Université Mohamed Khider de Biskra. Faculté des Sciences Exactes et de la Vie

Département : Biologie **Module: Biostatistiques.**

Année Univ: 2021-2022 **Prof Chala Adel**

Master 1 LMD Biologie (Tous les spécialités)

Protocole de TP 03 :

Exécution De la Régression Linéaire Simple

Pour valider une modèle de la régression linéaire simple, il faut que les conditions suivantes soient satisfaites :

a) Deux variables quantitatives mesurées : X=Explicative (indépendante), et la variable à expliquer Y (Dépendante).
b) La loi des erreurs est Gaussienne.

Il faut comprendre bien que la **réponse** d'une question posée s'appelle une variable aléatoire.

Tout réponse qui termine par une unité des mesures, est appelé une variable quantitative mesurable.

Tout réponse qui termine par des lettres, est appelé une variable qualitative.

Alors, notre objectif dans la curent TP c'est

1) Etablir la valeur numérique du coefficient de corrélation entre X et Y (est noté par r).
2) Tester la signification du modèle.
3) Etablir l'intervalle de confiance pour le coefficient A.
4) Etablir le coefficient de détermination de la régression R.
5) Etudier la normalité d'erreurs.

I) Brief sur le Modèle de la Régression Linéaire Simple

On peut poser les questions suivantes :

Existe-il une influence de la variable explicative X sur la variation de la variable à expliquer Y ?.

Peut-on assurer que cette influence s'écrit sous forme Linéaire ?.

Voici le tableau des données statistique, les mesures à été fait sur un échantillon de k individus.

X	x_1	x_2	x_3	x_4	x_5	……	…..	x_k	
Y	y_1	y_2	y_3	y_4	y_5	……	…..	y_k	

Pour cela on veut écrire la relation entre les deux variables X et Y sous la forme suivante :

$$Y = f(X).$$

Ou bien pour chaque couple mesurable (xi,yi), avec i$\in \{1,2,3,4,5, \dots k\}$, par la relation

$$y_i = Ax_i + B + \varepsilon_i,.$$

$avec\ \varepsilon_i\ sont\ les\ erreurs\ (les\ distences\ qui\ se\ trouvent\ entre\ le\ point\ (xi, yi)\ et\ la\ droite\ linéare,$

Remarque : la loi de la variable aléatoire qui exprime les erreurs suit la loi de Gauss. (Condition « b » pour valider modèle de la régression Linéaire Simple).

Pour cela on doit ordonner les réponses par les étapes suivantes

Proposition d'hypothèses

Hypothèse nulle H_0: (Il n'existe pas une liaison entre X et Y).

Hypothèse alternative H_1: (Il existe une liaison entre X et Y).

Conclusion (la décision)

Pour la décision, on utilise souvent la règle suivante :

Si Signification inférieure à α%. Alors on rejette H0.

Si Signification supérieure à α%. Alors on accepte H0.

Exemple:

Le tableau suivant, donne les résultats d'une expérience statistique porte sur 10 étudiants, nombre d'heures de la révision dans la semaine avant l'examen, et l'augmentation de notes obtenus sur module Mathématiques.

nombre d'heures	10	11	14	15	20	25	46	50	59	70
augmentation de notes	10	10	12	12	13	13	19	15	16	20

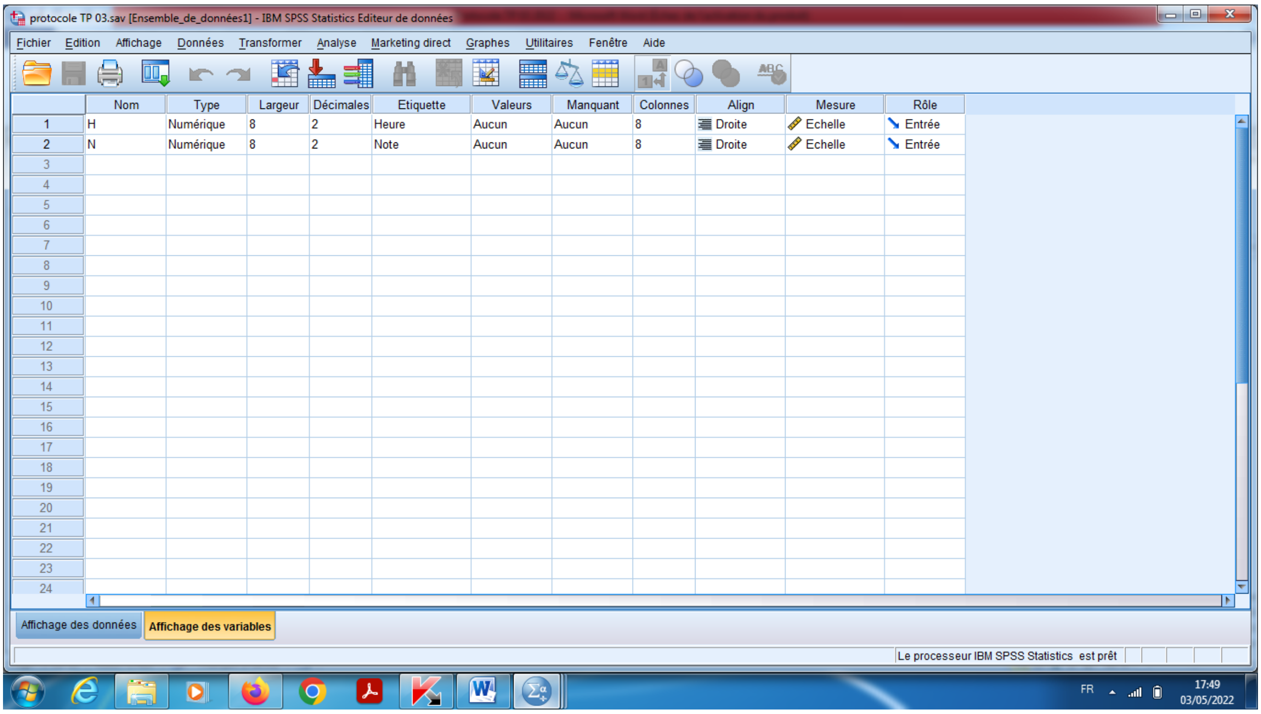
1) Etablir la valeur numérique du coefficient de corrélation entre X et Y (est noté par r).
2) Tester la signification du modèle.
3) Etablir l'intervalle de confiance pour le coefficient A.
4) Etablir le coefficient de détermination de la régression R.
5) Etudier la normalité d'erreurs.

