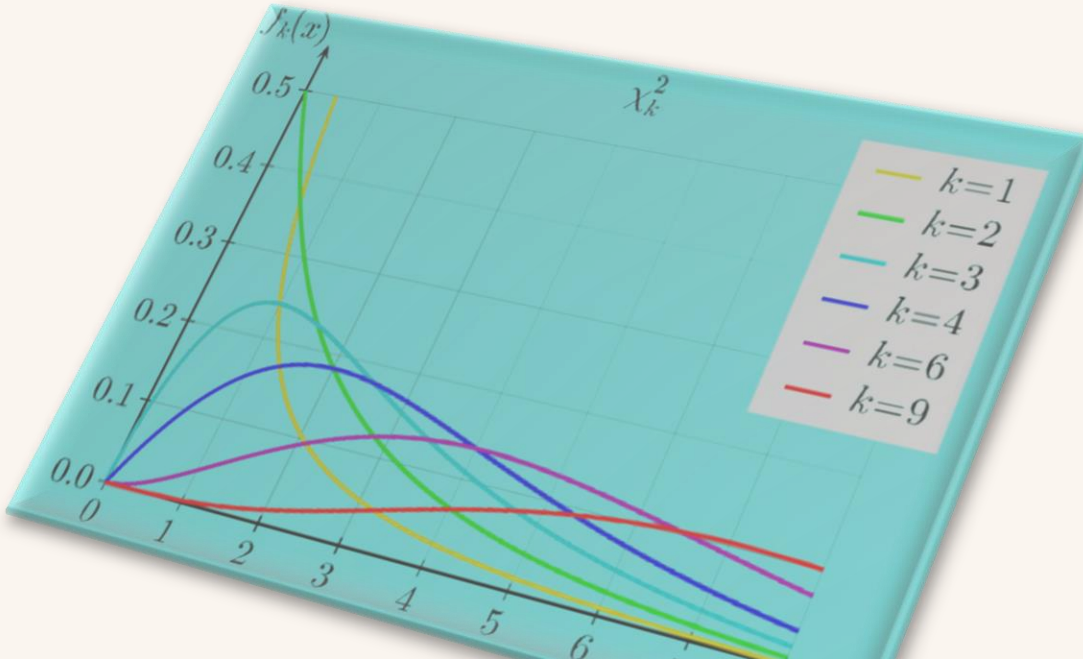
Statistique Inférentielle
Chapitre 22 : Test d'hypothèse

Concepts

- définition d'un test
- scénario : hypothèses, calcul, conclusion
- paramètres
- test Z normal
- test de Student
- test du χ^2 d'ajustement
- test du χ^2 d'indépendance



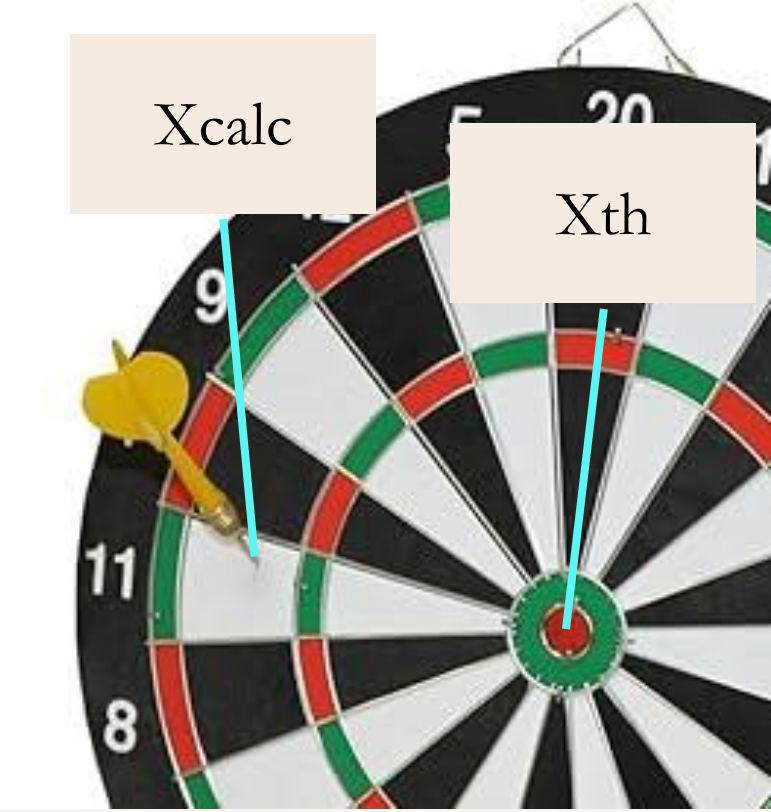
Définitions

Pourquoi faire un test statistique ?

- Dans le cas de phénomènes complexes ou aléatoires.
- La nature brouille ses messages.
- L'interprétation est difficile.

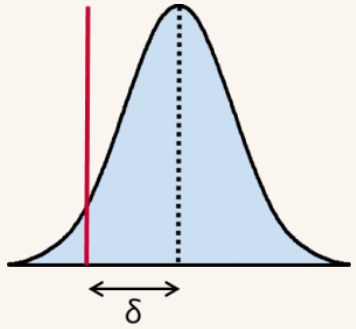
Qu'est-ce qu'un test statistique ?

- Sur base de données expérimentales aléatoires dont on tire un paramètre, le test donne une réponse à une question concernant ce paramètre.
  - paramètre : moyenne, variance, proportion, …
  - Xcalc
- Cette question porte sur la comparaison à une valeur de référence du paramètre de la population concernée.
  - Xth
- Xcalc est-il égal (ou inférieur ou supérieur) à Xth ?

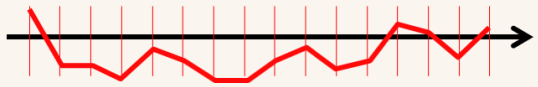


Test paramétrique vs non-paramétrique

- Test paramétrique
  - On suppose que les données suivent une loi de distribution connue (ex: Gauss).
  - Test fin et précis.
  - Il faut être sûr que les données suivent cette loi !
  - Attention aux valeurs aberrantes (extrêmes).



- Test non paramétrique
  - On ne suppose pas que les données suivent une loi de distribution connue.
  - Test basé uniquement sur les propriétés numériques des échantillons.
  - Peu sensible aux valeurs aberrantes.



