L'Importance de la Normalité en Statistique

RAHMANI Naceur

October 20, 2025

Introduction à la Normalité en Statistique

La **normalité** est un concept fondamental en statistique car elle est à la base de nombreuses méthodes statistiques. De nombreuses techniques, telles que les tests paramétriques, les régressions et les analyses de variance, font l'hypothèse que les données suivent une **distribution normale**.

Pourquoi la Normalité est-elle Importante ?

La normalité simplifie les calculs et garantit des propriétés statistiques optimales, telles que des **estimations non biaisées**, **efficaces**, et **consistantes**. La plupart des tests et méthodes statistiques reposent sur cette hypothèse.

Applications des Tests Paramétriques

De nombreux tests statistiques **paramétriques** reposent sur l'hypothèse de normalité des données. Ces tests incluent :

- Le **test t de Student** pour comparer deux moyennes,
- **L'ANOVA** pour comparer plusieurs moyennes,
- La **régression linéaire** pour modéliser les relations entre variables.

Hypothèse de Normalité

La validité de ces tests dépend de l'hypothèse que les données suivent une distribution normale, en particulier les résidus dans les modèles de régression.

Propriétés de la Distribution Normale

La **distribution normale** a plusieurs propriétés très utiles :

- Elle est **symétrique** autour de la moyenne.
- Elle est définie par deux paramètres : la **moyenne** μ et l'**écart-type** σ.
- Elle est caractérisée par la **courbe en cloche**, ce qui signifie que la majorité des données sont proches de la moyenne, et que les extrêmes (valeurs très grandes ou petites) sont rares.

Théorème Central Limite

Le **Théorème Central Limite** stipule que, même si les données ne suivent pas une distribution normale, la moyenne des échantillons suffisamment grands tend à suivre une distribution normale. Cela explique la **centralité** de la normale dans les statistiques.

Tests de Normalité

Pour tester la normalité des données, plusieurs tests existent :

- **Test de Shapiro-Wilk** : Très puissant pour tester la normalité des données.
- **Test de Kolmogorov-Smirnov** : Compare la fonction de répartition empirique à la fonction théorique d'une distribution normale.
- **Test d'Anderson-Darling** : Plus sensible aux écarts dans les **queues** de la distribution.
- **Test de Lilliefors** : Une version modifiée du test de Kolmogorov-Smirnov, adaptée aux paramètres inconnus.
- **Test de D'Agostino et Pearson** : Test de normalité basé sur l'asymétrie et la kurtose.

Ces tests comparent les données observées à une distribution normale théorique et permettent de décider si les données suivent cette distribution.

Test de Shapiro-Wilk

Le **test de Shapiro-Wilk** est un test de normalité basé sur la comparaison des quantiles empiriques et théoriques d'une distribution normale.

Hypothèses

- H_0 : Les données suivent une distribution normale.
- H_1 : Les données ne suivent pas une distribution normale.

La statistique de test W est calculée en comparant les quantiles empiriques des données aux quantiles théoriques de la distribution normale.

Test de Kolmogorov-Smirnov

Le **test de Kolmogorov-Smirnov (KS)** compare la fonction de répartition empirique (FRE) d'un échantillon à la fonction de répartition théorique d'une distribution normale.

Statistique de Test

La statistique de test est donnée par :

$$D_n = \sup_x \left| \hat{F}_n(x) - F(x) \right|$$

où $\hat{F}_n(x)$ est la fonction de répartition empirique et $F(x)$ est la fonction de répartition théorique de la normale.

Décision

Si D_n est supérieur à la valeur critique, on rejette H_0, indiquant que les données ne suivent pas une loi normale.

Tests Non Paramétriques

Lorsque les données ne suivent pas une distribution normale, on utilise des **tests non paramétriques**, qui ne font pas l'hypothèse de normalité. Ces tests incluent :

- **Test de Mann-Whitney** : Pour comparer deux groupes indépendants.
- **Test de Kruskal-Wallis** : Pour comparer plusieurs groupes indépendants.
- **Test de Wilcoxon** : Pour comparer deux groupes appariés.

