0) U … U (βq ≠ 0) 1≤ q ≤ k

(Photo avec les colonnes modèle 1 modèle 0 simplifié)

|  |  |  |
| --- | --- | --- |
|  | Modèle 1 (modèle complet) | Modèle 0 (modèle simplifié) |
| X1 | 1 |  |
| X2 | 2 |  |
| … |  |  |
| Xq | q |  |
| … |  |  |
| Xq+1 | q+1 | q+1 |
| … |  |  |
| Xk | k | k |
|  | SCreg1, SCR1 | SCreg0 ; SCR0 |

F = = = ∙ “on construit une statistique à partir de deux modèles »

Sous H0, F ~F(q, n-k-1)

#### Dans Jamovi : Model Builder

Block1:

Block 2: VI

On veut savoir si quand on rajoute une variable, on change les résultats (je pense)

Outputs : Model 1 = Modèle 0 dans la théorie Model 2 = Modèle 1 dans la théorie

df1 = nombre de variables explicatives pour chaque modèle

df11 = k-q

df12 = k

= 0.072

= 0. 126

n-k-1 = 998

q = 3

F = ….

Outputs de Model Comparisons

Cette fois, df1 = q

df2 = n-k-1

p= p valeur

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

#### Comparaison de modèles emboités

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

yi = β0 +β1x1i + … + βkxki + εi

= β0+β11+ … + βkk

yi - = β1(x1i-1) + … + βk(xki - k) + εi

= β1 + … + (photo de l’éq avec les couleurs là)

Il faut se rappeler que βj\* est sj/ sy x βj (en français, qu’est-ce que β\*j

Model coefficient, sous stand estimate = // cours de différentielle avec part de chocolat etc Plus la stand. Estimate est grande, + elle est prévalente sur les autres variables. Dans l’exemple, les 2 plus grands facteurs sont l’âge et le DEP

#### Dernière partie : on vérifie (quoi ?)

2.5 Diagnostic

yi = β0 + β1 x1i + β2 x2i + … + βjxji + … + βh xki + εi

εi ~ N(0, )

2.5.1 Hypthèse de linéarité

E (ε l X1, X2, … , Xk) = 0

(Dessin 1)

### Hypothèse de linéarité dans jamovi

On veut la courbe de lissage … ? (voir jeudi)

1. Calculer les scores ajustés
2. Calculer les résidus ε
3. Représenter ε
4. Copier qn

### 2.5.2 Hypothèse d’homoscédasticité = d’homogénéité mais pas pour une analyse de variance= dispersion des erreurs

Var (ε l X1, X2, … , Xk) =

(Dessin 2)

Jamovi :

Compute dvariable

RSQ.sqrt(res)

Fx(

* Si la droite n’est pas **horizontale**, l’hypothèse d’homoscédasticité n’est pas satisfaite.

### 2.5.3 Hypothèse de normalité

On utilise 2 outils : graphique et inférentiel (test)

1. 1er outil (graphique):

Représenter le QQ plot

Dessin 3

#### Dans Jamovi

Assumption checks > QQ truc

Test de Shapiro-Wilk: analyses> t -tessts > one sample t-test> Assumption checks>normality Shapiro Wilk

## 3. Analyse de variance à deux facteurs

Régression simple ou multiples : toutes variables numériques

Si elles sont toutes nominales : on fait une ANOVA. Ici, à 2 variables indépendantes. Quelle est l’influence de ces 2 variables sur la VD  (numérique) ?