Reponse

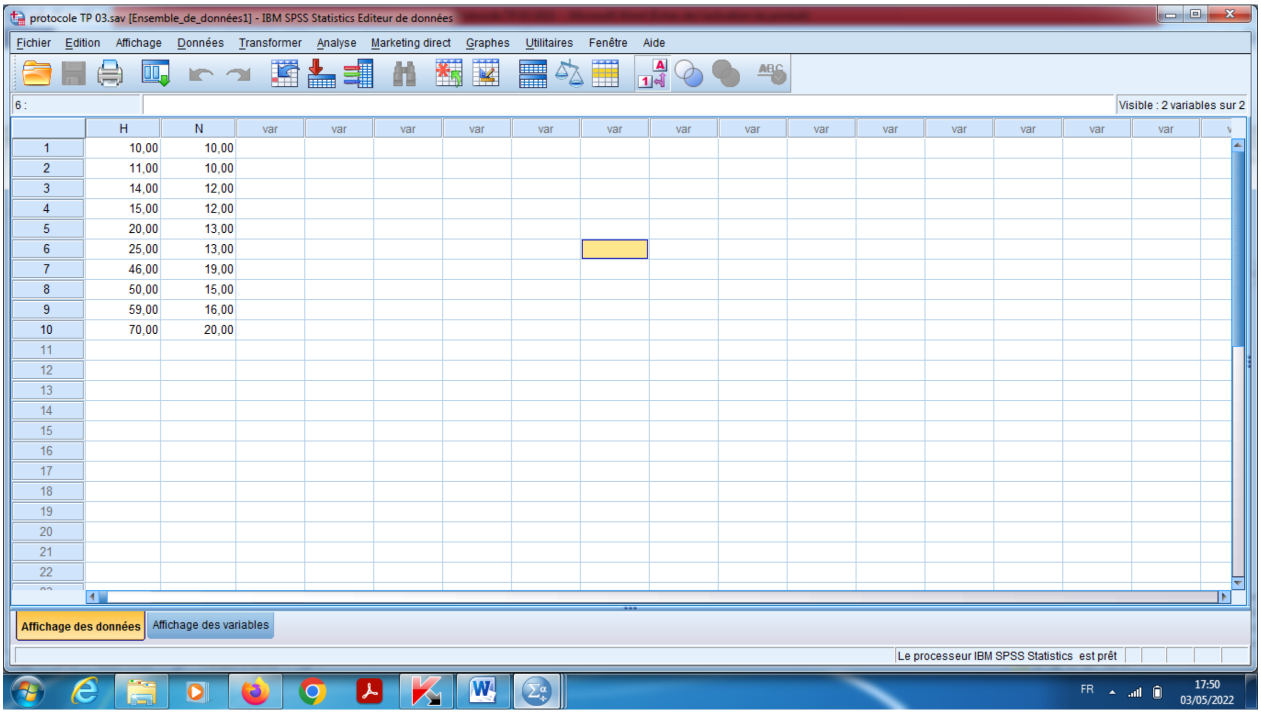
Pour cela, tout d'abord il faut entrer ces données dans Logiciel SPSS.

