Séance 8 : Tester la significativité

Tests statistiques, loi normale et p-value

Mattéo Lanoë

Printemps 2026

De la description à l'explication

Étape 1 : Décrire les variables isolées

Les outils de tendance centrale (moyenne, médiane) et de dispersion (variance, écart-type) offrent une image statique des données.

De la description à l'explication

Étape 1 : Décrire les variables isolées
Les outils de tendance centrale (moyenne, médiane) et de dispersion (variance, écart-type) offrent une image statique des données.

Étape 2 : Croiser les variables
Pour analyser un lien explicatif, il faut confronter une variable dépendante et une variable indépendante selon leur nature :

De la description à l'explication

Étape 1 : Décrire les variables isolées
Les outils de tendance centrale (moyenne, médiane) et de dispersion (variance, écart-type) offrent une image statique des données.

Étape 2 : Croiser les variables
Pour analyser un lien explicatif, il faut confronter une variable dépendante et une variable indépendante selon leur nature :

Les trois grands croisements

Quali x Quali : Tableaux croisés.

De la description à l'explication

Étape 1 : Décrire les variables isolées
Les outils de tendance centrale (moyenne, médiane) et de dispersion (variance, écart-type) offrent une image statique des données.

Étape 2 : Croiser les variables
Pour analyser un lien explicatif, il faut confronter une variable dépendante et une variable indépendante selon leur nature :

Les trois grands croisements

Quali x Quali : Tableaux croisés.
Quali x Quanti : Comparaison de moyennes par sous-groupes.

De la description à l'explication

Étape 1 : Décrire les variables isolées
Les outils de tendance centrale (moyenne, médiane) et de dispersion (variance, écart-type) offrent une image statique des données.

Étape 2 : Croiser les variables
Pour analyser un lien explicatif, il faut confronter une variable dépendante et une variable indépendante selon leur nature :

Les trois grands croisements

Quali x Quali : Tableaux croisés.
Quali x Quanti : Comparaison de moyennes par sous-groupes.
Quanti x Quanti : Corrélation et nuage de points.

Tester la robustesse d'un résultat

Le cas de l'expérience d'Esther Duflo

Lors de l'étude sur l'impact des bourses scolaires, une différence a été observée entre le groupe traité et le groupe de contrôle.

Tester la robustesse d'un résultat

Le cas de l'expérience d'Esther Duflo

Lors de l'étude sur l'impact des bourses scolaires, une différence a été observée entre le groupe traité et le groupe de contrôle.

La marge d'erreur statistique

Cet outil permet de vérifier si l'effet causal observé est statistiquement significatif. Il indique si la différence est suffisamment grande pour ne pas être simplement due au hasard. Le calcul complet repose sur la notion d'intervalle de confiance.

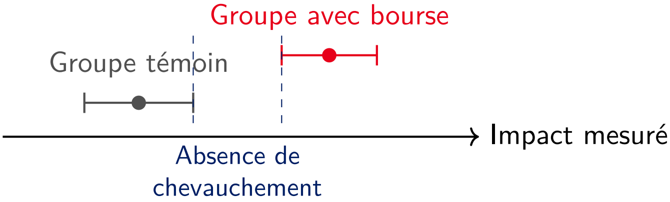
Tester la robustesse d'un résultat

Le cas de l'expérience d'Esther Duflo

Lors de l'étude sur l'impact des bourses scolaires, une différence a été observée entre le groupe traité et le groupe de contrôle.

La marge d'erreur statistique

Cet outil permet de vérifier si l'effet causal observé est statistiquement significatif. Il indique si la différence est suffisamment grande pour ne pas être simplement due au hasard. Le calcul complet repose sur la notion d'intervalle de confiance.



Un motif récurrent : la courbe en cloche

Pour construire ces marges d'erreur et ces intervalles de confiance, l'analyse s'appuie sur un modèle mathématique fondamental.

Un motif récurrent : la courbe en cloche

Pour construire ces marges d'erreur et ces intervalles de confiance, l'analyse s'appuie sur un modèle mathématique fondamental.

La distribution gaussienne

Dans la nature et dans la société, de nombreuses variables se distribuent spontanément selon un même schéma. La majorité des mesures se concentre autour de la moyenne, et les valeurs extrêmes deviennent de plus en plus rares.

Un motif récurrent : la courbe en cloche

Pour construire ces marges d'erreur et ces intervalles de confiance, l'analyse s'appuie sur un modèle mathématique fondamental.

La distribution gaussienne

Dans la nature et dans la société, de nombreuses variables se distribuent spontanément selon un même schéma. La majorité des mesures se concentre autour de la moyenne, et les valeurs extrêmes deviennent de plus en plus rares.

Exemples courants : La taille des individus, le poids, ou la répartition des notes à un examen.

