

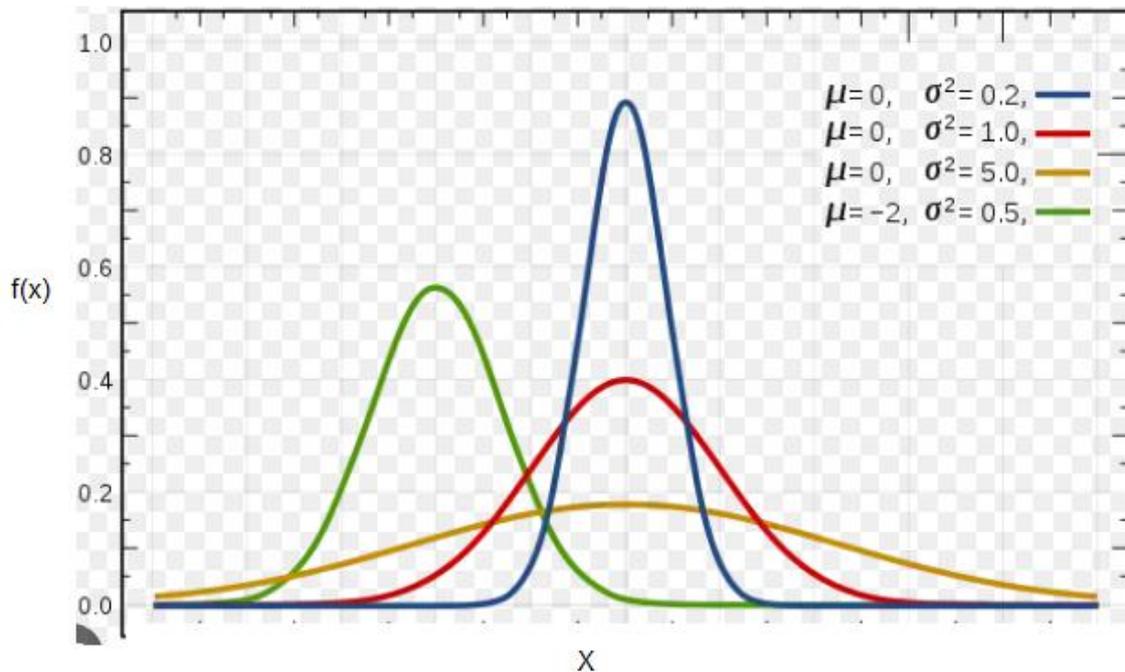
TP N°3 : Test de normalité

Objectif :

1. **Présenter les méthodes graphiques qui nous permet d'étudier la normalité.**
2. **Présenter les techniques des tests de normalité.**

La loi normale joue un rôle très important en statistique et en probabilité car elle a des propriétés pour calculer la probabilité et présente une condition nécessaire pour appliquer les tests (chapitre 3 et les cours de 3^{ème} année).

La loi normale ou loi de Gauss : C'est une loi des variables aléatoire continue. Elle est définie par deux paramètres : l'espérance noté μ et la variance noté σ^2 . Parmi les propriétés de la loi normale la symétrie par rapport à la moyenne.



La figure ci-dessus présente la courbe gaussienne pour différentes valeurs de la moyenne et la variance. Les courbes bleue, rouge et jaune sont symétriques par rapport à 0 et la courbe verte est symétrique par rapport à -2. Concernant l'écart type, plus sa valeur est élevée, plus la planéité est faible et vice versa. Pour les

courbes bleue et verte, on retrouve un fin aplatissement, pour la courbe rouge, un léger aplatissement, et pour la courbe jaune, on retrouve un aplatissement plat. La courbe en rouge présente la loi normale centrée réduite car sa moyenne égale à 0 et son écart type égale à 1.

Pour tester la normalité d'un échantillon c'est-à-dire s'il suit la loi normale ou non on a les méthodes graphiques comme l'histogramme et le diagramme Q-Q Gaussien (on va les expliquer dans la partie application) et les méthodes de calculs.

Techniques pour faire un test de normalité : On résume les étapes de ce test comme suit :

Etape 1 : On écrit les hypothèse H_0 et H_1 :

Hypothèse nulle, H_0 : « la distribution est gaussienne ».

Hypothèse alternative, H_1 : « la distribution n'est pas gaussienne ».

Etape 2 : On choisit quel test de normalité on utilise :

- On utilise test de Kolmogorov-Smirnov lorsque la taille des échantillons est très élevée ($n \geq 30$).
- On utilise test de Shapiro-Wilk lorsque la taille des échantillons est assez petit ($n < 30$).

Etape 3 : On calcule la signification (sig) avec SPSS

Etape 4 (décision): On compare le sig avec la valeur de α donnée :

- Si Signification (sig) inférieure à α , alors on rejette H_0 . C'est à dire la distribution n'est pas Gaussienne
- Si Signification supérieure (sig) à α , alors on accepte H_0 . C'est-à-dire la distribution est Gaussienne.