Exemples

- Est-ce que le processus de remplissage des bidons d'huile de moteur est bien réglé ?
- Est-ce que la nouvelle variété de maïs est plus productive que l'ancienne ?
- Une entreprise lance une campagne publicitaire pour promouvoir un produit et souhaite savoir si cette campagne a un impact significatif sur les ventes.
  - public visé par la campagne vs public non visé par la campagne

Exemple en marketing

- Une entreprise place des publicités sur les réseaux sociaux.
- Question : Le taux de clics moyen est-il genré (càd différent selon que l'utilisateur est un homme ou une femme) ?
- Si la réponse est "oui"
  - l'entreprise doit différencier la cible « femme » de la cible « homme » dans ses publicités digitales.
- Si la réponse est "non"
  - l'entreprise ne doit pas tenir compte du genre dans ses publicités digitales.

Scénario d'un test

ingrédients

recette

exemple

Analogie

- Similarité avec le tribunal de justice
  - L'accusé est supposé non-coupable (présomption d'innocence) = HO.
  - On tente de prouver que l'accusé est coupable = HA.
  - On rassemble les preuves contre HO (Xcalc).
  - On fixe un nombre de preuves jugé suffisant (Xth).
  - Tant qu'on ne rassemble pas assez de preuves (Xcalc < Xth), l'accusé reste "non-coupable".
  - Au final, soit HO soit HA l'emportera.
- Attention ! Si HO est acceptée
  - Alors c'est sans certitude absolue
  - Mais avec une forte présomption mesurée par une probabilité

Conditions de validation d'un test

- La question posée a deux réponses possibles.
  - oui ou non
- Être en possession d'un échantillon significatif.
  - nombre suffisamment important d'individus
  - (sans valeurs extrêmes)
- Ces variables aléatoires suivent une loi de probabilité connue.
  - loi normale (Laplace-Gauss)
- La question posée doit concerner un ou plusieurs paramètres de cette loi.

Scénario d'un test

1. On pose une hypothèse (HO).
2. A partir des données collectées, on tente de prouver que HO n'est pas vérifiée.
   - Un calcul est effectué, qui donne un résultat Xcalc.
3. On a besoin d'un seuil limite de rejet de HO
   - Un calcul est effectué, qui donne un résultat Xth.
4. Deux possibilités pour Xcalc :
   - soit Xcalc < Xth : Xcalc n'est pas en contradiction avec HO
     - on accepte HO
   - soit Xcalc > Xth : Xcalc s'écarte (trop) de HO
     - on rejette HO

"Ingrédients" d'un test

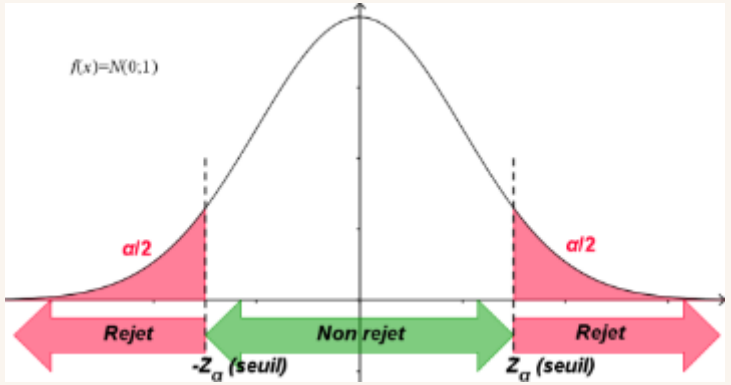
- Données
  - la réalisation d'un certain nombre de variables aléatoires
  - avec ses paramètres
- Modèle statistique
  - la loi de probabilité dépendant d'un paramètre (ou plusieurs)
- Hypothèse
  - une assertion concernant ce paramètre, appelée hypothèse nulle (HO)
- Règle de décision
  - A partir des données, calcul d'une statistique de test Xcalc.
  - A partir du modèle, calcul d'un seuil Xth.
- Conclusion
  - Si Xcalc > Xth alors on rejette HO, sinon on conserve HO

Scénario d'un test paramétrique : Exemple

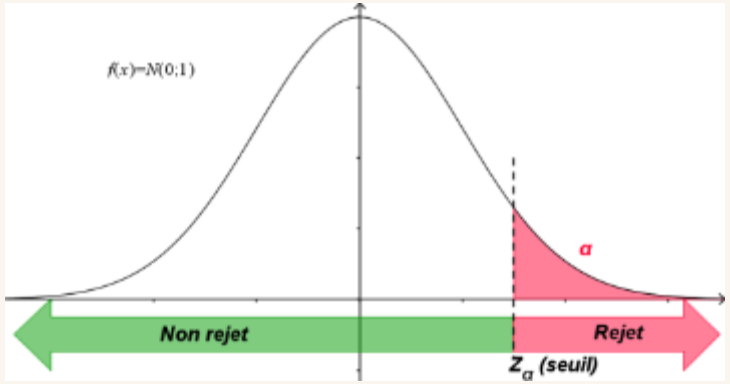
- Données
  - échantillon avec un ensemble de tailles d'individus de sexe masculin
- Modèle statistique
  - loi normale d'espérance M = 182 cm et d'écart-type E = 7 cm.
- Hypothèse
  - HO : L'échantillon est bien représentatif de la population masculine.
    - La taille moyenne de l'échantillon est "proche" de 182 cm.
- Règle de décision
  - Un calcul sur l'échantillon nous donne sa moyenne : 185 cm.
  - On en déduit Xcalc = | 185 - 182 | = 3 cm
  - Le modèle statistique nous définit Xth = 6 cm
  - Il y a rejet de HO si Xcalc > Xth, ou acceptation de HO si Xcalc < Xth
- Conclusion
  - Ici Xcalc < Xth => on conserve HO
  - L'échantillon est considéré valide.