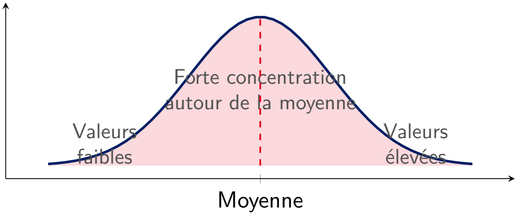
Un motif récurrent : la courbe en cloche

Pour construire ces marges d'erreur et ces intervalles de confiance, l'analyse s'appuie sur un modèle mathématique fondamental.

La distribution gaussienne

Dans la nature et dans la société, de nombreuses variables se distribuent spontanément selon un même schéma. La majorité des mesures se concentre autour de la moyenne, et les valeurs extrêmes deviennent de plus en plus rares.

Exemples courants : La taille des individus, le poids, ou la répartition des notes à un examen.



Le fondement théorique : La loi normale

Lorsqu'on extrait une infinité d'échantillons aléatoires d'une population pour en calculer la moyenne, ces moyennes se répartissent toujours selon une courbe géométrique stricte.

Le fondement théorique : La loi normale

Lorsqu'on extrait une infinité d'échantillons aléatoires d'une population pour en calculer la moyenne, ces moyennes se répartissent toujours selon une courbe géométrique stricte.

Le Théorème de la Limite Centrale

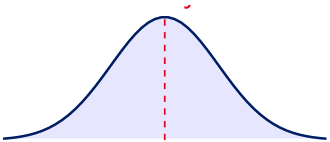
Peu importe la forme initiale des données, la distribution des moyennes d'échantillons forme une courbe en cloche (courbe de Gauss).

Le fondement théorique : La loi normale

Lorsqu'on extrait une infinité d'échantillons aléatoires d'une population pour en calculer la moyenne, ces moyennes se répartissent toujours selon une courbe géométrique stricte.

Le Théorème de la Limite Centrale

Peu importe la forme initiale des données, la distribution des moyennes d'échantillons forme une courbe en cloche (courbe de Gauss).

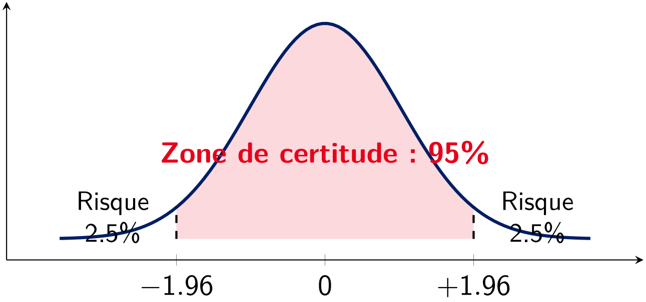


Mesurer l'incertitude sur la courbe

La géométrie de cette courbe permet de délimiter des zones de probabilité précises.
L'unité de mesure sur l'axe horizontal est l'erreur-type.

Mesurer l'incertitude sur la courbe

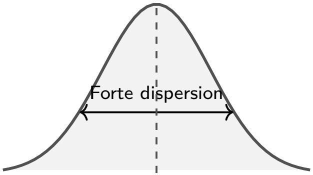
La géométrie de cette courbe permet de délimiter des zones de probabilité précises. L'unité de mesure sur l'axe horizontal est l'erreur-type.



Distinguer écart-type et erreur-type

L'écart-type (Individus)

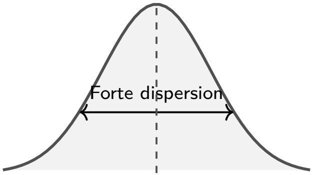
Il mesure la dispersion des **individus** au sein de l'échantillon. Il illustre l'éloignement des cas particuliers par rapport à la moyenne mesurée.



Distinguer écart-type et erreur-type

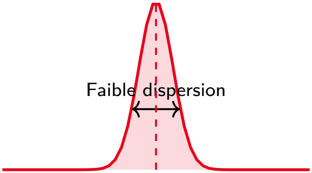
L'écart-type (Individus)

Il mesure la dispersion des **individus** au sein de l'échantillon. Il illustre l'éloignement des cas particuliers par rapport à la moyenne mesurée.



L'erreur-type (Moyennes)

Elle mesure la dispersion théorique des **moyennes** si l'on répétait l'enquête. Elle illustre le niveau d'incertitude de l'estimation statistique.



Les formules de calcul

Comment passe-t-on mathématiquement de la dispersion des individus à la dispersion des moyennes ?

L'écart-type (s)

Il se calcule à partir de la racine carrée de la variance (la moyenne des carrés des écarts à la moyenne).

