

Analyse sociologique de données d'enquête par questionnaire

Olivier Martin

Cours de l'UE « Méthodes d'investigation sociologique »

L3 de sciences sociales – Université Paris Descartes – 2011-2012

Séance 5 :

Comment étudier les relations
entre deux variables ?

→ Situation commune dans une démarche d'analyse quantitative

- Variables recodées, synthétisées
- Recherche des relations entre variables

→ La notion de « relation » est polysémique

- Corrélation : « Le salaire est corrélé à l'ancienneté »
- Causalité : « Le cancer est une des causes de décès »
- Dépendance : « le vote dépend du sexe »
- Interdépendance : « Revenu et profession sont interdépendants »
- ... lien, liaison...

→ **Ce n'est la statistique qui fournit le sens de la relation, sa signification...**

→ Nécessité de distinguer les cas selon la nature des variables

- Variables **toutes qualitatives**
- Variables toutes quantitatives (à venir)
- Variables mélangées (à venir)

→ Nécessité de distinguer les cas selon le nombre de variables

- Deux variables → **Méthodes bivariées**
- Trois variables et plus → Méthodes multivariées

Comment représenter les liens
entre deux variables qualitatives ?

Tableau croisé

Tableau croisé

= représentation de la distribution des réponses selon deux critères (en ligne, en colonne)

L'attachement aux livres selon le sexe

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	504	312	37	11	864
homme	186	146	30	7	369
Total	690	458	67	18	1233

(Rappel sur la construction des tableaux croisés et leur lecture...)

→ A la vue d'un tableau croisé, doit-on plutôt croire à l'hypothèse d'une indépendance ? Ou à l'hypothèse d'une dépendance entre les deux variables ?

indépendance = ?

→ situation où tel caractère en ligne (ici le caractère sexuel - h/f) n'a aucune relation avec le caractère en colonne (ici attachement aux livres)

Indépendance :

- = situation où avoir tel ou tel caractère n'influence pas tel ou tel autre caractère
- = être dans telle ou telle ligne ne prédispose pas à être dans telle ou telle colonne
- = la répartition de l'attachement aux livres ne dépend pas du sexe
- = la répartition de l'attachement aux livres est identique quel que soit le sexe

Dans l'exemple :

L'attachement aux livres selon le sexe

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	504	312	37	11	864
homme	186	146	30	7	369
Total	690	458	67	18	1233

Y a-t-il indépendance ?

Y a-t-il indépendance ?

Étape 1 :

Quelle est la répartition de l'attachement aux livres ?

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme					
homme					
Total	56,0 %	37,1 %	5,4 %	1,5 %	100,0 %

 = 690 / 1233

Y a-t-il indépendance ?

Étape 2 :

Quelle serait la distribution des pourcentages
s'il y avait indépendance ?

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	56,0 %	37,1 %	5,4 %	1,5 %	100,0 %
homme	56,0 %	37,1 %	5,4 %	1,5 %	100,0 %
Total	56,0 %	37,1 %	5,4 %	1,5 %	100,0 %

Y a-t-il indépendance ?

Étape 3 :

Quelle serait la distribution des effectifs
s'il y avait indépendance ?

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	483,5	320,9	46,9	12,6	864
homme	206,5	137,1	20,1	5,4	369
Total	690	458	67	18	1233

$= 56 \% \text{ de } 864 \text{ femmes} = (690 / 1233) \times 864$

Y a-t-il indépendance ?

Étape 3 (suite) :

Notons au passage la formule :

$$56 \% \text{ de } 864 \text{ femmes} = (690 / 1233) \times 864$$

$$\textit{Effectif d'indépendance} = \frac{\textit{Effectif total de la colonne}}{\textit{Effectif total de l'échantillon}} \times \textit{Effectif total de la ligne}$$

Y a-t-il indépendance ?

Étape 3 (suite)

Ce tableau est le **tableau d'indépendance**

= si les deux variables étaient indépendantes

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	483,5	320,9	46,9	12,6	864
homme	206,5	137,1	20,1	5,4	369
Total	690	458	67	18	1233

Y a-t-il indépendance ?

Étape 4 :

Comparer le **tableau d'indépendance**

avec le **tableau de l'enquête**

(tableau observé, tableau empirique)

Etape 4 : Comparaison

L'attachement aux livres selon le sexe

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	504	312	37	11	864
homme	186	146	30	7	369
Total	690	458	67	18	1233

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	483,5	320,9	46,9	12,6	864
homme	206,5	137,1	20,1	5,4	369
Total	690	458	67	18	1233

Etape 4 : Comparaison

Les effectifs sont différents... mais :

- 1) Cela résulte-t-il de non-indépendance (dépendance) entre les deux variables?
- 2) Cela résulte-t-il des aléas de l'échantillonnage?