Exemple de donnée :

On demande à des enfants en échec ou en réussite scolaire (VI1 ou facteur B modalités échec = b1/réussite= b2)de faire une tâche de géométrie ou de dessin (VI2 ou facteur A modalités géométrie =a1/dessin=a2). VD = nb de points.

iUne image contenant capture d’écran

Description générée avec un niveau de confiance très élevé

On a le même effectif dans chaque case

N=nab

### Structure des données dans Jamovi

Colonnes : ID, A(on entrera soit a1 soit a2), B(on entrera b1 ou b2), X(VD) Ligne : chaque sujet,

Une image contenant capture d’écran

Description générée avec un niveau de confiance très élevé (on mettra pas a1/b1/x221 mais les valeurs observées)

Ici, i = la modalité de la variable i (ici, i=A)

j=la modalité de la variable j (ici j=B)

k= le # de l’individu qui a les variables i et j. Par exemple, les ID 1 et 2 ont a1 et b1, du coup on précise que l’ID 2 est le 2ème dans ce cas en mettant k=2

### Le modèle :

xijk = µ + αi + βj + (αβ)ij + εijk

αi = effet/contribution du groupe

βj = effet dû au contexte (dessin/géométrie)

(αβ)ij = effet d’interaction

εijk ~ N(o, )

### Effet d’interaction

(je parle de variable x en abscisse et y en ordonnée et pour les segments)

On place 4 moyennes :

Les 2 moyennes marginales = (moyennes des 2 bouts des segments = moyenne des deux points en x1 et moyenne de deux points en x2)

La moyenne de y1 (point au milieu du segment y1)

La moyenne de y2 (point au milieu du segment y2)

Si parallèles = pas d’effet d’interaction

Si pas parallèles = effet d’interaction

Si les moyennes marginales sont à la même hauteur,

#### 8 exemples d’effet d’interaction

On représente la VD en ordonnée, et une des VI en abscisse. On a nos deux modalités a1 et a2 sur l’axe x.

Une image contenant tableau blanc, texte

Description générée avec un niveau de confiance très élevé

1. Aucun effet : Chaque point représente une moyenne. La VD a toujours la même valeur
2. Effet de A : Pas horizontaux mais toujours confondus. Moyennes marginales sont confondues. Pas d’effet de B.
3. Effet de B uniquement : moyennes marginales vertes sont les mêmes mais marginales de B sont différentes
4. Effets de A + B : Moyennes marginales ne sont pas au même niveau. Il y a effet de B et moyennes marginales pas au même endroit.
5. Effet d’interaction AB. Y at-il effet de A ? 🡪 on fait moyenne marginale, 🡪 pas d’effet de A. Ni d’effet de B. Mais effet de AB.
6. Effet de AB (car droites pas parallèles) + effet de A. Moyenne marginale de a1 et a2 pas à la même hauteur 🡪 effet de A.
7. AB + B ni parallèles ni confondues du coup effet d’interaction.
8. AB + A + B Les 2 pts verts ne sont pas à la même hauteur 🡪 effet de A

Effet de B car points blancs ne sont pas a la même hauteur

### Test statistique pour savoir si effet de A, B et/ou d’interaction AB

Test de somme des carrés

SCT : Somme des carrés totales à laquelle on soustrait la moyenne. Elle peut se répartir en « 4 morceaux » qu’on peut attribuer à A, B, AB et résidu.

SCA (somme carrés associés à A) = moyennes marginales associées à A.

SCB sommes marginales associées à B.

SCAB = ce qu’il reste quand on a fait A et B

Quand on les a calculés, on les met dans une table de l’ANOVA avec nos 4 souces de variabilité.

A : (Nb de modalités de a)-1

B : idem

Ddl associés au résidus : n-1 ∙ nb de cellules

F = variable empirique associé à A = CMA/CMR

Les trois varaibles empiriques F vont nous aider à prendre la décisier d’effet de A, de B et d’interaction

### Avec Jamovi

Analyse>ANOVA>ANOVA> voir slides

### Stratégie :

Une image contenant texte, tableau blanc

Description générée avec un niveau de confiance très élevé

Dans la sit. Illustration, on a des interactions. Dans la pratique, on doit tester les effets simples. Mais on doit savoir faire les tests post hoc (voir slides 3.3)

Installer GAMLj

### Puissance des tests (3.5)

En fonction de la présence ou non d’interaction, on fait soit les effets simples ou un test post-hoc.