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Les formules de calcul

Comment passe-t-on mathématiquement de la dispersion des individus à la dispersion des moyennes ?

L'écart-type (s)

Il se calcule à partir de la racine carrée de la variance (la moyenne des carrés des écarts à la moyenne).

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$\xrightarrow[\sqrt{n}]{\text{Divisé par}}$$

Les formules de calcul

Comment passe-t-on mathématiquement de la dispersion des individus à la dispersion des moyennes ?

L'écart-type (s)

Il se calcule à partir de la racine carrée de la variance (la moyenne des carrés des écarts à la moyenne).

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$\xrightarrow[\sqrt{n}]{\text{Divisé par}}$$

L'erreur-type (SE)

Elle se déduit directement de l'écart-type de l'échantillon, divisé par la racine carrée de sa taille (n).

$$SE = \frac{s}{\sqrt{n}}$$

Les formules de calcul

Comment passe-t-on mathématiquement de la dispersion des individus à la dispersion des moyennes ?

L'écart-type (s)

Il se calcule à partir de la racine carrée de la variance (la moyenne des carrés des écarts à la moyenne).

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

$$\xrightarrow[\sqrt{n}]{\text{Divisé par}}$$

L'erreur-type (SE)

Elle se déduit directement de l'écart-type de l'échantillon, divisé par la racine carrée de sa taille (n).

$$SE = \frac{s}{\sqrt{n}}$$

Conséquence mathématique : Si la taille de l'échantillon (n) augmente, le dénominateur grandit. L'erreur-type diminue donc mécaniquement, ce qui augmente la précision de l'enquête.

Que représente l'axe horizontal ?

Il faut adapter l'unité de mesure sur l'axe des abscisses en fonction de ce que la courbe représente.

1. La courbe des individus

Si la courbe montre la répartition des individus (ex : les tailles dans la population), l'unité est l'**écart-type**. Un score de 2 signifie qu'une personne se situe à 2 écarts-types au-dessus de la moyenne.

Que représente l'axe horizontal ?

Il faut adapter l'unité de mesure sur l'axe des abscisses en fonction de ce que la courbe représente.

1. La courbe des individus

Si la courbe montre la répartition des individus (ex : les tailles dans la population), l'unité est l'**écart-type**. Un score de 2 signifie qu'une personne se situe à 2 écarts-types au-dessus de la moyenne.

2. La courbe du théorème central limite

La courbe du théorème central limite montre la répartition des **moyennes** possibles. Son unité naturelle est donc l'**erreur-type**. Lorsqu'on teste notre échantillon empirique, on mesure sa distance au centre en erreurs-types.

Que représente l'axe horizontal ?

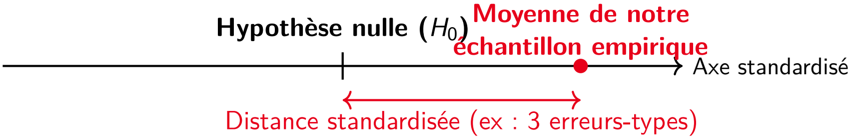
Il faut adapter l'unité de mesure sur l'axe des abscisses en fonction de ce que la courbe représente.

1. La courbe des individus

Si la courbe montre la répartition des individus (ex : les tailles dans la population), l'unité est l'**écart-type**. Un score de 2 signifie qu'une personne se situe à 2 écarts-types au-dessus de la moyenne.

2. La courbe du théorème central limite

La courbe du théorème central limite montre la répartition des **moyennes** possibles. Son unité naturelle est donc l'**erreur-type**. Lorsqu'on teste notre échantillon empirique, on mesure sa distance au centre en erreurs-types.

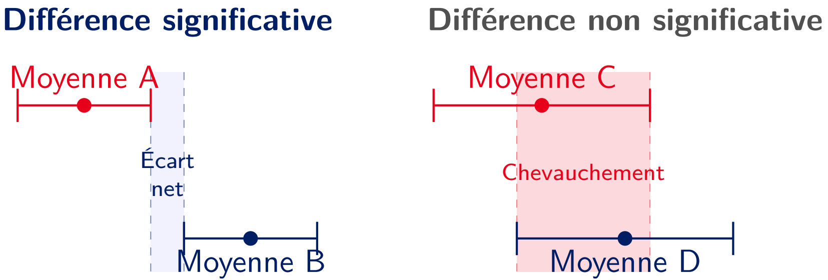


Intervalles de confiance et comparaison

L'intervalle de confiance applique cette théorie à un échantillon unique pour encadrer la vraie moyenne avec 95% de certitude.

