**Formulaire**

**Statistique multivariée**

Psychologie à l’université de Lausanne

2018/2019

# Eileen Rabel

Toutes les informations provenant de Professeur Jean-Philippe Antonietti

**Conseils JAMOVI:**

**Attention**: Entrer les données :

1. Toujours regarder s’il s’agit d’une **variable numérique** (continous), **nominal ou ordinal** !!
2. Si on a des **groupes indépendants** alors on peut superposer les résultats en une colonne avec en 2ième colonne le nom associé la donnée xi
3. Si les **groupes** sont **dépendantes** (mesure avant – après, VD et VI) alors faut une colonne par variable
4. Ecrire les **noms** des colonnes/groupes etc. **TOUJOURS** de la même manière !!
5. **CONTRÔLER les Hypothèses** !!

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Régression simple :**  **Statistique**  Modèle = Equation de la droite  Estimation des paramètres du modèle (= l’ordonnée à l’origine) et (= droite des moindres carrés/pente)  Prédiction d’un score  Variance  Variance expliquée  Variance totale  Ex : Série2, ex. 2d)-f)  Proportion/Part de variance expliquée  Evaluation de l’ajustement   1. Erreur standard de régression   Ex : Série2, ex. 1f)   1. Corrélation multiple rxy de Pearson   **Suite Régression simple :**  **Statistique**   1. Test de corrélation   Ex : Série2, ex. 1c)  Tests de valeur de β  Test de la valeur de l’ordonnée à l’origine  Test de la valeur de la pente  Intervalle de confiance 95%  Ex : Série2, ex. 1g)   1. de l’ordonnée à l’origine 2. de la pente   Diagramme de Dispersion avec droite de régression | **Formule**  = +\*X    Avec :  =  = -  Avec  = +\*Xi          **Formule**  **Description**  ddl = N-2  avec N = nombre des sujets  ddl = N-2  ddl = N-2  IC95% = [lower; upper] de (= intercept)  IC95% = [lower; upper] de | Mesure du lien entre deux variables  **Description**  Y = Equation de droite  et sont des approches théoriques qui estiment au mieux et  Variables non-standardisées !  Si xi a une certaine valeur, ou se trouve-t-il sur la pente y ?  En élevant au carré la valeur de l'écart-type des scores prédits/observés des Y, on obtient la variance expliquée des scores prédits/observés des Y  **R2**= part de la variance expliquée = **Coefficient de détermination**  Indication de la corrélation entre les variables  **Jamovi**  Mesure de **significativité** du lien entre deux variables  (H1 : pxy > / ≠ / < 0)  L’ordonné à l’origine/la pente de la droite est-elle significativement différente/supérieur/inférieur de 0 ?    Représentation graphique des données | **Jamovi**  **, ,** et : Entrer les données en 2 colonnes **>** Analyses > Exploration > Descriptives > les deux variables en « variables »  Regression **>** linear regression **>** dependante variable (VD) et covariables (VI)    Calculer le score prédit y de tous les individus : nouvelle colonne « scorespredy » avec « **fx = β0+β1\*X** »  Ex : Série2, ex. 2c)  Analyses > Exploration > Descriptives > Y et scorespredy dans « variables » > Statistics > Dispersion : Variance  Obtenir infos sur l’ajustement :  Linear Regression > Model fit **>** RMSE, R2 , Pearson    (**Attention : rxy n’est pas donné en négatif, faut regarder le diagramme**)  **APA/Réponse**  Regression > Correlation matrice > Pearson; Report significance  Observer t et p dans le tableau    Regression > Linear regression > Dep. variable (Y) ; cova. (X) > Model coefficents > Estimate: Confidence interval    Analyses > Scatterplot > x (abscisse), y (ordonnée) > regression line:  linear    Ex : Série2, ex. 1e) | **APA/Réponse**  **L’Ordonné à l’origine β0 est la valeur que prend théoriquement la variable Y lorsque tous les variables X prennent la valeur 0.