Natures de HO

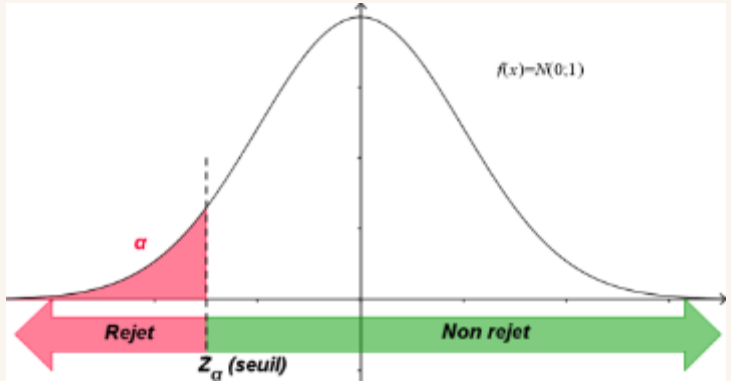
- On note
  - X_{th} : la valeur théorique (modèle)
  - X_{calc} : la valeur calculée (éch.)
- Test bilatéral
  - HO : $\{X_{calc} = X_{th}\}$
  - HA : $\{X_{calc} \neq X_{th}\}$
    - $X_{calc} > X_{th}$ ou $X_{calc} < X_{th}$



- Test unilatéral droit
  - HO : $\{X_{calc} \leq X_{th}\}$
  - HA : $\{X_{calc} > X_{th}\}$



- Test unilatéral gauche
  - HO : $\{X_{calc} \geq X_{th}\}$
  - HA : $\{X_{calc} < X_{th}\}$



Exemple de test paramétrique



- Considérons une population en surpoids, dont on connait la loi de répartition du poids des individus.
- On sélectionne un échantillon d'individus représentatifs (cf exemple précédent). On note le poids de ces individus.
- On teste un traitement amincissant sur cet échantillon.
- Après le traitement, on mesure à nouveau le poids des individus et on calcule l'écart avec les mesures initiales.
- Si cet écart est plausible sans traitement, alors le traitement est inefficace.
- Si cet écart est improbable, alors le traitement est efficace.

Risque et puissance d'un test

risque

niveau

puissance

erreur de première espèce

erreur de deuxième espèce

p-valeur

intensité

coefficient de Cramer

Risque d'un test

		Décision (suite au test)	
		HO	HA
Réalité (inconnue)	HO	décision correcte	erreur (1ère espèce) ou α
	HA	erreur (2ème espèce) ou β	décision correcte

Erreur de 1ère espèce
On considère HA (et on rejette HO)
alors qu'en réalité HO est vérifiée
Similarité avec le "faux positif" en diagnostic médical

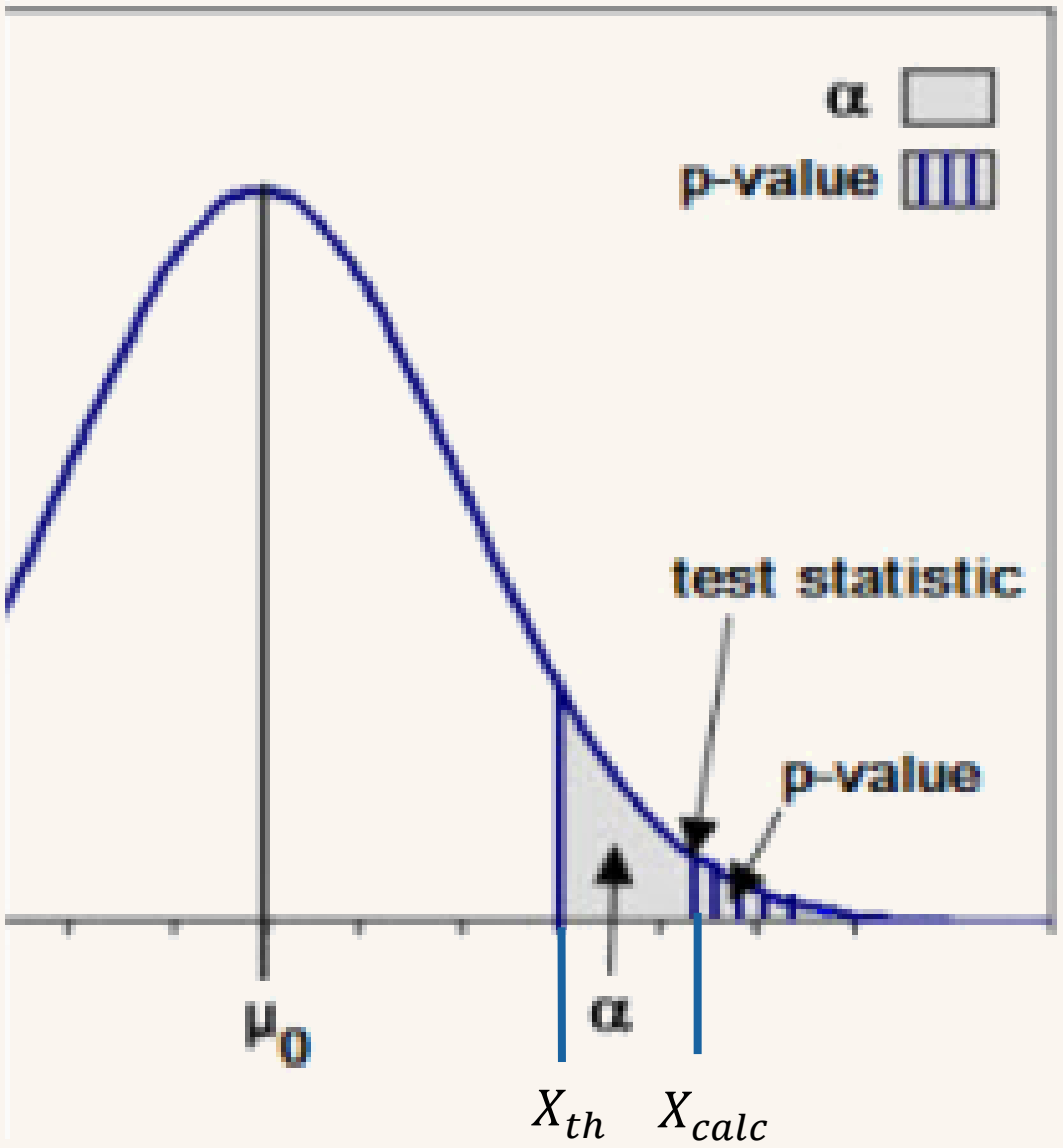
Erreur de 2ème espèce
On considère HO (et on rejette HA)
alors qu'en fait HA est vérifiée
"faux négatif"