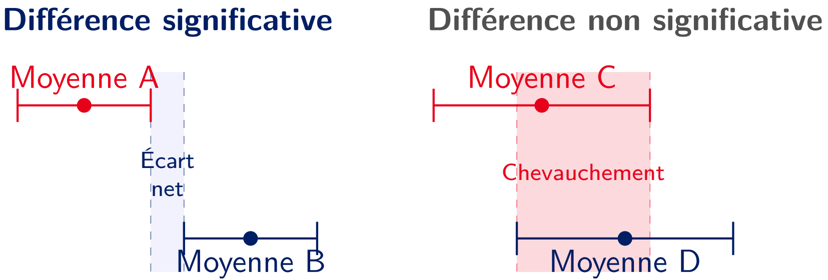
Intervalles de confiance et comparaison

L'intervalle de confiance applique cette théorie à un échantillon unique pour encadrer la vraie moyenne avec 95% de certitude.



Intervalles de confiance et comparaison

L'intervalle de confiance applique cette théorie à un échantillon unique pour encadrer la vraie moyenne avec 95% de certitude.



Lorsque la lecture visuelle des intervalles ne suffit pas pour trancher avec certitude, il faut recourir à un test statistique formel.

La logique de l'inférence : Les hypothèses

Un test statistique oppose systématiquement deux visions du monde. L'objectif est de mesurer la probabilité que les données observées soient le simple fruit du hasard.

La logique de l'inférence : Les hypothèses

Un test statistique oppose systématiquement deux visions du monde. L'objectif est de mesurer la probabilité que les données observées soient le simple fruit du hasard.

Hypothèse Nulle (H_0)
Il n'y a **aucune différence**.
Les variables sont indépendantes.
L'écart est dû au hasard.

Rejet de H_0 →

Hypothèse Alternative (H_1)
Il y a **une différence**.
Les variables sont liées.
L'effet est réel.

Prendre une décision : La valeur p

Le logiciel calcule un score de distance (ex : Z-score) entre la théorie (H_0) et l'échantillon. Ce score est traduit en probabilité : la **p-value**.

Prendre une décision : La valeur p

Le logiciel calcule un score de distance (ex : Z-score) entre la théorie (H_0) et l'échantillon. Ce score est traduit en probabilité : la **p-value**.

La règle de décision (Seuil de 5%)

Si la probabilité d'obtenir ce résultat par hasard est inférieure à 5% ($p < 0.05$), la théorie du hasard devient trop improbable.

Prendre une décision : La valeur p

Le logiciel calcule un score de distance (ex : Z-score) entre la théorie (H_0) et l'échantillon. Ce score est traduit en probabilité : la **p-value**.

La règle de décision (Seuil de 5%)

Si la probabilité d'obtenir ce résultat par hasard est inférieure à 5% ($p < 0.05$), la théorie du hasard devient trop improbable.

$\rightarrow$ **On rejette l'hypothèse nulle (H_0).** La différence est jugée statistiquement significative.

Un exemple de test : Comparaison de tailles

Afin d'illustrer la démarche d'un test statistique, prenons un cas pratique. L'objectif est de vérifier si la taille moyenne diffère entre deux populations.

Théorie (Hypothèse nulle)

Taille moyenne de la population de référence (Françaises) :

1,62 mètre

Un exemple de test : Comparaison de tailles

Afin d'illustrer la démarche d'un test statistique, prenons un cas pratique. L'objectif est de vérifier si la taille moyenne diffère entre deux populations.

Théorie (Hypothèse nulle)

Taille moyenne de la population de référence (Françaises) :

1,62 mètre

Réalité empirique (Échantillon)

Tirage aléatoire de 100 personnes en Scandinavie :

1,70 mètre

Un exemple de test : Comparaison de tailles

Afin d'illustrer la démarche d'un test statistique, prenons un cas pratique. L'objectif est de vérifier si la taille moyenne diffère entre deux populations.

Théorie (Hypothèse nulle)

Taille moyenne de la population de référence (Françaises) :

1,62 mètre

Réalité empirique (Échantillon)

Tirage aléatoire de 100 personnes en Scandinavie :

1,70 mètre

Question : Cet écart de 8 centimètres est-il dû au hasard de l'échantillonnage, ou traduit-il une véritable différence structurelle ?

Le calcul de la distance (Score Z)

Pour comparer cet écart aux lois du hasard, il faut le convertir en une distance standardisée, appelée le score Z.

- **1. Mesurer l'écart brut**
  Écart constaté $= 1,70 - 1,62 =$ **0,08 mètre**.

Le calcul de la distance (Score Z)

Pour comparer cet écart aux lois du hasard, il faut le convertir en une distance standardisée, appelée le score Z.

- **1. Mesurer l'écart brut**
  Écart constaté $= 1,70 - 1,62 =$ **0,08 mètre**.

- **2. Connaître l'erreur-type**
  Pour cet échantillon, l'erreur-type calculée est de **0,006**.