Nous suivons les étapes suivantes:

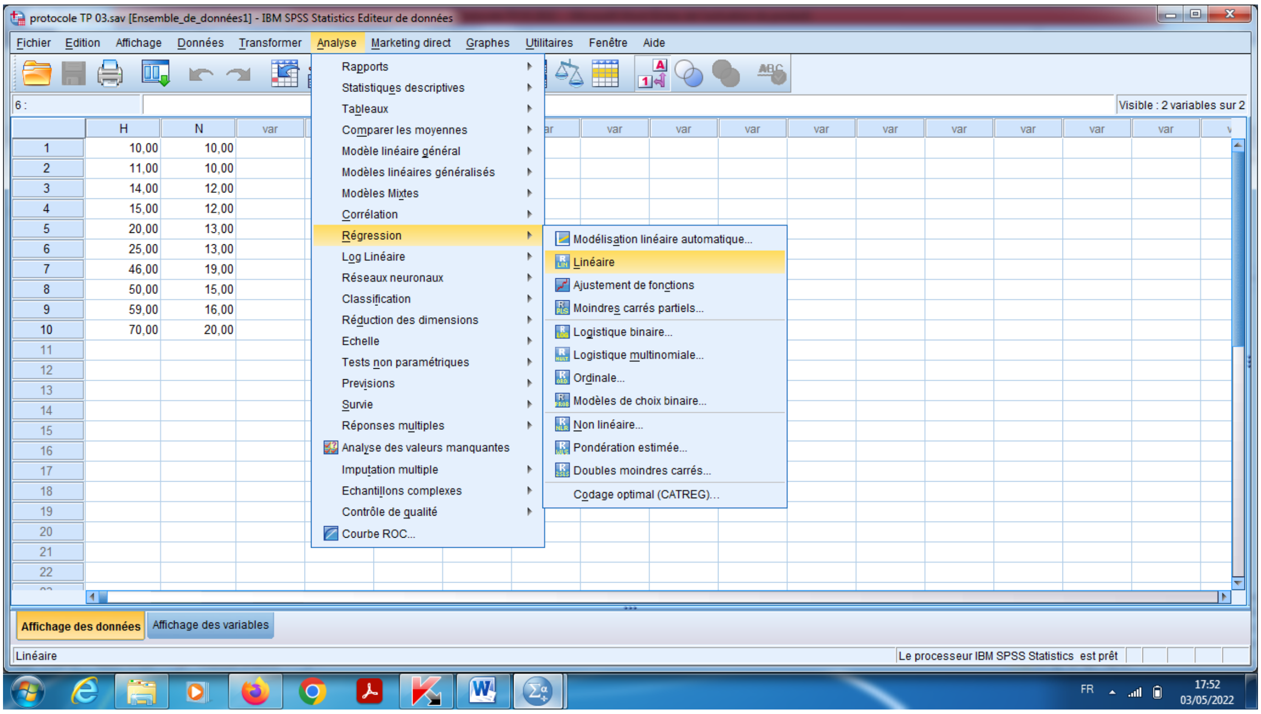
a) Il faut définir dans la barre en bas « Affichage des variables » : les variables (deux quantitatives de type numérique, et de mesure Echelle) suivantes: X=Volume Horaire, Y=Augmentation des notes.



b) On introduit les données dans la barre « Affichage des données ».



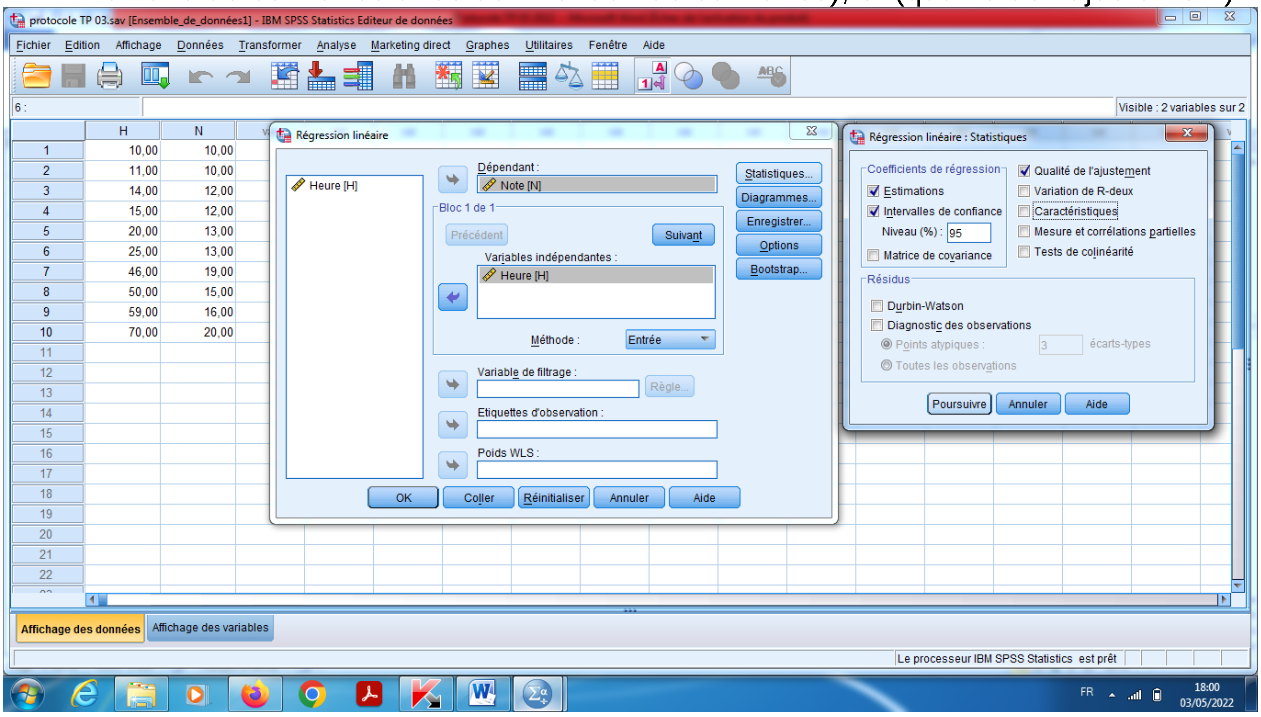
<u>Pour obtenir la régression Linéaire simple, on suivre les étapes : Analyse, Régression, puis Linéaire.</u>