Les tests non paramétriques sont plus robustes car ils ne dépendent pas de la forme de la distribution des données.

Conséquences de la Non-Normalité

Si l'hypothèse de normalité est violée et que des tests paramétriques sont utilisés, les résultats peuvent être :

- **Biaisés** : Les estimations de la moyenne et de la variance peuvent être incorrectes.
- **Moins puissants** : Les tests paramétriques deviennent moins efficaces pour détecter les différences réelles.
- **Infiables** : Les intervalles de confiance et les tests de significativité peuvent être incorrects.

Dans ces cas, des tests non paramétriques doivent être privilégiés.

Conclusion

La normalité est un principe fondamental en statistique, car de nombreux tests et méthodes reposent sur cette hypothèse. Si les données ne suivent pas une distribution normale, il est essentiel de tester la normalité et d'utiliser des **tests non paramétriques** lorsque cela est nécessaire.
La vérification de la normalité permet d'assurer la fiabilité et la validité des résultats statistiques.

Tests basés sur les quantiles empiriques

RAHMANI Naceur

October 20, 2025

Test de Shapiro-Wilk: Principe

Le test de Shapiro-Wilk teste la normalité des données en comparant les quantiles empiriques des données observées avec ceux d'une distribution normale théorique.

Principe du Test

Plus le coefficient W calculé est proche de 1, plus les données suivent une distribution normale.

Statistique de Test

La statistique de test W est donnée par :

$$W = \frac{\left(\sum_{i=1}^{n} a_i \hat{q}_i\right)^2}{\sum_{i=1}^{n} (\hat{q}_i - \bar{X})^2}$$

Où :

- a_i sont des coefficients calculés à partir des quantiles théoriques,
- $\hat{q}_i$ sont les quantiles empiriques des données observées,
- $\bar{X}$ est la moyenne des données.

Test de Shapiro-Wilk: Hypothèses et Décision

Hypothèses

- H_0 : Les données suivent une distribution normale.
- H_1 : Les données ne suivent pas une distribution normale.

Décision

Si W est proche de 1, on accepte H_0, indiquant que les données suivent une distribution normale. Si W est significativement inférieur à 1, on rejette H_0.

Test de Shapiro-Wilk: Exemple

Soit un échantillon de données $X = [3.2, 2.9, 4.1, 3.5, 3.8]$.

Exemple de Calcul

- Calculons le coefficient W pour ces données. - Si W est proche de 1, nous acceptons que les données suivent une loi normale.

Calcul détaillé

- Moyenne $\bar{X} = 3.5$, - Quantiles empiriques $\hat{q}_i$ calculés, - Application de la formule pour W.

Test de Wilks: Principe

Le test de Wilks est utilisé dans les analyses multivariées pour tester l'égalité des moyennes entre plusieurs groupes.

Principe

Il compare la matrice de variance-covariance des groupes avec la matrice de variance-covariance totale.

Statistique de Test

La statistique de test de Wilks Λ est définie comme suit :

$$\Lambda = \prod_{i=1}^{k} \frac{1}{1 + \frac{\lambda_i}{n}}$$

Où :

- λ_i sont les valeurs propres de la matrice de variance-covariance des groupes,
- n est la taille de l'échantillon.

Test de Wilks: Hypothèses et Décision

Hypothèses

- H_0 : Les moyennes des groupes sont égales.
- H_1 : Les moyennes des groupes sont différentes.

Décision

Si Λ est proche de 1, on accepte H_0, indiquant que les moyennes des groupes sont égales. Si Λ est significativement plus petit que 1, on rejette H_0.

Test de Wilks: Exemple

Supposons trois groupes de données avec des valeurs propres calculées $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, et $\lambda_3 = 0.2$.

Exemple de Calcul

- Calculons Λ et comparons-le avec une valeur critique. - Si Λ est plus petit que la valeur critique, nous rejetons H_0 et concluons que les moyennes des groupes sont différentes.