**  **β1 = coefficient de régression partiel**  Exemple de Equation de la droite:    Exemple de prédiction :  Si xi = 16 alors :  Si prédiction de y :    Attention à l’écriture de x! Doit être la même que dans la colonne  R2 = part de la variance expliquée (ex : si R2 de X = 0.321 alors X explique 32,1% de la variance)  Ex : Le coefficient de corrélation de B.-P. vaut rxy = -0.8, ce qui indique une forte correl. négative entre les variables  APA: r (N-2) = …; p = …  Ex: 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 X et Y est différente de zéro  L’ordonné à l’origine est significativement différente de 0 au seuil de 5% (t(10) = 7.356 ,  p < .001).  La pente est significativement différente de 0 au seuil de 5% (t(10) = -4.219, p = 0.002).   1. L'intervalle de confiance à 95% de l’ordonnée à l'origine vaut [lower ; upper ]. 2. L'intervalle de confiance 95% de la pente vaut [lower ; upper ]. |
| **Régression multiple :**  **Statistique**  Modèle théorique= Equation de la droite   1. 2 variables explicatives 2. k variables explicatives   Modèle linéaire   * Estimation des paramètres du modèle   Ex : Série3, ex.1b)  Scores ajustés/prédites  Scores résiduels  **Suite Régression multiple :**  **Formule**  **Statistique**  Corrélation entre residus et prédicteurs (Xk) / l’ensemble des variables  Erreur standard de régression  Ex : Série3, ex. 1c)   * Indice de séparation   Coefficients  Ex : Série3, ex. 1d)  **Test global** sur l’ensemble des coefficients β  « La régression multiple est-elle significative au seuil de α = 5% ? »  **Si p < 0.05 alors on va faire un test marginal : on teste la significativité des coefficients**  **Suite Régression multiple :**  **Statistique**  Ajustement du modèle  « Evaluez le modèle esquissé »  **Tests marginaux** (test des coefficients de régression partiels)  Pour voir laquelle des variables du test globale est significative.  **Suite Régression multiple :**  Comparaison de deux **modèles emboités**  **Statistique**  **Formule**  **Diagnostic : Condition d’application** nécessaires pour la régression multiple  Hypothèse de linéarité  Ex : Série4, ex. 2b) | **Description**  **Formule**    Avec :    Avec :  +  = +  –  **Description**          **Attention : Si l’échantillon n est très petit, on a le risque de commettre une erreur de 2ième espèce**  **Description**  **Formule**  1. Test Global  2. R2 + interpret  3. erreur st.+ interpret  4. Tests marginaux  t =  avec      F =  **Description**  Sous H0, F ~ F (q, n-k-1)  Avec  Et | Mesure de lien entre plusieurs variables  Variables non-standardisées !  Il faut remplacer les X par les *predictors* (nom des colonnes)  Il faut remplacer β par les chiffres dans *estimate* et X par les *predictors*  Nécessite de calculer les **scores prédites/ajustés** pour pouvoir calculer les residus*.* En suit distraire les scores ajustés de la variable à expliquer (VD = Y)  **Jamovi**  Sy = l’écart type de la variable dépendante Y  Avec k = nombre de variables X    R2 =r de détermination  r = coeff. de corré. 🡪 R de JAMOVI  R = coeff. de corrél. multiple (nécessite que dans la réponse on parle de r et R2)  Indice ajusté  Significativité de liaison globale entre Y et les VI (Xk)  On fait un test global de regression avec les Hypothèses suivantes :  H0 : R2 = 0  H1 : R2 > 0  **Si p < 0.05 alors on va faire un test marginal : on teste la significativité des coefficients !!**  Significativité des paramètres β  = facteur d’inflation de variances = Colinéarité statistique = VIF  Mesure de l’influence significative d’une variable sur les résultats  **Jamovi**  k = nombre de tous VI  q = nombre des VI du modèle simplifiée  = Coeff. De détermination du modèle 2 complet  = Coeff. de détermination du modèle 1 simplifié  Si **p < 0.05** alors la/les variables enlevées q ont une influence et il faut garder le modèle complet.  