

## **Introduction Générale**

L'objectif de cette analyse est d'examiner si le salaire est en relation avec les variables suivantes : âge, nombre d'années d'expérience, et note de l'entretien d'embauche. Nous utiliserons une régression linéaire multiple pour explorer ces relations.

### **Objectifs de l'Étude**

- ▶ Évaluer l'effet de l'âge sur le salaire : Analyser si les variations d'âge parmi les employés influencent leur niveau de rémunération.
- ▶ Examiner l'impact du nombre d'années d'expérience sur le salaire : Déterminer dans quelle mesure l'expérience professionnelle se traduit par une augmentation salariale.
- ▶ Étudier la relation entre la note de l'entretien d'embauche et le salaire : Comprendre comment les évaluations des performances lors des entretiens d'embauche affectent le salaire des employés.

### **Hypothèses de l'Étude**

Pour guider cette recherche, nous formulons les deux hypothèses principales suivantes :

Hypothèse 1 : Les caractéristiques individuelles (âge et nombre d'années d'expérience) ont un effet significatif sur le salaire.

Hypothèse 2 : La performance à l'entretien d'embauche, mesurée par la note, a un effet significatif sur le salaire.

Ces deux hypothèses permettront de structurer notre analyse et d'évaluer de manière exhaustive l'impact des variables d'intérêt sur le salaire.

### **Méthodologie**

Pour déterminer si le salaire a une relation avec l'âge, le nombre d'années d'expérience et la note de l'entretien d'embauche, nous adopterons une approche en trois étapes : une analyse univariée, une analyse bivariée et une analyse multivariée.

L'analyse univariée consistera à examiner la distribution de chaque variable indépendamment à travers des statistiques descriptives et des visualisations.

Ensuite, l'analyse bivariée évaluera les relations entre le salaire et chaque variable indépendante à l'aide de coefficients de corrélation de Pearson et de diagrammes de dispersion pour identifier les relations linéaires potentielles.

Enfin, l'analyse multivariée utilisera une régression multiple pour estimer les coefficients de régression, évaluer la significativité statistique de chaque variable et interpréter leur impact combiné sur le salaire. Chaque étape sera suivie d'une interprétation détaillée des résultats, et une conclusion générale résumera les découvertes principales, discutant des implications pratiques et des recommandations pour des recherches futures.

## 1 : Description des Données :

La base de données comprend les variables suivantes :

Âge : Âge des employés.

Nbre d'années d'expérience : Nombre d'années d'expérience des employés.

Note de l'entretien d'embauche : Note obtenue lors de l'entretien d'embauche.

Salaire : Salaire des employés.

## 2 : Importation à partir d'Excel :

Les données provenant d'Excel comptent parmi celles le plus souvent importées. Pour en permettre l'importation future, les données doivent être saisies sous forme de tableau en colonnes, où chaque colonne correspond à une variable. Il y a quelques règles de base à respecter pour importer les données à partir d'Excel. Les noms des variables doivent être sur la première ligne (ligne 1) et ils ne doivent pas être trop longs ni comporter de caractères spéciaux. Il est préférable de ne pas utiliser de caractères accentués dans les noms des variables. Les données doivent débiter à la deuxième ligne du chiffrier (ligne 2). Les données suivantes ont été saisies à partir d'un fichier Excel.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	age	Nbre d'ex	note Exam de recrutement	salaire												
2	41	12	86,5	14128												
3	35	9	95,71	15541												
4	25	8	102,79	16058												
5	32	12	108,92	16074												
6	30	8	116,38	16524												
7	23	9	120,81	16757												
8	44	14	1141	16923												
9	40	9	42,85	17047												
10	33	12	104,58	17349												
11	36	10	103,98	17450												
12	36	7	108,14	17709												
13	42	11	132,21	17725												
14	29	10	67,55	18149												
15	35	9	81,58	18250												
16	48	10	86,56	18312												
17	22	6	104,57	18381												
18	47	7	102,58	18621												
19	53	8	115,14	18862												
20	58	4	106,32	19242												
21	41	13	124,94	19330												
22	54	9	944	19484												

## 3 : Analyse Univariée : (Etude descriptives des variables)

### 3.1 : Descriptives salaire

Statistiques descriptives							
	N	Plage	Minimum	Maximum	Moyenne	Ecart type	Variance
salaire	18624	39849	14128	53977	34961,38	4998,597	24985971,374
N valide (liste)	18624						