Test de Shapiro-Francia

Rahmani Naceur

October 20, 2025

Introduction au Test de Shapiro-Francia

Le test de **Shapiro-Francia** est une version modifiée du test de **Shapiro-Wilk** pour tester la normalité des données. Il est particulièrement adapté aux échantillons de petite taille.

Principe

Le test de Shapiro-Francia mesure la **déviation** des quantiles empiriques des données par rapport aux quantiles théoriques d'une distribution normale. Il utilise une statistique de test W' qui est calculée sur la base des moments de l'échantillon.

Hypothèses du Test de Shapiro-Francia

Hypothèses

- H_0 : Les données suivent une distribution normale.
- H_1 : Les données ne suivent pas une distribution normale.

Le test est basé sur l'idée qu'une petite valeur de la statistique de test W' indique que les données sont éloignées d'une distribution normale, tandis qu'une statistique proche de 1 indique que les données suivent une distribution normale.

Formule de la Statistique de Test

La statistique de test W' de Shapiro-Francia est calculée comme suit :

$$W' = \frac{\left(\sum_{i=1}^{n} a_i X_i\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

où :

- X_i sont les données observées triées,
- $\bar{X}$ est la moyenne des données,
- a_i sont les coefficients calculés à partir des quantiles théoriques d'une distribution normale.

Calcul des Coefficients a_i

Les coefficients a_i sont des valeurs calculées à partir de la **distribution normale** pour un échantillon donné. Ces coefficients sont utilisés pour **pondérer les quantiles empiriques** dans le calcul de la statistique W'.

Calcul des coefficients a_i

Les coefficients a_i dépendent de la taille de l'échantillon n et sont donnés par les valeurs théoriques des quantiles de la distribution normale standard.

Décision et Interprétation

Décision

- Si W' est **proche de 1**, on accepte H_0 (les données suivent une distribution normale). - Si W' est **significativement inférieur à 1**, on rejette H_0 (les données ne suivent pas une distribution normale).

Valeur critique

La valeur critique dépend de la taille de l'échantillon n et du niveau de signification α (souvent $\alpha = 0.05$).

Exemple de Calcul

Soit un échantillon de données :

$$X = [2.5, 3.1, 2.9, 3.4, 3.8]$$

Étapes

- Trier les données : $X_{\text{trié}} = [2.5, 2.9, 3.1, 3.4, 3.8]$,
- Calculer la moyenne $\bar{X} = 3.174$,
- Calculer les coefficients a_i,
- Appliquer la formule de W'.

Supposons que $W' = 0.98$. Puisque W' est proche de 1, on accepterait H_0, ce qui signifie que l'échantillon suit une distribution normale.

Avantages et Limites

Avantages

- Le test de Shapiro-Francia est efficace pour les **petits échantillons**. - Il est **simple à mettre en œuvre** et **interpréter**.

Limites

- Il peut être **sensible à l'échantillon**. Pour des échantillons très petits ou très grands, il peut donner des résultats biaisés. - Comme tout test de normalité, il peut **échec** dans le cas où les données présentent des **déviations extrêmes** mais non significatives.

Test de Shapiro-Francia : Calcul des Coefficients a_i

Votre Nom

October 20, 2025

Introduction au Test de Shapiro-Francia

Le test de **Shapiro-Francia** est une version modifiée du test de **Shapiro-Wilk** pour tester la normalité des données. Il est particulièrement adapté aux **petits échantillons**.

Principe

Le test mesure la déviation des quantiles empiriques des données par rapport aux quantiles théoriques d'une distribution normale. La statistique de test W' est utilisée pour évaluer cette déviation.

Calcul des Coefficients a_i

Les **coefficients a_i** sont utilisés pour pondérer les quantiles empiriques dans le calcul de la statistique de test W'.