Si **p 0.05** alors pas d’influence de la/les variables q et on peut opter pour le modèle simplifié.  Voilà notre modèle sur lequel on se base pour tester linéarité, homoscédasticité et normalité  Pour vérifier que la moyenne des résidus est égale à 0 conditionnellement à .  Calcule des moyennes des résidus et on obtiendra une courbe de lissage qui coïncide presque avec l’axe | Description numérique :  **Jamovi**  **, ,** et : Entrer les données en k+1 colonnes **>** Analyses > Exploration > Descriptives > les variables en « variables » (k = nombre des VI, 1 = VD)  Regression **>** linear regression **>** dependante variable (VD) et covariables (VI)    Scores ajustés : Nouvelle colonne « scoresajustes » avec fx =    Résidus : Nouvelle colonne « residus » avec fx =  –    **APA/Réponse**  Correlation Matrice > residus et tous X en variables > observer la ligne des residus puis interpréter les corrél.  Linear Regr. > Model fit **>** RMSE  Analyses > Exploration > Descriptives > VD en *variables* > **Dispersion** > Std. Deviation  Linear Regr. > R2 , R ,  Calcule pour l’ANOVA :  Linear Regr.> Model Fit > **Overall Model Test** : F test  df1 = k ; df2 = n-k-1  **F =**  Table de l’ANOVA :   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source | SC | ddl | CM | F | | Regression | SCReg | k |  |  | | Residus | SCR | n-k-1 |  |  |   Linear Regr.> Model Fit > **Overall Model Test** : F test  **Jamovi**  Avec df1 = k ; df2 = n-k-1  Sous Ho, t suit une loi de Student à df = n-k-1 🡪 k = nombre des VI, n = taille de l’échantillon N  Linear regression > entrer VI en *Covariates* et VD en *Dependent Variable* > regarder le tableau Model coefficents 🡪 SE, Estimate (= et t sont donnés.    Le VIF = linear regression > Assumption Checks > Collinearity statistics    Linear regression > entrer VI en *Covariates* et VD en *Dependent Variable* > Model Builder > Block 1 = tous les VI de q ; Block 2 = tous variables restant > regarder le tableau de *Model comparaisons*  **APA/Réponse**  Comparaison de modèle 1 et 2    Modèle réduit : Pour définir si la variable a un impact sur le modèle il faut regarder le tableau *Model Fit Measures* et observer p.    Modèle complet 2  Modèle simplifié 1  1. Calculer les scores ajustés : Ajouter colonne « scorespred » avec **fx =**  2. Calculer les résidus : Ajouter colonne « residus » avec **fx = Y – scorepred** (Y = Variable indépendant/ à expliquer)  3. Représenter les résidus en fonction des : Exploration > Scatterplot > X-axis = *scorespred* et Y-Axis = *residus*  4. Ajouter la courbe de lissage ***: Regression line****Smooth* | **APA/Réponse**  **Interprétation : L’Ordonné à l’origine β0 est la valeur que prend théoriquement la variable Y lorsque tous les variables X prennent la valeur 0.**  **βk = coefficient de régression partiel**  Ex : Série3, ex. 2c)  Modèle :  Après estimation, l'équation du modèle devient :  = 1.669 + 0.605 X1 – 0.334 X2 +  0.486 X3 + 0.07 X4  Moyen d’interprétation : Lorsque le degré de responsabilité augmente la satisfaction au travail augmente aussi (X1). Si nombre des personnes supervisées augment, la satisfaction diminue (X2). Si qualité de l’environnement augmente, alors sat augmente aussi (X3). Et comme r de X4 proche de 0, montre que nombre d’années de service n’a qu’une influence négligeable sur la sat (X4)  Exemple : Satisfaction - scoresajustes  Ex: = 2.243 🡪 Si les erreurs se distribuaient normalement, 68% des observations se situeraient dans une bande de 2.243 autour du plan définit par le modèle  Indication de la taille de l’indice de séparation : La séparation vaut  = …, elle est faible/modérée/forte    Interprétation de R2 : Si R2 = 0.486 : 48,6% de la variance totale est expliquée par les prédicteurs du modèle.  *Général : Plus on a des VI, plus il y a de différence entre R2 et*  **Hypothèses :**  **APA :**  F (df1, df2) = … ; p 🡪 si p < 0.05 on rejette H0  Ex : Globalement la liasion entre la variable critère Y et les varables prédictirces Xk n’est pas significative au seuil de 5% (F(4, 10) = 2.367, p=0.123). Aucune des variables prédictrices permet d’expliquer la variable Y.  Ex : Série3, ex. 1i)  **Test global de regression**  F(3, 131) = 21:84, p < 0:001. Y dépend significativement au seuil de 5% de X1, X2 et X3. La part de variance expliquée vaut 33.3%. Nous voyons cependant que l'erreur standard est assez importante ( = 8.427).  En suite : **Test de coefficient partiel** (marginal)  Exemple :  Hypothèses :  **APA/Réponse**  Ex : Série3, ex. 3g)  Ho : βj = 0  H1 : βj ≠ 0  *1. Test du coefficient de régression partiel associé à la variable X1* :  Résultat :  t (n-k-1) = 7.829, p < 0.001.  Conclusion :  Comme p < α, on rejette l’hypothèse nulle affirmant que β1 = 0 au seuil de 5%.  *2. Test du coefficient de régression partiel associé à la variable X2 :*  Résultat :  t (n-k-1) = 16.169, p < 0.001.  Conclusion :  Comme p < α, on rejette l’hypothèse nulle affirmant que β2 = 0 au seuil de 5%.  *3. Test du coefficient de régression partiel associé à la variable X3 :*  Résultat :  t (n-k-1) = 0.406, p = 0.685.  Conclusion :  Comme p > α, on accepte l’hypothèse nulle affirmant que β1 = 0 au seuil de 5%.  **Interprétation :**  Le modèle linéaire est significatif mais seul X1 et X2 affectent véritablement Y. (remplacer les variables par les mots utilisés dans l’exercice)  Ex: Critère d’emboîtement :  Si les modèles sont emboités : Toutes les variables du modèle réduit sont présentées dans le modèle complet :    = « n’est pas inclue »  = « est inclue »  Exemple : Comparaison du modèle complet à un modèle réduit d’un prédicteur :  *Suppression de la variable X3 :* Hypothèses :  H0 : β3 = 0  H1 : β3 ≠ 0  Conclusion :  **p 0.05 :** Le test de Fisher permettant de comparer deux modèles emboîtés **n'est pas significatif** au seuil de 5% (F(1, 69) = 1.304, **p = 0.257**). Le modèle réduit – sans la variable X3 – où β3 = 0 est meilleur que le modèle complet : le modèle réduit est plus parcimonieux et décrit les données presque aussi bien que le modèle complet. Dans le modèle complet, X3 n'influence pas Y.  **p < 0.05 :** Le test de Fisher permettant de comparer deux modèles emboités **est significatif** au seuil de 5% (F(1, 69) = 9.828, **p = 0.003**). Le modèle complet est meilleur que le modèle sans le test d'information où β3 = 0. Cela signifie que dans le modèle complet de régression, X3 de manière non négligeable Y.    Si les hypothèses sont satisfaites La moyenne des résidus vaut 0 pour grosso modo tous les scores ajustés |
| **Suite Régression multiple :**  **Statistique**  Hypothèse d’homoscédasticité  Dispersion des erreurs  Hypothèse de normalité  Distribution des résidus | **Formule** | **Description**  La variance des erreurs est-elle constante ? En cas d’hétéroscédasticité les estimations des paramètres sont toujours fiables par contre celles des erreurs standards peuvent être fausses.  **Homoscédasticité** = la droite est horizontale (en moyenne il faut que la racine de la valeur absolue des résidus(Y) reste stable en fonction de la valeur des scores ajustés/prédits (X))  **Hétéroscédasticité** =la droite est inclinée (positive ou négative) (en moyenne il faut que la variance résiduelle (Y) augmente ou baisse en fonction de la valeur des scores ajustés/prédits (X) pour que nous violons l’homoscédasticité)  Suivent-les résidus une loi normale ?  