Le calcul de la distance (Score Z)

Pour comparer cet écart aux lois du hasard, il faut le convertir en une distance standardisée, appelée le score Z.

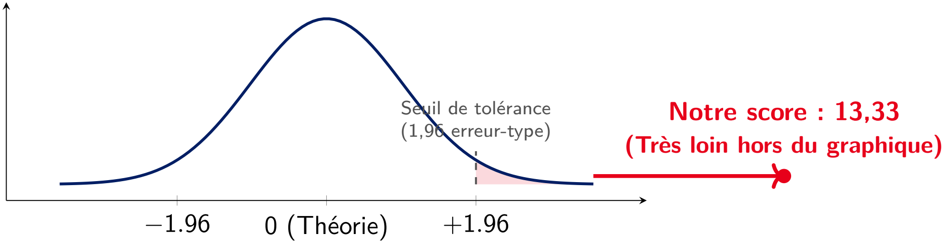
- **1. Mesurer l'écart brut**
  Écart constaté $= 1,70 - 1,62 =$ **0,08 mètre**.
- **2. Connaître l'erreur-type**
  Pour cet échantillon, l'erreur-type calculée est de **0,006**.
- **3. Calculer le score Z**
  Il s'agit de diviser l'écart brut par l'erreur-type pour savoir combien de "pas" ont été faits par rapport au centre.
  $ScoreZ = \frac{0{,}08}{0{,}006} = \mathbf{13{,}33}$

Où se situe notre échantillon ?

Plaçons ce résultat sur la courbe de distribution des moyennes, qui illustre ce que le hasard est capable de produire.

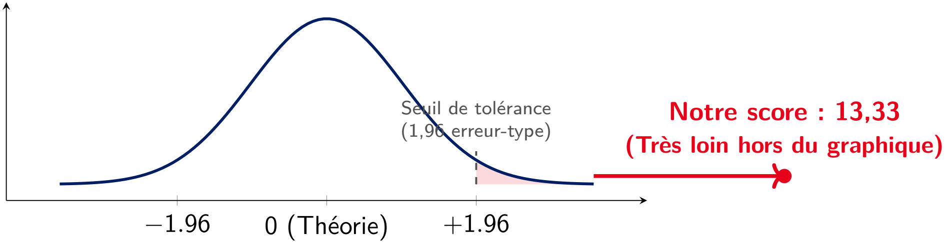
Où se situe notre échantillon ?

Plaçons ce résultat sur la courbe de distribution des moyennes, qui illustre ce que le hasard est capable de produire.



Où se situe notre échantillon ?

Plaçons ce résultat sur la courbe de distribution des moyennes, qui illustre ce que le hasard est capable de produire.



Le score de 13,33 se trouve massivement au-delà du seuil de 1,96. La distance franchie par notre échantillon est extrême.

La décision statistique

La position de l'échantillon sur la courbe dicte la décision finale de l'analyse.

La probabilité (Valeur p)

La probabilité d'obtenir un écart aussi grand par le simple fait du hasard est infinitésimale.

La décision statistique

La position de l'échantillon sur la courbe dicte la décision finale de l'analyse.

La probabilité (Valeur p)

La probabilité d'obtenir un écart aussi grand par le simple fait du hasard est infinitésimale.

La **valeur p** tombe donc à une chance sur plusieurs milliers (proche de 0).

La décision statistique

La position de l'échantillon sur la courbe dicte la décision finale de l'analyse.

La probabilité (Valeur p)

La probabilité d'obtenir un écart aussi grand par le simple fait du hasard est infinitésimale.

La **valeur p** tombe donc à une chance sur plusieurs milliers (proche de 0).

Conclusion

Puisque la valeur p est strictement inférieure à notre seuil de 5% (0,05) :

- On **rejette** l'hypothèse nulle.
- La différence observée est **statistiquement significative**.

Le test de Student (Comparaison de moyennes)

Ce test est très utilisé en sciences sociales pour évaluer l'inégalité entre deux moyennes (croisement d'une variable qualitative et d'une variable quantitative).

Le test de Student (Comparaison de moyennes)

Ce test est très utilisé en sciences sociales pour évaluer l'inégalité entre deux moyennes (croisement d'une variable qualitative et d'une variable quantitative).

L'objectif du test

Par un ensemble de calculs, le test évalue la différence brute entre les deux moyennes observées en la rapportant à l'erreur-type globale. Il produit une probabilité (la valeur p).

Le test de Student (Comparaison de moyennes)

Ce test est très utilisé en sciences sociales pour évaluer l'inégalité entre deux moyennes (croisement d'une variable qualitative et d'une variable quantitative).