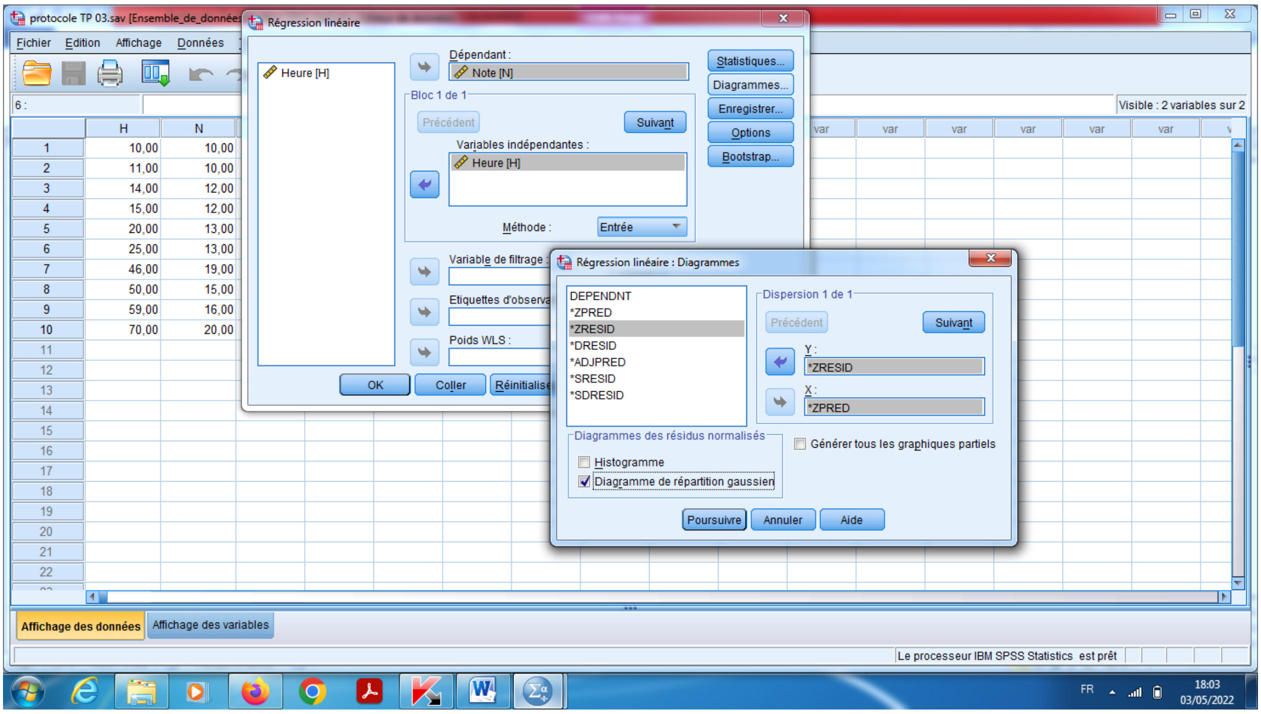
c) On pose la variable à expliqué Y =Note dans le choix « dépendante », et dans le « variables indépendantes» on pose la variable explicative X=Heure.

d) On garde la méthode utilisé (Entrée).

e) Dans le choix « Statistiques », on choisit sur « Coefficients de régression » (Estimation et intervalle de confiance avec 95% le taux de confiance), et (qualité de l'ajustement).



f) Dans le choix « Diagramme», on pose (ZPRED=X), et (ZRESID=Y), et sutrout n'oblier pas de cocher sur (Diagramme de répartition Gaussien), pour savoir est-ce que deuxième condition de validation est vrai pour assurer que le Modèle de la Régression Linéaire Simple est applicable.