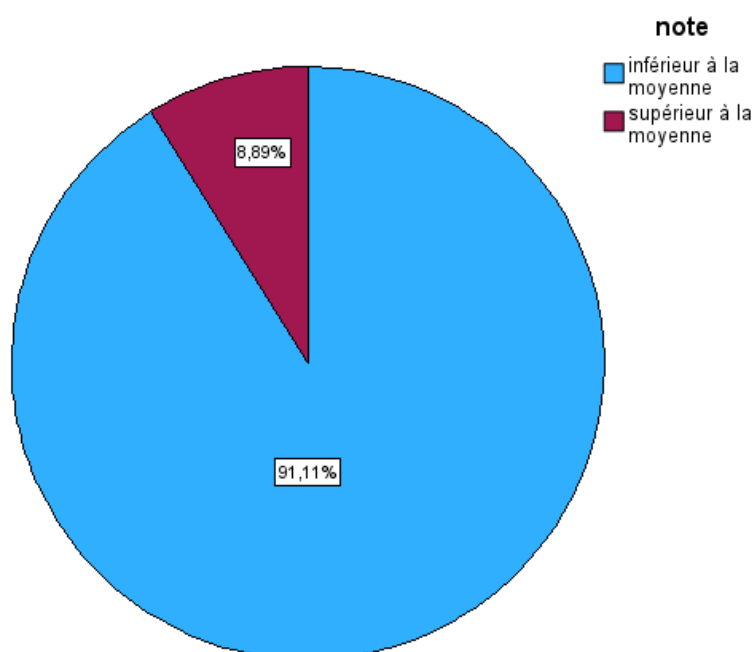
À partir de ces données, une hypothèse que l'on peut formuler est que la distribution des salaires dans cet ensemble de 18624 individus est assez étendue, avec une plage de 39849 entre le salaire minimum et le salaire maximum. De plus, l'écart type de 4998,597 indique une dispersion relativement importante des salaires par rapport à la moyenne de 34961,38. Cela suggère qu'il existe une variabilité significative dans les salaires des individus de cet ensemble de données.

### 3.2 : Descriptives note

Statistiques descriptives						
	N	Minimum	Maximum	Moyenne	Ecart type	Variance
note	18606	8,6800	1224,0000	135,222415	198,7305866	39493,846
N valide (liste)	18606					

Ces résultats indiquent que les salariés ont obtenu des notes variées lors de l'épreuve d'embauche. Certains ont peut-être obtenu des notes élevées, tandis que d'autres ont obtenu des notes plus basses. En outre, D'après le graphique ci-dessous il est établi que 8.89% des individus ont obtenu une note supérieure à la moyenne, tandis que le reste des individus ont obtenu une note inférieure à la moyenne.

#### 3.2.1 :Graphique note :



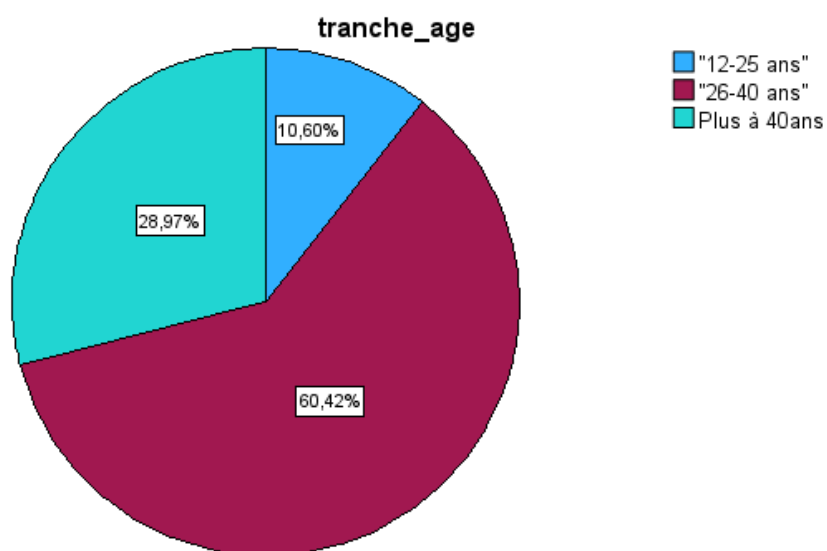
### 3.3 : Descriptives Age :

Statistiques descriptives					
	N	Minimum	Maximum	Moyenne	Ecart type
age	18710	12	65	36,15	8,507
N valide (liste)	18710				

Le tableau ci-dessus donnent une vue d'ensemble sur l'âge de la population étudiée. Ils fournissent des statistiques descriptives telles que le nombre total de personnes (N), l'âge minimum, l'âge maximum, l'âge moyen et l'écart type. En somme, cette population étudiée présente une diversité d'âges, avec une moyenne autour de 36 ans. L'écart type indique que les âges varient relativement peu autour de cette moyenne.