Remarque : Le α prend des valeurs entre 0 et 1. La valeur de α est déterminée au début du test, telles que 0.05, 0.1, 0.01,

Exemple 1 : On a relevé le nombre de bactéries contenues respectivement dans 20 milieux biologique, où les résultats sont les suivants :

31	32	32	33	33	33	34	34	34	34
34	35	35	35	35	35	35	36	36	36
36	36	37	37	37	37	38	38	39	39

Questions :

- 1- Déterminer l'objectif pour cette expérience.
- 2- Déterminer la variable étudiée
- 3- Déterminer l'hypothèse nulle et alternative pour la normalité.
- 4- Tracer le tableau de de la statistique descriptive.
- 5- Avec un risque de signification de 5%, que peut-on dire pour la normalité ?

Réponse :

1- **Objectif :** Tester si l'échantillon suit la loi normale

2- **Variable étudiée :** nombre de bactéries. Nature : quantitative.

3- **Hypothèse :**

L'hypothèse nulle, H_0 : « la distribution est gaussienne ».

Hypothèse alternative, H_1 : « la distribution n'est pas gaussienne ».

Pour répondre aux questions 4 et 5 il faut entrer les données.

Pour saisir ces données, vous pouvez utiliser la méthode 1 ou 2 que nous avons déjà vue dans le premier et le deuxième TP.

Etape 1 : On choisit l'affichage des variables, et on entre le nom de la variable et ses caractéristiques. Car la variable est quantitative on choisit type :

numérique et Mesure : Echelle.

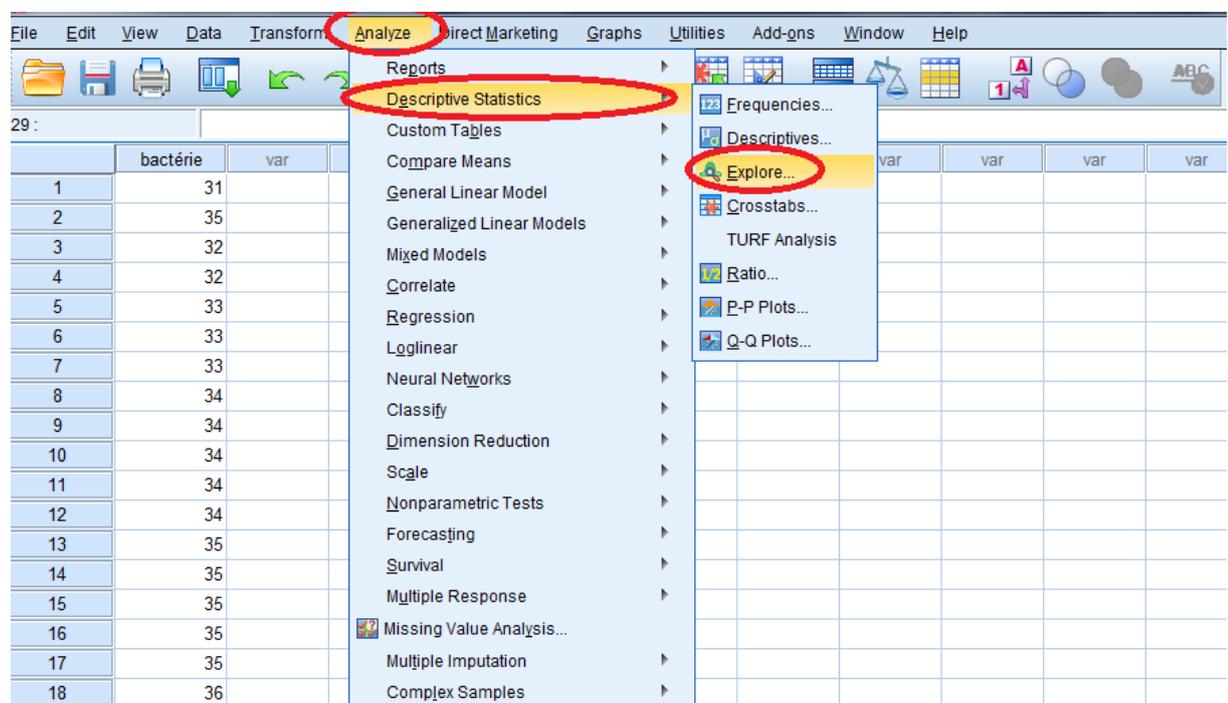


Etape 2 : Dans l'affichage des données, on remplit la colonne Bactérie par les modalités de cette variable, comme l'indique la figure suivante :

