**Fin Statistiques descriptive + début statistiques déductives.**

# I/ Statistiques descriptives !

## Rappel cours 3a :

- Pour **α (risque de 1ere espèce)= 5 % => ε (écart réduit) =1,96**
- Pour **α = 1 % => ε=2,6**

$$\boxed{i = \varepsilon \frac{s}{\sqrt{n}}}$$

Pour retrouver la formule du nombre de sujets nécessaires :

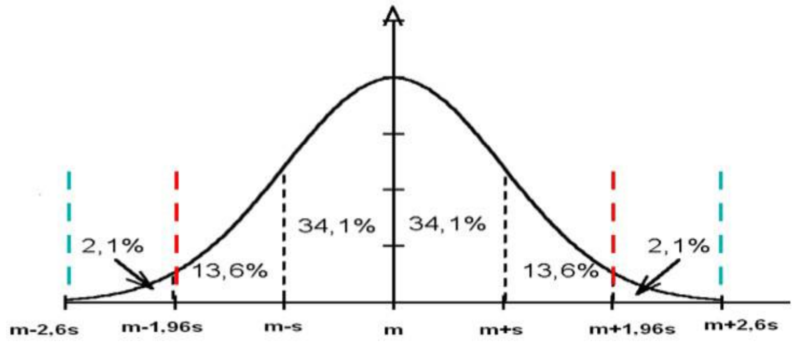
J'élève toute la formule au carré : $i^2 = \varepsilon^2 \frac{s^2}{n}$

Puis j'isole n et j'obtiens la formule : $n = \varepsilon^2 \frac{s^2}{i^2}$

## A) La loi normale ou de Gauss

La loi de gauss n'est applicable que sur des échantillons de **plus de 30 personnes !!**

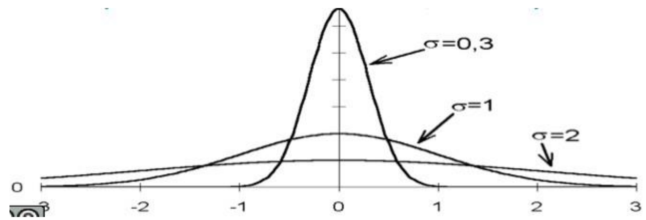
Voilà à quoi ressemble la courbe de Gauss :



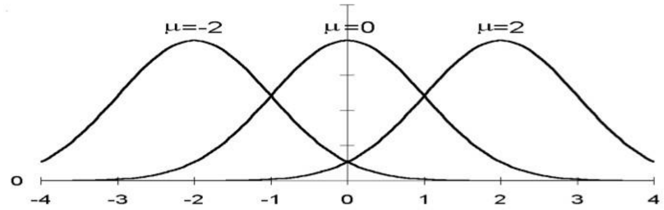

On retrouve un aspect en cloche avec au centre de la cloche la moyenne, s l'écart type.

- Entre (m-s) et (m+s) on a 68.2% de la population.
- Entre (m-1.96s) et (m+1.96s) on a 95.4% de la population. (D'où risque **α = 5 % => ε=1.96 :** Pour un écart réduit de 1.96, on a un IC d'environ 95% donc on a un risque de 5% de se tromper)
- Entre (m-2.6s) et (m+2.6s) on a 99.6% de la population. D'où risque **α = 1 % => ε=2.6 :** Pour un écart réduit de 2.6, on a un IC d'environ 99% (mal arrondi je vous l'accorde) donc on a un risque de 5% de se tromper

Plus la **dispersion est grande,** plus la courbe est large. La dispersion autour de la moyenne est donnée grâce à **l'écart type σ**.