### 3.3.1 :Fréquences âge :

tranche_age					
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	"12-25 ans"	1984	10,0	10,6	10,6
	"26-40 ans"	11305	56,8	60,4	71,0
	Plus à 40ans	5421	27,2	29,0	100,0
	Total	18710	94,0	100,0	
Manquant	Système	1193	6,0		
Total		19903	100,0		

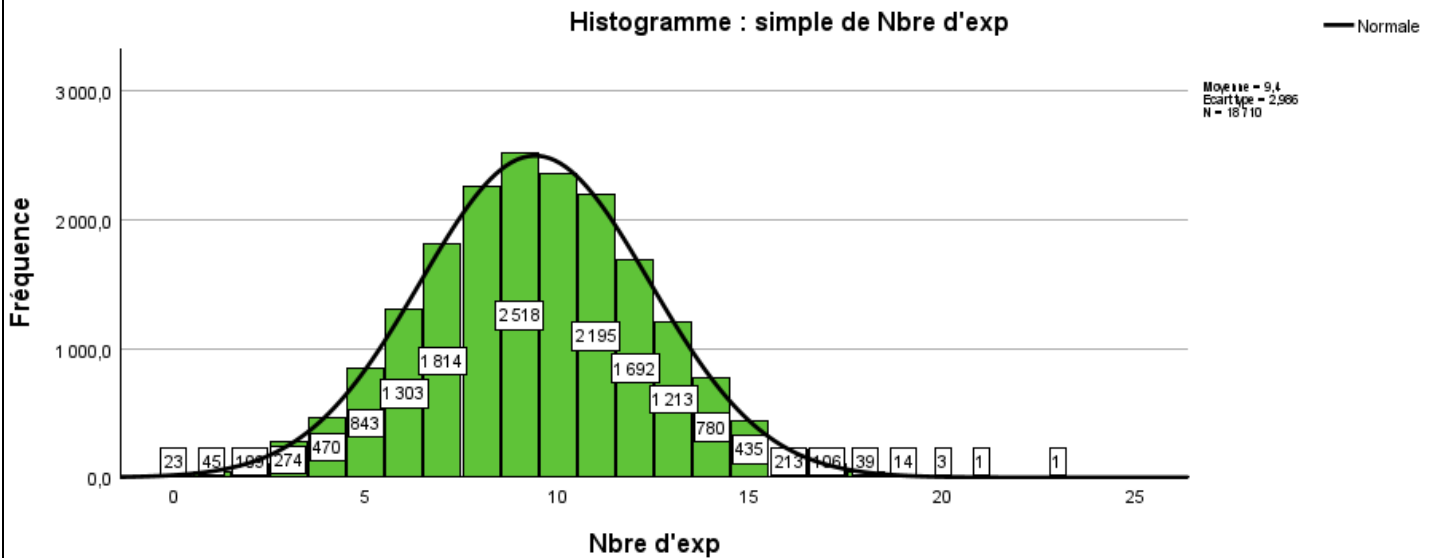


La répartition en trois tranches d'âge montre que la majorité des individus (60,42%) ont un âge qui varie entre 26-40ans, suivis par ceux de la tranche 40 ans et plus (28,97%), et enfin les plus jeunes de 12 à 25 ans (10,60%).

### 3.4 : Descriptives Année d'expérience :

Statistiques descriptives						
	N	Plage	Minimum	Maximum	Moyenne	Ecart type
Nbre d'exp	18710	23	0	23	9,40	2,986
N valide (liste)	18710					

### 3.4.1 : Graphique expérience :



La moyenne d'expérience dans la population étudiée est de 9,40, ce qui peut être interprété comme la valeur centrale autour de laquelle les données sont distribuées. L'écart type de 2,986 nous indique la dispersion des données d'expérience par rapport à la moyenne. Une valeur d'écart type plus élevée suggère une plus grande variabilité dans les niveaux d'expérience au sein de la population.

Ces résultats nous permettent de comprendre que la population étudiée présente une diversité en termes d'expérience, avec une moyenne de 9,40 et une dispersion de 2,986.

### 4 : Analyse bivariée :

L'analyse bivariée consiste à vérifier l'existence (et l'intensité) d'une possible relation entre deux variables. Une relation bivariée implique l'effet d'une variable sur une autre et non l'inverse : on parle alors de relation d'indépendance. La variable subissant l'influence est appelée « variable dépendante ». La variable qui produit l'effet est appelée « variable indépendante », mais on peut étudier l'absence de causalité malgré l'existence d'une relation<sup>1</sup>.

