

## Situations Problèmes & Documents autocorrectifs

### Situation Problème n°1

#### Document n° 1 : situation de l'étude

La présente situation s'inspire des travaux d'une thèse de doctorat en sciences de l'éducation, intitulée CONSTRUÇÃO DE CONHECIMENTOS EM SITUAÇÕES DE EXCLUSÃO SOCIAL E QUESTÕES DE GÊNERO, préparée en cotutelle par Valdir Pretto (*voir site cyberthèses Lyon2*) et soutenue en août 2009 au Brésil. La thématique centrale porte les connaissances alternatives construites par des sujets confrontés à des situations d'exclusion sociale, pour survivre. Un échantillon de 103 unités a été extrait d'une population habitant dans une zone périphérique d'une ville du Sud du Brésil. Cet échantillon a été constitué en deux étapes : tirage aléatoire des rues, tirage aléatoire des foyers. Une unité est alors le ou les sujets qui ont répondu au questionnaire qui a été administré par le chercheur. Dans un premier temps une enquête par questionnaire a été réalisée afin de recueillir des informations relatives aux individus mais aussi pour identifier les problèmes qu'ils estimaient être ceux qu'ils rencontraient et dont les résolutions nécessitaient des connaissances. D'autres méthodes de construction ont été mises en oeuvre fondées sur des entretiens audio-vidéographiés individuels puis collectifs. En ce qui concerne les problèmes, 14 catégories furent identifiées. Le questionnaire comportait une vingtaine de grandes questions portant sur les caractéristiques individuelles et sociales des sujets, la situation familiale, la composition de la famille, les ressources, le logement, la formation de chacun des membres de la famille, etc.. Dans le cadre de cette étude, nous ne retenons ici que quelques questions que nous présentons ci-dessous. Les tableaux statistiques fournis ensuite contiennent des valeurs réelles mais aussi des valeurs fictives adaptées à la situation d'étude présente.

| <i>code de la variable</i> | <i>descriptif succinct</i>   |
|----------------------------|--|
| [V1]                       | Sexe : 1_homme 2_femme   |
| [V2]                       | Age à la date de l'enquête du sujet qui répond aux questions   |
| [V3]                       | Lieu de naissance  |
| [V4]                       | Ancienneté de résidence dans la ville  |
| [V5]                       | Ancienneté de résidence dans le quartier   |
| [V6]                       | Etat civil : 1_Célibataire – 2_Marié(e) – 3_Séparé(e) – 4_Divorcé(e) – 5_Vit avec les parents – 6_Veuf(ve)   |
| [V7]                       | Nombre de personnes qui vivent dans le foyer   |
| [V8]                       | Ressources du foyer  |
| [V9]                       | <i>Savez-vous lire et écrire ?</i><br>1_Pas du tout ; 2_Un peu ; 3_Moyennement ; 4_Asez facilement ; 5_Très facilement   |
| [V10]                      | Niveau de scolarité : 1_Analphabète – 2_Semi-analphabète – 3_Enseignement fondamental incomplet – 4_Enseignement fondamental complet – 5_Enseignement moyen (lycée) incomplet – 6_Enseignement moyen (lycée) complet – 7_Enseignement supérieur incomplet – 8_Enseignement supérieur complet |
| [V11]                      | <i>Quels sont les problèmes que vous rencontrez actuellement là où vous habitez ?</i>  |
| [V12]                      | <i>Comment vous affrontez, résolvez les problèmes que vous rencontrez là où vous habitez ?</i>   |
| [V13]                      | Nombre de catégories de problèmes évoquées   |

Nous rapportons ci-dessous les tableaux statistiques des variables

**Tableau statistique de la variable V1**

| (Tableau n° T1)<br>V1=Sexe | Homme | Femme |
|----------------------------|-------|-------|
| Échantillon                | 34    | 69    |

**Tableau statistique de la variable V4**

| (Tableau n° T2)<br>V4= Temps de<br>résidence dans la<br>ville (années) | [1 ; 10[ | [10 ; 19[ | [19 ; 30[ | [30; 35[ | [35; 50[ |
|--|----------|-----------|-----------|----------|----------|
| Échantillon  | 19       | 44        | 22        | 9        | 9        |

**Tableau statistique de la variable V6**

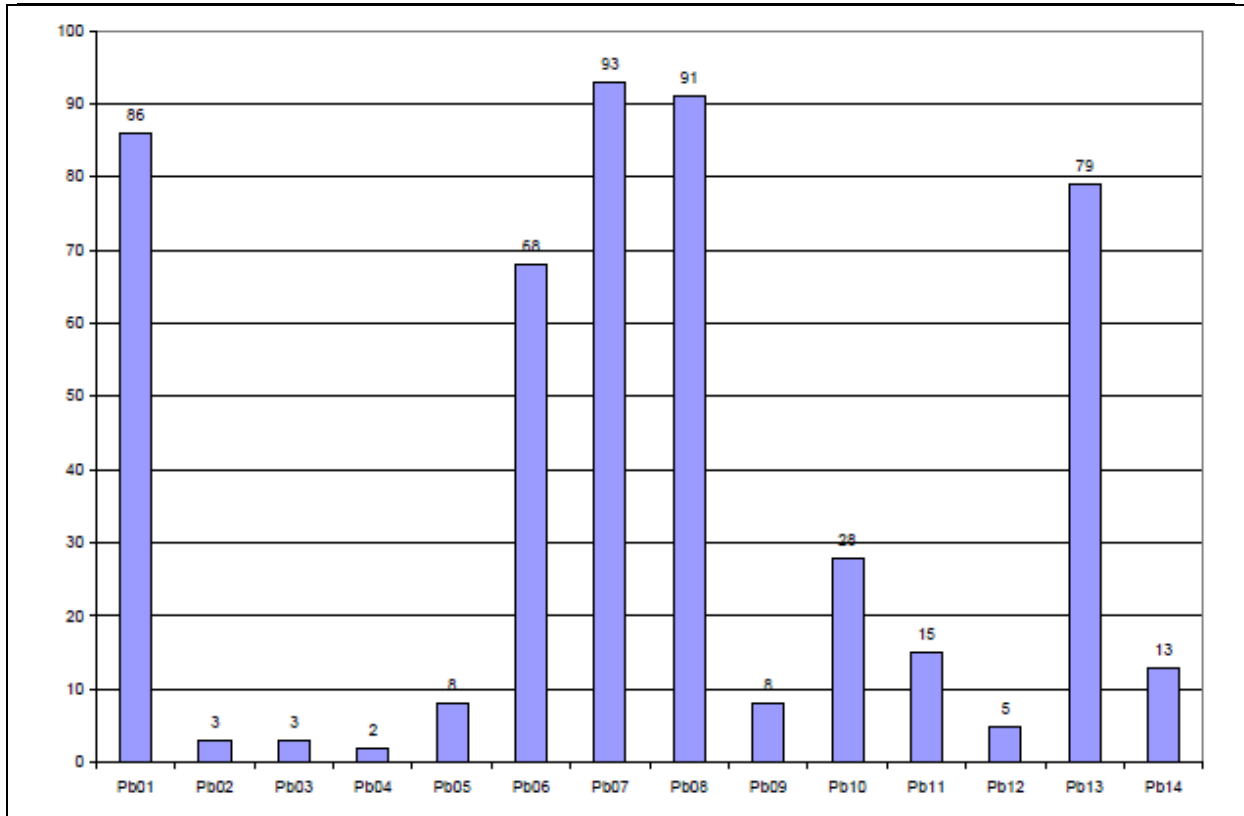
| (Tableau n° T3)<br>V6=Etat civil | 1_Célibataire | 2_Marié(e) | 3_Séparé(e) | 4_Divorcé(e) | 5_Vit<br>avec les<br>parents | 6_Veuf(ve) |
|----------------------------------|---------------|------------|-------------|--------------|------------------------------|------------|
| Échantillon                      | 6             | 82         | 1           | 8            | 1                            | 5          |

**Tableau statistique de la variable V9**

| (Tableau n° T4)<br>V9=Lire et Écrire<br>(valeurs estimées) | 1_Pas du tout | 2_Un peu | 3_Moyennement | 4_Asez facilement | 5_Très facilement. |
|--|---------------|----------|---------------|-------------------|--------------------|
| Échantillon  | 8             | 5        | 54            | 24                | 12                 |

**Tableau des catégories de problèmes identifiés**

| (Tableau n° T5)<br>codages<br>des<br>problèmes | Nombre<br>d'évocations | <i>Catégories</i>                |                                   |
|--|------------------------|----------------------------------|-----------------------------------|
| Pb01   | 86                     | Santé                            | Saúde                             |
| Pb02   | 3                      | Alimentation                     | Alimentação                       |
| Pb03   | 3                      | Économie familiale               | Economia familiar                 |
| Pb04   | 2                      | Situation familiale              | Situação familiar                 |
| Pb05   | 8                      | Travail                          | Trabalho                          |
| Pb06   | 68                     | Éducation, formation<br>scolaire | Educação                          |
| Pb07   | 93                     | Sécurité                         | Segurança                         |
| Pb08   | 91                     | Loisir                           | Lazer                             |
| Pb09   | 8                      | Questions religieuses            | Religiosidade                     |
| Pb10   | 28                     | Transport                        | Transporte                        |
| Pb11   | 15                     | Recyclage                        | Reciclagem                        |
| Pb12   | 5                      | Habitation                       | Habitação                         |
| Pb13   | 79                     | Assainissement                   | Saneamento                        |
| Pb14   | 13                     | Relations de voisinage           | Relacionamento com os<br>vizinhos |



**Tableau statistique de la variable V13**

| (Tableau n° T6)<br>V13=nombre de<br>catégories évoquées | 0 | 1 | 2 | 3  | 4  | 5  | 6  | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|----|----|----|----|---|---|---|----|----|----|----|----|
| Échantillon   | 0 | 1 | 5 | 11 | 22 | 31 | 24 | 5 | 1 | 2 | 1  | 0  | 0  | 0  | 0  |

| (Tableau n° T7) | Pb08 cité | Pb08 n'est pas cité |     |
|-----------------|-----------|---------------------|-----|
| <b>Homme</b>    | 27        | 7                   | 34  |
| <b>Femme</b>    | 64        | 5                   | 69  |
|                 | 91        | 12                  | 103 |

**Document n° 2 : traitement demandé**

**TQ1- De quoi s'agit-il ?**

Q101- Après avoir pris connaissance des informations fournies dans le document n°1, identifier et caractériser “succinctement” mais très clairement, les éléments de cette étude statistique : individus, population(s), échantillon(s), variables.

**Population :**

Ensemble des habitant dans une zone périphérique d'une ville du Sud du Brésil (**individus**)

**Échantillons:**

Sous-ensemble des 103 individus extrait de la population

**Variables.**

| Variables quantitatives |           | Variables qualitatives |           |            |
|-------------------------|-----------|------------------------|-----------|------------|
| discrètes               | continues | ordinales              | nominales | Textuelles |
| [V7]                    | [V2]      | [V9]                   | [V1]      | [V11]      |
| [V8]                    | [V4]      |                        | [V3]      | [V12]      |
| [V13]                   | [V5]      |                        | [V6]      |            |
|                         |           |                        | [V10]     |            |

**TQ2-Etude de la variable V4 « Temps de résidence dans la ville (années) »**

Q201 Sur l'échantillon, quelle est la valeur m de la durée moyenne de résidence dans la ville et quelle est la valeur de l'écart-type ?

| Intervalles | valeurs centrales<br>(xi) | effectifs<br>(ni)     | $n_i x_i$      | $n_i(x_i-m)^2$     |
|-------------|---------------------------|-----------------------|----------------|--------------------|
| [1 ; 10[    | 5,5                       | <b>19</b>             | 104,5          | 3460,260156        |
| [10 ; 19[   | 14,5                      | <b>44</b>             | 638            | 889,0787068        |
| [19 ; 30[   | 24,5                      | <b>22</b>             | 539            | 666,6752757        |
| [30 ; 35[   | 32,5                      | <b>9</b>              | 292,5          | 1641,429824        |
| [35 ; 50[   | 42,5                      | <b>9</b>              | 382,5          | 4972,30361         |
|             |                           | <i>effectif total</i> | <i>somme</i>   | <i>somme</i>       |
|             |                           | <b>103</b>            | <b>1956,5</b>  | <b>11629,74757</b> |
|             |                           |                       | <i>moyenne</i> | <i>variance</i>    |
|             |                           |                       | <b>18,9951</b> | <b>112,9101706</b> |
|             |                           |                       |                | <i>écart-type</i>  |
|             |                           |                       |                | <b>10,62591975</b> |

Q202- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

Ainsi la durée moyenne de résidence sur l'échantillon est d'environ 19 ans avec un fluctuation moyenne donnée par l'écart-type de 10 ans 8 mois environ

**TQ3- Étude du lien entre l'évocation de la catégorie Pb08 – Loisir et la variable V01 Sexe**

Q301- A partir de l'échantillon des 103 réponses, peut-on considérer qu'il y a ou non un lien entre ces deux variables?

En exploitant les données du tableau T7, tester les hypothèses correspondantes (Ho et H1 qui devront être explicitées) au seuil de risque  $\alpha=0.05$  puis  $\alpha=0.01$  à l'aide du test d'indépendance du Khi-deux.