Niveau et puissance d'un test

- Niveau α
  - Lié à l'erreur de 1ère espèce
  - Risque (probabilité) de rejeter HO alors que HO est vrai
  - Valeur la plus élevée du risque pris, par ex. 5%
- Niveau β
  - Lié à l'erreur de 2ème espèce
  - Risque (probabilité) d'accepter HO alors que HO est fausse
  - Non considéré dans ce cours
- Puissance π
  - Probabilité de rejeter HO alors que HO est fausse
  - $\pi = 1 - \beta$

p-value

- α : probabilité acceptée de commettre une erreur de première espèce
  - "rejetter HO alors qu'en réalité elle est vérifiée"
- *p-value* : probabilité d'observer un résultat plus extrême
- *p-value* est à X_{calc}
  ce que α est à X_{th}
- $X_{calc} > X_{th} \equiv p\text{-}value < \alpha$
- plus X_{calc} est élevé,
  plus *p-value* est faible.
- usage
  - *p-value* < 0,01 :
    très forte présomption contre HO
  - 0,01 < *p-value* < 0,05 :
    forte présomption contre HO
  - 0,05 < *p-value* < 0,10 :
    faible présomption contre HO



Test Z

Test Z bilatéral sur l'espérance, avec σ connu

Test Z unilatéral à droite sur l'espérance, avec σ connu

Modèle statistique utilisé

Les variables $X_1, X_2, \ldots X_n$ sont i.i.d., $\mathcal{N}(\mu, \sigma^2)$,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \mathcal{T}_{n-1}$$

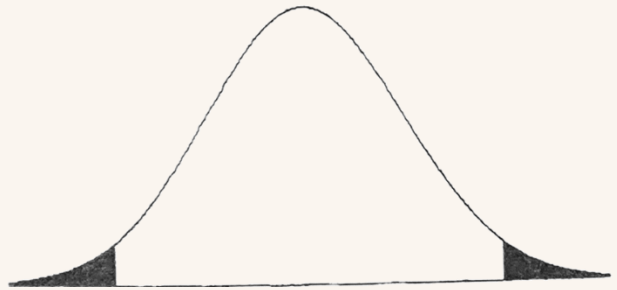
Utilité

- Comparaison d'une valeur à une loi normale
  - quand la variance est connue
- Comparaison (des moyennes) de deux échantillons
  - issues de deux lois normales
  - quand leurs variances sont égales et connues
- Echantillon de grande taille
  - $n > 30$

Conditions

- Pour toutes les comparaisons
  - Les données sont continues.
  - La distribution des données est normale.
  - Les échantillons ont été prélevés de manière aléatoire.
  - L'espérance μ_0 est connue.
  - L'écart-type σ est connu.
- En plus, dans le cas de la comparaison de deux échantillons
  - Les deux échantillons doivent être indépendants.
  - Il y a homogénéité de la variance entre les populations.

Principe du test Z bilatéral



- On veut tester $HO = \{ \mu = \mu_0 \}$ (test bilatéral)
  - contre $HA = \{ \mu \neq \mu_0 \}$
- Test de niveau α : Rejeter HO si $Z_{calc} > Z_{th}$
  - $Z_{calc} = \frac{|\bar{X} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}$
  - $Z_{th} = Z_{1-\alpha/2}$ = quantile $1 - \propto/2$ de la loi normale centrée réduite
- Explication
  - rejet si $|\bar{X} - \mu_0| > l$
    - avec l tel que le niveau du test égale α
    - $l = \frac{\sigma}{\sqrt{n}} Z_{1-\propto/2}$
  - test bilatéral, donc il y a deux évènements de même prob. $\alpha/2$:

$$\bar{X} - \mu_0 > l \quad \text{et} \quad \bar{X} - \mu_0 < -l$$

Tableur
LOI.NORMALE.INVERSE.N

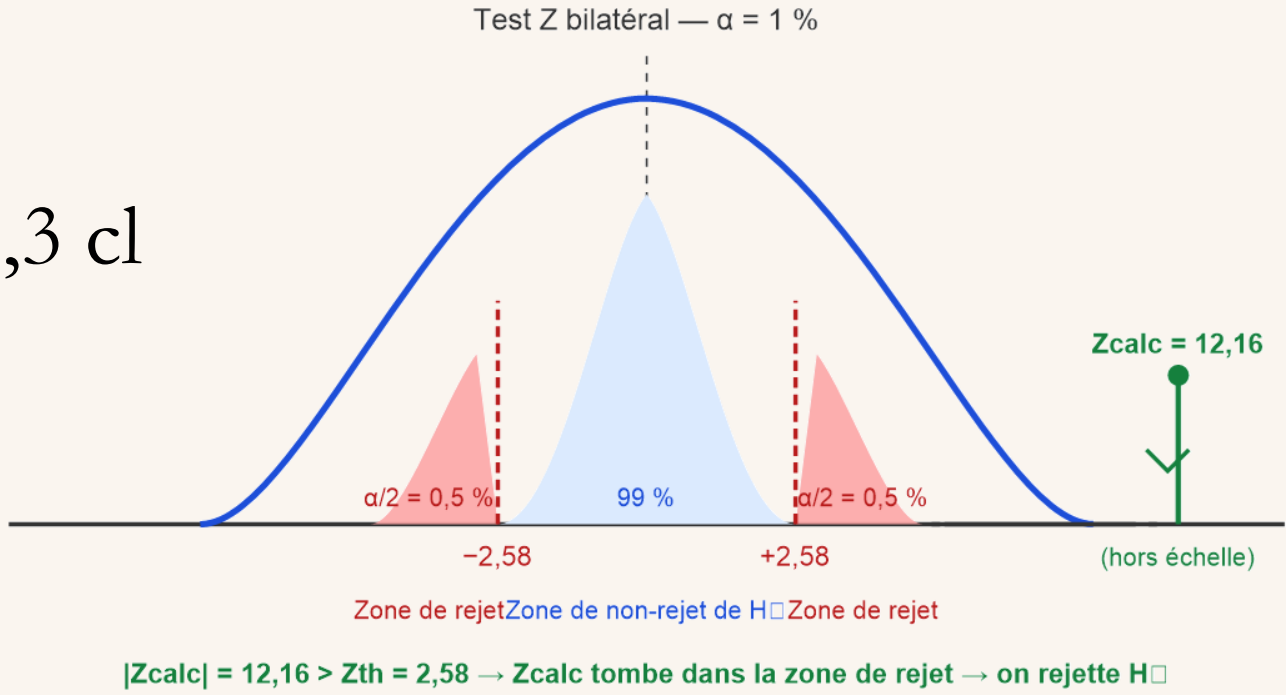
Exemple test Z bilatéral (cf complément)