Dans ce contexte, nous allons examiner les relations entre les variables suivantes : Âge et Salaire, Nombre d'années d'expérience et Salaire, et Note de l'entretien d'embauche et Salaire. Cette analyse peut inclure des visualisations et des mesures statistiques pour comprendre la force et la direction des relations.

**Quel test choisir ?** Le test à choisir varie selon la nature des variables. Le tableau suivant résume l'application des tests dans l'analyse bivariée : Pour notre cas nous avons quatre mesures avec échelles métriques.

<sup>1</sup> Zheng, Plaisent, et Zuccaro, *L'analyse des données de sondage avec SPSS*.

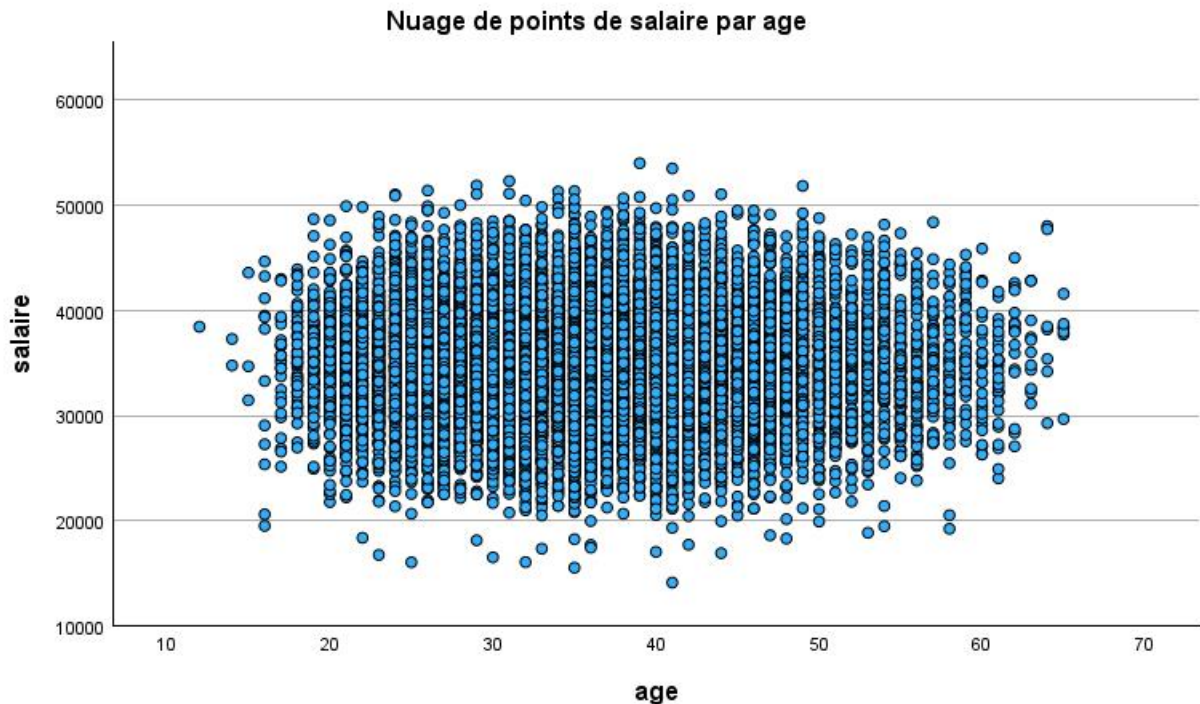
Type de mesure	Deux mesures avec échelle non métrique	Une mesure avec échelle non métrique et une mesure avec échelle métrique	Deux mesures avec échelles métriques
Type d'analyse	Tableau croisé	Comparaison de moyennes	Corrélation ou régression
Tests statistiques	Chi-deux	T de <i>student</i> (deux moyennes) F (ANOVA) (deux moyennes ou plus)	I (corrélation) ou F (régression)
Force de la relation	V de Cramer	n	r ou R
Commande SPSS	Analyse ( <i>analyze</i> )	Analyse ( <i>analyze</i> )	Corrélation : Analyse ( <i>analyze</i> )
	Statistiques descriptives ( <i>descriptive statistics</i> )	Comparer les moyennes ( <i>compare means</i> )	Corrélation : Bivariée ( <i>bivariate</i> ) OU Régression : Analyse ( <i>analyze</i> )
	Tableaux croisés ( <i>crosslabs</i> )	Test T pour échantillons indépendants ( <i>independent sample T-test</i> ) OU ANOVA à un facteur ( <i>one way ANOVA</i> )	Régression : Linéaire ( <i>linear</i> )

**Source :** Zheng, Plaisent, et Zuccaro, *L'analyse des données de sondage avec SPSS*.

#### 4.1 : Âge et Salaire :

##### 4.1.1 : Nuage de points :

Un nuage de points permet de visualiser si une tendance linéaire existe. Par exemple, une pente ascendante suggère que le salaire augmente avec l'âge, tandis qu'une pente descendante suggère le contraire. Une dispersion aléatoire sans tendance claire indique qu'il n'y a pas de relation linéaire évidente.