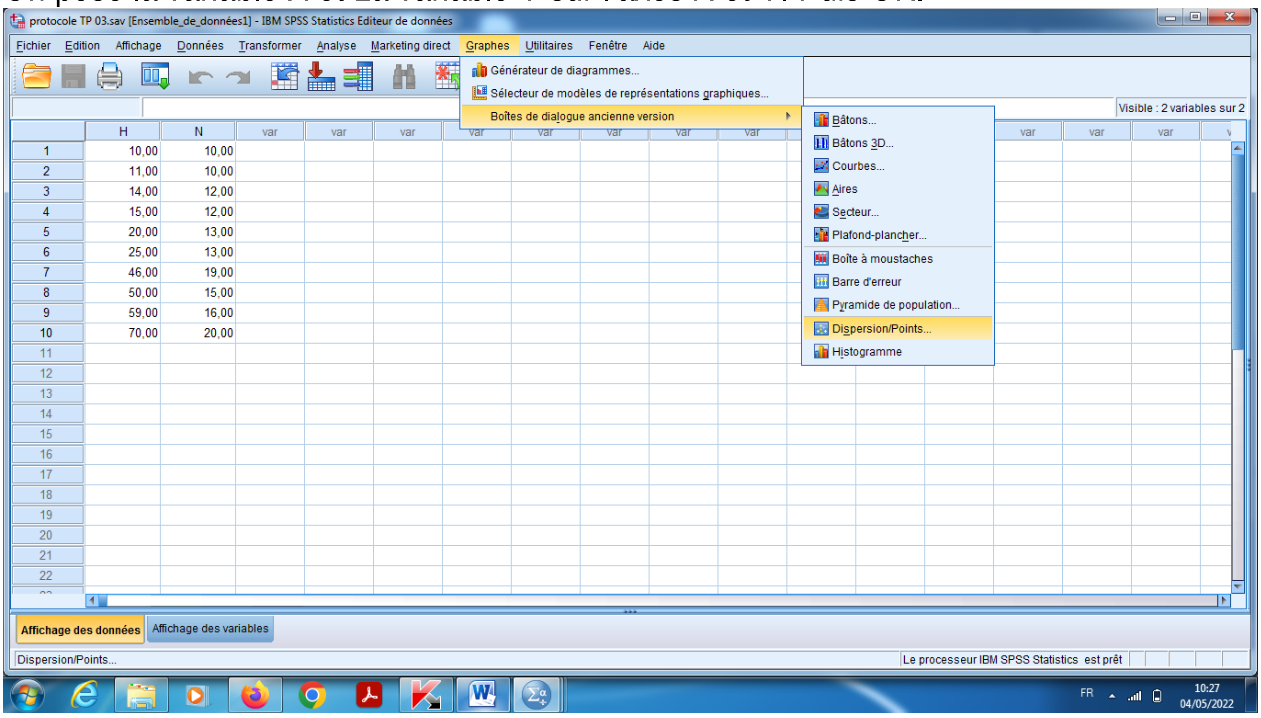


g) En fin OK.

h) Pour voir Nuage des points, en suivant les étapes :

Graphs, puis Boite de dialogue anciennes version, pui choisir Disperssion/Points, et puis Disperssion Simple.

On pose la variable X et La variable Y sur l'axes X et Y. Puis OK.



Interprétation des résultats

1) On constat ici que les résultats sont organiser selon les tableaux :

Dans le tableau 1 « Variables introduits », on remarque qu'il y 'a une seule variable exlicative X=Volume Horraires (c'est pour ca notre modèle prend son nom), et modèle « 1 ».