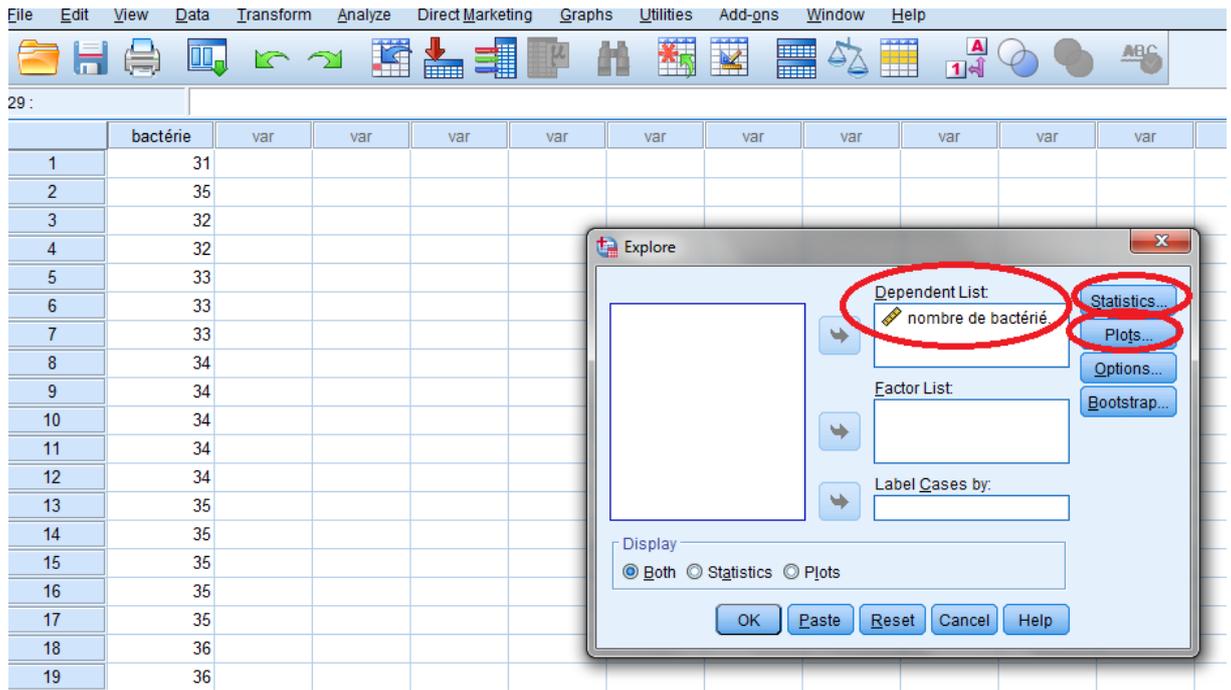
	bactérie	var									
1	31										
2	35										
3	32										
4	32										
5	33										
6	33										
7	33										
8	34										
9	34										
10	34										
11	34										
12	34										
13	35										
14	35										
15	35										
16	35										
17	35										
18	36										
19	36										
20	36										

Etape 3 : Dans cette étape on calcule les mesures de tendance centrale de dispersion et de forme ainsi que le test de normalité.

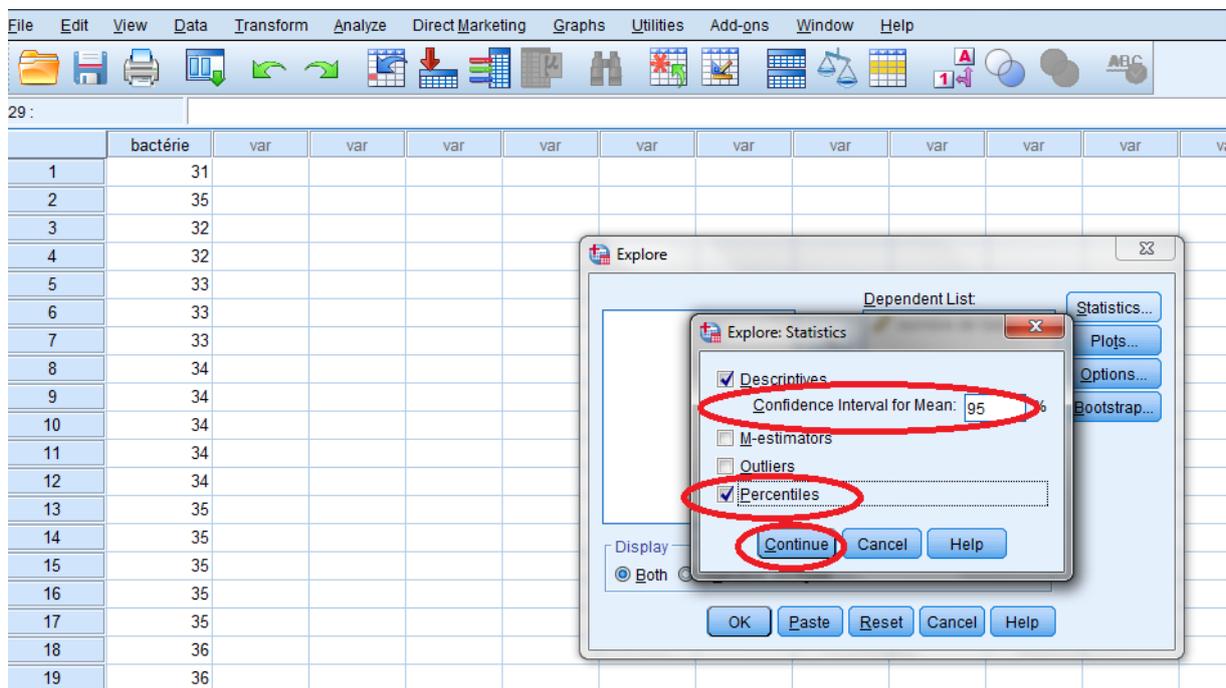
Pour obtenir le test de normalité, en cliquant sur le bouton « Analyze » qui se trouve dans la barre des outils, et choisir « Statistiques descriptives », et puis « Explorer ».