+ Test de la puissance des tests

ηÊta carré = part de variance expliquée. Comme le R carré pour la régression multiple

ηÊta carré partielle = part de var expliquée comparée à l’effet du résidu uniquement

ηÊta carré = SCEffet/SCT

η

Êta carré partiel = SCEffet / SCEffet+SCR

### Dans Jamovi

On coche « taille d’effet » dans le menu anova 🡪 rajoute 2 colonnes à la fin (voir slides)

On voit que l’interaction est significative et la taille d’effet est de 0.194 = la part de **var expliquée par l’interaction**.

## Chapitre 4 : Plans complexes

### Différents plans d’expérience (4.1)

#### Plan S<A> (4.1.1)

Un plan intersujets

S <A> = les sujets sont emboîtés dans le facteur A

Blocs : si les carrés se touchent = les sujets appartiennent au même groupe. Le facteur A = un groupe. Repérés par la variable S

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | a1 |  | a2 |  | a3 |
| s1 |  | s4 |  | s7 |  |
| s2 |  | s5 |  | s8 |  |
| s3 |  | s6 |  | s9 |  |

SCTotal 🡪 soit SCintergrupoe soit SCintragroupe

Table de l’ANOVA

Source du à A ou S qui se trouve dans A

#### Plan S \* A  - Plan à un facteur intrasujet (plu emboité mais croisé) (4.1.2)

Plan croisé : sujets x modalités d’un facteur A

|  |  |  |  |
| --- | --- | --- | --- |
|  | a1 | a2 | a3 |
| s1 |  |  |  |
| s2 |  |  |  |
| s3 |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table de l’ANOVA

|  |  |  |  |
| --- | --- | --- | --- |
|  | CM | Ddl | F |
| 1. | S | n-1 |  |
| 2 | A | a-1 | 2/3 |
| 3 | SxA | (n-1)(a-1) |  |

On ne calcule qu’un F, c’est celui qui divise les CMA par CMrésidus ( ou interaction)

Il y a des facteurs fixes et des facteurs aléatoires. Les facteurs fixes et modalités de A nous intéressent. Il y en a peu et on les connait toutes (les modalités). On veut les comparer.

Exemple de modalités de A (Analgésique): Aspirine, Ibuprofen et …

Si on veut savoir s’ils ont tous la même efficacité, on sél. Au hasard 3 modalités pour le facteur A. mais échantillonnage du facteur A : il n’est pas fixe mais aléatoire car vient d’un échantillonnage.

#### Plan à deux facteurs intergroupes S < A \* B > : sujets emboités dans 2 facteurs A et B (4.1.3)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | a1 |  | a2 |  | a3 |  | a2  b1 |  | a2  b2 |  | a2  b3 |
| s1 |  | s4 |  | S7 |  | S10 |  | S13 |  | S16 |  |
| s2 |  | s5 |  | S8 |  | S11 |  | S14 |  | S17 |  |
| s3 |  | s6 |  | S9 |  | S12 |  | S15 |  | s18 |  |

Table de l’ANOVA

|  |  |  |  |
| --- | --- | --- | --- |
|  | CM | Ddl | F |
| 1. | A | a-1 | ¼ |
| 2 | B | b-1 | 2/3 |
| 3 | AxB | (a-1)(b-1) | ¾ |
| 4. | S/AxB | ab(n-1) \* |  |

\* car plan équilibré. Ici n=3

Total : abn-1 = nb de ddl total (

Ddl associé à la somme des carrés totale

n=nombre de cellules du plan

Savoir si on a un effet principal de A : on divise la première ligne par la 4ème.