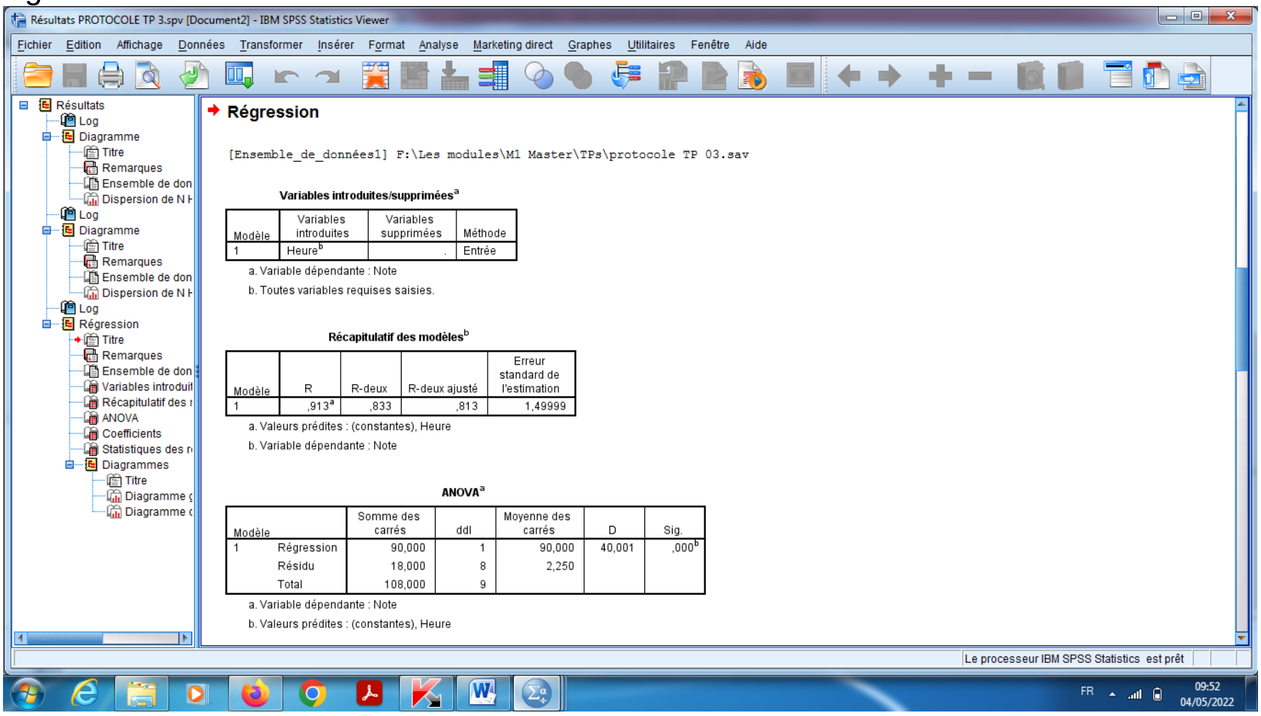
Dans le tableau 2 « Récapitulatif des modèles », on remarque que la valeur de coefficient de correlation « r : notation de cours » ici dans SPSS est noté par « R =0,913 » ,c'est-à-dire qu'il y a une très bonne corrélation entre Volume Horraire et l'augmentation des notes, de plus la valeur de coefficient de determitation de la régression c'est « R :notation de cours » ici dans SPSS est noté par « R^2=0,833 » c'est-à-dire que le taux de corrélation c'est 83,3% (parmi 100 points il y a 83 qui se trouvent dans la droite ».

D'autre part, on propose l'hypothèses suivantes :

H0 : « Notre modèle de régression n'est pas corrélé ».

H1 : « Notre modèle de régression est bien corrélé ».

Alors, on peut remarquer dans l tableau 3 de l'Analyse de la Variance que la valeur « Sig=0,000<0,05 », donc on acceptte H1, c'est-à-dire qu'il ya une liaison siginicative entre X=Heure et Y=Notes, ou bien notre modèle est globalement significative, avec une taux de signification c'est 100%.



Régression

[Ensemble_de_données1] F:\Les modules\M1 Master\TPs\protocole TP 03.sav

Variables introduites/supprimées[a]

Modèle	Variables introduites	Variables supprimées	Méthode
1	Heure[b]	.	Entrée

a. Variable dépendante : Note

b. Toutes variables requises saisies.

Récapitulatif des modèles[b]

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,913[a]	,833	,813	1,49999

a. Valeurs prédites : (constantes), Heure

b. Variable dépendante : Note

ANOVA[a]

Modèle		Somme des carrés	ddl	Moyenne des carrés	D	Sig.
1	Régression	90,000	1	90,000	40,001	,000[b]
	Résidu	18,000	8	2,250		
	Total	108,000	9			

a. Variable dépendante : Note

b. Valeurs prédites : (constantes), Heure

2) Dans le tableau 4, on peut voir les valeurs à éstimer pour la droite de la régression linéaire, avec « A=0,143 et B=9,436 » (dans le cours sont notés respectivements par α,β), c'est-à-dire que sur l'augmentation de 10 heurs de la révision est donne une augmentation de 0,143 sur les notes.