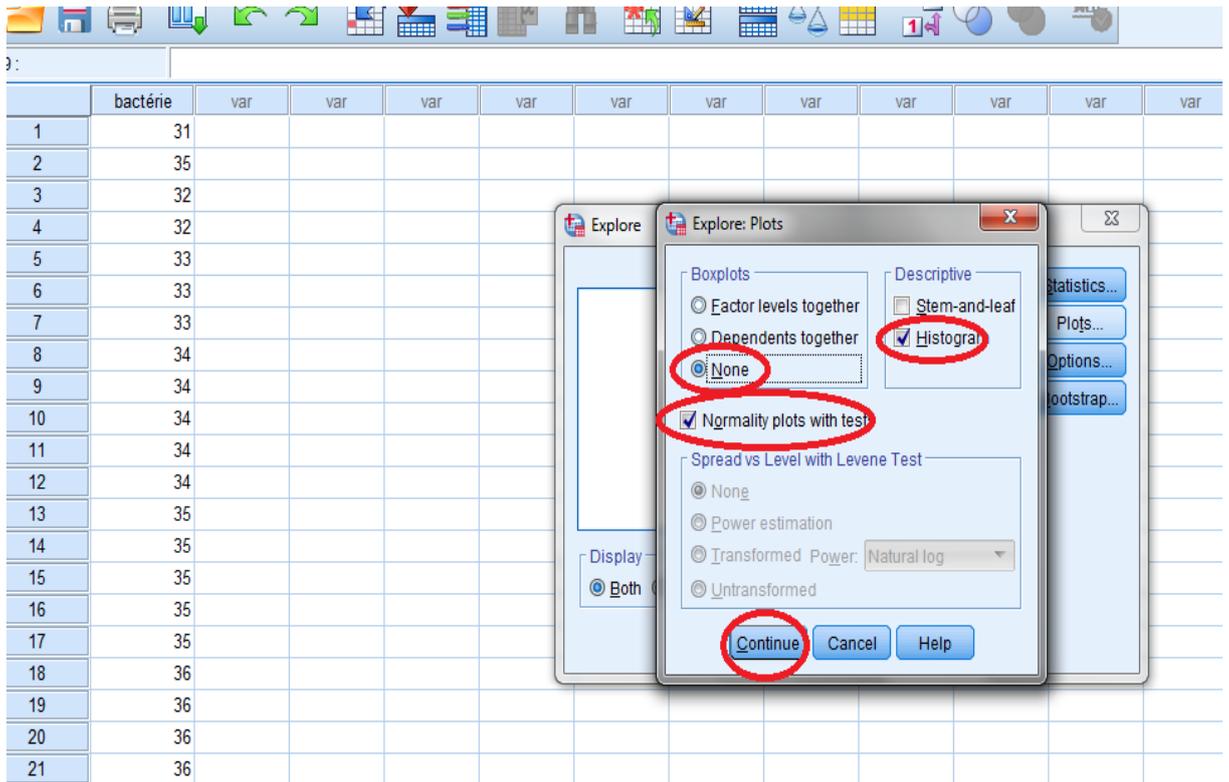
- On choisit dans « liste variables dépendantes » la variable quantitative mesurable « nombre de bactéries ».



- a- Dans le Bouton « Statistiques » on garde le pourcentage pour l'intervalle de confiance le niveau de signification 95%.
- b- On choisit « Centiles » pour visualiser les quartiles (Q1, Q2 et Q3).
- c- En cliquant sur « poursuivre ».



- Sur le bouton « Diagrammes », dans « Caractéristiques » en cliquant sur le choix « Histogramme », et aussi on coche sur le choix « Graphes de répartition gaussiens avec tests » pour validé test de normalité.
- En cliquant sur « poursuivre ».