Nous mettons en œuvre un test d'indépendance (Test du Khi-deux) pour décider laquelle des deux hypothèses est la plus plausible :

Ho Les hommes et les femmes ont des points de vue homogènes relatifs à la question de l'évocation de la catégorie Pb08 – Loisir

H1 Les hommes et les femmes ont des points de vue hétérogènes relatifs à la question de l'évocation de la catégorie Pb08 – Loisir

Tableau des effectifs observés :

| <b>(Tableau n° T7a) Pb08 cité Pb08 n'est pas cité</b> |    |    |     |
|---|----|----|-----|
| <b>Homme</b>  | 27 | 7  | 34  |
| <b>Femme</b>  | 64 | 5  | 69  |
|   | 91 | 12 | 103 |

Tableau des effectifs théoriques sous l'hypothèse Ho :

| <b>(Tableau n° T7b) Pb08 cité Pb08 n'est pas cité</b> |        |       |     |
|---|--------|-------|-----|
| <b>Homme</b>  | 30,039 | 3,961 | 34  |
| <b>Femme</b>  | 60,961 | 8,039 | 69  |
|   | 91     | 12    | 103 |

Tableau des contributions absolues :

| <b>(Tableau n° T7c) Pb08 cité Pb08 n'est pas cité</b> |        |        |               |
|---|--------|--------|---------------|
| <b>Homme</b>  | 0,3074 | 2,3313 | 2,6387        |
| <b>Femme</b>  | 0,1515 | 1,1487 | 1,3002        |
|   | 0,4589 | 3,4800 | <b>3,9389</b> |

Q302- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

Au seuil de  $\alpha = 5\%$  la valeur critique lue dans la table du Khi-deux ddl = 1 est de  $k = 3,84$ . la valeur empirique est de 3,93 et donc supérieure à la valeur théorique sous Ho. On rejette l'hypothèse d'homogénéité/indépendance à ce niveau de risque de 1<sup>ère</sup> espèce.

Au seuil de  $\alpha = 1\%$  la valeur critique lue dans la table du Khi-deux ddl = 2 est de  $k = 6,63$ . la valeur empirique est de 3,93 et donc inférieure à la valeur théorique sous Ho. On ne rejette pas l'hypothèse d'homogénéité à ce niveau de risque de 1<sup>ère</sup> espèce. Cette conservation de Ho contre H1 est alors prise avec un risque de seconde espèce de niveau  $\beta$  inconnu

#### **TQ4 - Étude de la variable V9**

Q401 Donner la proportion des individus considérés comme ayant atteint un niveau « au moins moyen » pour lire et écrire dans l'échantillon.

Nombre d'individus ayant répondu « 3\_Moyennement », ou « 4\_Asez facilement » ou « 5\_Très facilement » est de  $54+24+12=90$

Proportion dans l'échantillon :

$$f = \frac{90}{103} = 0,873786407... \approx 87,38\%$$

Q402 Donner une estimation ponctuelle de la proportion des individus pouvant être considérés comme ayant atteint un niveau « au moins moyen » pour lire et écrire dans la population.

Estimation ponctuelle de la proportion dans la population :

**87,38%**

Q403 Donner une estimation par intervalle de confiance à 99% de la proportion des individus pouvant être considérés comme ayant atteint un niveau « au moins moyen » pour lire et écrire dans la population.

En ce qui concerne cet intervalle de confiance à 99%, nous l'obtenons en utilisant la formule ci-dessous :

$$\left[ f - \kappa \sqrt{\frac{f(1-f)}{n-1}}; f + \kappa \sqrt{\frac{f(1-f)}{n-1}} \right]$$

$$\left[ 0,8738 - 2,57 \sqrt{\frac{0,8738(1-0,8738)}{103-1}}; 0,8738 + 2,57 \sqrt{\frac{0,8738(1-0,8738)}{103-1}} \right]$$

or  $2,57 \left( \sqrt{\frac{0,8738(1-0,8738)}{102}} \right) \approx 2,57(0,032881802) \approx 0,084506233$

d'où [ 0,789280 ; 0,958292] C'est à dire entre 78,9% et 95,8%

Avec un niveau de confiance de 99%, nous estimons que la proportion se situe entre 78,9% et 95,8%. Ce qui laisse à penser qu'ils sont la majorité.

Q404- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

### TQ5 - Étude de l'homogénéité des points de vue relatifs à l'ordre d'importance des problèmes rencontrés

Nous rapportons les rangements des 14 catégories de problèmes produits par les 8 sujets qui ont constitué le groupe des sujets auprès desquels ont été réalisés les entretiens. Les résultats qui utilisent les rangs *ex aequo*, sont reportés dans le tableau n°T8 ci-dessous :

| <b>Tableau T8</b> |             | Pb_01        | Pb_02        | Pb_03        | Pb_04        | Pb_05     | Pb_06      | Pb_07     | Pb_08        | Pb_09     | Pb_10      | Pb_11        | Pb_12      | Pb_13     | Pb_14 |
|-------------------|-------------|--------------|--------------|--------------|--------------|-----------|------------|-----------|--------------|-----------|------------|--------------|------------|-----------|-------|
| <b>individus</b>  | Sujet_01    | 4            | 10           | 10           | 10           | 10        | 5          | 2         | 3            | 10        | 10         | 10           | 10         | 1         | 10    |
|                   | Sujet_02    | 4            | 10,5         | 10,5         | 10,5         | 10,5      | 6          | 1         | 10,5         | 10,5      | 2          | 10,5         | 5          | 3         | 10,5  |
|                   | Sujet_03    | 4            | 10,5         | 10,5         | 10,5         | 10,5      | 6          | 1         | 10,5         | 10,5      | 2          | 10,5         | 5          | 3         | 10,5  |
|                   | Sujet_04    | 5            | 11,5         | 11,5         | 11,5         | 11,5      | 7          | 2         | 6            | 11,5      | 8          | 4            | 11,5       | 3         | 1     |
|                   | Sujet_05    | 1            | 3            | 9            | 9            | 9         | 9          | 9         | 9            | 2         | 9          | 9            | 9          | 9         | 9     |
|                   | Sujet_06    | 1            | 3            | 9            | 9            | 9         | 9          | 9         | 9            | 2         | 9          | 9            | 9          | 9         | 9     |
|                   | Sujet_07    | 2            | 1            | 9,5          | 9,5          | 9,5       | 9,5        | 3         | 4            | 9,5       | 9,5        | 9,5          | 9,5        | 9,5       | 9,5   |
|                   | Sujet_08    | 2            | 1            | 9,5          | 9,5          | 9,5       | 9,5        | 3         | 4            | 9,5       | 9,5        | 9,5          | 9,5        | 9,5       | 9,5   |
| <b>Totaux</b>     | <b>23</b>   | <b>50,5</b>  | <b>79,5</b>  | <b>79,5</b>  | <b>79,5</b>  | <b>61</b> | <b>30</b>  | <b>56</b> | <b>65,5</b>  | <b>59</b> | <b>72</b>  | <b>68,5</b>  | <b>47</b>  | <b>69</b> |       |
| <b>Ho</b>         | <b>60</b>   | <b>60</b>    | <b>60</b>    | <b>60</b>    | <b>60</b>    | <b>60</b> | <b>60</b>  | <b>60</b> | <b>60</b>    | <b>60</b> | <b>60</b>  | <b>60</b>    | <b>60</b>  | <b>60</b> |       |
| <b>Ecart</b>      | <b>1369</b> | <b>90,25</b> | <b>380,3</b> | <b>380,3</b> | <b>380,3</b> | <b>1</b>  | <b>900</b> | <b>16</b> | <b>30,25</b> | <b>1</b>  | <b>144</b> | <b>72,25</b> | <b>169</b> | <b>81</b> |       |

Q501 - Quelle est la valeur empirique du coefficient W de concordance de Kendall sans faire la correction liée à l'existence des ex aequo ?

Q502- Peut-on considérer qu'il existe un classement ordonné significatif correspondant à une certaine homogénéité des points de vue de ces étudiants? Réaliser le test W de Kendall de confrontation des deux hypothèses Ho et H1 qui devront être explicitées, à niveau de risque de 1<sup>ère</sup> espèce  $\alpha=0.01$ .

Ho Les rangements déterminés par l'ordre d'évocation des problèmes sont hétérogènes au sein de l'échantillon,

H<sub>1</sub> Il existe une tendance à l'homogénéité des rangements déterminés par l'ordre d'évocation des problèmes au sein de l'échantillon,

Nous utilisons le test W de Kendall.

|              |         |                      |                                   |
|--------------|---------|----------------------|-----------------------------------|
| $S_{Kmax} =$ | $S_K =$ | $W = S_K / S_{Kmax}$ | $W_{critique} \text{ seuil } 1\%$ |
| 14560        | 4014,5  | <b>0,27572115</b>    | 0,26623254                        |

Ho = hypothèse d'hétérogénéité des rangements et d'absence de concordance est rejetée au seuil de 1%. Nous acceptons à ce niveau de risque le rangement suivant comme représentant celui avec lequel l'ensemble des individus serait plutôt en accord.

Q503- Donner le rangement qui serait celui le plus concordant.

| objectif                      | n°1 | n°2 | n°3 | n°4 | n°5 | n°6 | n°7 | n°8 | n°9 | n°10 | n°11 | n°12 | n°13 | n°14 |
|-------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|------|------|
| <b>RANGEMENT significatif</b> | 1   | 4   | 12  | 12  | 12  | 7   | 2   | 5   | 8   | 6    | 11   | 9    | 3    | 10   |

Q504- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

**TQ6- Comparaison des rangements réalisés par les hommes et les femmes relativement aux catégories de problèmes évoqués.**

Nous considérons les deux rangements obtenus à partir du test de concordance W de Kendall explicitant les ordres concordants respectivement du groupe des femmes et du groupe des hommes. Les résultats sont consignés dans le tableau T9 ci-dessous (avec présence d'ex aequo comme cela se présente dans la thèse).

|                                       | Pb_01 | Pb_02 | Pb_03 | Pb_04 | Pb_05 | Pb_06 | Pb_07 | Pb_08 | Pb_09 | Pb_10 | Pb_11 | Pb_12 | Pb_13 | Pb_14 |
|---------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Rangement concordant du groupe Hommes | 1     | 12    | 13,5  | 13,5  | 9     | 3     | 2     | 5     | 8     | 6     | 11    | 10    | 4     | 7     |
| Rangement concordant du groupe Femmes | 2     | 13    | 12    | 14    | 9     | 5     | 1     | 3     | 10    | 6     | 7     | 11    | 4     | 8     |
| Somme des écarts au carré             | 1     | 1     | 2,25  | 0,25  | 0     | 4     | 1     | 4     | 4     | 0     | 16    | 1     | 0     | 1     |

$$\Delta^2 = 35,5$$

Q601-Peut-on considérer qu'il existe une tendance à l'accord significative entre ces deux groupes ?

Tester les hypothèses correspondantes (Ho et H1) au seuil de 5% à l'aide du test de Rhô de Spearman.

Ho Les rangements déterminés par l'ordre d'évocation des problèmes sont indépendants au sein de l'échantillon entre les hommes et les femmes

H<sub>1</sub> Il existe un lien entre les rangements déterminés par l'ordre d'évocation des problèmes au sein de l'échantillon par les hommes et par les femmes.

Calcul de la valeur empirique du coefficient ρ de Spearman

$$R_s = 1 - \frac{\Delta^2}{n(n^2 - 1)} = 1 - \frac{35,5}{14(14^2 - 1)} = 1 - \frac{35,5}{455} = \frac{419,5}{455} \approx 0,922$$

La valeur critique de Rs au seuil α=0.05 (test bilatéral) est de 0,538.

Q602- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

Nous rejetons l'hypothèse Ho au profit de H1. C'est-à-dire que nous retenons l'idée d'un lien entre l'ordre déterminé par les évocations des problèmes chez les hommes et celui chez les femmes.



Nous rapportons ci-dessous les tableaux statistiques des variables :

**(Tableau n° T1) Tableau statistique de la variable V01 = sexe**

|                  | SEXE    |         |
|------------------|---------|---------|
| Modalités de V01 | 1_femme | 2_homme |
| Effectifs        | 70      | 55      |

**(Tableau n° T2) Tableau statistique de la variable V02 = âge à la date de l'enquête**

| Valeurs de V02 | [23; 30[ | [30; 36[ | [36; 42[ | [42; 48[ | [48; 53[ |
|----------------|----------|----------|----------|----------|----------|
| Effectifs      | 23       | 46       | 19       | 23       | 14       |

**(Tableau n° T5) Tableau statistique de la variable V05 =**

Degré d'utilité attribué à la discipline : Hypermédia en Éducation

| Modalités de V05 | 0  | 1 | 2 | 3 | 4 | 5 | 6 | 7  | 8  | 9  | 10 |
|------------------|--|---|---|---|---|---|---|----|----|----|----|
|                  | le plus faible <span style="float: right;">le plus fort</span> |   |   |   |   |   |   |    |    |    |    |
| Effectifs        | 0  | 4 | 0 | 0 | 0 | 7 | 0 | 11 | 26 | 22 | 55 |

**(Tableau T7) Tableau statistique de la variable V07 :**

Auto-évaluation du niveau de compétences...

| Modalités de V07 | 1_Tout à fait faible | 2_Plutôt faible | 3_Plutôt fort | 4_Tout à fait fort |
|------------------|----------------------|-----------------|---------------|--------------------|
| Effectifs        | 25                   | 30              | 42            | 28                 |

**(Tableau T8 fictif) Tableau statistique de la variable V08= Connaissances**

**Nombre de réussites obtenues à un QCM de 20 items portant sur des connaissances scientifiques et mathématiques**

| scores/20 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|-----------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| effectifs | 0 | 0 | 2 | 5 | 1 | 0 | 0 | 3 | 7 | 7 | 10 | 12 | 8  | 12 | 3  | 15 | 3  | 12 | 10 | 8  | 7  |

**Document n° 2 : résultats du traitement demandé**

**TQ1- De quoi s'agit-il ?**

Q101- Après avoir pris connaissance des informations fournies dans le **document n°1**, identifier et caractériser "succinctement" et clairement, les éléments de cette étude statistique : population, échantillon, individus, variables.