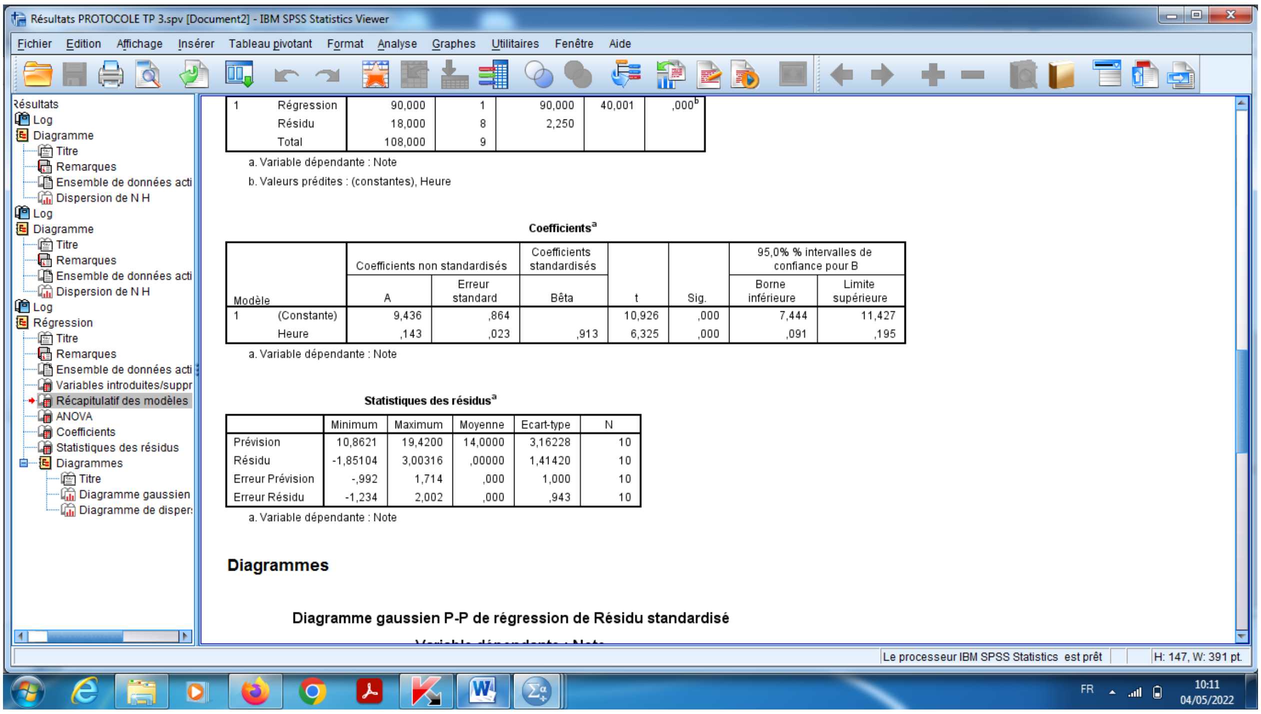
Alors la droite de régression s'écrit sous la forme :

$$y = 0{,}143x + 9{,}436.$$

De plus pour l'intervalle de confiance pour le coefficient

$\alpha \in [0{,}091; 0{,}195], et\ \beta \in [7{,}444; 11{,}427]$. « Voir les formules exlicites dans le cours ».

Autrement dit que le pente de la droite peut être varié de 0,091 jusqu'à 0,195.



Modèle						
1	Régression	90,000	1	90,000	40,001	,000[b]
	Résidu	18,000	8	2,250		
	Total	108,000	9			

a. Variable dépendante : Note
b. Valeurs prédites : (constantes), Heure

Coefficients[a]

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Sig.	95,0% % intervalles de confiance pour B	
		A	Erreur standard	Bêta			Borne inférieure	Limite supérieure
1	(Constante)	9,436	,864		10,926	,000	7,444	11,427
	Heure	,143	,023	,913	6,325	,000	,091	,195

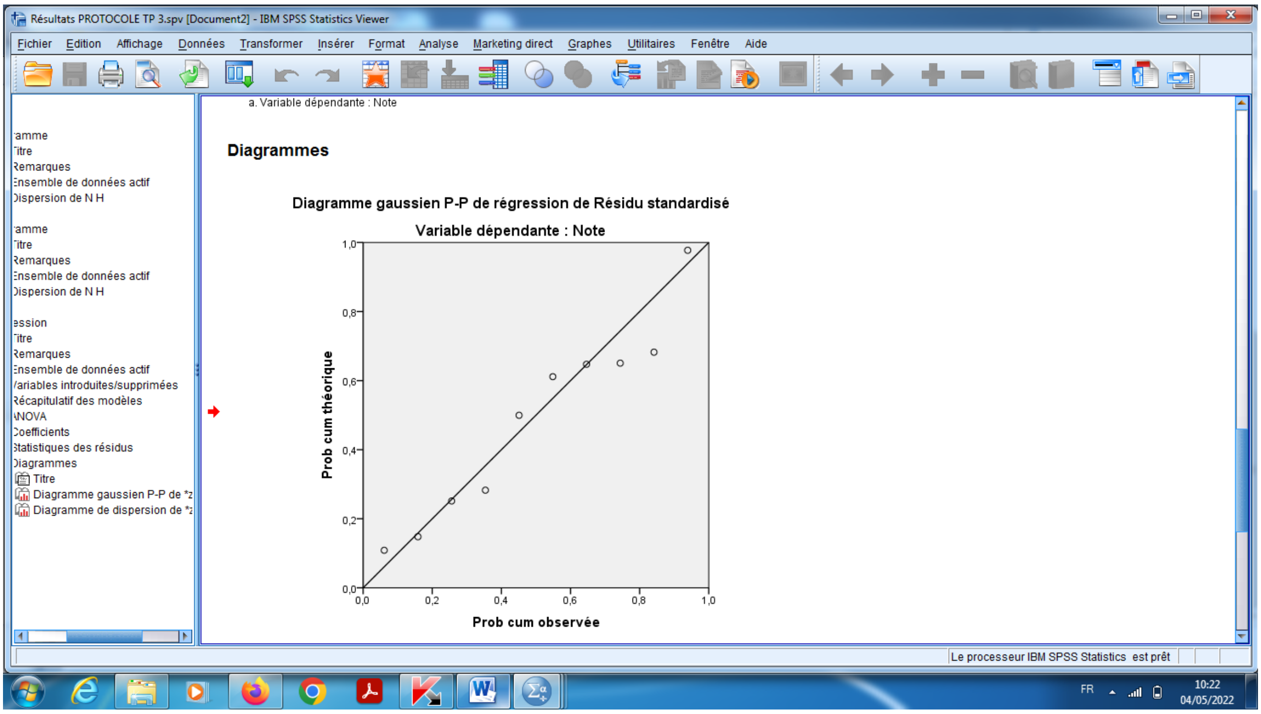
a. Variable dépendante : Note

Statistiques des résidus[a]