- Une société pharmaceutique produit des flacons médicaux dont la contenance est supposée être contractuellement de 100 cl avec un écart-type 1 cl.
- Un échantillon de 85 flacons est mesuré. Cf tableau.
- HO : Le processus de production est correct. Les effectifs observés devraient tourner autour de la valeur espérée de 100 cl.
  - HO = { $\mu = 100$ cl }
- HA : Le processus de production est déréglé. Les effectifs s'écartent trop de la valeur espérée.
  - HA = { $\mu \neq 100$ cl }

102,2	100,5	103,1	100,5	101,7
99,5	99,9	102,1	103,4	100,8
101,4	98,8	99,2	102,8	102,2
102,3	101,2	103,2	99,7	101,0
103,1	100,4	100,9	103,2	103,5
102,5	102,2	101,5	102,3	102,3
100,3	100,9	102,0	101,0	…..

Exemple test Z bilatéral

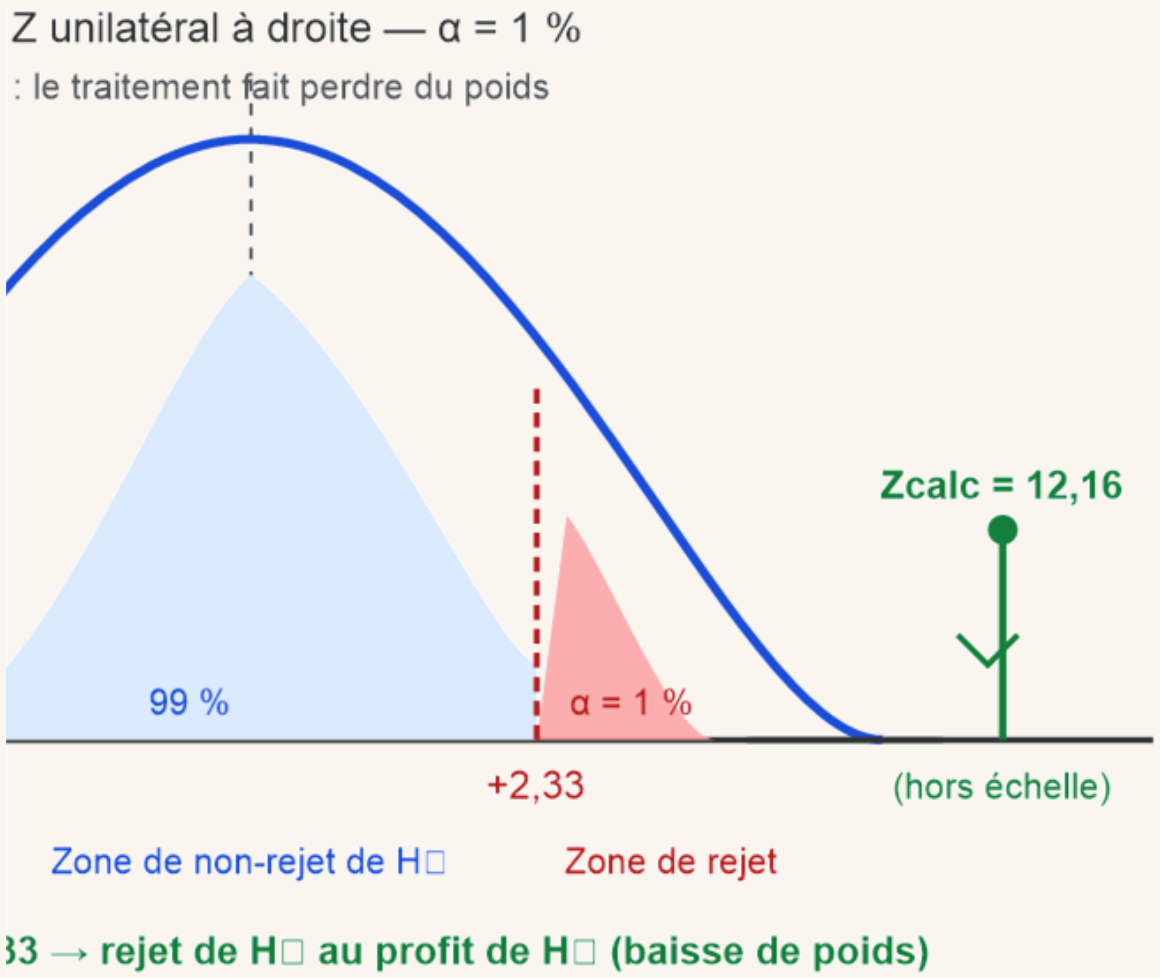
- espérance $\mu_0 = 100{,}0$ cl
- moyenne échantillon $\bar{X} = 101{,}3$ cl
- $Z_{calc} = \dfrac{|\bar{X} - \mu_0|}{\dfrac{\sigma}{\sqrt{n}}} = 12{,}16$
- $\propto = 1\%$
- $1 - \dfrac{\propto}{2} = 0{,}995$
- $Z_{th} = 2{,}58$
- $Z_{calc} > Z_{th} \rightarrow HO\ rejetée$



102,2	100,5	103,1	100,5	101,7
99,5	99,9	102,1	103,4	100,8
101,4	98,8	99,2	102,8	102,2
102,3	101,2	103,2	99,7	101,0
103,1	100,4	100,9	103,2	103,5
102,5	102,2	101,5	102,3	102,3
100,3	100,9	102,0	101,0	…..