2 moyens :   1. Représentation graphique par diagramme des quantiles-quantiles des résidus 2. Test de normalité de Shapiro-Wilk   3. Représentaiton de l’hisotgramme et densité sous : Descriptivec > Plot > Histogram, Density | **Jamovi**  **APA/Réponse**  1. Calculer les scores ajustés : Ajouter colonne « scorespred » avec **fx =**  2. Calculer les résidus : Ajouter colonne « residus » avec **fx = Y – scorepred** (Y = Variable indépendant/ à expliquer)  3. Ajouter colonne « varianceresiduelle » avec :  **Fx = SQRT(ABS(residus))**  4. Représenter les résidus en fonction des : Exploration > Scatterplot > X-axis = *scorespred* et Y-Axis = *varianceresiduelle*  5. Ajouter la courbe de lissage ***: Regression line****linear*   1. Graphique: Linear regression > Residus en *Dependant*, Scorepred en *Covariates* > Assumption Checks > Q-Q plot of residuals      1. Shapiro-Wilk: T-tests > One sample T-test > Residus in *Dependant Variables* > Contrôler hypotheses > Assumption Checks > Normality (Shapiro-Wilk) | Dans notre exemple : 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. (attention, fonctionne aussi dans l’autre sens !)    *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*.*982. *p* = 0*.*354. 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. |
| **Suite Régression multiple :**  **Statistique**  Variables standardisées  Ex : Série4, ex. 3  Comparaison de modèle standardisé avec modèle non-standardisé | **Description**  **Formule**  Le modèle d’estimation des paramètres de p. 4 devient :  Avec    (  ) | **Jamovi**  Suppression de β0  Influence respective d’une variable  Permet de comparer l’influence respective d’une des variables  (Idée des ingrédients d’un gâteau au chocolat de différentiel) | **APA/Réponse**  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 | Apr`es avoir standardis´e toutes les variables, le mod`ele devient :  OUV∗ = 0.114ARI∗ + 0.312INF∗ +0.384VOC∗+ 0.220COM∗+0.000SPA∗   * + Lorsque les variables INF, VOC, COM et SPA sont maintenues constantes, une hausse d’un ´ecart-type de la variable ARI se traduit sur la variable OUV, en moyenne, par une hausse de 0.114 ´ecart-type.   + Lorsque les variables ARI, VOC, COM et SPA sont maintenues constantes, une hausse d’un ´ecart-type de la variable INF se traduit sur la variable prestige, en moyenne, par une hausse de 0.312 ´ecart-type.   L’impact de l’´VOC (0.384) sur L’OUV est trois fois plus important que ARI (0.114) et pr`es de 4 fois plus important que la SPA (0.000). (Faire cela pour tous les variables X)  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. |
| **Analyse de variances à 2 facteurs de classification :**  **Statistique**  **Attention : Nécessite une Condition d’application !!!**  **Test des effets principaux et de l’interaction** | **Description**  **Formule**  Avec ~ N(0, )  Pour faire une analyse sur 2 facteurs :   1. Conditions d’application 2. Table de l’ANOVA 3. Effets de A,B et A\*B 4. Diagramme d’interaction   Calcule pour l’ANOVA : | Deux VI nominales comme variables explicatives, appelés *Facteur A* et *Facteur B*  Ex. :  Facteur A = Forme de question  Avec a1 = dessin, a2 = géométrie  Facteur B = Examen  Avec b1 = réussite ; b2 = échec  On cherche à savoir si les 2 facteurs ont des effets sur la VD (effets principaux) et/ou en interaction (effet combiné des VI sur la VD)  Pour Effet de A :  F(a-1, ab(n-1)) = … ; p >/=/< 0.05  Pour Effet de B :  F(b-1, ab(n-1)) = … ; p >/=/< 0.05  Pour Effet d’interaction A\*B  F(((a-1)(b-1)), ab(n-1)) = … ; p >/=/< 0.05 | **Jamovi**  Analyse > ANOVA > ANOVA > VD = *dependant variables* ; VI = *fixed factor* 🡪 on obtient la table de l’ANOVA.  