#### Plan mixte à un facteur intrasujet et un facteur intersujets (4.1.4)

Plan S < A > \* B (4.1.4): sujets emboités dans A, puis croisés avec B. Plan mixte. A : intersujet, B : intrasujet

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | a1 | a1 | a1 |  | b1 | b1 | b1 |  |
| s1 |  |  |  | s1 |  |  |  |  |
| s2 |  |  |  | s2 |  |  |  |  |
| s3 |  |  |  | s3 |  |  |  |  |

n=3

#### Plan à 2 facteurs intrasujet (4.1.5)

SxAxB = plan où on croise 3 facteurs

On n’a qu’un groupe 🡪 une colonne

6 conditions expérimentales (par le croisement de A et B)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | a1  b1 | a1  b2 | a1  b3 | a2  b1 | a2  b2 | a2  b3 |  |
| s1 |  |  |  |  |  |  |  |
| s2 |  |  |  |  |  |  |  |
| s3 |  |  |  |  |  |  |  |

### Analyse de variance sur mesures répétées (4.2) SxA

### Jamovi et résumé du cours

* Slides 181022

Etape 1 : On réalise une ANOVA en supposant que l’hypothèse de spléricité est satisfaite (ce qui n’est pas forcément vrai). Permet de trancher entre les hypothèses suivantes :

H0 : la durée µ1 = µ2 = µ3 = µ4

H1 : Il exite un j ou un k pour lequel µj ≠ µk

(manque des notes)

Retour slides

#### Etape 2 : Vérification de l’hypothèse de sphéricité

= « Variance des différences des variables est-elle toujours la même ? »

H0 : la variance entre B1-B2) = Var (B1-B) … = Var (B3-B4) sphéricité

H1 : [Var(B1-B2) ≠ Var (B1-B3)] U … U [Var(B2-B3) ≠ Var (B3 – B4)]

#### Etape 3 : Tester la sphricité: Test de Mauchly

Voir slides (7/21)

Si la sphéricité n’est pas satisfaite, la loi à laquelle nous nous référerons pour tester l’effet de notre facteur intrasujet sera une loi de Fisher – Suedecor ( ?) à ε(a-1), ε (n-1)(a-1) degrés de liberté (ddl).

D’autant plus petit que nos données sont non-sphériques.

F~F(ε(a-1), ε(n-1)(a-1))

#### Résumé du cours :

On fait des anova selon les variables intersujets ou intrasujet (répété)

Jamovi

On fait un test selon lhypothèse que sphéricité,

On se rend compte que non ? 🡪

On corrige par Greenhoues Ghesser (voir sl. 9/20)

Levene

Diagramme d’interaction

#### Etape 4 : Comparaison multiples

Etape 5 dans les slides

### Analyse de variance sur plan mixte (4.3)

Slides 14/21

## Analyse de covariance (5)

PAS DANS ANCOVA : DANS GLM

Variables muettes : si poly : k-1 (nombre de modalité de la VI – 1). Elle seront = 1

### Une variable numérique et une variable norminale (5.1)

1. Amélioration de la précision du modèle

[2 premiers tab]

Dans le 2 e, on crée 2 droites de régression et on s’aperçoit que l’erreur standard du modèle est bcp plus petite (car on a 2 groupes qui semblent significativement différents). 🡪 on peut créer une meilleure prédiction selon les groupes.

1. Eviter les biais

[tableau post test – pre test]

2 nuages permettent de positionner…

Double flèche à gauche : On pourrait attribuer la différence à l’intervention. Mais les groupes de base n’étaient pas équivalents. Donc la différence est la double flèche à droite

Créons un seul modèle contenant à la fois la variable numérique

Niveau de la variable Y pour i qui sera = à

Yi = a + βXi + γDi + εi

Εi ~N(o, …)

Avec D une vairalbe nominale doichtomise (variable « dummy » ou « muette »)

ON veut donner la valeur de 0 à… par exemple la variale sexe

Di =

### Jamovi – faire les droites et effet du sexe (γ)

PISA dans les données sur moodle.

On a Sexe, Age, SEI (niveau socioeco, nmérique), Voie (VSG/VSO) et Note en maths. Pour la variable dichotomique, on s’intéressera au sexe. Les filles est le groupe auquel on va comparer les autres ( on soustrait fille à garcon ou le contraire ja’i pas suivi)

D.G = variable indicatrice (voir slide 2)

Si Di = 0 (Sexe == « F »)

Yi = a + βX + γ0 (il a enlevé ε ( ?))