Principe du test Z, unilatéral droit

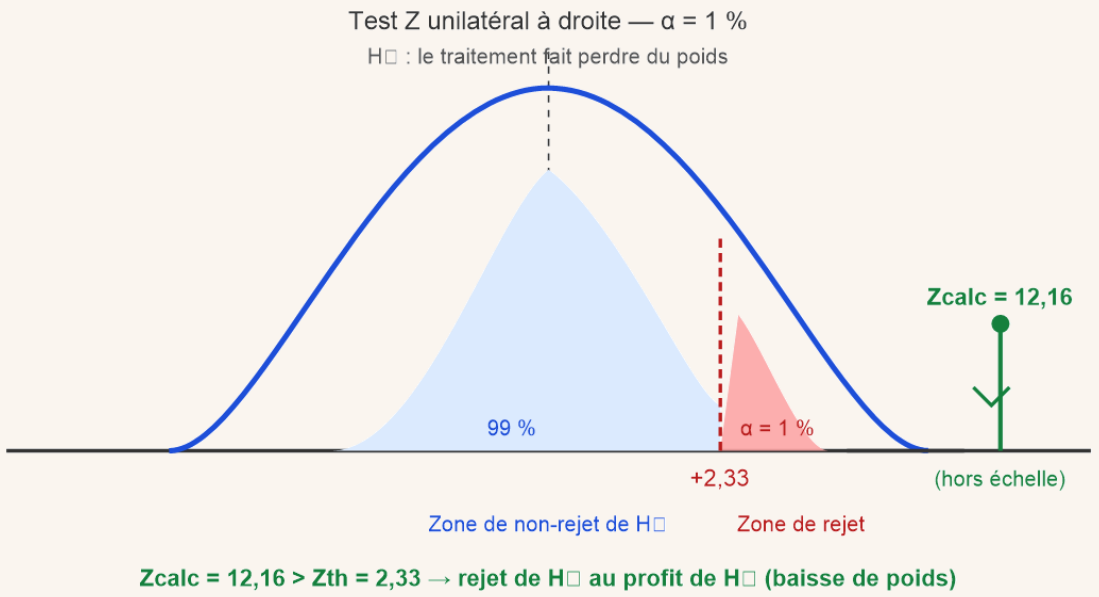
- On veut tester
  $HO = \{ \mu \leq \mu_0 \}$
  (test unilatéral)
  - contre $HA = \{ \mu > \mu_0 \}$
- Test de niveau α :
  Rejeter HO si
  $$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > u_{1-\alpha}$$



Principe du test Z, unilatéral droit

- Explication
  - rejet si $\bar{X} - \mu_0 > l$
    - avec l tel que le niveau du test égale α
  - test unilatéral, donc il y a un seul évènement :

    $$\bar{X} - \mu_0 > l$$
  - $l = \frac{\sigma}{\sqrt{n}} u_{1-\alpha} = \frac{\sigma}{\sqrt{n}} Z_{1-\alpha}$
    - $Z_{1-\alpha}$: quantile $1-\alpha$ de la loi normale centrée réduite



Test T, dit de "Student"

Test bilatéral sur l'espérance, avec σ inconnu

Test unilatéral sur l'espérance, avec σ inconnu

Histoire

- Origine : biostatistique
- Inventé par William Gosset, employé chez Guinness à Dublin.
- Contrôle de la qualité de la production de bière stout
  - taille des grains d'orge
  - petits échantillons
- Publié sous le pseudonyme de Student, 1908

Utilité

- Surnom : la "petite normale"
- Comparaison d'une valeur à une loi normale
  - quand la variance est inconnue
- Comparaison des moyennes de deux échantillons
  - issues de deux lois normales
  - quand leurs variances sont égales et inconnues
- Echantillon de petite taille
  - $n < 30$

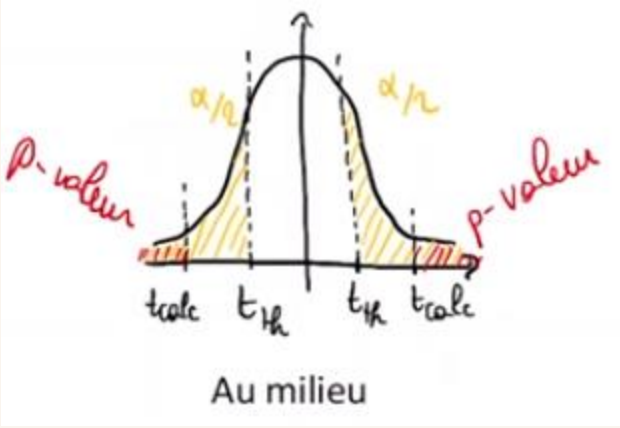
Conditions

- Pour toutes les comparaisons
  - Les données sont continues.
  - La distribution des données est normale.
  - Les échantillons ont été prélevés de manière aléatoire.
  - L'espérance μ_0 est connue.
  - L'écart-type σ est inconnu et/ou $n < 30$
- En plus, pour la comparaison de deux échantillons
  - Les deux échantillons doivent être indépendants.
  - Il y a homogénéité de la variance entre les populations.

Principe

- approche similaire au test Z
- σ remplacé par S
- loi Normale Z remplacée par la loi de Student T
- paramètres
  - α
  - degré de liberté k = n - 1
- remarque : quand n > 30, T devient très proche de Z

Principe du test T bilatéral



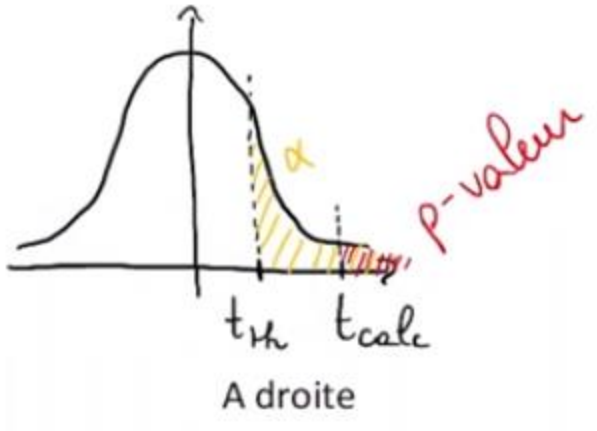
- On veut tester $HO = \{ \mu = \mu_0 \}$ (test bilatéral)
  - contre $HA = \{ \mu \neq \mu_0 \}$
- Test de niveau α : Rejeter HO si $t_{calc} = \frac{|\bar{X} - \mu_0|}{\frac{S}{\sqrt{n}}} > t_{n-1;1-\alpha/2} = t_{th}$
- Explication
  - rejet si $|\bar{X} - \mu_0| > l$
    - avec l tel que le niveau du test égale α
  - test bilatéral, donc il y a deux évènements de même prob. $\alpha/2$:

$$\bar{X} - \mu_0 > l \quad \text{et} \quad \bar{X} - \mu_0 < -l$$

  - $l = \frac{S}{\sqrt{n}} t_{n-1;1-\alpha/2} = \frac{S}{\sqrt{n}} t_{th}$
    - t_{th} = quantile $1 - \alpha/2$ de la loi de Student à n-1 degrés de liberté