	Minimum	Maximum	Moyenne	Ecart-type	N
Prévision	10,8621	19,4200	14,0000	3,16228	10
Résidu	-1,85104	3,00316	,00000	1,41420	10
Erreur Prévision	-,992	1,714	,000	1,000	10
Erreur Résidu	-1,234	2,002	,000	,943	10

a. Variable dépendante : Note

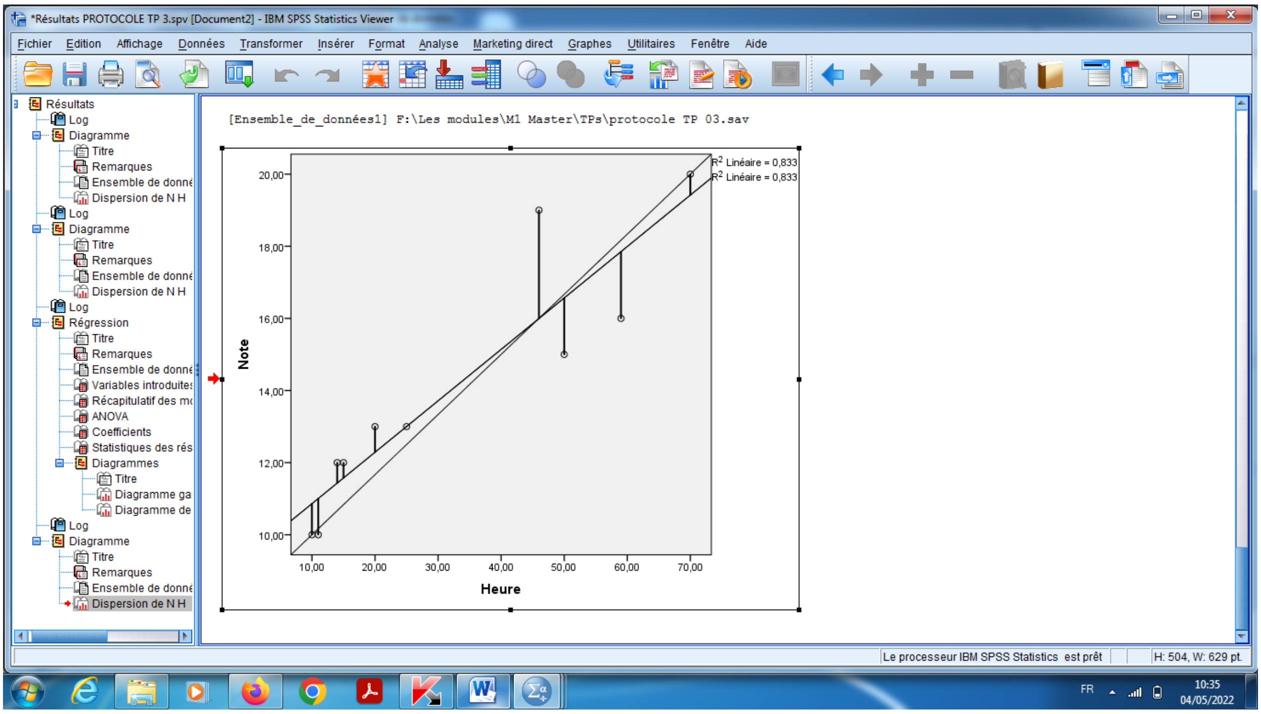
3) Dans le diagramme ci-dessous, on peut établir que condition de validation de la régression linéaire simple est valide, c'est-à-dire que les erreurs sont gaussiennes.



4) **<u>Pour Nuage des point et caractéristiques.</u>**

Double clic sur le graph, et on clique eur « Element » qui se trouve dans la barre de Menu, et puis choisir « Ajouter une courbe d'ajustement ».

R^2=0,833.


*Résultats PROTOCOLE TP 3.spv [Document2] - IBM SPSS Statistics Viewer
Fichier Edition Affichage Données Transformer Insérer Format Analyse Marketing direct Graphes Utilitaires Fenêtre Aide
[Ensemble_de_données1] F:\Les modules\M1 Master\TPs\protocole TP 03.sav
R² Linéaire = 0,833
R² Linéaire = 0,833
Note
Heure
Le processeur IBM SPSS Statistics est prêt
H: 504, W: 629 pt.