Q102 – Identifier clairement la nature des 9 variables retenues

**Population :**

Ensemble de ceux ou celles (**individus**) qui suivent ou ont suivi la formation mais aussi ceux qui la suivront

**Échantillons:**

Échantillon aléatoire de 125 individus qui ont été soumis à l'enquête

**Variables.**

| Variables quantitatives |           |       | Variables qualitatives  |                |            |
|-------------------------|-----------|-------|-------------------------|----------------|------------|
| discrètes               | continues | rangs | ordinales               | nominales      | textuelles |
| [V08]                   | [V02]     | [V09] | [V05]<br>[V06]<br>[V07] | [V01]<br>[V03] | [V04]      |

La question V09 est un vecteur-variable à 10 composantes dont chaque réalisation est une permutation des nombres de 1 à 10.

**TQ2-Étude de la variable V02 sur l'échantillon**

Q201 Quel est l'**âge moyen** des individus avec quel **écart-type** ?

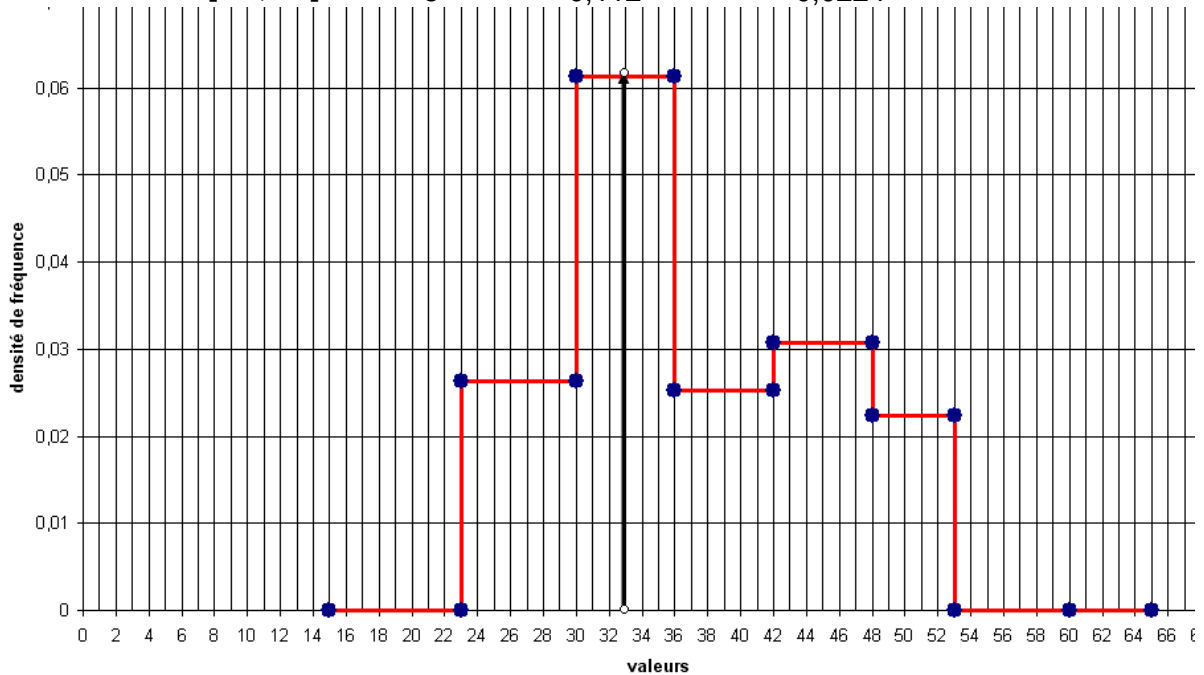
| Intervalles      | valeurs centrales<br>(xi) | effectifs<br>(ni)     | $n_i x_i$      | $n_i(x_i-m)^2$    |
|------------------|---------------------------|-----------------------|----------------|-------------------|
| <b>[23; 30[</b>  | 26,5                      | <b>23</b>             | 609,5          | 2480,031488       |
| <b>[30 ; 36[</b> | 33                        | <b>46</b>             | 1518           | 693,930976        |
| <b>[36 ; 42[</b> | 39                        | <b>19</b>             | 741            | 85,071664         |
| <b>[42 ; 48[</b> | 45                        | <b>23</b>             | 1035           | 1514,997488       |
| <b>[48 ; 53[</b> | 50,5                      | <b>14</b>             | 707            | 2595,536384       |
|                  |                           | <i>effectif total</i> | <i>somme</i>   | <i>somme</i>      |
|                  |                           | <b>125</b>            | <b>4610,5</b>  | <b>7369,568</b>   |
|                  |                           |                       | <i>moyenne</i> | <i>variance</i>   |
|                  |                           |                       | <b>36,884</b>  | <b>58,956544</b>  |
|                  |                           |                       |                | <i>écart-type</i> |
|                  |                           |                       |                | <b>7,678316</b>   |

L'âge moyen est de 36,88 années soit environ 36 ans 11 mois avec un écart-type de l'ordre de 7 ans et 8 mois.

**Q202 Construire un histogramme de la distribution des fréquences**

Tableau T11

| Intervalles      | amplitude | fréquences | densité de fréquences |
|------------------|-----------|------------|-----------------------|
| <b>[23 ; 30[</b> | 7         | 0,184      | 0,026285714           |
| <b>[30 ; 36[</b> | 6         | 0,368      | 0,061333333           |
| <b>[36 ; 42[</b> | 6         | 0,152      | 0,025333333           |
| <b>[42 ; 48[</b> | 6         | 0,184      | 0,030666667           |
| <b>[48 ; 53[</b> | 5         | 0,112      | 0,0224                |



**Q203 Quel est l'âge modal ?**

Nous pouvons situer la valeur modale dans l'intervalle [30 ; 36[ sur lequel la densité de fréquence est maximale : 0,0613. Nous pouvons fournir comme estimation ponctuelle de l'âge modal, le centre de l'intervalle, c'est-à-dire 33 ans.

**Q204 Quel est l'âge médian ?**

En prenant appui sur l'historgramme et recourant à un raisonnement par interpolation linéaire, déterminer une estimation de l'âge médian des bacheliers.

On peut déjà estimer que la durée médiane est située dans l'intervalle [30 ; 36[.

On pose l'équation suivante pour obtenir la médiane Me :

$$(Me - 30) \frac{46}{125} \frac{1}{6} + \frac{23}{125} = 0,5$$

$$(Me - 30) \frac{46}{125} \frac{1}{6} = 0,5 - \frac{23}{125} \text{ d'où } (Me - 30) \frac{46}{6} = (0,5)(125) - 23$$

$$\text{ou encore } (Me - 30) = \frac{6}{46} [62,5 - 23] = \frac{6}{46} (39,5) \approx 5,152$$

$$Me = 30 + \frac{237}{46} \text{ soit } Me \approx 35,152$$

On peut estimer que l'âge médian est de l'ordre de 35 ans et 2 mois

**Q205- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.**

**TQ3 - Étude de l'homogénéité des points de vue relatifs à l'ordre décroissant d'importance de 10 disciplines constituant le curriculum de la formation**

Nous rapportons ici les réponses fournies (fictivement) par un sous-échantillon de 14 individus extrait de l'échantillon global des diplômés du Mestrado de la spécialité mathématiques. Chaque répondant ignore le rangement proposé par les autres. Les résultats concernant les rangs sont reportés dans le tableau T9 ci-dessous :

**Tableau T9**

| Disciplines |     |     |     |     |     |     |     |     |     |     |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| individus   | D01 | D02 | D03 | D04 | D05 | D06 | D07 | D08 | D09 | D10 |
| Mestr_001   | 6   | 1   | 2   | 3   | 8   | 7   | 4   | 10  | 9   | 5   |
| Mestr_122   | 6   | 9   | 1   | 4   | 5   | 7   | 10  | 8   | 3   | 2   |
| Mestr_073   | 1   | 2   | 8   | 7   | 5   | 4   | 3   | 6   | 9   | 10  |
| Mestr_014   | 3   | 2   | 10  | 8   | 1   | 5   | 6   | 7   | 4   | 9   |
| Mestr_105   | 3   | 10  | 7   | 5   | 6   | 8   | 4   | 2   | 1   | 9   |
| Mestr_026   | 7   | 1   | 6   | 3   | 2   | 5   | 4   | 8   | 9   | 10  |
| Mestr_017   | 6   | 1   | 3   | 4   | 8   | 5   | 2   | 7   | 10  | 9   |
| Mestr_038   | 9   | 1   | 2   | 3   | 6   | 8   | 10  | 4   | 5   | 7   |
| Mestr_019   | 7   | 5   | 2   | 4   | 6   | 8   | 1   | 9   | 10  | 3   |
| Mestr_015   | 7   | 1   | 10  | 3   | 6   | 8   | 2   | 4   | 5   | 9   |
| Mestr_111   | 5   | 2   | 3   | 8   | 1   | 6   | 4   | 7   | 10  | 9   |
| Mestr_012   | 6   | 9   | 4   | 2   | 3   | 7   | 1   | 8   | 5   | 10  |
| Mestr_003   | 1   | 9   | 3   | 4   | 7   | 6   | 5   | 8   | 10  | 2   |
| Mestr_064   | 3   | 4   | 6   | 10  | 5   | 9   | 2   | 8   | 1   | 7   |

**Énoncés des disciplines**

|   |
|---|
| D01 Hypermédia en éducation                                 |
| D02 Théories de l'apprentissage et de l'enseignement        |
| D03 Théorie et systématisation de l'enseignement            |
| D04 Méthodologie de la recherche                            |
| D05 Séminaires communs                                      |
| D06 Accompagnement de l'élaboration du mémoire              |
| D07 Thèmes de méthodologie d'enseignement des mathématiques |
| D08 Fondements de Calcul                                    |
| D09 Fondements de Mathématiques                             |
| D10 Fondements de Géométrie                                 |

Q301 Quel nom donne-t-on à ce tableau T9 ? Quelles informations contient-il ?

C'est un tableau de rangs qui donne les rangements réalisés par les  $p = 14$  individus relativement à  $n = 10$  disciplines. Chaque ligne est donc une permutation de  $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ .

| Disciplines                                     | D01 | D02 | D03 | D04 | D05 | D06 | D07 | D08 | D09 | D10 | Somme totale de tous rangs attribués |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------------------------------------|
| <b>Sommes des rangs attribués à chaque item</b> | 70  | 57  | 67  | 68  | 69  | 93  | 58  | 96  | 91  | 101 | 770                                  |

Q302 Peut-on considérer qu'il existe un classement ordonné significatif correspondant à une certaine homogénéité des points de vue des étudiants encore en formation ou qui en sont issus? Avec le test W de Kendall tester respectivement au seuil de 1% et au seuil de 5% afin de choisir laquelle des deux hypothèses  $H_0$  et  $H_1$  qu'il faut bien expliciter auparavant, est la plus vraisemblable.

Deux hypothèses sont en concurrence :

$H_0$  Les rangements attribués par les bacheliers interrogés sont indépendants et constituent un tout hétérogène.

$H_1$  Il existe une tendance à l'accord entre les rangements attribués.

Pour mettre en place la procédure de test fondée sur celle proposée par Kendall à partir de la statistique  $W$ , il nous faut obtenir la somme totale des  $p=14$  rangs obtenus par chacun des  $n=10$  items.

Sous l'hypothèse  $H_0$  ces sommes seraient toutes identiques et seraient égales à

$$\frac{p \frac{n(n+1)}{2}}{n} = \frac{14(55)}{10} = 77$$

On mesure ensuite un écart global  $S_K$  entre ces valeurs obtenues et les valeurs espérées sous  $H_0$ .

$$S_K = (70-77)^2 + (57-77)^2 + (67-77)^2 + (68-77)^2 + (69-77)^2 + (93-77)^2 + (58-77)^2 + (96-77)^2 + (91-77)^2 + (101-77)^2$$

$$S_K = 49+400+100+81+64+256+361+361+196+576$$

$$S_K = 2444$$

On mesure l'écart global maximum qui correspond à l'accord parfait c'est-à-dire le cas où déca-uplet des sommes serait une permutation de (14, 28, 42, 56, 70, 84, 98, 112, 126, 140) :

$$S_{K \max} = (14-77)^2 + (28-77)^2 + (42-77)^2 + (56-77)^2 + (70-77)^2 + (84-77)^2 + (98-77)^2 + (112-77)^2 + (126-77)^2 + (140-77)^2$$

$$S_{K \max} = 16170$$

L'écart maximum peut être obtenu directement par la formule :

$$S_{K \max} = \frac{1}{12} p^2 (n^3 - n)$$

$$S_{K \max} = \frac{1}{12} 14^2 (10^3 - 10) = 16170$$

On calcule ensuite la valeur empirique de la statistique  $W$  de Kendall :

$$W = \frac{S_K}{\frac{1}{12} p^2 (n^3 - n)}$$

Pour information, retournons aux propriétés de la variable  $W$ . On connaît un certain nombre de caractéristiques :

D'une part :  $0 \leq W \leq 1$

D'autre part :

| Espérance            | Variance                                | Moment centré d'ordre 3                     | Moment centré d'ordre 4  |
|----------------------|---|---|--|
| $E(W) = \frac{1}{p}$ | $\sigma^2(W) = \frac{2(p-1)}{p^3(n-1)}$ | $\mu_3(W) = \frac{8(p-1)(p-2)}{p^5(n-1)^2}$ | $\mu_4(W) = \frac{12(p-1)^2}{p^6(n-1)^2} + \frac{48(p-1)(p-2)(p-3)}{p^7(n-1)^3} - \frac{48(p-1)}{p^7(n+1)(n-1)^2}$ |

Notons que  $w = 1$  correspond à la situation d'accord parfait sur le rangement des  $n$  objets pour les  $p$  critères.