Interprétation des résultats : On obtient une nouvelle fenêtre, c'est la fenêtre des résultats, elle nous donne : les tableaux et les graphs

1- Tableau de la statistique descriptive :

Descriptives

		Statistic	Std. Error	
nombre de bactéries	Mean	35.2000	.37263	
	95% Confidence Interval for Mean	Lower Bound	34.4379	
		Upper Bound	35.9621	
	5% Trimmed Mean	35.2037		
	Median	35.0000		
	Variance	4.166		
	Std. Deviation	2.04096		
	Minimum	31.00		
	Maximum	39.00		
	Range	8.00		
	Interquartile Range	3.00		
	Skewness	-.003	.427	
	Kurtosis	-.439	.833	

On remarque que la moyenne (mean) est égale à 35.20 est presque égale à la médiane 35. Pour la symétrie (Skewness) égale à -0.003 est presque égale à 0, donc on peut dire que la distribution est presque symétrique.

2- **Tableau des Quantiles :** Ce tableau nous donne les quantiles où Q1 (25%) égale à 34, Q2 (50%) égale à 35 et Q3 (75%) égale à 37:

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	nombre de bactéries	31.5500	32.1000	34.0000	35.0000	37.0000	38.0000	39.0000
Tukey's Hinges	nombre de bactéries			34.0000	35.0000	37.0000		

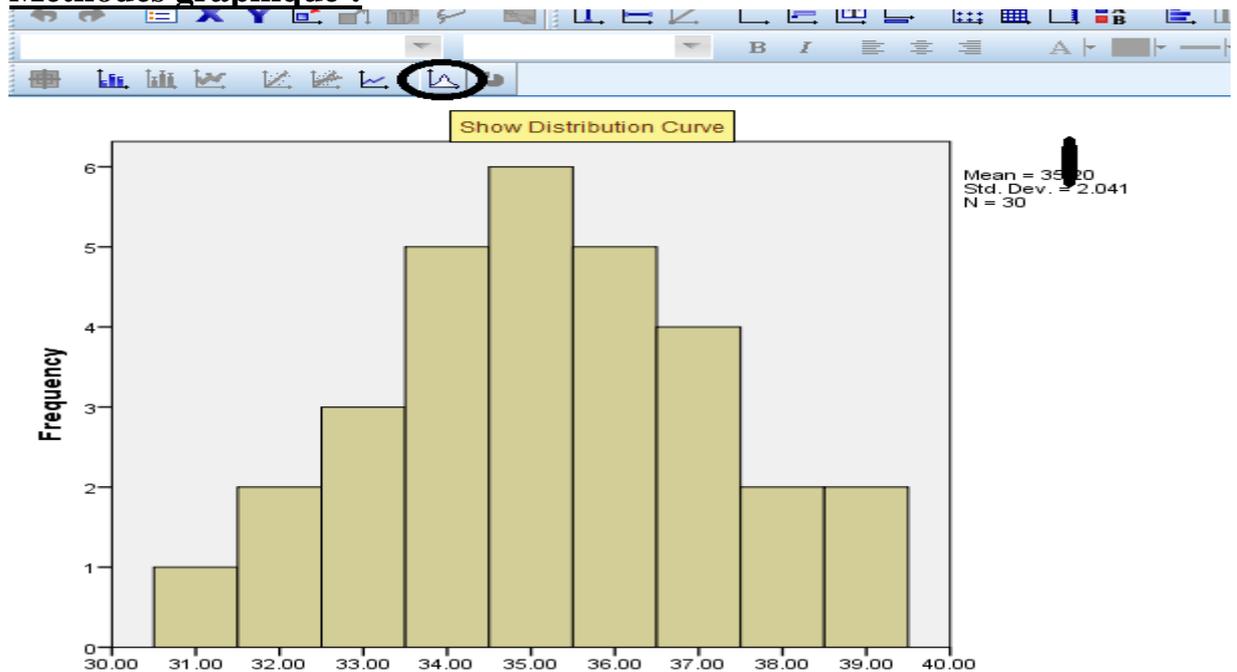
3- **Tableau de tests de normalité:**

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
nombre de bactéries	.106	30	.200 [*]	.974	30	.653