Table de l’ANOVA :   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source | SC | ddl | CM | F | | A | SCA | a-1 |  |  | | B | SCB | b-1 |  |  | | AB | SCAB | (a-1)(b-1) |  |  | | Residus | SCR | ab(n-1) |  |  |   **Diagramme d’interaction** :  Analyse > ANOVA > ANOVA > VD = *dependant variables*; VI = *fixed factor* > estimated marginal means > Mettre les 2 VI en Test 1 > **Output** > Meansplot | **APA/Réponse**  Ex : Série5, ex. 2  *Effet du facteur A (Age)*  1.Hypothèses : H0 : µ1·· = µ2··  H1 : µ1·· ≠ µ2··  2.Seuil : α = 0.05  3. Valeur empirique : Femp = 1.192  4. Probabilité critique : p = Prob(F (a − 1, ab(n − 1)) ≥ Femp) = Prob(F (1, 28) ≥ 1.192) = .284  5.Conclusion :  On ne rejette pas l’hypothèse H0 car p > α. Le niveau de stress perçu n’est pas influencé par l’âge. 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’·  2.Seuil : α = 0.05  3.Valeur empirique : Femp = 4.960  4.Probabilité critique : p = Prob(F (b − 1, ab(n − 1)) ≥ Femp) = Prob(F (1, 28) ≥ 4.960) = .034  5.Conclusion :  On rejette l’hypothèse H0 car p < α. Le stress perçu est influencé par le statut professionnel. Le niveau de stress des managers est plus élevé que celui des équipiers.  *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’·  2.Seuil : α = 0.05  3.Valeur empirique : Femp = 0.002  4.Probabilité critique : p = Prob(F ((a − 1)(b − 1)), ab(n − 1) ≥ Femp) = Prob(F (1, 28) ≥ 0.002) = .965  5.Conclusion :  On ne rejette pas l’hypothèse H0 car p > α. Il n’y a pas d’effet d’interaction |
| **Condition d’application** d’une Analyse de variances à deux facteurs !!   1. Il faut tout d’abord que les distributions suivent une loi normale. Ceci peut ˆêtre vérifié en appliquant un **test de Shapiro-Wilk** sur la distribution des résidus : Linear Models > General Linear Model > Y en dependent variables, X en Factors > Assumption Checks > Q-Q Plots of residuals ; Test normality of residuals. On peut aussi les tester avec le QQ-Plot en interprétant le graphique : ANOVA > Assumption Checks > Les facteurs A et B dans la case > Q-Q Plots of residuals  1. Test d’homogénéité des variances : Test de Levene : ANOVA > Assumption Checks > Homogeneity tests. On le trouve aussi sous General linear model > Assumption Checks      * ***Uniquement si ces deux hypothèses sont satisfaites on peut réaliser une analyse de variance !!*** | | | | |
| Par la suite aux tests des effets principaux et interaction il faut regarder les **p-valeur des effets**.   * Si on n’a pas d’effet d’interaction entre les facteurs significatif, mais on un effet des facteurs A et/ou B * On réalise un **Post Hoc Test** pour effectuer une comparaison multiple * S’il y a une **interaction significative** il faut faire un **test des effets simples ou un Post Hoc Test**. Le choix de ce test dépend de nos variables.   + Est-ce qu’on n’a que des variables **intergroupes** ?     - Il y a que des groupes indépendants (ex ***: Factor Sexe*** (A)= femme (a1) + homme (a2) ***¦ Factor Cigarette (B)*** = fumeur (b1) + non-fumeur (b2)) et chaque individu ne peut faire partie que d’un groupe (hommefumeur, femmefumeur ; Hommenonfumeur, femmenonfumeur) * On fait un **général linear model** (GLM) pour tester les effets simples   + Tous autres variables     - **Intragroupe** : Mesure répétée : Un sujet apparait dans plusieurs mesures (après – avant)     - **Mixtes intra et inter** : ***Inter***: Femmes + hommes et ***Intra*** : à l’âge de 20 et 40 ans * On réalise un **Post Hoc Test** pour effectuer une comparaison multiple   **APA/Réponse** | | | | |
| **Statistique**  Test des **effets simples** d’une interaction  (Uniquement avec des VI inter-groupes !!!) | **Formule**  Calcule pour l’ANOVA : | **Description**  On s’intéresse à lequel des modérateurs (=modalités) d’un facteur a réellement un effet sur l’autre facteur (ex : l’effet de a1 sur B = quel lien entre une femme et les cigarettes.) | **Jamovi**  Analyses > linear model > General Linear Model > VD = *dependant variable*, VI = *Factors* > Simple effects > *Simple effect variable* = A **ou** B, *Moderator* = soit a1 soit a2 ou b1/b2  Table de l’ANOVA :   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Source | SC | ddl | CM | F | | A(b1) | SCA(b1) | a-1 |  |  | | A(b2) | SCA(b2) | a-1 |  |  | | B(a1) | SCB(a1) | b-1 |  |  | | B(a2) | SCB(a2) | b-1 |  |  | | Residus | SCR | ab(n-1) |  |  | | 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 à 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 |