Si  $w=0$  correspond au cas où chaque colonne a même total, c'est-à-dire que les objets seraient tous *ex æquo*. De faibles valeurs de  $W$  suggèrent encore cette indiscernabilité entre les objets, c'est-à-dire l'idée d'une indépendance des rangements.

Pour décider du rejet ou non de l'hypothèse nulle  $H_0$  d'indépendance des  $p$  rangements, on compare la valeur empirique obtenue avec la valeur critique correspondant à un niveau  $\alpha$  du risque de 1<sup>ère</sup> espèce :

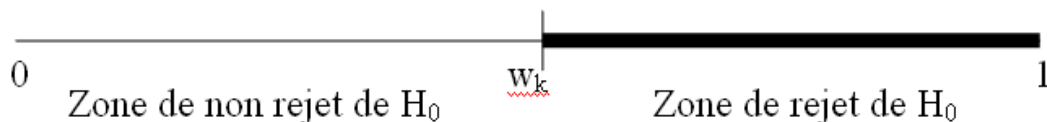
a) Pour  $n \leq 7$  on utilise la table ci-dessous qui fournit les valeurs critiques  $w_k$  telles que  $P(W \geq w_k) = \alpha = 0,05$  :

| p   | n     | 3 | 4     | 5     | 6     |
|-----|-------|---|-------|-------|-------|
| 3   | 1     |   | 0,750 | 0,600 | 0,500 |
| 4   | 0,822 |   | 0,619 | 0,500 | 0,421 |
| 5   | 0,716 |   | 0,553 | 0,449 | 0,377 |
| 6   | 0,660 |   | 0,512 | 0,418 | 0,351 |
| 7   | 0,626 |   | 0,484 | 0,395 | 0,332 |
| 8   | 0,595 |   | 0,461 | 0,378 | 0,319 |
| 9   | 0,576 |   | 0,447 | 0,365 | 0,307 |
| 10  | 0,560 |   | 0,434 | 0,354 | 0,299 |
| 11  | 0,548 |   | 0,425 | 0,346 | 0,287 |
| 12  | 0,535 |   | 0,415 | 0,336 | 0,287 |
| 13  | 0,527 |   | 0,409 | 0,332 | 0,280 |
| 14  | 0,520 |   | 0,402 | 0,327 | 0,275 |
| 15  | 0,514 |   | 0,395 | 0,322 | 0,272 |
| 20  | 0,49  |   | 0,37  | 0,30  | 0,25  |
| 40  | 0,43  |   | 0,33  | 0,26  | 0,22  |
| 60  | 0,41  |   | 0,31  | 0,25  | 0,21  |
| 100 | 0,38  |   | 0,29  | 0,24  | 0,20  |
| ∞   | 0,33  |   | 0,25  | 0,20  | 0,17  |

b) Pour  $n \leq 7$  et  $2 < p \leq 20$  la variable  $\frac{(p-1)W}{1-W}$  est distribuée comme la variable de Fisher-Snédecor  $F(n_1 = n-1-\frac{2}{p}; n_2 = (p-1)(n-1-\frac{2}{p}))$

c) Pour  $n > 7$  on utilise la variable  $p(n-1)W$  qui est une variable de Pearson  $\chi^2_{n-1}$   
La table fournie en annexe qui donne les valeurs critiques  $w_k$  telles que  $P(W > w_k) = \alpha$  est déduite de cette propriété.

Globalement la règle de décision s'appuie sur la configuration suivante de l'espace de décision :

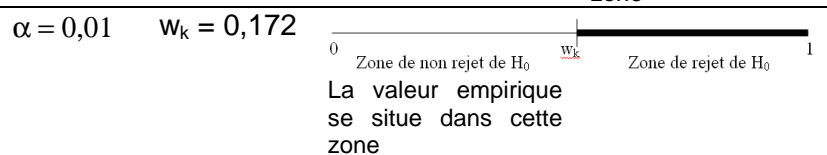
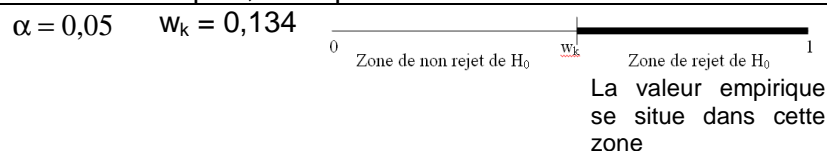


Dans le cas où l'on est conduit à rejeter l'hypothèse d'indépendance des rangements, on utilise souvent la règle de classement suivante:

**Les objets sont classés dans l'ordre défini par les totaux des colonnes.**

Retour à la situation étudiée : 
$$W = \frac{S_K}{\frac{1}{12} p^2 (n^3 - n)} = \frac{2444}{16170} \approx 0,15114$$

Dans la table de valeurs critiques, nous pouvons lire :



Il ressort qu'au niveau de risque  $\alpha = 0,05$  nous rejetons l'hypothèse  $H_0$  c'est-à-dire que nous pouvons retenir l'idée d'une tendance à l'accord qui pourrait être réalisé autour du rangement suivant :

|      | D01 | D02 | D03 | D04 | D05 | D06 | D07 | D08 | D09 | D10 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Rang | 6   | 1   | 3   | 4   | 5   | 8   | 2   | 9   | 7   | 10  |

En revanche, au niveau  $\alpha = 0,01$ , nous ne rejetons pas l'hypothèse  $H_0$  c'est-à-dire que nous ne pouvons retenir l'idée d'une tendance à l'accord qui pourrait être réalisé autour du rangement obtenu. Nous prenons alors un risque de 2<sup>nd</sup>e espèce, de niveau  $\beta$ , dont nous pouvons préciser la valeur.

Q303 S'il existe un rangement significatif pour un niveau de risque donné, quel est ce rangement ?

| Rang        | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Disciplines | D02 | D07 | D03 | D04 | D05 | D01 | D09 | D06 | D08 | D10 |

Ou encore :

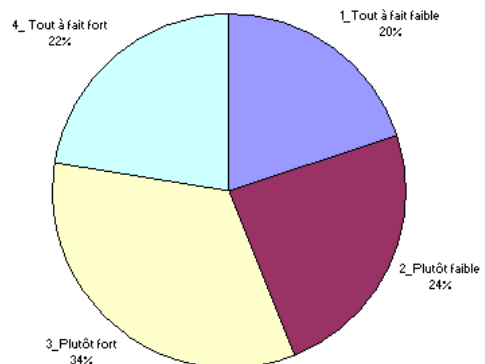
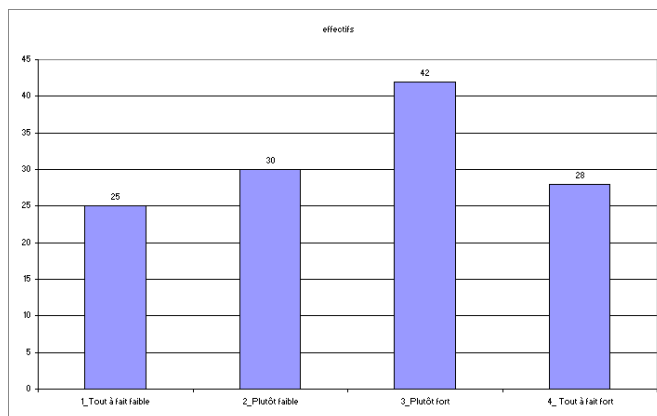
- 1 D02 Théories de l'apprentissage et de l'enseignement
- 2 D07 Thèmes de méthodologie d'enseignement des mathématiques
- 3 D03 Théorie et systématisation de l'enseignement
- 4 D04 Méthodologie de la recherche
- 5 D05 Séminaires communs
- 6 D01 Hypermédia en éducation
- 7 D09 Fondements de Mathématiques
- 8 D06 Accompagnement de l'élaboration du mémoire
- 9 D08 Fondements de Calcul
- 10 D10 Fondements de Géométrie

Q304- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

**TQ4 - Étude de la variable V07 = Auto-évaluation du niveau de compétences**

Q401 De quoi s'agit-il dans cette étude guidée par les items Q402 et Q403 ? Quel est l'objectif de cette procédure ? Quelles précisions sont requises dans les énoncés des résultats ?

| Modalités de V07 | 1_ Tout à fait faible | 2_ Plutôt faible | 3_ Plutôt fort | 4_ Tout à fait fort |
|------------------|-----------------------|------------------|----------------|---------------------|
| effectifs        | 25                    | 30               | 42             | 28                  |
| fréquences       | 0,2                   | 0,24             | 0,34           | 0,22                |
| %                | 20%                   | 24%              | 33,6%          | 22,4%               |



Q402 Donner la proportion des individus qui se considèrent comme «3\_plutôt fort» dans l'échantillon.

Nombre d'individus ayant répondu « 3\_Plutôt fort » est 42 parmi les 125 qui ont répondu.  
Proportion dans l'échantillon :

$$f = \frac{42}{125} = 0.336 = 33.6\%$$

Q403 Donner une estimation ponctuelle de la proportion des individus qui se considèrent comme «3\_plutôt fort» dans la population.

Estimation ponctuelle de la proportion dans la population :

**33.6%**

Q404 Donner une estimation par intervalle de confiance à 95% ou à 99% (selon votre choix) de la proportion des individus qui se considèrent comme «3\_plutôt fort» dans la population.

$$\left[ f - \kappa \sqrt{\frac{f(1-f)}{n-1}}; f + \kappa \sqrt{\frac{f(1-f)}{n-1}} \right]$$

Au seuil de confiance  $1-\alpha=0,95$  la valeur critique  $k=1,96$

$$\left[ 0.336 - 1.96 \sqrt{\frac{0.336(1-0.336)}{125-1}}; 0.336 + 1.96 \sqrt{\frac{0.336(1-0.336)}{125-1}} \right]$$

or  $1,96 \left( \sqrt{\frac{0.336(0.664)}{124}} \right) \approx 1,96(0,042417281) \approx 0,083137872$

d'où [ 0,2528 ; 0,4192] c'est à dire entre 25,28% et 41,92%

Avec un niveau de confiance de 95%, nous estimons que la proportion se situe environ entre 25,28% et 41,92%

Au seuil de confiance  $1-\alpha=0,99$  la valeur critique  $k=2,57$

$$\left[ 0.336 - 2.57 \sqrt{\frac{0.336(1-0.336)}{125-1}}; 0.336 + 2.57 \sqrt{\frac{0.336(1-0.336)}{125-1}} \right]$$

or  $2,57 \left( \sqrt{\frac{0,336(0,664)}{124}} \right) \approx 2,57(0,0424417281) \approx 0,10901214$

d'où [ 0,2269 ; 0,4451] C'est à dire entre 22,69% et 44,51%

Avec un niveau de confiance de 99%, nous estimons que la proportion se situe environ entre 22,69% et 44,51%

Q405 Quelle incidence a le niveau de confiance ?

Accroître le niveau de confiance conduit à accroître l'amplitude de l'intervalle de confiance, c'est-à-dire diminuer le niveau de précision.

Q406- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

**TQ5- Étude du lien entre sexe et les spécialités mathématiques et sciences physiques (croisement des variables V01 et V03)**

| Tableau T10 |       | V03 = Spécialité   |               |     |
|-------------|-------|--------------------|---------------|-----|
|             |       | Sciences Physiques | Mathématiques |     |
| V01<br>Sexe | Femme | 18                 | 52            | 70  |
|             | Homme | 33                 | 22            | 55  |
|             |       | 51                 | 74            | 125 |

Q501 Quel nom donne-t-on à ce tableau ? Quelles informations contient-il ?

C'est un tableau de contingence. Tableau de distribution des effectifs de la conjonction de deux variables. Tableau des effectifs conjoints.

Q502 A partir de ces données issues de l'échantillon des répondants, peut-on considérer qu'il y a une différence selon le sexe ?

Il s'agit de tester les hypothèses correspondantes (Ho et H1 qu'il faut expliciter) au seuil de 5% ou au seuil de 1% selon votre choix, à l'aide du test d'indépendance du Khi-deux.