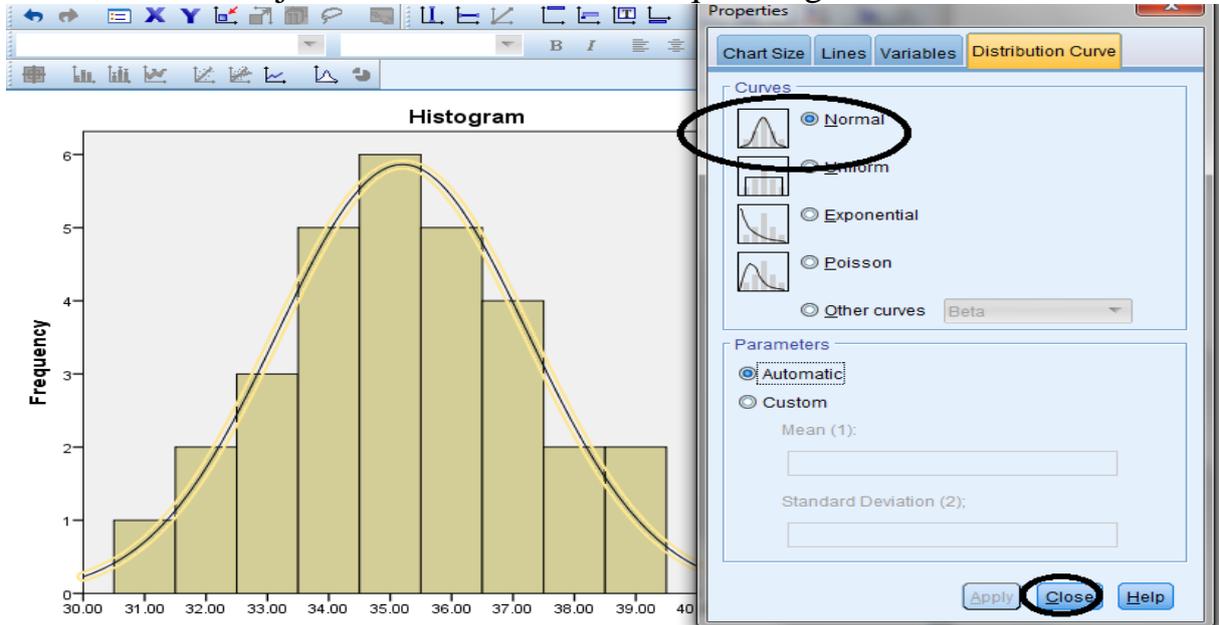
On a la taille de l'échantillon est égale à 30 donc on choisit le test de Kolomogorov-Smirnov et on lit la valeur de sig qui est égale à 0.200, elle est supérieur à 0.05.

Conclusion : Car le sig est supérieur à 0.05 alors on accepte H_0 c'est-à-dire la distribution est Gaussienne.

4- **Méthodes graphique :**



- **La première méthode c'est l'histogramme** : Pour ajouter la courbe Gaussienne à l'histogramme on clique deux fois sur le graph et on choisit ajouter courbe comme l'indique la figure

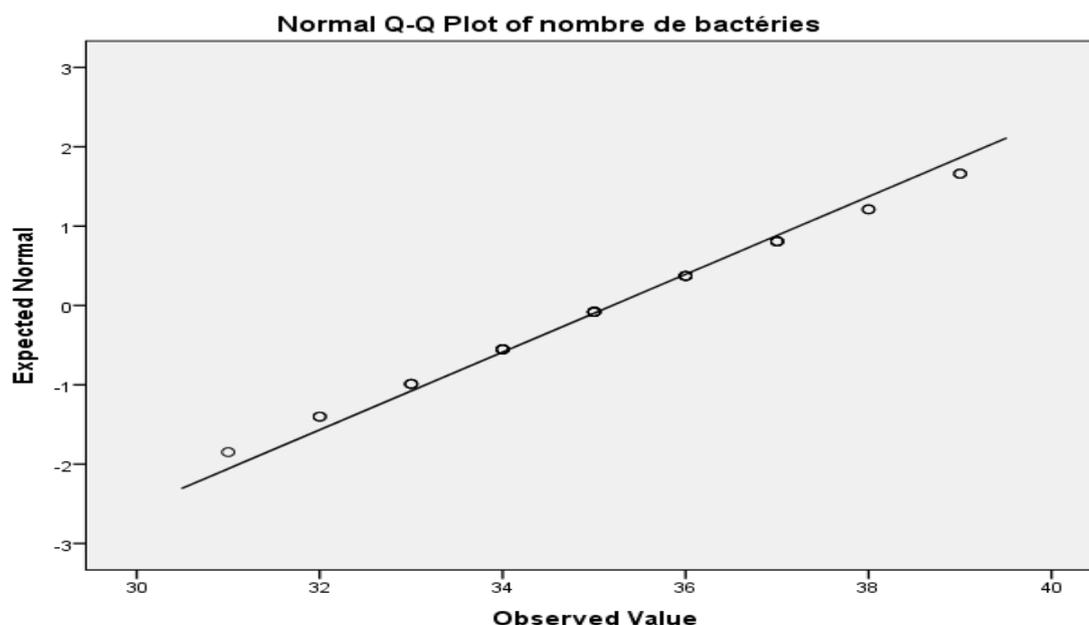


On remarque que la courbe et l'histogramme sont presque identiques

- **Deuxième méthode c'est la méthode de diagramme Gaussien Q-Q :**
Les abscisses présentent les valeurs de la variables (nombre de bactéries) et les ordonnées présentes les valeurs de la loi normale. On observe que la majorité des points se trouvent sur la droite ou plus proche de la droite donc on peut dire que la distribution est suit la loi normale ou Gauss.