L'objectif du test

Par un ensemble de calculs, le test évalue la différence brute entre les deux moyennes observées en la rapportant à l'erreur-type globale. Il produit une probabilité (la valeur p).

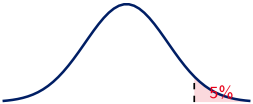
Cette valeur p permet de décider si l'on peut rejeter l'hypothèse nulle. Si la valeur p franchit le seuil d'exigence (généralement 5%), la différence entre les deux moyennes est jugée statistiquement significative.

Test de Student : Répartir l'incertitude

Selon la question de recherche, la zone d'incertitude de 5% n'est pas placée au même endroit sur la distribution de référence.

Le test unilatéral (1 queue)

On teste une différence dans une seule direction. L'incertitude de 5% est concentrée d'un seul côté.

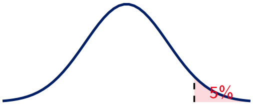


Test de Student : Répartir l'incertitude

Selon la question de recherche, la zone d'incertitude de 5% n'est pas placée au même endroit sur la distribution de référence.

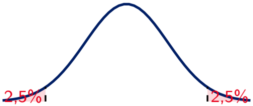
Le test unilatéral (1 queue)

On teste une différence dans une seule direction. L'incertitude de 5% est concentrée d'un seul côté.



Le test bilatéral (2 queues)

On teste une différence dans les deux sens. Le test est plus exigeant : l'incertitude est divisée en deux (2,5% à chaque extrémité).



Le test du Khi-deux (Tableaux croisés)

Pour analyser le croisement de deux variables qualitatives, la lecture des proportions d'un tableau permet d'avoir une première intuition sur leur dépendance.

Le test du Khi-deux (Tableaux croisés)

Pour analyser le croisement de deux variables qualitatives, la lecture des proportions d'un tableau permet d'avoir une première intuition sur leur dépendance.

La nécessité d'un test rigoureux

Toutefois, une simple lecture descriptive ne suffit pas pour affirmer que les variables sont liées. Il faut s'assurer que les écarts de proportions observés ne sont pas le simple fruit de fluctuations aléatoires.

Le test du Khi-deux (Tableaux croisés)

Pour analyser le croisement de deux variables qualitatives, la lecture des proportions d'un tableau permet d'avoir une première intuition sur leur dépendance.

La nécessité d'un test rigoureux

Toutefois, une simple lecture descriptive ne suffit pas pour affirmer que les variables sont liées. Il faut s'assurer que les écarts de proportions observés ne sont pas le simple fruit de fluctuations aléatoires.

C'est ici qu'intervient le test du Khi-deux, dont la logique repose sur la confrontation directe entre la réalité observée et une situation purement théorique.

La mécanique du Khi-deux : Mesurer la distance

Pour vérifier l'indépendance de deux variables qualitatives, le logiciel informatique construit un tableau artificiel de référence.

- **Le tableau d'indépendance parfaite :** Il représente une situation théorique où les variables n'auraient strictement aucun lien.

La mécanique du Khi-deux : Mesurer la distance

Pour vérifier l'indépendance de deux variables qualitatives, le logiciel informatique construit un tableau artificiel de référence.

- **Le tableau d'indépendance parfaite :** Il représente une situation théorique où les variables n'auraient strictement aucun lien.
- **L'étape de comparaison :** À partir de ce tableau parfait, on calcule la distance avec les vrais chiffres de notre enquête empirique.

La mécanique du Khi-deux : Mesurer la distance

Pour vérifier l'indépendance de deux variables qualitatives, le logiciel informatique construit un tableau artificiel de référence.

- **Le tableau d'indépendance parfaite :** Il représente une situation théorique où les variables n'auraient strictement aucun lien.
- **L'étape de comparaison :** À partir de ce tableau parfait, on calcule la distance avec les vrais chiffres de notre enquête empirique.

La mécanique du Khi-deux : Mesurer la distance

Pour vérifier l'indépendance de deux variables qualitatives, le logiciel informatique construit un tableau artificiel de référence.

- **Le tableau d'indépendance parfaite :** Il représente une situation théorique où les variables n'auraient strictement aucun lien.
- **L'étape de comparaison :** À partir de ce tableau parfait, on calcule la distance avec les vrais chiffres de notre enquête empirique.