Tableur
`LOI.STUDENT.INVERSE.BILATERALE`

Principe du test T unilatéral



- On veut tester $HO = \{ \mu \leq \mu_0 \}$ (test unilatéral)
  - contre $HA = \{ \mu > \mu_0 \}$
- Test de niveau α : Rejeter HO si $t_{calc} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} > t_{n-1;1-\alpha} = t_{th}$
- Explication
  - rejet si $\bar{X} - \mu_0 > l$
    - avec l tel que le niveau du test égale α
  - test unilatéral, donc il y a un seul évènement : $\bar{X} - \mu_0 > l$
  - $l = \frac{S}{\sqrt{n}} t_{n-1;1-\alpha} = \frac{S}{\sqrt{n}} t_{th}$
    - t_{th} = quantile $1 - \propto$ de la loi de Student à n-1 degrés de liberté

Tableur

`LOI.STUDENT.INVERSE.N`

Test du χ^2 d'ajustement

χ^2 ou khi-carré ou khi-deux

Test d'ajustement à une loi



- Une variable en jeu
- La variable suit-elle une loi donnée ?
  - loi uniforme, loi normale, de Poisson, etc.
- Exemple
  - Est-ce que les sacs de bonbons ont le même nombre d'éléments de chaque saveur ?
    - répartition de loi uniforme
  - HO : Les proportions de saveurs de bonbons sont égales.
  - HA : Les proportions de saveurs ne sont pas les mêmes.

Conditions

- Les données sont continues.
- La distribution des données suit une loi connue a priori.
- Les échantillons ont été prélevés de manière aléatoire.
- Au moins cinq éléments observés.

Principe

- On veut tester HO = { les valeurs suivent la loi }
  - contre HA = { les valeurs s'écartent significativement de la loi }
- Test de niveau α : Rejeter HO si $\chi^2_{calc} > \chi^2_{th}$
- Calcul
  - Soit
    - n : le nombre de valeurs observées
    - $N_{i\,obs}$: les valeurs observées et $N_{i\,th}$: les valeurs données par la loi
    - DDL = n - 1
    - α : le niveau du test
  - $\chi^2_{calc} = \sum_{i=1}^{n} \frac{(N_{i\,obs} - N_{i\,th})^2}{N_{i\,th}}$
  - $\chi^2_{th} = fonction\ de\ \propto et\ DDL$
    - cf Wikipedia

tableur :
`LOI.KHIDEUX.INVERSE.DROITE()`

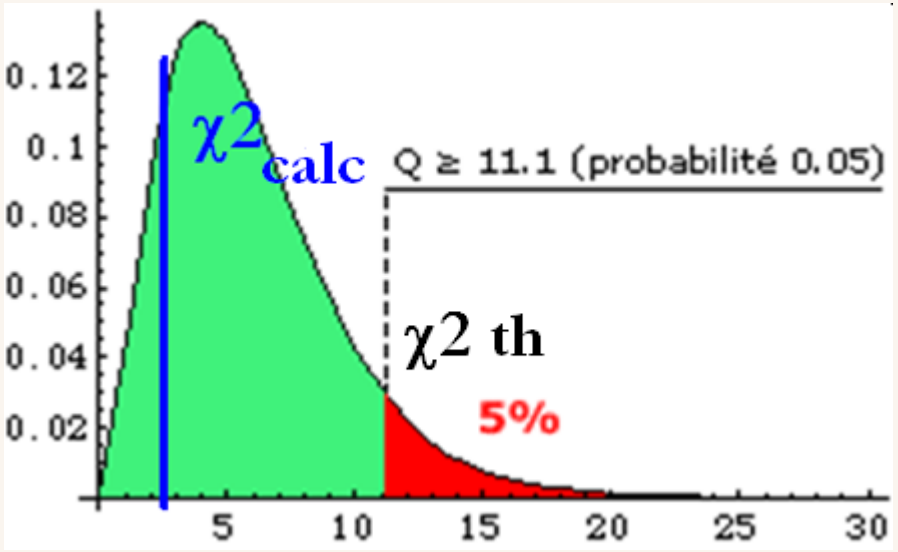
Exemple (cf complément)

- On lance 900 fois un dé à six faces. Sur base des résultats ci-dessous, peut-on dire que le dé est pipé ?
- HO : Le dé n'est pas pipé. Les effectifs observés devraient être autour de 150 par valeur.
  - 150 = 600 * 1/6
- HA : Le dé est pipé. Les effectifs s'écartent trop de la valeur théorique de 150.

valeur du dé	1	2	3	4	5	6	total
eff. observé	141	156	157	140	160	146	900
eff. théorique	150	150	150	150	150	150	900

Exemple (cf complément)

- $\chi^2_{calc} = \sum_{i=1}^{n} \frac{(N_{i\,obs} - N_{i\,th})^2}{N_{i\,th}} = 2{,}55$
- $DDL = n - 1 = 5$
- $\propto = 5\%$
- $\chi^2_{th} = 11{,}07$
- $\chi^2_{calc} < \chi^2_{th} \rightarrow HO\ acceptée$

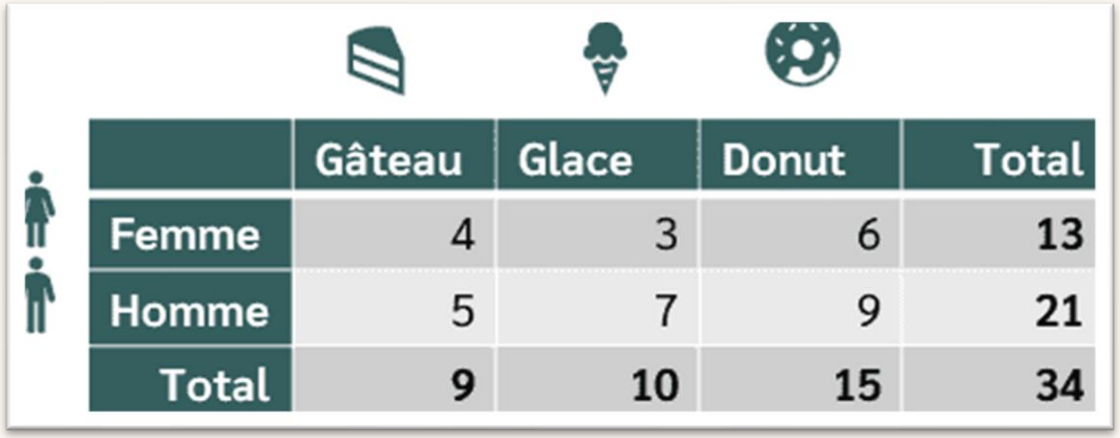


valeur du dé	1	2	3	4	5	6	total
eff. observé	141	156	157	140	160	146	900
eff. théorique	150	150	150	150	150	150	900