- **Conclusion :**

On conclut que la distribution de la variable quantitative « nombre de bactéries » est gaussienne au taux de confiance 95%.



Exemple 2 : La série statistique suivante présente les notes de 20 étudiants :

7	7	8	10	10	10	10	10	10	10
10	11	11	11	11	12	12	12	13	15

Questions :

- 1- Déterminer l'objectif pour cette expérience.
- 2- Déterminer la variable étudiée.
- 3- Déterminer l'hypothèse nulle et alternative pour la normalité.
- 4- Tracer le tableau de de la statistique descriptive.
- 5- Avec un risque de signification de 10%, que peut-on dire pour la normalité ?

Réponse :

1- Objectif : Tester si l'échantillon suit la loi normale

2- Variable étudiée : note des étudiants. Nature : quantitative.

3- Hypothèse :

L'hypothèse nulle, H_0 : « la distribution est gaussienne ».

Hypothèse alternative, H_1 : « la distribution n'est pas gaussienne ».

4- Tableau statistique :

Descriptives

		Statistic	Std. Error
notes des étudiants	Mean	10.5000	.41990
	94% Confidence Interval for Mean	Lower Bound 9.6602 Upper Bound 11.3398	
	5% Trimmed Mean	10.4444	
	Median	10.0000	
	Variance	3.526	
	Std. Deviation	1.87785	
	Minimum	7.00	
	Maximum	15.00	
	Range	8.00	
	Interquartile Range	1.75	
	Skewness	.132	.512
	Kurtosis	1.133	.992

- **Tableau des Quantiles :**

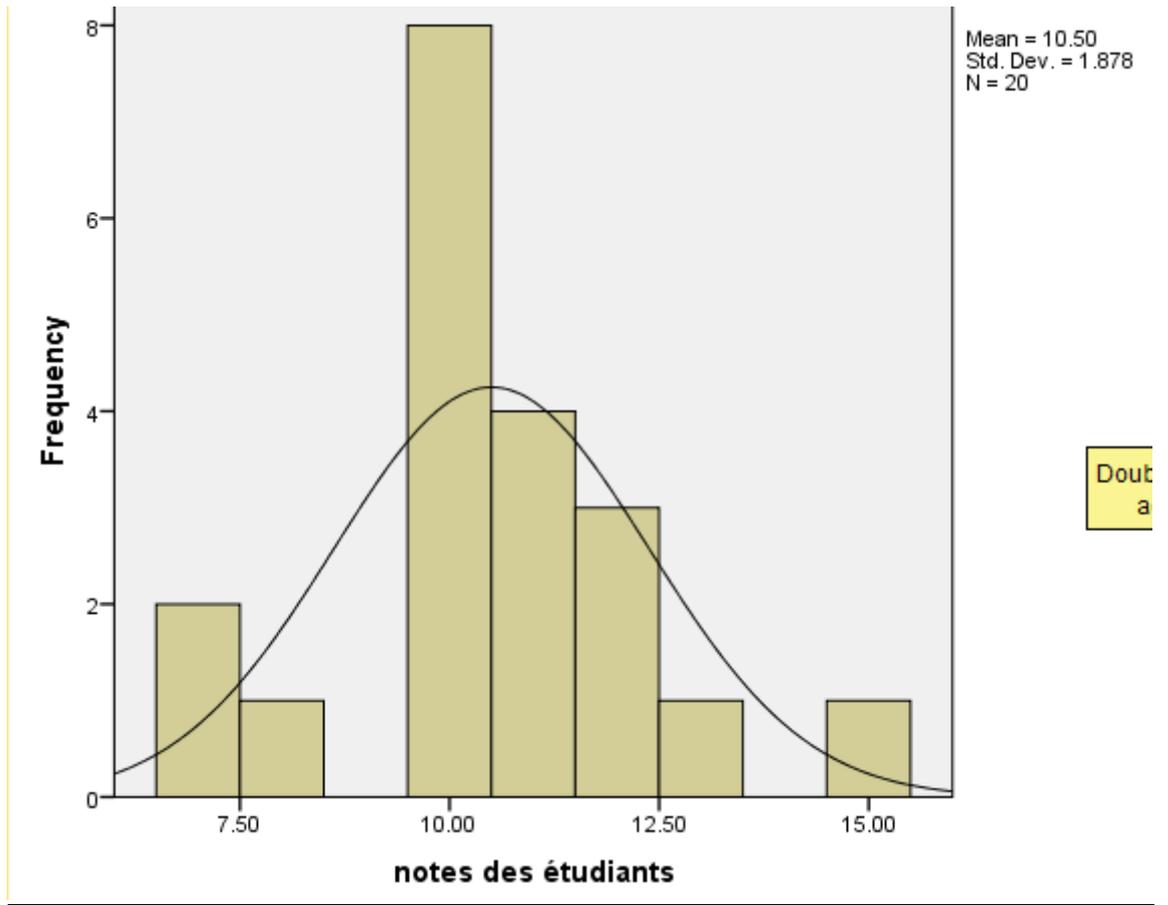
		Percentiles						
		Percentiles						
		5	10	25	50	75	90	95
Weighted Average (Definition 1)	notes des étudiants	7.0000	7.1000	10.0000	10.0000	11.7500	12.9000	14.9000
Tukey's Hinges	notes des étudiants			10.0000	10.0000	11.5000		