Nous mettons en œuvre un test d'indépendance (Test du Khi-deux) pour décider laquelle des deux hypothèses est la plus plausible :

Ho Les hommes et les femmes ont des points de vue homogènes relatifs au choix en mathématiques et sciences physique

H1 Les hommes et les femmes ont des points de vue hétérogènes relatifs à relatifs au choix en mathématiques et sciences physique

Tableau des effectifs observés :

| Tableau T10 |       | V03 = Spécialité   |               |     |
|-------------|-------|--------------------|---------------|-----|
|             |       | Sciences Physiques | Mathématiques |     |
| V01<br>Sexe | Femme | 18                 | 52            | 70  |
|             | Homme | 33                 | 22            | 55  |
|             |       | 51                 | 74            | 125 |

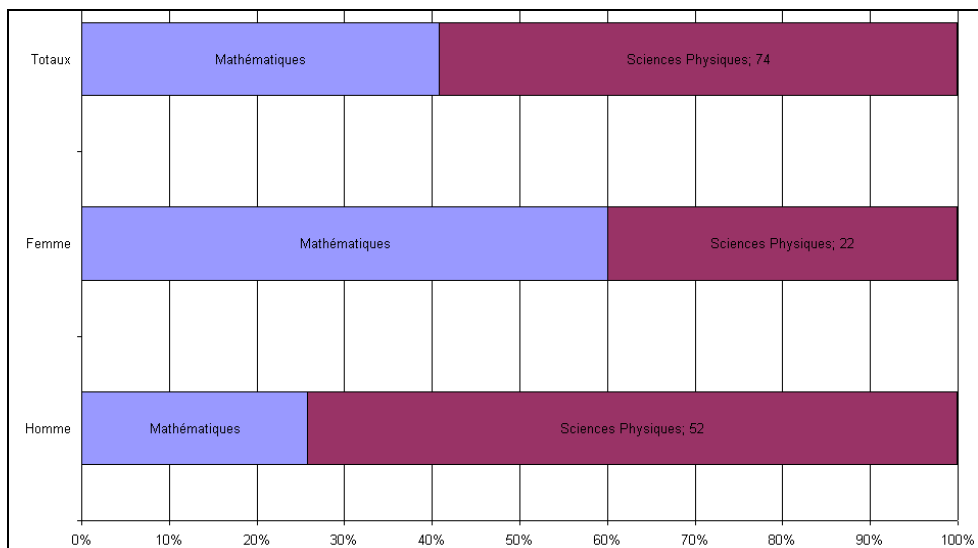


Tableau des effectifs théoriques sous l'hypothèse  $H_0$  :

| Tableau T10bis |       | V03 = Spécialité   |               |     |
|----------------|-------|--------------------|---------------|-----|
|                |       | Sciences Physiques | Mathématiques |     |
| V01<br>Sexe    | Femme | 28,56              | 41,44         | 70  |
|                | Homme | 22,44              | 32,56         | 55  |
|                |       | 51                 | 74            | 125 |

Tableau des contributions absolues :

| Tableau T10 |       | V03 = Spécialité   |               |                |
|-------------|-------|--------------------|---------------|----------------|
|             |       | Sciences Physiques | Mathématiques |                |
| V01<br>Sexe | Femme | 3,9045             | 2,6909        | 6,5954         |
|             | Homme | 4,9694             | 3,4249        | 8,3943         |
|             |       | 8,8739             | 6,1158        | <b>14,9897</b> |

Au seuil de  $\alpha = 5\%$  la valeur critique lue dans la table du Khi-deux ddl = 1 est de  $k = 3,84$ .

Au seuil de  $\alpha = 1\%$  la valeur critique lue dans la table du Khi-deux ddl = 1 est de  $k = 6,64$ .

Ici la valeur empirique est de **14,98** :

Elle est donc supérieure à la valeur théorique sous  $H_0$  au niveau de risque 0.05. On rejette l'hypothèse d'homogénéité/indépendance à ce niveau de risque de 1<sup>ère</sup> espèce.

Mais elle est aussi, au niveau de risque 0.01, supérieure à la valeur théorique sous  $H_0$ . On rejette l'hypothèse d'homogénéité/indépendance à ce niveau de risque de 1<sup>ère</sup> espèce.

Q503- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**

**TQ6-Étude de la variable V08 sur l'échantillon**

Q601 Quel est le **score moyen** des individus avec quel **écart-type** ?

Q602 Construire un **diagramme** de la distribution des fréquences

Q603 Quel est le **score modal** ?

Q604 Quel est le **score médian** ?

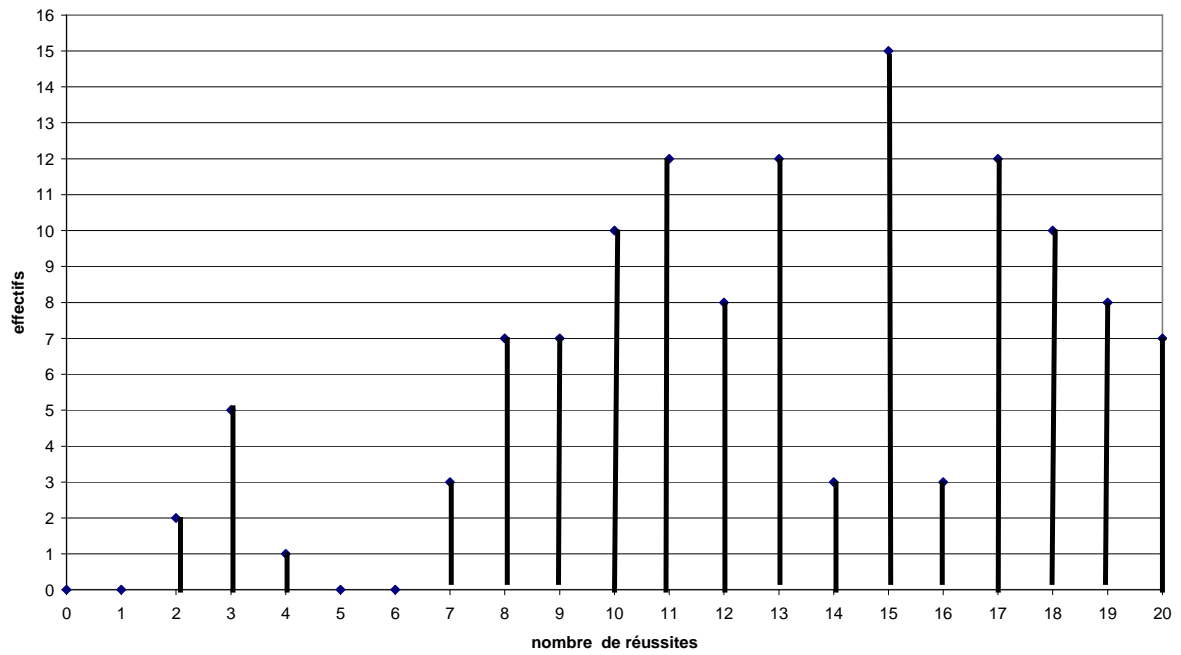
Q605- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

|                                      |            |
|--------------------------------------|------------|
| Moyenne                              | 13,104     |
| Variance                             | 20,205184  |
| Ecart-Type                           | 4,49501769 |
| Minimum                              | 2          |
| 1 <sup>er</sup> quartile Q1          | 10         |
| 2 <sup>ème</sup> quartile Q2=médiane | 13         |
| 3 <sup>ème</sup> quartile Q3         | 17         |
| Maximum                              | 20         |
| Mode                                 | 15         |

Cours de Jean-Claude Régnier  
Situations Problèmes & Documents autocorrectifs  
19/34

---

scores/20



### Document n° 1 : situation de l'étude

La présente situation s'inspire d'une recherche fondée pour partie sur une enquête par entretien auprès de 20 individus et pour le reste sur une enquête par questionnaire impliquant 430 individus visant la construction de données sur les représentations sociales des « maîtres généralistes » et des « maîtres E » à l'égard de l'enfant « aidé ». Cette recherche a fait l'objet d'une thèse réalisée par Jean-Bruno Bernard et soutenue le 24 avril 2007 à l'Université Lyon2. Elle est consultable sur le site des thèses en ligne de Lyon2. La population d'étude est celle des « enseignants du premier de degré », plus particulièrement les deux sous-populations constituées par les « maîtres généralistes » et par les « maîtres E ». Les données ont été construites par le moyen d'un questionnaire comportant 28 questions pour les « maîtres généralistes » et 23 pour les « maîtres E ». Le questionnaire a été soumis à deux échantillons qui furent considérés comme représentatifs des sous-populations respectives : 300 « maîtres généralistes » et 130 « maîtres E ».

Parmi toutes les variables étudiées, nous ne considérons ici que quelques-unes réellement étudiées auxquelles nous ajoutons d'autres dont nous suggérons la mise en œuvre dans une poursuite d'exploration.

Les deux questionnaires que nous nommons QMG et QME, utilisés pour cette recherche portaient, entre autre, sur les variables suivantes :

| <i>code de la variable</i> | <i>descriptif succinct</i>  |
|----------------------------|---|
| S01                        | <b>Sexe</b> : 1 = homme 2 = femme   |
| S02                        | <b>Ancienneté</b> dans la fonction (années)   |
| S03                        | <b>Statut</b> : Instituteur (P1) Professeur d'école (P2).   |
| S04                        | <b>Zone géographique d'intervention</b> : rurale (Z1) ; urbain (Z2) ; ZEP (Z3)  |
| V02                        | <b>Différence entre difficulté d'apprentissage et échec scolaire</b> : 1=oui ; 2=non ; 3=je ne sais pas   |
| V 04                       | <b>Certaines matières sont-elles plus représentatives que d'autres quand on parle de difficultés scolaires ?</b> 1=oui ; 2=non ; 3=je ne sais pas |
| V07                        | <b>Considérez-vous qu'il existe un profil type d'enfants prédisposés à être en difficulté d'apprentissage ?</b> 1=oui ; 2=non ; 3=je ne sais pas  |
| V08 QME                    | <b>Vous est-il arrivé des sentir des tensions entre vous et le « maître généraliste » ?</b> 1=oui ; 2=non ; 3=je ne sais pas                      |
| V12 QMG                    | <b>Vous est-il arrivé de sentir des tensions entre vous et le « maître E » ?</b> 1=oui ; 2=non ; 3=je ne sais pas                                 |
| V12 QME                    | <b>Vous considérez-vous en tant que « maître E » comme un enseignant ?</b> 1=oui ; 2=non ; 3=je ne sais pas                                       |
| V16 QMG                    | <b>Considérez-vous le « maître E » comme un enseignant ?</b> 1=oui ; 2=non ; 3=je ne sais pas   |
| V21QMG<br>V18QME           | <b>De quoi dépend l'efficacité du travail en équipe ?</b>   |

Nous rapportons ci-dessous les tableaux statistiques des variables

**Tableaux statistiques de la variable S01**

| <b>Tableau n° T1</b>     | S01=Sexe  | homme | femme |
|--------------------------|-----------|-------|-------|
| « Maîtres généralistes » | Effectifs | 100   | 200   |
| « Maîtres E »            | Effectifs | 76    | 44    |

**Tableaux statistiques de la variable S02**

| <b>Tableau n° T2</b>     | S02= Ancienneté | [1 ; 10[ | [10 ; 20[ | [20 ; 30[ | [30; 40[ |
|--------------------------|-----------------|----------|-----------|-----------|----------|
| « Maîtres généralistes » | effectifs       | 95       | 68        | 66        | 71       |
| « Maîtres E »            | effectifs       | 27       | 40        | 43        | 20       |

**Tableaux statistiques de la variable S03**

| <b>Tableau n° T3</b>     | S03=Statut | Instituteur | Professeur des écoles |
|--------------------------|------------|-------------|-----------------------|
| « Maîtres généralistes » | Effectifs  | 88          | 212                   |
| « Maîtres E »            | Effectifs  | 31          | 99                    |

**Tableaux statistiques de la variable S04**

| <b>Tableau n° T4</b>     | S04=Zone  | rurale | urbaine | ZEP |
|--------------------------|-----------|--------|---------|-----|
| « Maîtres généralistes » | Effectifs | 166    | 85      | 49  |
| « Maîtres E »            | Effectifs | 44     | 45      | 41  |

**Tableaux statistiques de la variable V02**

| <b>Tableau n° T5</b>     | V02=différence | oui | non | Je ne sais pas |
|--------------------------|----------------|-----|-----|----------------|
| « Maîtres généralistes » | Effectifs      | 262 | 29  | 5              |
| « Maîtres E »            | Effectifs      | 129 | 0   | 1              |

**Tableaux statistiques de la variable V04**

| <b>Tableau n° T6</b>     | V04=matières | oui | non | Je ne sais pas |
|--------------------------|--------------|-----|-----|----------------|
| « Maîtres généralistes » | Effectifs    | 244 | 46  | 9              |
| « Maîtres E »            | Effectifs    | 125 | 4   | 0              |

**Tableaux statistiques de la variable V07**

| <b>Tableau n° T7</b>     | V07=profil-type | oui | non | Je ne sais pas |
|--------------------------|-----------------|-----|-----|----------------|
| « Maîtres généralistes » | Effectifs       | 108 | 160 | 30             |
| « Maîtres E »            | Effectifs       | 52  | 57  | 14             |

**Tableaux statistiques de la variable « Tensions ressenties »**

| <b>Tableau n° T8</b>     | V08QME/V12QMG | oui | non | Je ne sais pas |
|--------------------------|---------------|-----|-----|----------------|
| « Maîtres généralistes » | Effectifs     | 61  | 221 | 11             |
| « Maîtres E »            | Effectifs     | 101 | 26  | 0              |

**Tableaux statistiques de la variable « Maître E est un enseignant »**

| <b>Tableau n° T9</b>     | V12QME/V16QMG | oui | non | Je ne sais pas |
|--------------------------|---------------|-----|-----|----------------|
| « Maîtres généralistes » | Effectifs     | 286 | 4   | 2              |
| « Maîtres E »            | Effectifs     | 123 | 4   | 0              |

**Document n° 2 : traitement demandé**

**TQ1- De quoi s'agit-il ?**

Q101- Après avoir pris connaissance des informations fournies dans le document n°1, identifier et caractériser "succinctement" et clairement, les éléments de cette étude statistique : population, échantillon, individus, variables.