Test du χ^2 d'indépendance

χ^2 ou khi-carré ou khi-deux

Tableau de contingence



	Gâteau	Glace	Donut	Total
Femme	4	3	6	13
Homme	5	7	9	21
Total	9	10	15	34

- Aperçu de deux variables, le plus souvent catégorisées.
- On peut y lire la fréquence de la combinaison des deux valeurs.

Test d'indépendance

- Deux variables en jeu
- Ces variables sont-elles liées entre elles ?
- Exemple
  - Au cinéma, les achats de snacks par les spectateurs sont-ils liés au type de film qu'ils vont voir ?
  - HO : La proportion de personnes qui achètent des snacks est indépendante du type de film
  - HA : La proportion de personnes qui achètent des snacks diffère suivant le type de films

coul vs pays	Belgique	France	Suisse	Pérou	**Total / coul**
Rouge	48	69	210	11	**338**
Jaune	27	79	141	16	**263**
Vert	5	28	48	5	**86**
Total / pays	**80**	**176**	**399**	**32**	**687**

Conditions

- Les échantillons ont été prélevés de manière aléatoire.
- Au moins cinq éléments observés par cellule du tableau de contingence.

Principe

- HO = { les deux observations sont indépendantes}
  - HA = { les deux observations sont significativement liées }
- Test de niveau α : Rejeter HO si χ^2 calc > χ^2 th
- Paramètres du calcul
  - N_{tot} : l'effectif total observé
  - i : les valeurs de la 1e observation
  - j : les valeurs de la 2e observation
  - N_{ij} : l'effectif observé pour les valeurs i et j
  - N'_{ij}: l'effectif théorique pour les valeurs i et j
  - $N_{i.}$ et $N_{.j}$ les sous-totaux par valeur
  - DDL = (nb val i - 1) * (nb val j - 1)
  - α, le niveau du test

Principe

- Calcul
  - $\chi^2_{calc} = \sum_i \sum_j \frac{(N_{ij} - N'_{ij})^2}{N'_{ij}}$, avec $N'_{ij} = \frac{N_{i.} * N_{.j}}{N_{tot}}$
  - $\chi^2_{th} = fonction\ de\ \propto et\ DDL$
    - cf Wikipedia

tableur :
`LOI.KHIDEUX.INVERSE.DROITE()`

Exemple (cf complément)

Ce tableau de contingence indique le relevé statistique de 687 pommes différenciées par leur couleur et par leur pays d'origine. Y a-t-il un lien entre ces deux observations ?

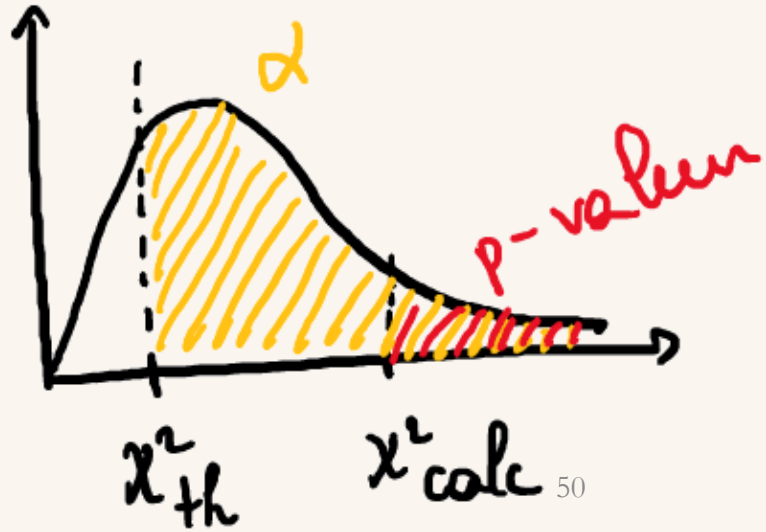
tableau de contingence (Nij)					
coul vs pays	Belgique	France	Suisse	Pérou	T. / c.
Rouge	48	69	210	11	338
Jaune	27	79	141	16	263
Vert	5	28	48	5	86
Total / pays	80	176	399	32	687

Exemple

$$\chi^2 = \sum_i \sum_j \frac{(N_{ij} - N'_{ij})^2}{N'_{ij}}$$

$$N'_{ij} = \frac{N_{i.} * N_{.j}}{N_{tot}}$$

- $\chi^2_{calc} = 16{,}79$
- $\text{DDL} = (\text{ nb coul } - 1) * (\text{ nb pays } - 1) = 6$
- $\propto = 5\%$
- CCL : Les couleurs et les pays sont liées.
  - enfin, 95% de chance pour que …



Intensité

- Quand HO est rejetée, càd quand les valeurs observées sont liées (HA), on veut pouvoir mesurer l'intensité de cette liaison.
- Coefficient de Kramer
  - entre 0 (faible) et 1 (fort)
  - rem : k et p sont les ddl

$$C = \sqrt{\frac{\chi^2}{N_{..} * Min(k-1;p-1)}}$$

Autres tests

Anova

test F de Fisher

test de Student pour deux groupes

Anova & test F de Fisher

- Données liées à des catégories
- Exemple : la taille des feuilles de saule est-elle liée à leur localisation (ville, campagne, etc.) ?
- Rem : ANOVA = *"analysis of variances"*

Corrélation linéaire & test de Bravais-Pearson

- Relation linéaire entre deux variables
- Exemple : On a mesuré la taille et l'âge de 36 arbres. Ces deux variables ont-elles un lien au seuil de 5% ?

Test de Student pour deux groupes

- Comparaison de deux groupes
- La moyenne de ces groupes est-elle semblable ou significativement différente ?