

Chapitre 9

STATISTICA

Test d'hypothèses

Ajustement de distribution

Ajustement de distribution

But des tests d'hypothèse:

Les hypothèses sont des prévisions *a priori* que nous faisons sur la structure des résultats.

But des ajustement de distribution:

Je souhaite savoir mon échantillon est semblable à une fonction mathématique connue (loi normal, par exemple).

Ajustement de distribution

A quoi cela sert?

Le module Ajustement de Distributions est utilisé pour évaluer l'ajustement des données observées à des distributions théoriques.

Dans certains domaines de recherche, il est possible de formuler des hypothèses quant à la distribution spécifique de la variable étudiée. Par exemple, des variables dont les valeurs sont déterminées par un nombre infini d'événements aléatoires indépendants suivent la distribution normale : on peut penser que la taille d'un individu résulte d'une infinité de facteurs indépendants comme des prédispositions génétiques, des maladies infantiles, la nutrition, etc...

En revanche, si les valeurs d'une variable résultent d'événements très rares, la variable sera distribuée selon la distribution de Poisson (parfois appelée distribution des événements rares). Par exemple, les accidents industriels peuvent être considérés comme la conjonction d'une suite d'événements malheureux (et peu probables), et leur fréquence tend à être distribuée selon la distribution de Poisson.

Les procédures d'ajustement de distributions sont également utilisées lorsque vous souhaitez vérifier l'hypothèse de normalité avant d'utiliser des tests non paramétriques. Par exemple, vous pouvez utiliser le test de normalité de Kolmogorov-Smirnov ou le test W de Shapiro-Wilk afin de tester la normalité.

Voir dans l'aide les différents 'type de distribution'

Ajustement de distribution

The screenshot displays the STATISTICA software interface. The menu bar includes 'Statistiques', 'Graphiques', 'Outils', 'Données', 'Fenêtre', and 'A'. The 'Statistiques' menu is open, showing various statistical analysis options. The 'Ajustement de Distributions' option is highlighted, and a dialog box titled 'Ajustement de Distributions : Exercice.07.sta' is open. The dialog box has a 'Base' tab and two radio buttons: 'Distributions Continues' (selected) and 'Distributions Discontinues'. Under 'Distributions Continues', the 'Normale' distribution is selected. Other options include 'Uniforme', 'Exponentielle', 'Gamma', 'Log-Normale', and 'Chi-deux'. Under 'Distributions Discontinues', options include 'Binomiale', 'Poisson', 'Géométrique', and 'Bernoulli'. The dialog box also features buttons for 'Annuler', 'Options', 'Ouvrir', and 'SELECT CRUES'.

Ajustement de distribution normale

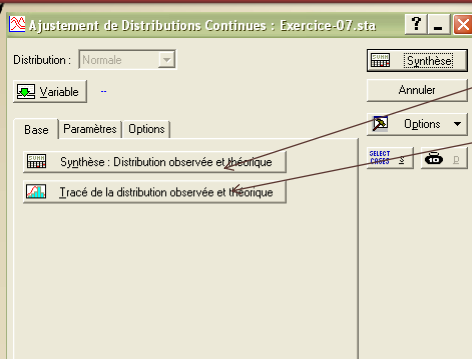
Normal

La distribution normale (la courbe en "cloche", qui est symétrique autour de la moyenne) est une fonction théorique fréquemment utilisée dans les statistiques inférentielles comme approximation des distributions d'échantillonnage (voir également la rubrique concepts élémentaires). En général, la distribution normale fournit un bon modèle pour une variable aléatoire, lorsque :

1. La variable a une forte tendance à prendre la valeur centrale ;
2. Des écarts positifs et négatifs autour de cette valeur centrale ont la même probabilité ;
3. La fréquence des écarts diminue rapidement à mesure que les écarts deviennent plus grands.

Tout comme un mécanisme sous-jacent produit une distribution normale, on peut penser à un nombre infini d'événements aléatoires indépendants (binomiaux) induisant les valeurs d'une variable particulière. Par exemple, il y a probablement un nombre infini de facteurs qui déterminent la taille d'une personne (des milliers de gènes, la nutrition, les maladies, etc.). Par conséquent, on peut s'attendre à ce que la taille soit distribuée normalement dans la population.

Ajustement de distribution normale



•Vous donne la table des fréquences

•Vous donne le graphique des répartitions de fréquence

Ajustement de distribution normale

•Spécifiez le nombre de catégories pour la table de fréquence ou l'histogramme. Remarque : le nombre maximum de catégories autorisées est de 50.

•Déterminer les limites

•Paramètre de votre échantillon

Ajustement de distribution normale

•STATISTICA va calculer la statistique du D-max à partir des données tabulées de la feuille de données.

•STATISTICA va passer une revue les données observées triées une à une, et calculer en chaque point la fréquence cumulée théorique ; la statistique du D-max de Kolmogorov est la plus grande différence absolue entre la distribution cumulée observée et théorique.