Q102- Quel traitement proposeriez-vous pour le tableau T7 ?

**TQ2-Étude de la variable S02 « ancienneté »**

Q201 Quelle est la valeur de l'ancienneté moyenne et celle de l'écart-type ?

Q202- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

**TQ3- Étude de la question de l'existence de profil-type**

Q301- A partir de l'échantillon des 321 réponses, peut-on considérer qu'il y a ou non une différence de point de vue entre les deux sous-populations ?

En exploitant les données du tableau T7, tester les hypothèses correspondantes ( $H_0$  et  $H_1$ ) au seuil de risque 5% à l'aide du test d'homogénéité du Khi-deux.

Q302- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

**TQ4 - Étude de la variable V07 « Existence de profil type»**

Q401 Donner la proportion des individus déclarant « oui, il existe un profil type... » dans l'échantillon des « maîtres généralistes ».

Q402 Donner une estimation ponctuelle de la proportion des individus déclarant « oui, il existe un profil type... » dans la population. des « maîtres généralistes »

Q403 Donner une estimation par intervalle de confiance à 95% de la proportion des individus « oui, il existe un profil type... » dans la population des « maîtres généralistes »

Q404- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

**TQ5 - Étude de l'homogénéité des points de vue relatifs l'efficacité du travail en équipe**

À partir d'un sous-échantillon aléatoire sans remise de 12 enseignants extrait de l'échantillon des 300 « maîtres généralistes », nous supposons avoir pu recueillir les rangements par ordre d'importance décroissante de 8 items caractérisant ce dont dépend un travail en équipe efficace

|               |  |
|---------------|--|
|               | L'efficacité du travail en équipe dépend avant tout... |
| <i>Item 1</i> | De l'âge des enseignants                               |
| <i>Item 2</i> | De l'ancienneté des enseignants                        |
| <i>Item 3</i> | Du sexe des enseignants                                |
| <i>Item 4</i> | De la formation des enseignants                        |
| <i>Item 5</i> | Du statut des enseignants                              |
| <i>Item 6</i> | Du lieu géographique d'enseignement                    |
| <i>Item 7</i> | D'autres facteurs                                      |
| <i>Item 8</i> | De facteurs que j'ignore                               |

Les résultats (fictifs) concernant les rangs sont reportés dans le tableau n°T10 ci-dessous :

| <b>Tableau T10</b> | <i>Item</i> | <i>Item</i> | <i>Item</i> | <i>Item</i> | <i>Item</i> | <i>Item</i> | <i>Item</i> | <i>Item</i> |
|--------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| <b>individus</b>   | <b>1</b>    | <b>2</b>    | <b>3</b>    | <b>4</b>    | <b>5</b>    | <b>6</b>    | <b>7</b>    | <b>8</b>    |
| MG18               | 2           | 1           | 3           | 4           | 7           | 6           | 5           | 8           |
| MG32               | 3           | 7           | 4           | 5           | 2           | 1           | 8           | 6           |
| MG15               | 8           | 2           | 6           | 1           | 4           | 5           | 3           | 7           |
| MG06               | 2           | 1           | 3           | 8           | 5           | 4           | 7           | 6           |
| MG65               | 4           | 7           | 1           | 6           | 5           | 2           | 8           | 3           |
| MG152              | 3           | 2           | 1           | 4           | 7           | 6           | 5           | 8           |
| MG251              | 1           | 5           | 7           | 6           | 8           | 2           | 3           | 4           |
| MG99               | 6           | 2           | 3           | 4           | 5           | 1           | 8           | 7           |
| MG63               | 6           | 1           | 3           | 2           | 7           | 4           | 5           | 8           |
| MG25               | 3           | 2           | 1           | 4           | 7           | 6           | 5           | 8           |
| MG17               | 3           | 1           | 5           | 6           | 2           | 4           | 8           | 7           |
| MG211              | 2           | 4           | 8           | 1           | 5           | 7           | 3           | 6           |

Q501 - Quelle est la valeur empirique du coefficient W de concordance de Kendall ?

Q502- Peut-on considérer qu'il existe un classement ordonné significatif correspondant à une certaine homogénéité des points de vue de ces enseignants? Tester cette hypothèse à niveau de risque de 1<sup>ère</sup> espèce  $\alpha=0.01$

Q503- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

**TQ6- Représentations des difficultés d'apprentissage et de l'échec scolaire.**

Nous considérons l'échantillon des 20 enseignants ayant participé à l'enquête par entretien. Dans les discours produits, nous avons dénombré pour chaque individu, les mots qui caractérisent les difficultés scolaires en les différenciant nettement de l'échec scolaire. Ensuite un rangement de ces 20 enseignants a été réalisé à partir de la variable "nombre de mots utilisés" Les résultats sont consignés dans le tableau T11 (étude et résultats fictifs)

| <b>Tableau T11</b> | <b>Catégorie</b> | <b>Rang</b> | <b>Tableau T11</b> | <b>Catégorie</b> | <b>Rang</b> |
|--------------------|------------------|-------------|--------------------|------------------|-------------|
| <b>individus</b>   |                  |             | <b>individus</b>   |                  |             |
| Ind01              | MG               | 1           | Ind11              | MG               | 11          |
| Ind02              | MG               | 2           | Ind12              | MG               | 12          |
| Ind03              | MG               | 3           | Ind13              | ME               | 13          |
| Ind04              | ME               | 4           | Ind14              | MG               | 14          |
| Ind05              | MG               | 5           | Ind15              | ME               | 15          |
| Ind06              | MG               | 6           | Ind16              | ME               | 16          |
| Ind07              | ME               | 7           | Ind17              | ME               | 17          |
| Ind08              | MG               | 8           | Ind18              | MG               | 18          |
| Ind09              | MG               | 9           | Ind19              | MG               | 19          |
| Ind10              | MG               | 10          | Ind20              | MG               | 20          |

Q601-Peut-on considérer qu'il existe une différence significative entre ces deux groupes « maîtres généralistes »[MG] et « maîtres E » [ME] ?

Tester les hypothèses correspondantes (Ho et H1) au seuil de 5% à l'aide du test de Mann et Whitney.

Q602- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

---

# Vocabulaire élémentaire de statistique

## *Cours de Jean-Claude Régnier*

### **Amplitude d'une classe (intervalle d'une variable continu)**

C'est la longueur d'un intervalle. L'amplitude de la classe [20 ;55[ est de 55-20=30

### **Caractère qualitatif nominal**

*Voir* variable qualitative nominale

### **Caractère qualitatif ordinal**

*Voir* variable qualitative ordinale

### **Caractère quantitatif continu**

*Voir* variable quantitative continue

### **Caractère quantitatif discret**

*Voir* variable quantitative discrète

### **Caractère statistique**

*Voir* variable statistique

C'est ce qui est observé sur les individus d'une population ou d'un échantillon

### **Centile**

Les centiles C1, C2, ..., C99 partagent une série statistique en 100 parties d'effectifs égaux.

### **Khi-deux**

La variable du Khi-deux est utilisée par l'intermédiaire de tables statistiques pour le test du Khi-deux (*voir* test du Khi-deux)

### **Classe (intervalle) modale**

C'est la classe ou intervalle d'une variable continue d'effectif ou de fréquence maximum. Elle correspond sur l'histogramme de la distribution d'un caractère quantitatif continu au « rectangle de surface maximale ».

### **Classes ou intervalles d'une variable continue**

Pour un caractère (une variable) quantitatif continu, il convient de découper l'ensemble des valeurs possibles en intervalles ou classes d'amplitudes égales ou inégales. Par exemple si la distribution des âges de la population étudiée et comprise entre 0 et moins de 100 ans, on pourra construire des classes : [0 ;15[ , [15 ;25[, [25 ;35[, [35 ;45[ etc....

### **Coefficient de corrélation linéaire de Bravais-Pearson**

Le coefficient de corrélation entre deux variables quantitatives X et Y est le nombre noté r

$$\text{tel que : } r = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$$

On peut montrer que :  $-1 \leq r \leq +1$

*Interprétation* : Le coefficient de corrélation linéaire est un nombre sans unité ; Si r est nul ou proche de zéro, les deux variables étudiées sont non-corrélées linéairement. Le nuage de points n'a pas de direction privilégiée. Si r est proche de 1 ou -1, les deux variables X et Y sont bien corrélées linéairement, c'est-à-dire qu'il existe entre elles une liaison de type linéaire. Le nuage de points est presque aligné le long d'une droite (ascendante si r proche de 1, et descendante si r proche de -1)

### **Coefficient de variation**

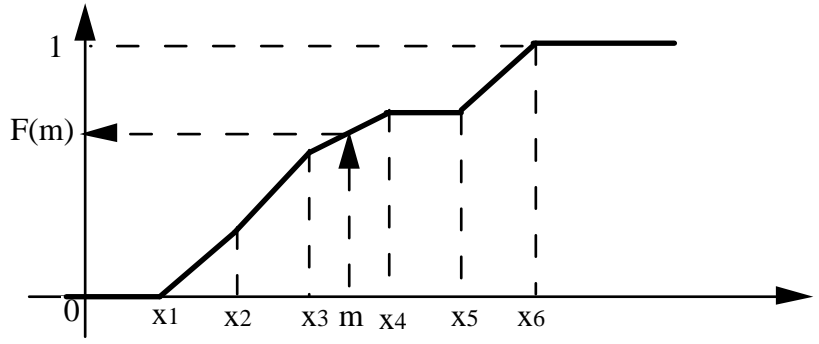
Ce coefficient est obtenu en calculant le rapport de l'écart-type à la moyenne.

$$CV = \frac{\text{écart - type}}{\text{moyenne}} = \frac{\sigma}{\mu}$$

Il est parfois exprimé en % et il permet de s'affranchir des unités de mesure et des ordres de grandeur de la variable. Il permet ainsi de comparer plusieurs distributions de variables.

**Courbe cumulative croissante**

La courbe cumulative croissante est du type :



Elle représente la fréquence cumulée associée à une valeur x, c'est à dire le % d'observation dont la valeur est strictement inférieure à x.

Cette courbe permet de déterminer graphiquement les fractiles. La médiane correspond à l'antécédent de la valeur 0,5 (50%) sur la courbe des fréquences cumulées. La médiane est le fractile d'ordre 0,5

**Covariance entre deux variables quantitatives X et Y**

La covariance est un indicateur de co-variation entre deux séries numériques X et Y. Rappelons-nous que la variance d'une variable statistique X peut s'exprimer comme la moyenne des carrés des écarts à la moyenne. La covariance est définie comme la moyenne des produits des écarts à la moyenne. On peut alors calculer cette covariance de la manière suivante :

$$\text{cov}(X, Y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

On peut vérifier que cette formule est équivalente à la formule suivante :

$$\text{cov}(X, Y) = \frac{\sum_i x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y}$$

Cette expression s'interprète comme étant : la moyenne des produits moins le produit des moyennes

**Déciles**

Les déciles D1, D2, ..., D9 partagent une série statistique en 10 parties d'effectifs égaux.

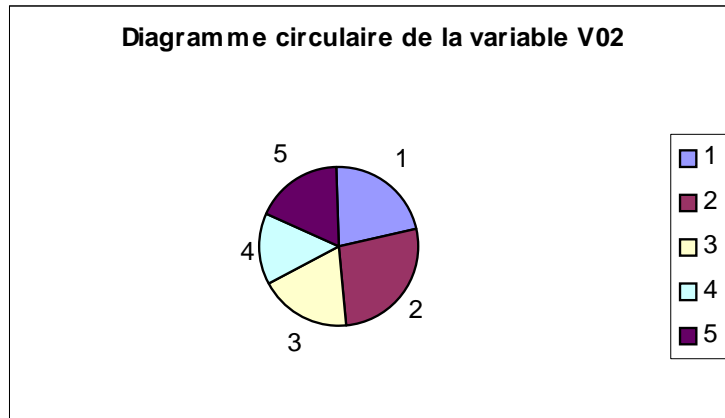
**Densité de fréquence**

La densité de fréquence intervient pour exprimer la distribution des fréquences d'une variable continue. Dans le cas des variables statistiques que nous étudions empiriquement, la densité de fréquence s'obtient en divisant la fréquence relative à l'intervalle par l'amplitude cet intervalle. Il s'agit alors de la fréquence par unité d'intervalle et la densité d'effectif s'obtient en divisant l'effectif par l'amplitude.

**Diagramme circulaire**

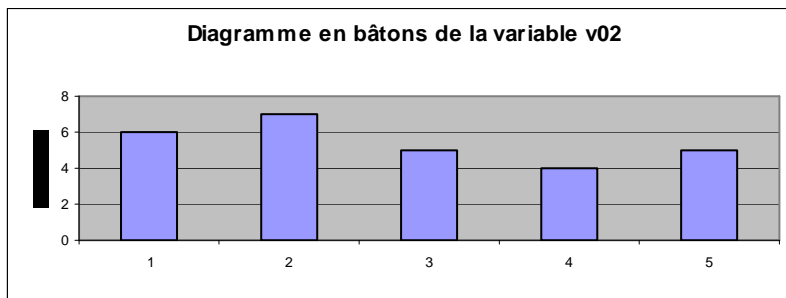
Cette représentation graphique fournit une visualisation de la distribution des fréquences ou de celle des effectifs. On obtient une telle représentation graphique en établissant une correspondance entre la mesure en degré des angles et la valeur des effectifs ou des fréquences. Ainsi, pour la modalité  $x_k$  de fréquence  $f_k\%$  exprimée en pourcentage :

mesure en degrés de l'angle de la portion représentant la modalité  $x_k$  vaut  $\frac{360 \times f\%_k}{100}$



### Diagramme en bâtons

Diagramme représentant une variable qualitative ou quantitative discrète. Cette représentation graphique fournit une visualisation de la distribution des fréquences ou de celle des effectifs. Chaque segment nommé « bâton », a une longueur proportionnelle à la fréquence ou à l'effectif correspondant à la modalité représentée. La largeur de ce « bâton » n'a aucun sens statistique.