D'après l'analyse du nuage de points entre l'âge et le salaire de la population, nous observons une dispersion aléatoire des points sur le graphique. Cela suggère qu'il n'y a pas de relation linéaire claire entre l'âge et le salaire. Les points sont répartis de manière irrégulière, indiquant une diversité de salaires pour différentes tranches d'âge.

#### 4.1.2 : Coefficient de corrélation :

Un coefficient de corrélation proche de +1 indique une forte relation positive.

Un coefficient proche de -1 indique une forte relation négative.

Un coefficient proche de 0 indique une absence de relation linéaire.

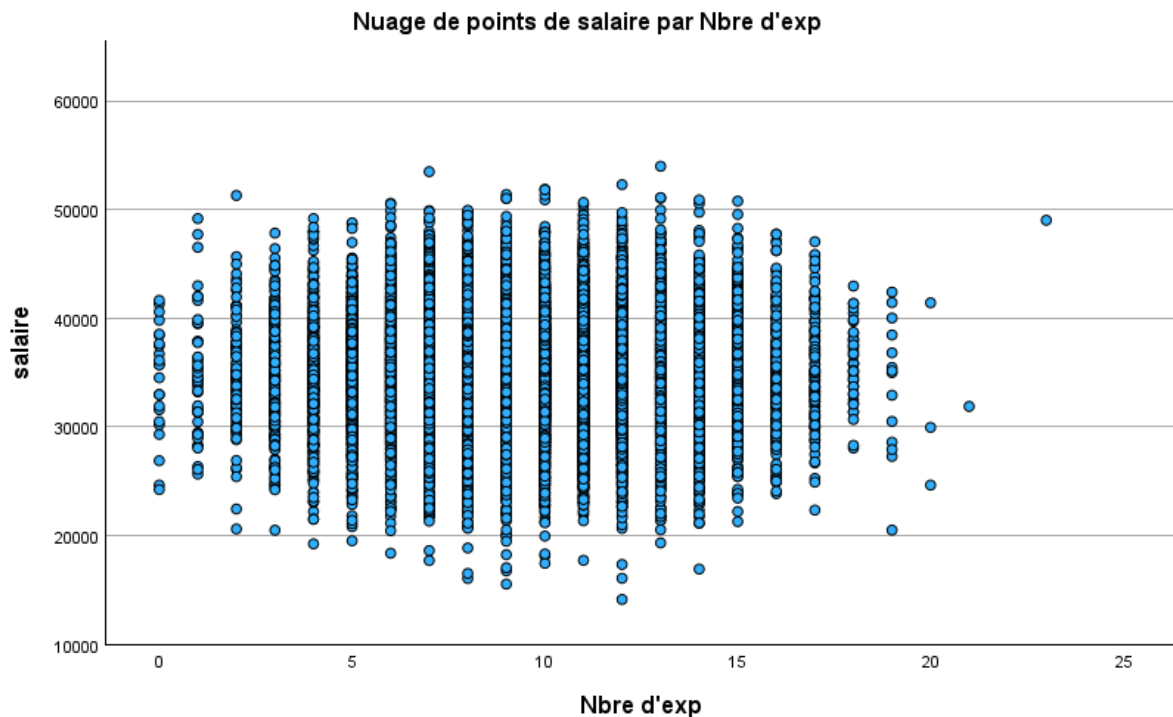
Corrélations			
		salaire	age
salaire	Corrélation de Pearson	1	,007
	Sig. (bilatérale)		,343
	N	18624	18624
age	Corrélation de Pearson	,007	1
	Sig. (bilatérale)	,343	
	N	18624	18710

La corrélation de Pearson entre l'âge et le salaire de la population étudiée est de 0,007. Cette corrélation très faible indique une relation linéaire pratiquement inexistante entre l'âge et le salaire. La moyenne d'âge est de 9,40 avec un écart type de 2,986. L'analyse du nuage de points entre l'âge et le salaire confirme cette absence de relation claire, avec une dispersion aléatoire des points sur le graphique.