Formule des Coefficients a_i

Les coefficients a_i sont calculés à partir des quantiles théoriques de la **distribution normale** standard. La formule est la suivante :

$$a_i = \frac{\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)}{\sqrt{\sum_{i=1}^{n}\left(\Phi^{-1}\left(\frac{i-0.375}{n+0.25}\right)\right)^2}}$$

Où :

- Φ^{-1} est l'inverse de la fonction de répartition de la loi normale standard,
- i est l'indice de l'observation dans l'échantillon trié,
- n est la taille de l'échantillon.

Exemple de Calcul des Coefficients a_i

Soit un échantillon de données $X = [2.5, 3.1, 3.7, 4.0, 4.2]$, de taille $n = 5$.

Étapes

- Trier les données : $X_{\text{trié}} = [2.5, 3.1, 3.7, 4.0, 4.2]$,
- Calculer les probabilités cumulatives $p_i = \frac{i-0.375}{n+0.25}$ pour chaque valeur i,
- Calculer les quantiles théoriques $\Phi^{-1}(p_i)$,
- Calculer les coefficients a_i en utilisant la formule donnée.

Calcul des Probabilités Cumulatives p_i

Nous calculons les probabilités cumulatives pour chaque observation de l'échantillon. Étant donné $n = 5$, les probabilités sont calculées comme suit :

$$p_1 = \frac{1 - 0.375}{5 + 0.25} = 0.123, \quad p_2 = \frac{2 - 0.375}{5 + 0.25} = 0.231, \quad p_3 = \frac{3 - 0.375}{5 + 0.25} =$$

$$p_4 = \frac{4 - 0.375}{5 + 0.25} = 0.446, \quad p_5 = \frac{5 - 0.375}{5 + 0.25} = 0.554$$

Calcul des Quantiles Théoriques $\Phi^{-1}(p_i)$

Nous calculons les quantiles théoriques associés aux probabilités cumulatives p_i en utilisant l'inverse de la fonction de répartition de la loi normale :

$$\Phi^{-1}(0.123) \approx -1.160, \quad \Phi^{-1}(0.231) \approx -0.737, \quad \Phi^{-1}(0.338) \approx -0.426$$

$$\Phi^{-1}(0.446) \approx -0.125, \quad \Phi^{-1}(0.554) \approx 0.125$$

Calcul des Coefficients a_i

Nous calculons maintenant les coefficients a_i en normalisant les quantiles théoriques :

$$a_1 = \frac{-1.160}{\sqrt{2.1013}} \approx -0.779, \quad a_2 = \frac{-0.737}{\sqrt{2.1013}} \approx -0.493$$

$$a_3 = \frac{-0.426}{\sqrt{2.1013}} \approx -0.281, \quad a_4 = \frac{-0.125}{\sqrt{2.1013}} \approx -0.083, \quad a_5 = \frac{0.125}{\sqrt{2.1013}} \approx$$

Ainsi, les **coefficients a_i** sont :

$$a_1 \approx -0.779, \quad a_2 \approx -0.493, \quad a_3 \approx -0.281, \quad a_4 \approx -0.083, \quad a_5 \approx 0.08$$

Utilisation des Coefficients a_i dans la Statistique de Test

Une fois les **coefficients a_i** calculés, on peut les utiliser pour déterminer la statistique de test W' du test de Shapiro-Francia. La statistique W' est donnée par :

$$W' = \frac{\left(\sum_{i=1}^{n} a_i X_i\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Où :

- X_i sont les données observées triées,
- $\bar{X}$ est la moyenne des données,
- a_i sont les coefficients calculés à partir des quantiles théoriques.

Conclusion

Le test de **Shapiro-Francia** est utilisé pour tester la normalité des données. Les **coefficients a_i** jouent un rôle crucial dans la pondération des quantiles empiriques pour calculer la statistique de test W'.
Ce test est particulièrement adapté aux **petits échantillons** et permet de déterminer si les données suivent une distribution normale en comparant les quantiles empiriques aux quantiles théoriques d'une distribution normale.