### Ecart-type d'une variable quantitative

L'écart-type est la racine carrée de la variance. Pour un variable quantitative discrète :  
 sur la population d'effectif N:

$$\sigma = \sqrt{\frac{1}{N} \sum_{k=1}^{k=p} n_k (x_k - \mu)^2}$$

sur l'échantillon d'effectif n :

$$\sigma_{ech} = \sqrt{\frac{1}{n} \sum_{k=1}^{k=p} n_k (x_k - m)^2}$$

Pour une variable quantitative continue, les formule sont identiques en prenant pour valeur  $x_k$ , les centres des intervalles  $c_k$ .

### Échantillon

L'échantillon est un sous-ensemble de la population.

### Effectif

L'effectif est la quantité d'individus relative à un résultat (valeur ou modalité). Nous noterons  $n_1, n_2, \dots, n_p$ , les effectifs correspondant respectivement aux résultats  $x_1, x_2, \dots, x_p$ . La distribution des effectifs de la variable X n'est autre que la suite des  $(x_k, n_k)$  avec  $k = 1, \dots, p$ .

**Effectif total**

L'effectif total, noté N ou n, est la somme des effectifs relatifs à chaque résultat distinct. Nous avons

avec  $N = \sum_{k=1}^{k=p} n_k$  si nous travaillons sur la population et  $n = \sum_{k=1}^{k=p} n_k$  si l'étude est restreinte à l'échantillon.

**Effectifs cumulés**

Voir courbe cumulée croissante

**Estimation ponctuelle ou par intervalle de confiance d'une moyenne  $\mu$  d'une variable sur une population**

*Estimation ponctuelle*

Pour estimer la moyenne  $\mu$ , nous mettons en œuvre la statistique  $\bar{X}$  nommée moyenne empirique ou échantillonnale (le nom de la variable, ici X, surmonté d'une barre, on lit « X barre ») suivante qui consiste à calculer la moyenne des n résultats réalisés sur l'échantillon. L'opération est représentée symboliquement par la formule suivante:

$$m = \bar{X} = \frac{1}{n} \sum_{i=1}^{i=n} X_i$$

*Estimation par intervalle de confiance au niveau  $1-\alpha$*

La fourchette d'estimation de la moyenne s'obtient avec la formule suivante appliquée aux résultats obtenus sur un échantillon aléatoire avec remise de taille n :

$$m - \kappa \frac{s}{\sqrt{n}} < \mu < m + \kappa \frac{s}{\sqrt{n}}$$

avec

|  |  |  |
|--|--|--|
| $m = \bar{X} = \frac{1}{n} \sum_{i=1}^{i=n} X_i$ | $s^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - m)^2$ | $s = \sqrt{\frac{n}{n-1} \sigma_{\text{échantillon}}^2}$ |
|--|--|--|

Ainsi construit, cet intervalle  $[m - \kappa \frac{s}{\sqrt{n}} ; m + \kappa \frac{s}{\sqrt{n}}]$  a une probabilité de  $1-\alpha$  de contenir la vraie valeur  $\mu$  qui nous est inconnue. La valeur critique  $\kappa$  est à chercher dans la table de Student ddl=n-1 ou dans la table de Laplace-Gauss, si n est supérieur à 50

**Estimation ponctuelle et par intervalle de confiance d'une proportion  $\pi$  sur une population**

*Estimation ponctuelle*

Elle est le résultat d'une réalisation de la fréquence empirique sur l'échantillon choisi. Elle est tout simplement obtenue à partir de la proportion calculée avec les valeurs observées sur l'échantillon :  $\pi = f_n$  où  $f_n$  est la valeur de  $F_n$  calculée à partir de l'échantillon obtenu à l'issue de n tirages. Notons que  $\frac{\pi(1-\pi)}{n}$  la variance de  $F_n$  est estimée ponctuellement soit par

$\frac{f_n(1-f_n)}{n-1}$ , soit par sa valeur maximale  $\frac{1}{4(n-1)}$

*Estimation par intervalle de confiance*

Les deux situations usuelles sont:

- Intervalle de confiance à 95% :  $k=1,9600$
- Intervalle de confiance à 99% :  $k=2,5758$

Ensuite on utilise les résultats obtenus précédemment dans le cadre de l'estimation ponctuelle à savoir  $f_n$ . La fourchette d'estimation de la proportion  $\pi$  s'obtient avec la formule suivante :

$$f_n - \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}} < \pi < f_n + \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}}$$

Ainsi construit, cet intervalle  $\left[ f_n - \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}}; f_n + \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}} \right]$  a une probabilité de  $(1-\alpha)$  de contenir la vraie valeur  $\pi$  qui nous est inconnue. La valeur critique  $\kappa$  est à chercher dans la table de Laplace-Gauss, si  $n$  est supérieur à 100

### Estimation ponctuelle de l'écart-type $\sigma$ d'une variable sur une population

Elle est tout simplement obtenue à partir de la racine carrée de la variance calculée avec les valeurs observées sur l'échantillon à laquelle on applique le coefficient multiplicateur

$\frac{n}{n-1}$  corrigeant le biais

$$s = \sqrt{\frac{n}{n-1} \sigma_{\text{échantillon}}^2} = \sqrt{\frac{n}{n-1}} \sigma_{\text{échantillon}}$$

### Estimer

Estimer, c'est attribuer une valeur (ou une modalité) à un paramètre, ou une caractéristique inconnue tel qu'en particulier une moyenne, une variance, un écart type, une proportion, une médiane, un mode, un effectif. L'estimation peut être ponctuelle ou par intervalle de confiance, dite fourchette d'estimation.

### Étendue

L'étendue est l'intervalle entre la borne inférieure de rang 1 prise par X, et la borne supérieure de rang p prise par X c'est à dire  $[x_1; x_p]$

### Fractiles d'ordre $\alpha$ d'une variable quantitative

Le nombre  $f_\alpha$  tel que la fréquence cumulée associée à  $f_\alpha$  est égale à  $\alpha$ . La médiane est le fractile d'ordre 0,5. Les quartiles sont les fractiles d'ordre 0,25 ; 0,5 ; 0,75.

### Fréquence

La fréquence est définie par  $f_k = \frac{n_k}{\text{effectif total}}$  où l'effectif total est N si le référent est la

population et n si le référent est l'échantillon.  $(x_k, f_k)$  avec  $k = 1, \dots, p$  la distribution des fréquences de la variable X. Cette fréquence est souvent ramenée en pourcentage en multipliant  $f_k$  par 100.

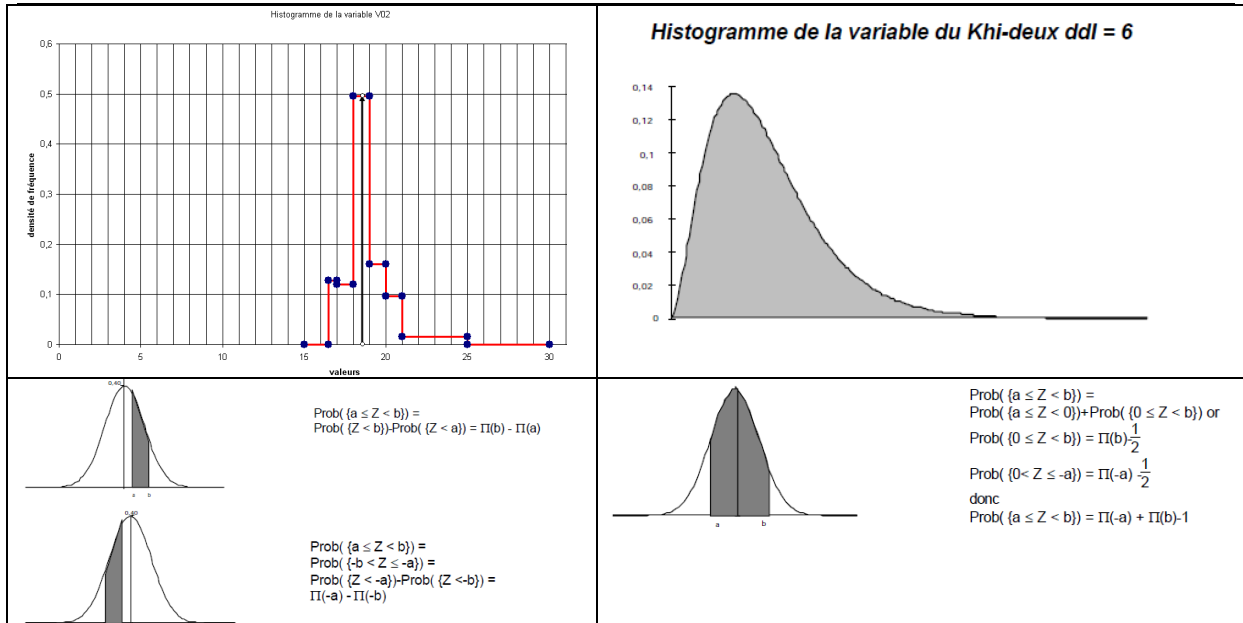
### Fréquences cumulées

Voir courbe cumulée croissante.

### Histogramme d'une variable continue

Représentation graphique délimitée par la courbe de la densité de fréquence, dont la surface représente la fréquence.

Exemples :



### Hypothèses statistiques

Elles correspondent à un couple d'énoncés conjecturaux par rapport auxquels le chercheur prend une décision de choix au vu d'une observation sur un échantillon. Le premier énoncé est désigné par  $H_0$  (hypothèse nulle) et le second, désigné par  $H_1$  (hypothèse alternative) est obtenu par la négation logique de l'hypothèse nulle. Les tests statistiques ont pour objet d'aider le chercheur dans cette prise de décision en contrôlant les risques d'erreur encourus : [erreur de première espèce = rejet à tort de  $H_0$  ; erreur de seconde espèce = conservation à tort de l'hypothèse  $H_0$ ]

### Indépendance de deux événements

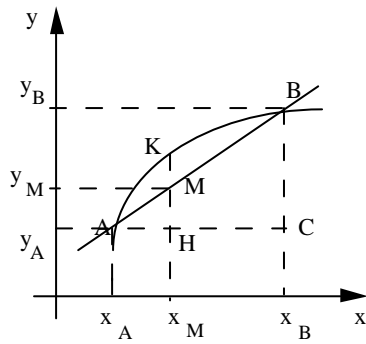
Deux événements A et B sont dit aléatoirement indépendants si la probabilité de la réalisation de l'événement A et B est égal au produit de la probabilité de A par celle de B. Deux événements A et B sont dit statistiquement indépendants si la fréquence des individus qui réalisent de l'événement A et B est égal au produit des fréquences de A et de B. Si nous considérons deux variables statistiques X et Y, par exemple X= « genre » et Y= « attitude à l'égard des mathématiques, l'événement A= « être une femme » et B= « aimer les mathématiques » peut être étudié du point de vue de l'indépendance ou de la dépendance qui est la négation de l'indépendance.

#### Individu ou unité

L'individu ou l'unité statistique est un élément de la population.

#### Interpolation linéaire

Méthode d'estimation qui s'appuie sur la proportionnalité.



Supposons connues  $x_M$ , l'abscisse du point M et  $x_K$ , l'abscisse du point K, avec  $x_M = x_K$ , nous pouvons calculer  $y_M$ , l'ordonnée de M, qui fournit une valeur approchée de  $y_K$ , par ordonnée de K, grâce à la relation qui se fonde sur l'alignement des points A, M et B:

$$\frac{x_M - x_A}{x_B - x_A} = \frac{y_M - y_A}{y_B - y_A}$$

ou encore par la relation

$$\boxed{\frac{y_M - y_A}{x_M - x_A} = \frac{y_B - y_A}{x_B - x_A}}$$

Si c'est  $y_K$  qui est connue, nous pouvons alors calculer  $x_M$  en tant que valeur approchée de  $x_K$ .

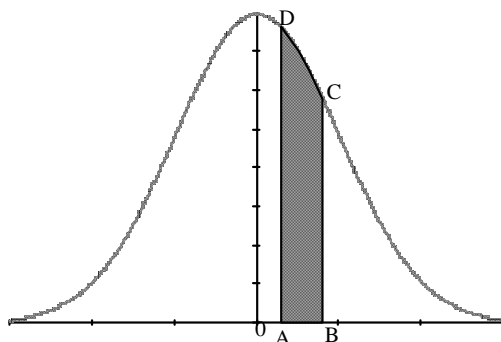
### Intervalle interquartile

L'intervalle interquartile est l'intervalle  $[Q1 ; Q3]$ . Il représente approximativement les 50 % des réponses qui encadrent la médiane  $Q2$ . son amplitude, calculée par la différence  $Q3 - Q1$ , mesure la dispersion autour de la médiane  $Q2$ .