## 4.2 : Nombre d'années d'expérience et Salaire :

### 4.2.1 :Nuage de points :

En observant le graph, il n'y a pas de tendance claire ou de motif évident qui se dégage. Cela suggère qu'il n'y a pas de relation linéaire forte entre le salaire et le nombre d'années d'expérience. Au lieu de cela, les points sont dispersés de manière aléatoire, indiquant une diversité de salaires pour différents niveaux d'expérience. Il est possible qu'il y ait d'autres facteurs qui influencent le salaire en dehors de l'expérience professionnelle uniquement.



### 4.2.2 :Coefficient de corrélation :

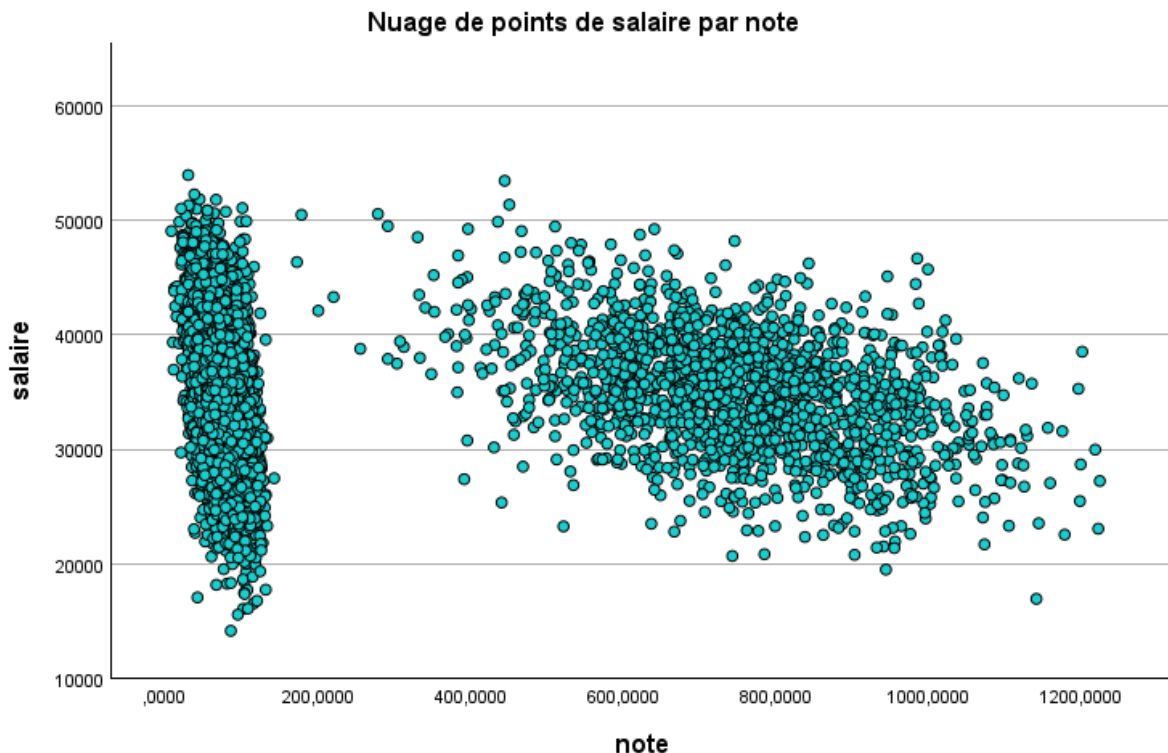
Corrélations			
		salaire	Nbre d'exp
salaire	Corrélation de Pearson	1	,008
	Sig. (bilatérale)		,292
	N	18624	18624
Nbre d'exp	Corrélation de Pearson	,008	1
	Sig. (bilatérale)	,292	
	N	18624	18710

La corrélation de Pearson entre le salaire et le nombre d'années d'expérience est de 0.008, ce qui indique une très faible corrélation linéaire entre les deux variables.



#### 4.3 : Note de l'entretien d'embauche et Salaire :

##### 4.3.1 :Nuage de points :



D'après l'analyse du nuage de points entre la note de l'entretien d'embauche et le salaire, nous pouvons observer une dispersion aléatoire des points sur le graphique. Cela suggère qu'il n'y a pas de relation linéaire claire entre la note de l'entretien d'embauche et le salaire. Les points sont répartis de manière irrégulière, indiquant une diversité de salaires pour différentes notes d'entretien. Il n'y a pas de tendance ou de motif évident qui se dégage de ce nuage de points.