En changeant la moyenne, on change la courbe de place. La moyenne est au centre de celle-ci :




## B) Estimation de données qualitatives :

Dans l'estimation de données qualitatives, on s'intéresse à la proportion de la population présentant une caractéristique quelconque

**ECHANTILLON**
n = effectif
Po = pourcentage observe
s = ecart type

**POPULATION TOTALE**
N = effectif
p = pourcentage réel
σ = ecart type

**L'estimation** assure la correspondance entre l'échantillon et la population (avec un intervalle de confiance)

Lorsque l'étude a été correctement menée, on pose ainsi que :

$$\mu \in IC_{(1-\alpha)}$$

$$IC_{(1-\alpha)} = [\ po - (\varepsilon.s)\ ;\ po + (\varepsilon.s)\ ]$$

*ou encore*

$$IC_{(1-\alpha)} = [\ po - \varepsilon.\sqrt{po.qo/n}\ ;\ po + \varepsilon.\sqrt{po.qo/n}\ ]$$

*avec*

$$s = \sqrt{po.qo/n} \quad \textit{et} \quad qo = 1-po$$

| | |
|---|---|
| po : pourcentage observé dans l'échantillon<br>s : écart type de l'échantillon<br>n : effectif de l'échantillon | p : pourcentage réel de la population |

**<u>Exemple :</u>** Pour décider si oui ou non des télés seront mises en place les chambres d'un hôpital, un sondage est effectué auprès d'un échantillon représentatif de 200 malades. 110 personne souhaitent leur mise en place, les 90 autres ne trouve pas d'intérêt à avoir des télés. L'hôpital peut-il commencer à acheter les tv ?

P0= 110/200 = 0.55

IC95% = [0,55 - 1,96$\sqrt{0,55*0,45/200}$ ; 0,55 + 1,96$\sqrt{0,55*0,45/200}$]
IC95% =[0,48 ; 0, 62]

L'intervalle de confiance à 95% indique des possibilités de valeurs en dessous de la moyenne. L'hôpital ne peut donc rien conclure de ce sondage.

**=> D'où l'importance de TOUJOURS FAIRE ATTENTION AUX CONCLUSIONS D'UNE ENQUETE.**

***NB : Une NON-REPONSE à un sondage provenant de l'échantillon interroge constitue toujours un BIAIS !***

L'indice i permet de calculer la précision de l'estimation :

$$i = \varepsilon.s = \varepsilon.\sqrt{po.qo/n}$$

i : indice de précision po : pourcentage observe s : écart type de l'échantillon
n : effectif de l'échantillon ε : écart réduit => *facteur dépendant du risque α*

**NB : la précision AUGMENTE lorsque i DIMINUE.**

-Les **deux règles essentielles** décrites pour l'estimation d'une variable quantitative s'appliquent ici également.

Calcul du nombre de sujets nécessaires pour une précision donnée :

$$n = \varepsilon_2(po.qo) / i_2$$

i : indice de précision voulu po : pourcentage observe s : écart type de l'échantillon
n : effectif nécessaire de l'échantillon ε : écart réduit => *facteur dépendant du risque α*

# II/ Statistiques déductives

**H0= hypothèse nulle :**
Il n'y a pas de différence observée entre les deux groupes

**H1 = hypothèse alternative.**
Il y a une différence significative entre les deux groupes

On choisit toujours pour H0 l'hypothèse qu'il serait **le plus grave de rejeter à tort.**

**Tableau à connaitre par cœur sur les risques :**

| Réalité \ Décision du statisticien | Rejet H0 | Non rejet H0 |
|---|---|---|
| H0 Vraie | Erreur 1ère espèce α | 1 - α |
| | | |
| H1 Vraie | Puissance 1 - β | Erreur 2ème espèce β |

**<u>Les étapes d'un test d'hypothèse :</u>**
Pour mettre en œuvre un test d'hypothèse, on suivra TOUJOURS les étapes suivantes :

1. **Définir H0 et H1**. Les deux hypothèses jouent des rôles **symétriques**.

2. Déterminer le caractère des données à étudier/comparer :
- <u>qualitative/qualitative</u>
- qualitative/quantitative
- quantitative/quantitative

3. Choisir le test en fonction du type de données On nomme **Z** le paramètre qui sera calculé.

4. Choisir le seuil d'erreur de 1 ère espèce $\alpha$ généralement fixe à **5%.**

5. Recueillir les données, calculer Z et utiliser la règle de rejet (définie à partir de H0 et de $\alpha$) : Il s'agit de **comparer Z par rapport à une valeur théorique** de référence

6. Interpréter des résultats :
- Au niveau de l'échantillon : Accepte-t-on H0 ?
- Au niveau de la population : peut-on extrapoler les résultats à la population générale ?