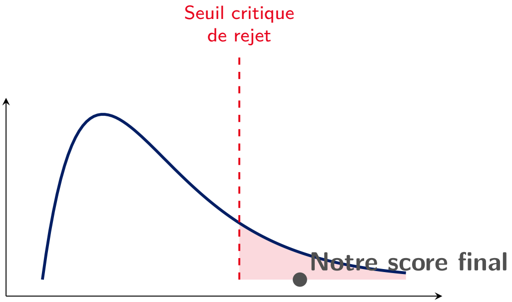
Le calcul du score global

Pour chaque case du tableau, on observe l'écart entre la réalité et la théorie. On met cet écart au carré (pour éviter les annulations), puis on le divise par le nombre théorique attendu.

$$\frac{(\text{Effectif observé} - \text{Effectif théorique})^2}{\text{Effectif théorique}}$$

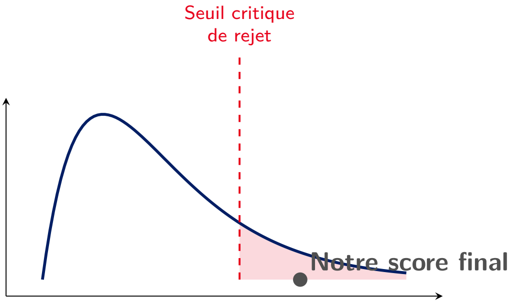
La loi du Khi-deux et la décision statistique

Si l'on reproduisait des tirages aléatoires une infinité de fois, l'accumulation de ces scores d'écart formerait une courbe asymétrique : la loi du Khi-deux.



La loi du Khi-deux et la décision statistique

Si l'on reproduisait des tirages aléatoires une infinité de fois, l'accumulation de ces scores d'écart formerait une courbe asymétrique : la loi du Khi-deux.



L'interprétation

Le score global obtenu par l'addition de toutes les cases est placé en abscisse sur cette courbe.

S'il dépasse le seuil critique, la distance avec la théorie est trop grande. On rejette alors l'hypothèse nulle d'indépendance (exemple : on prouve que le type de logement influence bien la possession d'un chien ou d'un chat).

Application : Type de logement et animal de compagnie

Nous allons utiliser un exemple pour illustrer le mécanisme du test du Khi-deux. Imaginons une enquête auprès de 100 ménages pour savoir si leur type de logement influence de manière significative la possession d'un animal de compagnie.

Application : Type de logement et animal de compagnie

Nous allons utiliser un exemple pour illustrer le mécanisme du test du Khi-deux. Imaginons une enquête auprès de 100 ménages pour savoir si leur type de logement influence de manière significative la possession d'un animal de compagnie.

1. Distribution réelle (Observée)

Ménages	**Chien**	**Chat**	**Total**
Maison	**30**	**10**	40
Appartement	**20**	**40**	60
Total	50	50	100

Application : Type de logement et animal de compagnie

Nous allons utiliser un exemple pour illustrer le mécanisme du test du Khi-deux. Imaginons une enquête auprès de 100 ménages pour savoir si leur type de logement influence de manière significative la possession d'un animal de compagnie.

1. Distribution réelle (Observée)

Ménages	**Chien**	**Chat**	**Total**
Maison	**30**	**10**	40
Appartement	**20**	**40**	60
Total	50	50	100

Nous avons interrogé 100 ménages, répartis équitablement entre 50 propriétaires de chiens et 50 propriétaires de chats. Le type de logement est également connu pour chacun d'eux (40 en maison, 60 en appartement). Ce sont nos données de terrain.

Khi-deux : Construire la table d'indépendance

Pour effectuer le test statistique formel, nous devons comparer nos données observées à une situation théorique d'indépendance parfaite, où les variables n'auraient strictement aucun lien.

Khi-deux : Construire la table d'indépendance

Pour effectuer le test statistique formel, nous devons comparer nos données observées à une situation théorique d'indépendance parfaite, où les variables n'auraient strictement aucun lien.

2. Distribution théorique (Indépendance)

Ménages	**Chien**	**Chat**	**Total**
Maison	20	20	40
Appartement	30	30	60
Total	50	50	100

Khi-deux : Construire la table d'indépendance

Pour effectuer le test statistique formel, nous devons comparer nos données observées à une situation théorique d'indépendance parfaite, où les variables n'auraient strictement aucun lien.

2. Distribution théorique (Indépendance)

Ménages	**Chien**	**Chat**	**Total**
Maison	20	20	40
Appartement	30	30	60
Total	50	50	100

Le principe de calcul (Exemple de la cellule Maison x Chien)

On multiplie le total de sa ligne par le total de sa colonne, puis on divise par le grand total (ici, (40×50) / $100 = 20$).