##### 4.3.2 :Coefficient de corrélation :

Corrélations			
		salaire	note
salaire	Corrélation de Pearson	1	-,066**
	Sig. (bilatérale)		<,001
	N	18624	18520
note	Corrélation de Pearson	-,066**	1
	Sig. (bilatérale)	<,001	
	N	18520	18606
**. La corrélation est significative au niveau 0.01 (bilatéral).			

la corrélation entre la note de l'entretien d'embauche et le salaire est de -0,66, ce qui indique une corrélation linéaire modérée et inverse entre ces deux variables. Cela suggère qu'une note d'entretien d'embauche plus élevée est généralement associée à un salaire plus bas, et vice versa. La valeur de Sig inférieure à 0,001 indique une corrélation statistiquement significative, ce qui renforce la validité de l'association.

#### 4.4 : Synthèse :

L'analyse bivariée des variables âge et salaire, nombre d'années d'expérience et salaire, ainsi que la note de l'entretien d'embauche et le salaire a permis de mettre en évidence certaines relations entre ces variables.

Cependant, il est important de noter que les corrélations observées sont relativement faibles, ce qui indique une absence de relations linéaires claires entre ces variables. Les nuages de points ont montré une dispersion aléatoire des données, suggérant une diversité de salaires pour différentes tranches d'âge, niveaux d'expérience et notes d'entretien.

En d'autres termes, Suite aux résultats de l'analyse bivariée, nous constatons que les relations entre chaque variable indépendante (Âge, Nombre d'années d'expérience, Note de l'entretien d'embauche) et le salaire sont généralement faibles ou inexistantes, à l'exception de la note de l'entretien d'embauche qui montre une relation inverse modérée avec le salaire. La corrélation de Pearson est de -0,66, ce qui indique une corrélation linéaire modérée et inverse entre la note de l'entretien d'embauche et le salaire. Une note d'entretien plus élevée est généralement associée à un salaire plus bas, et cette relation est statistiquement significative ( $p < 0,001$ ).

#### 5 : Analyse multivariée :

Afin de mieux comprendre les déterminants du salaire, il est nécessaire de procéder à une analyse multivariée en utilisant la régression multiple.

La régression multiple nous fournira une équation linéaire qui modélise le salaire en fonction des variables explicatives, permettant ainsi de prédire le salaire et de comprendre les relations complexes entre les différentes variables.

Nous chercherons à comprendre l'importance relative de chaque variable et à établir un modèle prédictif pour estimer le salaire en fonction de ces facteurs.

##### 5.1 : Récapitulatif des modèles :

Récapitulatif des modèles				
Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,066 <sup>a</sup>	,004	,004	4988,815
a. Prédicteurs : (Constante), note , age, Nbre d'exp				

##### 5.1.1 :R :

La valeur de R (**0,066**) représente le coefficient de corrélation multiple entre les prédicteurs et la variable dépendante. Cette valeur proche de 0 indique une très faible corrélation entre les variables indépendantes et le salaire. En d'autres termes, les prédicteurs inclus dans le modèle expliquent très peu de la variation du salaire.

##### 5.1.2 :R-deux (R<sup>2</sup>) :

R<sup>2</sup> (**0,004**) représente la proportion de la variance dans la variable dépendante (salaire) qui est expliquée par les prédicteurs (âge, nombre d'années d'expérience, note de l'entretien d'embauche). Une valeur de 0,004 indique que seulement **0,4%** de la variance du salaire est expliquée par ce modèle de régression multiple, ce qui est extrêmement faible.

### 5.1.3 :R-deux ajusté :

Le R-deux ajusté (**0,004**) est une version ajustée de  $R^2$  qui tient compte du nombre de prédictors dans le modèle. Comme  $R^2$ , cette valeur est également très faible, confirmant que le modèle n'explique pas bien la variance du salaire.

### 5.2 : Erreur standard de l'estimation :

L'erreur standard de l'estimation (**4988,815**) mesure la précision des prédictions du modèle. Une valeur élevée de l'erreur standard suggère que les prédictions du modèle sont largement dispersées autour des valeurs réelles du salaire, indiquant une précision faible du modèle.

### 5.3 : Synthèse :

Ces résultats montrent que les variables **Âge, Nombre d'années d'expérience et Note de l'entretien d'embauche**, lorsqu'elles sont prises ensemble, n'expliquent que très faiblement la variation du salaire. Le modèle global a une capacité prédictive très limitée, comme en témoignent les faibles valeurs de R,  $R^2$  et  $R^2$  ajusté. Cela suggère qu'il y a probablement d'autres facteurs non inclus dans ce modèle qui influencent le salaire de manière significative.