Si Di = 1 (Sexe == G )

) a + βX+ 1γ

= a + γ + βX 🡪 on peut faire les droites qui caractérisent les 2 groupes !

Graphiquement :

[dess 1]

Effet de la variable quali

H0 : γ = 0

H1 : γ ≠ 0

Pour chaque paramètre, on peut faire un test marginal (sur γ ?) (j’ai pas compris ce qu’il voulait dire)

Step 1 : On calcule t =

Slide 3

### Comparer 2 modèles emboités

Y dépendante X = numérique

Modèle 0 : Y ~ X

Mod¨èle 1 : Y ~ X + D

On les compare avec un test de fisher : slide 6

Model Builder on introduit uniquement la var numérique X dans le blcok 1 , et la var dichotomique dans le block 2.

On compare les modèles et on obtient un F

APA : F(1, 1582) = 30.105, p<.001 (1582 = ddl)

### Variable nominale polytomique (5.2)

+ de 2 modalités (pas dichotomique)

Exemple : y a-t-il une différence entre les filières VSO, VSG et VSB?

D1 permet de repérer les individus en VSG et D2 ceux en VSB : implicitement, la catégorie de référence sera VSO, à laquelle on va comparer les autres. (on soustrait vso aux autres je crois)

|  |  |  |  |
| --- | --- | --- | --- |
|  | D1 | D2 |  |
| VSO |  |  |  |
| VSG |  |  |  |
| VSB |  |  |  |

Jamovi : computed slide 8

Le modèle de régression est le suivant

Yi = α + βXi + γ1D1i+ γ2D2i + εi

D1i D2i

0 0 Y = α + βX + 0γ1+ 0γ2 + εi

Y = α + βX + εi

1 0 Y = α + βX + 1γ1+ 0γ2 + εi

Y = (α + γ1)+ βX +εi

…

Graphiquement

[dessin 2]

### Test de l’effet de la variable nominale

Savoir si différence ou non entre les 3 grupes de la var nominale. On va faire des modèles emboîtés.

H0 : (γ1 = 0) ∩ (γ2 = 0)

H1 : (γ1 ≠ 0) U (γ2 ≠ 0)

Les modèles à construire sont :

Modèle 0 : Y ~ X

Modèle 1 : Y ~ X + D1 + D2

Construire les modèles dans Jamovi :

Prmier block : var X

Dans 2ème block : toutes les variables (voir slide 11)

On teste l’impotance de la variable filière

### Modélisation des interactions (5.3)

[dessin 3]

..

L’effet de D est plus important en x2 dans le 2ème tableau

On peut modifier le modèle général de l’interaction de la manirère suivante :

Yi = a + βXi + γDi + δ(DiXi) + εi

Cette équation se décline ainsi :

Lorsque D = 0, Y vaut :

D= 0 Y= = a + βX + 0γ + δ(0X) = a + βX

D = 1 Y = a + βX + 1γ…

[dessin 4]

Pour tester l’interaction :

H0 : δ =0 Modèle 0 : Y ~ X + D

H1 : δ ≠ 0 Modèle 1 : Y ~ X + D + DX

### Représentation graphique des résultats de l’ANCOVA sur Jamovi

Variable nominale comme facteur.

1. On ouvre une nouvelle page jamovi.
2. On utilise la modalité DG comme première modalité (mais pas encore codée en 1 0)
3. Linear models>GLM>VD : VD, Factors: DG, Covariates: Vis
4. Les variables sont centrées. Pour les décentrer :

Note glm bullet points : vérifier la variable qui est mise en premier qui doit être DG (la dummy).

1. GLM horiz : VI, Separate : DG (//S72m)

### Conclusion

Le GLM ne permet pas de comparer les modèles emboités.

# Questions

Autre tableau : on voit que les différences numériques sont faibles, mais fortes en matière de corrélation

Doit-on utiliser ces valeurs pour l’examen ?

# Examen

1h30 (check salle)

Dans un doc word dans lequel on mettra les réponses en normes APA

1 pt pour la conclusion (phrase)

5pts par exo normalement

En-tête : nom prénom et n° de la copie