Khi-deux : Le calcul du score d'écart

Pour obtenir le score global, on calcule l'écart pour chacune des quatre cases du tableau, puis on additionne le tout.

$$\chi^2 = \sum \frac{(\text{Obs} - \text{Thé})^2}{\text{Thé}}$$

Khi-deux : Le calcul du score d'écart

Pour obtenir le score global, on calcule l'écart pour chacune des quatre cases du tableau, puis on additionne le tout.

$$\chi^2 = \sum \frac{(\text{Obs} - \text{Thé})^2}{\text{Thé}}$$

$$\chi^2 = \frac{(30-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(20-30)^2}{30} + \frac{(40-30)^2}{30}$$

Khi-deux : Le calcul du score d'écart

Pour obtenir le score global, on calcule l'écart pour chacune des quatre cases du tableau, puis on additionne le tout.

$$\chi^2 = \sum \frac{(\text{Obs} - \text{Thé})^2}{\text{Thé}}$$

$$\chi^2 = \frac{(30-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(20-30)^2}{30} + \frac{(40-30)^2}{30}$$

$$\chi^2 = 5 + 5 + 3,33 + 3,33$$

Khi-deux : Le calcul du score d'écart

Pour obtenir le score global, on calcule l'écart pour chacune des quatre cases du tableau, puis on additionne le tout.

$$\chi^2 = \sum \frac{(\text{Obs} - \text{Thé})^2}{\text{Thé}}$$

$$\chi^2 = \frac{(30-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(20-30)^2}{30} + \frac{(40-30)^2}{30}$$

$$\chi^2 = 5 + 5 + 3,33 + 3,33$$

Score final obtenu

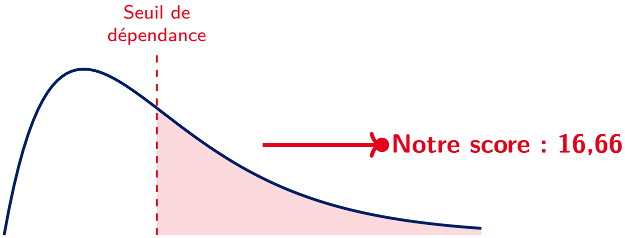
$$\chi^2 \approx 16,66$$

Khi-deux : La décision statistique

Le score final doit être replacé sur la distribution théorique du hasard pour prendre une décision objective.

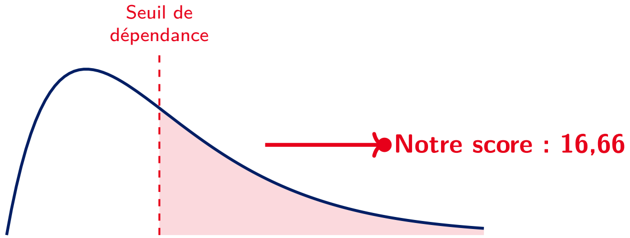
Khi-deux : La décision statistique

Le score final doit être replacé sur la distribution théorique du hasard pour prendre une décision objective.



Khi-deux : La décision statistique

Le score final doit être replacé sur la distribution théorique du hasard pour prendre une décision objective.



Conclusion

Le score (16,66) dépasse très largement le seuil critique de la courbe. La probabilité que cet écart soit dû au hasard est quasi nulle. On rejette l'hypothèse nulle : le type de logement influence bien le type d'animal possédé.

Traduire le score en décision : la valeur p

Le score de distance que nous avons calculé (16,66) est converti par les logiciels statistiques en une probabilité : la **valeur p** (ou *p-value*).

Traduire le score en décision : la valeur p

Le score de distance que nous avons calculé (16,66) est converti par les logiciels statistiques en une probabilité : la **valeur p** (ou *p-value*).

La règle de décision (Seuil de 5%)

En sciences sociales, on fixe généralement le seuil de tolérance au hasard à 5% (soit $0,05$).

Traduire le score en décision : la valeur p

Le score de distance que nous avons calculé (16,66) est converti par les logiciels statistiques en une probabilité : la **valeur p** (ou *p-value*).

La règle de décision (Seuil de 5%)

En sciences sociales, on fixe généralement le seuil de tolérance au hasard à 5% (soit $0,05$).

Si la probabilité d'obtenir les données de notre enquête par hasard est **inférieure à 5% ($p < 0,05$)**, on estime que la théorie du hasard devient beaucoup trop improbable.

Traduire le score en décision : la valeur p

Le score de distance que nous avons calculé (16,66) est converti par les logiciels statistiques en une probabilité : la **valeur p** (ou *p-value*).

La règle de décision (Seuil de 5%)

En sciences sociales, on fixe généralement le seuil de tolérance au hasard à 5% (soit $0,05$).

Si la probabilité d'obtenir les données de notre enquête par hasard est **inférieure à 5% ($p < 0,05$)**, on estime que la théorie du hasard devient beaucoup trop improbable.

Le rejet de l'hypothèse nulle (H_0)

Dans notre exemple, un score de 16,66 correspond à une probabilité (valeur p) très proche de 0, ce qui est largement inférieur à 0,05.

Par conséquent, **on rejette** H_0. On conclut qu'il n'y a pas d'indépendance totale entre les deux variables : elles sont liées et ce lien est statistiquement significatif.