### 5.4 : Coefficients du modèle :

Coefficients <sup>a</sup>								
		Coefficients non standardisés		Coefficients standardisés			Intervalle de confiance à 95,0% pour B	
Modèle		B	Erreur standard	Bêta	t	Sig.	Borne inférieure	Borne supérieure
1	(Constante)	34937,022	194,180		179,921	<,001	34556,411	35317,633
	age	3,281	4,313	,006	,761	,447	-5,172	11,734
	Nbre d'exp	13,373	12,305	,008	1,087	,277	-10,746	37,491
	note	-1,648	,185	-,065	-8,930	<,001	-2,009	-1,286

a. Variable dépendante : salaire

Le tableau ci-dessus présente les coefficients de régression pour le modèle qui prédit le salaire en fonction de trois variables indépendantes : **l'âge, le nombre d'années d'expérience et la note de l'entretien d'embauche**.

#### 5.4.1 :Constante :

**B = 34937,022** : Cela représente le salaire de base lorsque toutes les variables indépendantes (âge, nombre d'années d'expérience, et note de l'entretien d'embauche) sont égales à zéro. Cela suggère un salaire de base de 34 937,022 unités monétaires.

#### 5.4.2 :Âge :

**B = 3,281** : Chaque année supplémentaire d'âge est associée à une augmentation du salaire de 3,281 unités monétaires, en tenant compte des autres variables dans le modèle. Toutefois, cette relation n'est pas statistiquement significative ( $p = 0,447 > 0,05$ ).

#### 5.4.3 :Nombre d'années d'expérience :

**B = 13,373** : Chaque année supplémentaire d'expérience est associée à une augmentation du salaire de 13,373 unités monétaires, en tenant compte des autres variables dans le modèle. Cette relation n'est pas non plus statistiquement significative ( $p = 0,277 > 0,05$ ).

#### **5.4.4 :Note de l'entretien d'embauche :**

**B = -1,648** : Chaque point de plus dans la note de l'entretien d'embauche est associé à une diminution du salaire de 1,648 unités monétaires, en tenant compte des autres variables dans le modèle. Cette relation est statistiquement significative ( $p < 0,001$ ).

#### **5.4.5 :Équation de la Droite de Régression :**

L'équation de régression permet de prédire le salaire en fonction de l'âge, du nombre d'années d'expérience et de la note de l'entretien d'embauche. Cependant, il est important de noter que, dans ce modèle, seules les relations avec la note de l'entretien d'embauche sont significatives, tandis que les relations avec l'âge et le nombre d'années d'expérience ne le sont pas. Par conséquent, les prédictions basées sur cette équation doivent être interprétées avec prudence, et il peut être nécessaire d'inclure d'autres variables pour améliorer l'exactitude des prédictions.

$$\text{Salaire} = 34937,022 + (3,281 * \text{Age}) + (13,373 * \text{Nbre d'exp}) + (-1,648 * \text{Note})$$

## **Conclusion Générale**

L'analyse de régression multiple visait à déterminer l'influence de trois variables indépendantes (l'âge, le nombre d'années d'expérience et la note de l'entretien d'embauche) sur le salaire.

Les résultats montrent que, dans ce modèle, l'âge et le nombre d'années d'expérience ne sont pas des prédicteurs significatifs du salaire. Par contre, une note élevée à l'entretien d'embauche est associée à un salaire plus bas, ce qui est une relation inattendue et pourrait indiquer des biais ou des facteurs confondants non pris en compte dans l'analyse.

### **Recommandations**

#### Exploration de Variables Supplémentaires :

Pour améliorer la compréhension des déterminants du salaire, il serait bénéfique d'explorer d'autres variables potentielles, telles que le niveau d'éducation, les compétences spécifiques, le secteur d'activité, la localisation géographique, et d'autres facteurs contextuels ou personnels.

#### Validation des Résultats :

Il serait utile de valider ces résultats avec des ensembles de données plus larges et diversifiés pour vérifier la robustesse des conclusions et identifier toute variabilité contextuelle.

En conclusion, cette analyse de régression multiple a révélé que, parmi les variables étudiées, seule la note de l'entretien d'embauche a une relation significative avec le salaire, et cette relation est négative. D'autres variables et facteurs doivent être explorés pour mieux comprendre les déterminants du salaire et améliorer les modèles prédictifs dans ce domaine. Les résultats soulignent l'importance d'une approche holistique et multidimensionnelle dans l'analyse des facteurs influençant les salaires.