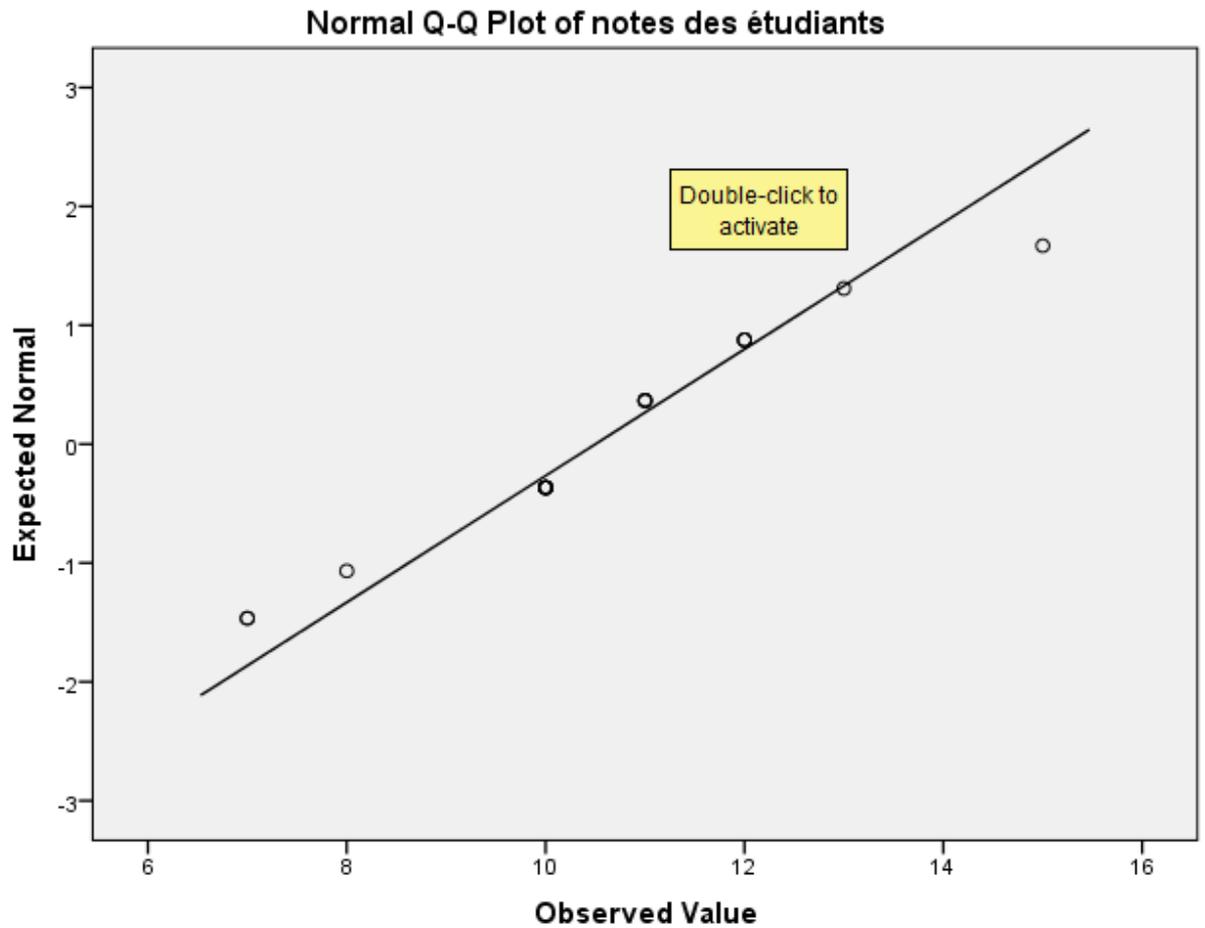
- **Tableau de test de normalité :**

		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
notes des étudiants		.245	20	.003	.918	20	.091

On lit la valeur de sig dans le test de shapiro- Wilk car la taille de l'échantillon égale à 20 inférieur de 30. On observe que le sig égale à 0.091 est inférieur à 0.1 donc on rejette H_0 c'est-à-dire la distribution n'est pas Gaussienne.

Méthode graphique :





On remarque que la courbe Gaussienne n'est pas identique avec l'histogramme de même on a trois points sur la droite, donc on peut dire que la distribution n'est pas Gaussienne.