| **Statistique**  Puissance / taille d’effet  **Plans complexes**  Plan S <A>  Plan S\*A  Plan S<A\*B> | **Formule**  *η*2 =  *η*2 =  *p*    avec | **Description**  **Jamovi**  Parts de variance expliquée : Mesure la taille d’effet d’une variable sur la variance expliquée  Plan intersujet. Analyse de variance à 1 facteur de classification intersujet ayant x modalités, chaque individu ne fait partie d’un seul groupe  Plan intrasujet, Analyse de variance sur mesures répétées. Plan à mesure répétée. Chaque sujet est soumis à chaque modalité/conditions de A  Plan à deux facteurs intergroupes**, Analyse de variance à 2 facteurs intergroupes** 🡪 a\*b sous-groupes, chaque individu n’appartient qu’à un seul groupe | **APA/Réponse**  ANOVA > Effect Size > *η*2  *η*2  *p*    Table de l’ANOVA :   |  |  |  | | --- | --- | --- | | CM | ddl | F | | A | a-1 |  | | S/A | a(n -1) |  | | Total | an-1 |  |   S/A = intragroupe (tous sujets se trouvant en A mais en différents modalités a)  Table de l’ANOVA :   |  |  |  | | --- | --- | --- | | CM | ddl | F | | S | n-1 |  | | A | a-1 |  | | S\*A | (a-1)(n -1) |  | | Total | an-1 |  |   Table de l’ANOVA :   |  |  |  | | --- | --- | --- | | CM | ddl | F | | A | a-1 |  | | B | b-1 |  | | A\*B | (a-1)(b -1) |  | | S/A\*B | ab(n-1) |  | | Total | abn-1 |  | | **Décision sur la taille de l’effet :**  Effet de petite taille : 0.10  Effet de moyenne taille : 0.25  Effet de grande taille : 0.40 |
| **Statistique**  **Formule**  Plan S <A>\*B  Plan S <A\*B> | **Description** | Plan mixte, **Analyse de variances à 1 facteur intrasujet et 1 facteur intersujet**  Plan à 2 facteurs intrasujet, **Analyse de variance à 2 facteurs intrasujets** | **Jamovi**  **APA/Réponse**   |  |  |  | | --- | --- | --- | | CMintergroupe | ddl | F | | B | b-1 |  | | A\*B | (a-1)(b -1) |  | | B\*S/A | a(b-1)(n-1) |  | | Total | abn-1 |  |   Table de l’ANOVA :   |  |  |  | | --- | --- | --- | | CMintragroupe | ddl | F | | A | a-1 |  | | S/A | a(n -1) |  |   Table de l’ANOVA :   |  |  |  | | --- | --- | --- | | CM | ddl | F | | S | n-1 |  | | A | a-1 |  | | B | b-1 |  | | S\*A | (n-1)(a-1) |  | | S\*B | (n-1)(b-1) |  | | A\*B | (a-1)(b -1) |  | | S/A\*B | ab(n-1) |  | | Total | abn-1 |  | |  |
| Condition d’application d’une analyse de variance: Test de sphéricité de Mauchly  **Statistique**  Correction de Greenhouse Geiser  **Analyse de variance sur mesures répétées S\*A**  Le facteur A a-t-il changé significativement durant les mesures ?  **Formule**  **Statistique**  **Analyse de variance sur un plan mixte S <A>\*B** | avec  **Formule** | L’hypothèse de sphéricité doit être corroborée  **Description**  Si la sphéricité n’est pas corroborée (p < α) il faut faire une **correction de Greenhouse Geiser, pour tester l’effet de notre facteur intrasujet nous utilisons la loi de Fisher-Suedecon.**  Plan à x mesures  **Jamovi**  **Description**  Rappelons que a correspond au nombre de modalités du facteur inter-sujets et que b correspond au nombre de modalités du facteur intrasujet (i.e. nombre de mesures répétées). | ANOVA > Repeated Measures ANOVA> Entrer les données comme décrit en dessous > Assumption Checks > Sphericity tests  **Jamovi**    Greenhouse Geiser :    **Conditions d’application satisfaites ?** (Levene, QQ-Plot, Shapiro-Wilk, Maulchy (Greenhouse))   1. Entrer les données    1. 1 colonne par modalité en autant de colonnes qu’il y a de modalités.    