La comparaison... et ses difficultés

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	485	320	45	14	864
homme	205	138	22	4	369
Total	690	458	67	18	1233

Proches...

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	483,5	320,9	46,9	12,6	864
homme	206,5	137,1	20,1	5,4	369
Total	690	458	67	18	1233

La comparaison... et ses difficultés

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	480	325	50	9	864
homme	210	133	17	9	369
Total	690	458	67	18	1233

Un peu moins proches...

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	483,5	320,9	46,9	12,6	864
homme	206,5	137,1	20,1	5,4	369
Total	690	458	67	18	1233

La comparaison... et ses difficultés

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	580	256	10	18	864
homme	110	202	57	0	369
Total	690	458	67	18	1233

Très différents...

	tu aimes beaucoup les livres	tu aimes plutôt les livres	tu es indifférent aux livres	tu n'aimes pas les livres mais tu dois t'en servir malgré tout	Total
femme	483,5	320,9	46,9	12,6	864
homme	206,5	137,1	20,1	5,4	369
Total	690	458	67	18	1233

Problème :
parvenir à évaluer les effets des aléas
lors de l'échantillonnage.

.....

La science statistique
permet de « résoudre » cette difficulté

**La science statistique
permet de « résoudre » cette difficulté**

→ Les tests statistiques

**→ En l'occurrence, le test du khi-deux,
(chi-deux, khi², chi², χ^2)**

Raisonnement :

- 1) Faire l'hypothèse que les deux variables n'entretiennent aucune relation de dépendance.
- 2) En déduire quelle est la forme probable d'un tableau d'individus issus d'une population choisis au sein d'une population où il aurait indépendance entre les deux variables.
- 3) Voir si cette hypothèse est compatible avec les données de notre échantillon empirique.
- 4) Conclure sur la validité ou, plus exactement, la plausibilité (crédibilité) de cette hypothèse.

Raisonnement :

- 1) Faire l'hypothèse que les deux variables n'entretiennent aucune relation de dépendance.
- 2) En déduire quelle est la forme probable d'un tableau d'individus issus d'une population choisis au sein d'une population où il aurait indépendance entre les deux variables.
- 3) Voir si cette hypothèse est compatible avec les données de notre échantillon empirique.
- 4) Conclure sur la validité ou, plus exactement, la plausibilité (crédibilité) de cette hypothèse.

Raisonnement :

- 1) Faire l'hypothèse que les deux variables n'entretiennent aucune relation de dépendance.
- 2) En déduire quelle est la forme probable d'un tableau d'individus issus d'une population choisis au sein d'une population où il aurait indépendance entre les deux variables.
- 3) Voir si cette hypothèse est compatible avec les données de notre échantillon empirique.
- 4) Conclure sur la validité ou, plus exactement, la plausibilité (crédibilité) de cette hypothèse.

Raisonnement :

- 1) Faire l'hypothèse que les deux variables n'entretiennent aucune relation de dépendance.
- 2) En déduire quelle est la forme probable d'un tableau d'individus issus d'une population choisis au sein d'une population où il aurait indépendance entre les deux variables.
- 3) Voir si cette hypothèse est compatible avec les données de notre échantillon empirique.
- 4) Conclure sur la validité ou, plus exactement, la plausibilité (crédibilité) de cette hypothèse.

...

- 1) L'hypothèse d'indépendance entre les deux variables permet de déterminer à quoi doivent ressembler les divers tableaux possibles.
 - quels types de tableaux peut-on obtenir et avec quelle probabilité peut-on les obtenir ?

- 2) Il suffit alors de déterminer quel est le type du tableau dit « tableau observé » résultant de l'échantillon.

... Deux cas se présentent :

Cas 1 : Si notre tableau observé est semblable à un tableau qu'il est relativement probable d'obtenir lorsque les deux variables sont indépendantes, alors il est raisonnable de penser que l'hypothèse d'indépendance doit être acceptée.

... Deux cas se présentent :

Cas 2 : Si notre tableau observé s'apparente à un type de tableau peu ou très peu probable lorsque les deux variables sont indépendantes, alors il est raisonnable de penser que l'hypothèse d'indépendance doit être écartée.

On en « conclut » que les deux variables sont dépendantes.