Paramètre du graphique

Si l'effectif théorique s'avère inférieur ou égal à 5, STATISTICA va ajouter l'effectif théorique pour la catégorie respective à la classe adjacente. Le fait de combiner les effectifs va se poursuivre jusqu'à atteindre un effectif cumulé théorique supérieur à 5.

Ajustement de distribution normale

Interprétation des résultats:

- Si le test de Kolmogorov-Smirnov est statistiquement significatif, l'hypothèse que les données observées suivent la distribution supposée est rejetée.
- Si le test du χ^2 est statistiquement significatif, l'hypothèse que les données observées suivent la distribution supposée est rejetée.

Parfois, les résultats de ces deux tests sont contradictoires:

Ce type de résultat n'est pas rare parce que le test de Kolmogorov-Smirnov n'est pas une procédure aussi précise qu'une technique de détection de grosses déviations par rapport à une distribution supposée. Fréquemment, la valeur du *Chi-deux* est une valeur grandement affectée par la façon par laquelle la distribution est "découpée", c'est-à-dire, par le nombre de catégories, les valeurs minimum, et maximum que vous avez choisi.

Ajustement de distribution normale

Interprétation des résultats:

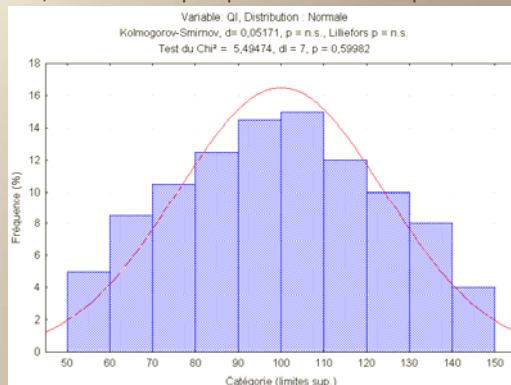
Ouvrez le fichier *exemple-01.sta*

Nous avons un échantillon de 200 personnes dont on a testé le QI.

La distribution s'ajuste-t-elle à une distribution normale?

Faites 10 catégories entre [50-150].

Ni K-S ni χ^2 ne sont significatifs, alors on accepte que cet échantillon peut être ajusté à une loi normale.



Ajustement de distribution normale

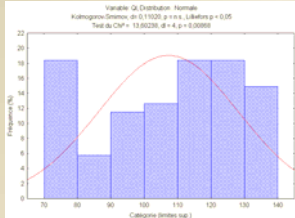
Exercice 1:
 Ouvrez le fichier **exercice-08.sta**
 Nous avons testé le QI d'un échantillon de 87 personnes, dont 39 femmes.
 Déterminer si l'échantillon suit une distribution normale. (catégoriser entre 70 et 140; en 7 catégories puis en 14 catégories)

Ajustement de distribution normale

Correction Exercice 1:

Variable: QI, Distribution : Normale (Exercice-08.sta)
 Kolmogorov-Smirnov, d = 0,11020, p = n.s., Lilliefors p < 0,05
 Chi-Deux = 13,60238, dl = 4, p = 0,00868

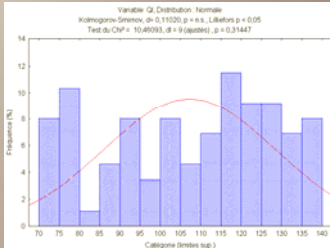
K-S est non significatif → la distribution est normale
 Chi² significatif → la distribution n'est pas normale



Variable: QI, Distribution : Normale
 Kolmogorov-Smirnov, d = 0,11020, p = n.s., Lilliefors p < 0,05
 Test du Chi² = 13,60238, dl = 4, p = 0,00868

Si vous faites 14 catégories au lieu de 7:

Variable: QI, Distribution : Normale (Exercice-08.sta)
 Kolmogorov-Smirnov, d = 0,11020, p = n.s., Lilliefors p < 0,05
 Chi-Deux = 10,46093, dl = 9 (ajustés), p = 0,31447



→ La distribution est normale

Ajustement de distribution normale

Exercice 2:

Ouvrez le fichier *irisdat.sta*

Ce fichier comporte des données rapportées par Fisher (1936) sur les longueurs et largeurs des sépales (Lonsepal, Larsepal) et des pétales (Lonpetal, Larpetal) pour 50 fleurs de trois types d'iris différents.

Les distributions des quatre variables décrivant les longueurs et largeurs des sépales et pétales vont à présent être examinées. Plus particulièrement, on s'attend à ce que ces mesures suivent la distribution normale.

Reportez-vous à 'Ajustement de Distributions – Exemple' dans l'aide pour la correction.

Ajustement de distribution-Acquis

Ce que vous devez savoir faire à l'issue de ce chapitre pour les ajustements de distribution:

- Différencier les distributions suivantes: loi normal, binomiale, Chi^2 , Poisson, Bernoulli
- Effectuer un test d'ajustement en gérant les paramètres de catégories (nombre, limites)
- Interpréter les résultats du K-S et Chi^2 en fonction des graphiques.