### Laplace Gauss $LG(\mu ; \sigma)$ (Variable de Laplace-Gauss ou variable normale)

Il s'agit d'une famille de variables continues mathématiquement bien connue dont l'importance théorique est remarquable. Chaque membre de cette famille est caractérisé par deux paramètres :  $(\mu ; \sigma)$  qui ne sont autre que sa moyenne et son écart-type. La variable  $LG(0 ; 1)$  centrée réduite est celle à partir de laquelle sont effectués les calculs concernant les autres variables  $LG(\mu ; \sigma)$ . L'histogramme d'une variable de Laplace-Gauss a une forme bien connue, appelée fréquemment « courbe de Gauss » ou courbe en cloche, dont l'usage par une analogie abusive se rencontre fréquemment.

La courbe de Gauss est mathématiquement définie par les formules suivantes :



fonction densité de fréquence

$$f(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$$

dont la fonction de répartition

$$\text{est } F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

La variable  $LG(0 ; 1)$  joue un rôle très important pour d'autres variables dont les distributions peuvent être approchées par celle de  $LG(0 ; 1)$ .

### Médiane

La médiane (notée  $Q2$ ) est la modalité qui partage la population en deux parties de même effectif. La proportion des individus ayant choisi une modalité inférieure ou égale à  $Q2$  est de 50 %.

### Modalité

Les modalités sont les résultats possibles d'une variable qualitative.

### Mode

Pour une variable quantitative **discrète** ou qualitative, le mode est une valeur de la variable correspondant à l'effectif maximum ou la fréquence maximum.

Pour une variable quantitative **continue**, le mode est une valeur de la variable correspondant à la densité de fréquence maximum.

### Moyenne

La moyenne pour une variable quantitative discrète est la valeur de la variable X obtenue par l'une des deux procédures de calcul suivantes :

- sur la population d'effectif N, nous la désignons par la lettre grecque  $\mu$

$$\mu = \frac{1}{N} \sum_{i=1}^{i=N} o_i = \frac{1}{N} \sum_{k=1}^{k=p} n_k x_k = \sum_{k=1}^{k=p} f_k x_k$$

- sur l'échantillon d'effectif n, nous la désignons par la lettre latine  $m$

$$m = \frac{1}{n} \sum_{i=1}^{i=n} o_i = \frac{1}{n} \sum_{k=1}^{k=p} n_k x_k = \sum_{k=1}^{k=p} f_k x_k$$

La moyenne pour une variable quantitative continue est la valeur de la variable X obtenue par l'une des deux procédures de calcul suivantes où  $c_k$  désigne les centres des intervalles:

- sur la population d'effectif N, nous la désignons par la lettre grecque  $\mu$

$$\mu = \frac{1}{N} \sum_{k=1}^{k=p} n_k c_k = \sum_{k=1}^{k=p} f_k c_k$$

- sur l'échantillon d'effectif n, nous la désignons par la lettre latine  $m$

$$m = \frac{1}{n} \sum_{k=1}^{k=p} n_k c_k = \sum_{k=1}^{k=p} f_k c_k$$

### Nuage (statistique) de points

Quand on dispose des n résultats  $(x_i ; y_i)$  issus de l'étude conjointe de deux variables quantitatives  $(X ; Y)$ , on peut représenter graphiquement par n points  $M_i$  du plan repéré  $(O, \vec{i}, \vec{j})$ , dont les coordonnées sont  $(x_i ; y_i)$ . L'ensemble géométrique ainsi obtenu est appelé « nuage de points ». Il s'agit alors d'interpréter cette forme géométrique pour rechercher le type de lien entre X et Y. La forme de base est évidemment la droite indiquant une liaison linéaire entre X et Y, c'est à dire :  $Y=aX+b$

### Population (statistique)

La population (ou univers statistique) est l'ensemble de tous les individus (ou unité statistique) concernés par la recherche.

### Quartiles d'une variable quantitative ou d'une variable qualitative ordinale

Les Quartiles Q1, Q2, Q3 partagent une série statistique en 4 parties d'effectifs égaux. Q2 représente la médiane

### Recensement

Le recensement est la collecte des informations de l'ensemble de tous les individus d'une population. Cette collecte est dite exhaustive. Le recensement national organisé par l'INSEE tous les 7 ans en est un exemple.

### Région critique

La **région critique K** est l'ensemble des valeurs de la variable de décision qui conduisent à rejeter l'hypothèse nulle  $H_0$  au profit de l'hypothèse alternative  $H_1$ . La région critique est déterminée par la relation :  $\text{Prob}\{\text{rejeter } H_0 / H_0 \text{ vraie}\} = \alpha$ .

### **Région d'acceptation**

La **région d'acceptation A** est l'ensemble des valeurs de la variable de décision qui conduisent à ne pas rejeter l'hypothèse nulle  $H_0$  au détriment de l'hypothèse alternative  $H_1$ . La région critique  $K$  et la région d'acceptation  $A$  sont deux **ensembles complémentaires**. Dans cette perspective **construire un test** revient à construire une **variable de décision** explicite et à déterminer **une région critique** .

### **Représentativité**

Cette notion est fortement perturbée par l'imprégnation du sens commun. L'idée de base est clairement d'obtenir une partie d'information par un échantillon qui « représente » au mieux la population, c'est à dire une partie qui apporte des informations suffisantes pour bien connaître le tout. Il y a une forme de représentativité qui est définie par le « modèle réduit ». Celle sur laquelle nous nous appuyons pour procéder aux opérations d'estimation ou de tests d'hypothèse est essentiellement fondée sur le caractère aléatoire de l'appartenance de l'individu à l'échantillon. En général c'est l'équiprobabilité qui est la propriété de référence mais des échantillons peuvent aussi être constitués aléatoirement à partir d'une appartenance non équiprobable des individus — ce qui importe est la connaissance de la probabilité d'appartenance de chaque individu à l'échantillon.

### **Sondage**

Opération d'extraction d'un échantillon

### **Statistique descriptive bivariée**

Cadre théorique pour l'analyse des propriétés fondée sur la description simultanée de deux variables conjointes.

### **Statistique descriptive univariée**

Cadre théorique pour l'analyse des propriétés fondée sur la description d'une variable.

### **Statistique inférentielle**

Cadre théorique permettant de conduire des raisonnements inductifs et déductifs pour une extension contrôlée des propriétés observées sur un échantillon à toute la population..

### **(Variable de ) Student**

La famille des variables de Student notée  $T_{(n)}$  à  $n$  ddl est une famille bien connue mathématiquement tout comme la famille des variables de Laplace-Gauss. Chaque membre se distingue par son degré de liberté  $ddl = n$ . Dans la pratique elle est utilisée par l'intermédiaire de tables statistiques.

### **Tableau de contingence**

Il s'agit du tableau statistique correspondant à l'étude conjointe de deux variables statistiques. Ce tableau donne les effectifs des couples de résultats de  $(X ; Y)$ .

### **Test du Khi-deux d'adéquation**

Le test du Khi-deux d'adéquation est une procédure mise en pratique pour comparer la distribution de fréquence d'une variable qualitative ou d'une variable quantitative observée sur un échantillon, à une distribution théorique connue sur la population.

### **Test du Khi-deux d'homogénéité**

Le test du Khi-deux d'homogénéité est une procédure mise en pratique pour comparer des proportions relatives aux modalités d'une variable qualitative au sein de plusieurs populations

### **Test du Khi-deux d'indépendance**

Le test du Khi-deux d'indépendance est une procédure mise en pratique pour décider du choix entre les deux hypothèses alternatives :

$H_0$  les deux variables sont indépendantes

$H_1$  les deux variables sont dépendantes

### **Test statistique, test d'hypothèse statistique**

Procédure pour aider à la prise de décision entre deux hypothèses alternatives. Dans cette perspective, l'idée de vérité est remplacée par celle de vraisemblance. Un test statistique est alors une procédure logico-mathématique sur la quelle on s'appuie pour choisir entre deux hypothèses alternatives sur la base d'informations partielles issues d'un échantillon, celle qui est la plus vraisemblable.

|                       |                            | « Situation de nature »<br>inconnue   |                                     |
|-----------------------|----------------------------|---------------------------------------|-------------------------------------|
|                       |                            | Ho vraie<br>H1 fausse                 | H1 vraie<br>Ho fausse               |
| Décision du chercheur | Rejeter Ho<br>Conserver H1 | Erreur de première espèce<br>$\alpha$ | Correcte                            |
|                       | Conserver Ho<br>Rejeter H1 | Correcte                              | Erreur de seconde espèce<br>$\beta$ |

|                       |                            | « Situation de nature »<br>inconnue                |   |
|-----------------------|----------------------------|--|---|
|                       |                            | Ho vraie<br>H1 fausse                              | H1 vraie<br>Ho fausse                             |
| Décision du chercheur | Rejeter Ho<br>Conserver H1 | Prob{Rejeter Ho sachant Ho vraie}=<br>$\alpha$     | Prob{Conserver H1 sachant H1 vraie}=<br>$1-\beta$ |
|                       | Conserver Ho<br>Rejeter H1 | Prob{Conserver Ho sachant Ho vraie}=<br>$1-\alpha$ | Prob{Rejeter H1 sachant H1 vraie}=<br>$\beta$     |

### Variable statistique

La variable statistique qui modélise l'objet d'étude, est application mathématique qui à un individu (unité) statistique fait correspondre un résultat (quantité ou qualité) ; Ce concept fondamental renvoie au but principal de la science statistique qui est donné un cadre théorique à l'étude de la **variabilité**.

### Variable statistique qualitative nominale

Une variable statistique est qualitative nominale si les résultats ne sont pas des quantités. Ses modalités sont alors codées à l'aide de mots ou d'expressions courtes ou parfois à l'aide de nombres qui n'autorisent cependant aucunement une interprétation en terme de grandeur. Par exemple la variable « genre » possède deux modalités « Masculin » et « Féminin », codées respectivement « 1 » et « 2 ».

### Variable statistique qualitative ordinale

Une variable statistique est qualitative ordinale si les résultats qui ne sont pas des quantités, sont toutefois ordonnés. Ses modalités sont codées à l'aide de mots ou d'expressions courtes marquant une gradation ou parfois à l'aide de nombres. Dans ce dernier cas, l'ordre des nombres traduit la gradation. Par exemple la variable « goût pour la statistique » peut être modélisée avec les modalités suivantes :

| Expression verbale | Je n'aime pas du tout | J'aime un peu | J'aime beaucoup | J'aime à la folie |
|--------------------|-----------------------|---------------|-----------------|-------------------|
| Codage numérique   | 1                     | 2             | 3               | 4                 |

### Variable quantitative continue

Les variables quantitatives continues sont celles qui ne sont pas discrètes. L'ensemble des résultats possibles est l'ensemble des nombres réels ou un sous-ensemble. Par exemple : le « poids » des cartables, la taille des élèves de quatrième, la distance entre le domicile et l'école, la durée du trajet entre le domicile et l'école.

### Variable quantitative discrète

Les résultats sont alors de nature quantitative. L'ensemble des résultats possibles, que nous nommons valeurs est un ensemble de nombres qui possède une structure ayant des propriétés mathématiques remarquables habituelles autorisant l'usage de la comparaison, de la grandeur et des opérations arithmétiques ou algébriques. Ce que nous nommons variable quantitative discrète correspond à un ensemble fini ou dénombrable de résultats. Intuitivement chaque résultat est isolable. Par exemple, les variables statistiques suivantes : nombre d'enfants par famille, nombre d'élèves par classe de seconde

### Variable centrée réduite

Variable dont la moyenne est égale à 0 et l'écart-type est égal à 1. Si X est une variable de moyenne  $\mu$  et d'écart-type  $\sigma$  alors la variable Z obtenue par transformation de X par l'opération suivante

$$Z = \frac{X - \mu}{\sigma}$$
 est une variable centrée réduite. Cela revient à choisir la moyenne  $\mu$  comme nouvelle

origine des valeurs et l'écart-type  $\sigma$  comme nouvelle unité d'écart à la moyenne.

### Variance

La variance ou moment centré d'ordre 2 d'une variable quantitative discrète est la valeur obtenue par l'une des deux procédures de calcul suivante :

- sur la population d'effectif  $N$ , nous la désignons par la lettre grecque  $\sigma^2$

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^{k=p} n_k (x_k - \mu)^2 = \left( \frac{1}{N} \sum_{k=1}^{k=p} n_k x_k^2 \right) - \mu^2$$

- sur l'échantillon d'effectif  $n$ , nous la désignons par la lettre grecque  $\sigma_{ech}^2$

$$\sigma_{ech}^2 = \frac{1}{n} \sum_{k=1}^{k=p} n_k (x_k - m)^2 = \left( \frac{1}{n} \sum_{k=1}^{k=p} n_k x_k^2 \right) - m^2$$

La variance ou moment centré d'ordre 2 d'une variable quantitative continue est la valeur obtenue par l'une des deux procédures de calcul suivante où  $c_k$  sont les centres des intervalles :

- sur la population d'effectif  $N$ , nous la désignons par la lettre grecque  $\sigma^2$

$$\sigma^2 = \frac{1}{N} \sum_{k=1}^{k=p} n_k (c_k - \mu)^2 = \left( \frac{1}{N} \sum_{k=1}^{k=p} n_k c_k^2 \right) - \mu^2$$

- sur l'échantillon d'effectif  $n$ , nous la désignons par la lettre grecque  $\sigma_{ech}^2$

$$\sigma_{ech}^2 = \frac{1}{n} \sum_{k=1}^{k=p} n_k (x_k - m)^2 = \left( \frac{1}{n} \sum_{k=1}^{k=p} n_k c_k^2 \right) - m^2$$