**NB** : L'acceptation de H0 implique forcement le rejet de H1 et *vice versa*.

## <u>Etude de la liaison entre deux caractères qualitatifs :</u>

Dans le cas de deux caractères qualitatif, on peut choisir d'utiliser soit :

1. un test de comparaison des pourcentages
2. un test du $X_2$

### <u>A-Test de comparaison des pourcentages :</u>

Toutes les formules (sauf celles des ddl) fournies dans la suite de la fiche ne sont pas à connaître !

Soient **2 groupes A et B** et une caractéristique **qualitative x** (couleur des yeux etc.). On peut se demander si la proportion d'individus du groupe A présentant x coïncide avec la proportion d'individus du groupe B présentant x. (ex : groupe B possede-t'il plus de personne aux yeux bleu que le groupe A ?)

1) Définir H0 et H1.
- **H0** : il n'y a pas de différence observée entre les groupes A et B, c'est à dire que la proportion d'individus du groupe A présentant x coïncide avec la proportion d'individus du groupe B présentant x. (ex : Pas plus d'yeux bleus dans un groupe que dans l'autre)

- **H1** : il existe une différence significative entre les groupes A et B, c'est à dire que la proportion d'individus du groupe A présentant x est différente de celle du groupe B.

2) Déterminer le caractère des données à étudier/comparer : Les variables sont :
- des types d'individus → variable **qualitative**
- une caractéristique x **qualitative**

3) Choisir le test en fonction du type de données : En présence de données qualitatives, on peut choisir d'utiliser un **test de comparaison des pourcentages**.

4) Choisir le seuil d'erreur de 1 ère espèce α généralement fixe à **5%.**

5) Recueillir les données, calculer Z et utiliser la règle de rejet
La variable Z est ici représentée par **l'écart réduit ε**. On comparera :

- $\varepsilon_{th}$ donné par la table de l'écart réduit, en fonction de α
- $\varepsilon_{\text{calculée}} = \varepsilon_{exp} = \dfrac{p_A - p_B}{\sqrt{\dfrac{p_A q_A}{n_A} + \dfrac{p_B q_B}{n_B}}}$ (Formule pas à apprendre)

6) Interpréter les résultats : Au niveau de l'échantillon :

1. si $\varepsilon_{calculee} > \varepsilon_{th}$ alors **on rejette H0** et on accepte H1.
2. si $\varepsilon_{calculee} < \varepsilon_{th}$ alors **on accepte H0** et on rejette H1.

## <u>B Le test du $\chi^2$ :</u>

Soient **2 groupes A et B** et une caractéristique **qualitative x** (couleur des yeux etc.). On peut se demande si la proportion d'individus du groupe A présentant x coïncide avec la proportion d'individus du groupe B présentant x.

1) définir H0 et H1.

2) Déterminer le caractère des données à étudier/comparer : Les variables seront dans ce cas qualitatif. (Comme comparaison de pourcentages)

3) Choisir le test en fonction du type de données En présence de données qualitatives, on peut choisir d'utiliser un **test du $\chi^2$**.

4) Choisir le seuil d'erreur de 1 ère espèce α généralement fixe à **5%.**

5) Recueillir les données, calculer Z et utiliser la règle de rejet
La variable Z est ici représentée par le **$\chi^2$**. On comparera :

- $\chi^2_{th}$ est donnée par la table du $\chi^2$ en croisant :
  - α le risque de première espèce
  - **le nombre de degré de liberté. (Voir la diapo pour le tableau)**

Pour le test du $\chi^2$, le nombre de ddl vaut : (A connaitre par cœur !!!)

**Ddl : ($nb_{lignes}$ – 1) x ($nb_{colonnes}$ – 1)**

- $\chi^2_{\text{calculée}} = \chi^2_{exp} = \sum \dfrac{(Oi - Ci)^2}{Ci}$ (Pas à apprendre)

6) Interpréter les résultats : Au niveau de l'échantillon :

1. si $\chi^2_{calculee} > \chi^2_{th}$ alors **on rejette H0** et on accepte H1.
2. si $\chi^2_{calculee} < \chi^2_{th}$ alors **on accepte H0** et on rejette H1.

Si les échantillon(s) considérés sont **représentatifs** de la population, on pourra alors extrapoler le résultat obtenu à l'ensemble de la population.

**Fini !!!**