2. ANOVA > Repeated Measures ANOVA > Donner un nom aux ***RM Factor*** (=Factor A intragroupe ex. Temps) et aux *Levels* (a1,a2,… ex : Avant, Après) > Tirer nos modalités du facteur intragroupe dans *Repeated Measures Cells* > observer F et p avec F[(a-1), ((a-1)(n-))] 2. Vérification des hypothèses de sphéricité (voir plus haut) 3. Si besoin faire un correction de Greenhouse 4. Diagramme d’interaction : Estimated marginal Means > Les Factors en *Term 1* > Marginal means plot  1. Comparaisons multiples pour identifier les différences franchement significatives des effets simples: Post Hoc Test > cocher les corrections demandés dans la consigne (Tukey, Holm,etc) et regarder le Ptukey, Pholm, etc.    1. Analyser les la significativité des effets simples. 2. Conclusion   **APA/Réponse**    **Conditions d’application satisfaites ?**   1. Entrer les données 2. Le factore intergroupe (=groupes indépendants) en 1 colonne, les modalités du factor intragroupe (mesure répétée) en autant de colonnes qu’il y a des modalités. 3. ANOVA > Repeated Measures ANOVA > Donner un nom aux ***RM Factor*** (=Factor A intragroupe ex. Temps) et aux *Levels* (a1,a2,… ex : Avant, Après) > Tirer nos modalités du facteur intragroupe dans *Repeated Measures Cells* > Le Factor B intergroupe en **Between Subject Factors** 4. Vérification des hypothèses de sphéricité 5. Si besoin faire un correction de Greenhouse 6. Diagramme d’interaction : Estimated marginal Means > Les Factors en *Term 1* > Marginal means plot  1. Interpreter les resultats de l’ANOVA 2. Comparaisons multiples pour identifier les différences franchement significatives effets simples:    1. **Interaction oui ?** 🡪 General Linear Model : Analyses > linear model > General Linear Model > VD = *dependant variable*, VI = *Factors* > Simple effects > *Simple effect variable* = A **ou** B, *Moderator* = soit a1 soit a2 ou b1/b2    2. **Interaction non mais effet A ou B oui ?** 🡪 Post Hoc Test > cocher les corrections demandés dans la consigne (Tukey, Holm,etc) et regarder le Ptukey, Pholm, etc.    3. Analyser les la significativité des effets simples. 3. Conclusion | Si on a 4 modalités:  **APA/Réponse**  H0 : Var(B1-B2) = Var(B1-B3) = … = Var(B3-B4)  H1 : [Var(B1-B2)≠Var(B1-B3)] ˅ …˅ [Var(B2-B3) ≠(Var(B3-B4)]  W = 0.921, p = 0.959. Comme p > α, on ne rejette pas H0 au seuil α = 5%. L’hypothèse de sphéricité est corroborée. Nous n’aurons pas besoin de corriger les résultats de l’analyse de variance.  Greenhouse :  F ~ F**(**ε(a-1), ε[(a-1)(n-1)]**)**  Ex : Série6, ex. 1  H0 :  H1 : j  k  **1b)** :    F (3, 42) = 28.301, p < 0.001. Comme p < α, on rejette donc H0 au seuil α = 5%. Le poids moyen des personnes participant à l’expérience n’est pas toujours le même, il évolue.  **4.**    Comme le facteur intrasujet possède 4 modalités, nous allons réaliser la comparaison de 6 paires de variables.  Tableau des résultats :    Interprétation  Toutes les différences sont statistiquement significatives au seuil de 5% `a l’exception d’une seule : celle qui compare le poids mesuré en fin de traitement `a celui mesuré six mois plus tard (p = 0.709). Du début à la fin de l’intervention, les sujets perdent donc régulièrement du poids : la perte de poids est significative entre le début 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 terminé, leur poids ne change plus, il reste stable (p = 0.709).  Ex : Série6, ex. 2  ***5.*** *a* = 2 *b* = 3 *n* = 12  Effet du facteur inter-sujets A (Exposition `a la fum´ee) **Attention au 2ème df !**  *F* (1*,* 22) = 5*.*307, *p* = *.*031. On rejette l’hypoth`ese *H*0 car *p < α*. Le niveau d’hyperactivit´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.    **6b)**    *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. |

**APA/Réponse**

**Jamovi**

**Statistique**

**Statistique**

**Formule**

**Description**

**APA/Réponse**