Eléments d'analyses statistisques de données

Gilles Le Chenadec

ENSTA Bretagne

27 août 2018

Analyses statistiques bivariées

Introduction

- Les analyses à mener ne peuvent se limiter à l'analyse d'une variable :
  - La compréhension passe généralement par plusieurs variables
  - On comprend bien que pour développer un hélico qui vole longtemps on doit trouver le meilleur ensemble de paramètres
  - D'où une analyse multivariée ... qui commence par l'analyse bivariée.
- Questions : deux variables sont elles liées ou sont elles indépendantes ? Quelle est la nature de cette dépendance ?
- La réponse à ces questions (et les outils à utiliser) dépend de la nature des deux variables X et Y ; on distingue les cas suivants :
  - 2 variables quantitatives
  - 1 variable quantitative et variable qualitative
  - 2 variables qualitatives (ou groupées par intervalles)

Définitions et notations

Analyses statistiques bivariées

Considérons deux variables X et Y qui peuvent être soit qualitatives, quantitatives discrètes ou continues.
Le but de cette section est d'étudier les relations entre X et Y, ce qui revient, mathématiquement, à étudier les propriétés du couple (X, Y). On définit dans ce chapitre essentiellement deux types de quantités, celles dites :

- marginales qui ne dépendent que d'un seul critère mais pas des deux ;
- conditionnelles qui renseignent sur un critère en fonction des valeurs ou modalités de l'autre.

- On a la série statistique double sous forme brute pour un échantillon de taille n : $\{(X(1), Y(1)), \ldots, (X(n), Y(n))\}$:
- les modalités/valeurs/centre de classes de la variables X sont : $\{x_1, \ldots, x_p\}$
- les modalités/valeurs/centre de classes de la variables Y sont : $\{y_1, \ldots, y_q\}$
- l'indice i servira exclusivement à indexer des données relatives à X, et l'indice j exclusivement à indexer des données relatives à Y.

Liaison de deux variables qualitatives

Analyses statistiques bivariées

Variables concernées

Les outils présentés dans cette partie sont valables pour deux variables qualitatives et/ou quantitatives discrètes ou continues regroupées par intervalles...toute variable qui donne la possibilité de construire une table statistique des effectifs.

Exemple : temps de vol d'un hélicoptère

Numéro du lancer	Helico	Coupé	Temps
1	A	NON	1.97
2	A	NON	2.03
3	A	NON	2.1
4	A	NON	1.86
5	A	NON	2
6	A	NON	2.14
7	A	NON	2.07
8	A	NON	2.19
9	B	NON	2.52
10	B	NON	3.01
11	B	NON	2.69
12	B	NON	2.71
13	B	NON	2.93
14	B	NON	2.83
15	B	NON	2.7
16	B	NON	3.11
17	C	NON	2.52
18	C	NON	2.25
19	C	NON	2.47
20	C	NON	2.44
21	C	NON	2.44
22	C	NON	2.45
23	C	NON	2.36
24	C	NON	2.27
25	D	OUI	1.87
26	D	OUI	2.12

Table des effectifs de *Hélico* et *Coupé*

n_{ij}	A	B	C	D
Oui	8	8	8	0
Non	0	0	0	8

Table statistique des effectifs

Liaison de deux variables qualitatives

Soient $i \in \{1, \dots, p\}$ et $j \in \{1, \dots, q\}$

Definition (Table statistique des effectifs)

L'**effectif du couple** (x_i, y_j), noté n_{ij}, est le nombre d'individus associés à la fois à x_i pour X et à y_j pour Y et sa **fréquence relative**, noté f_{ij}, est le quotient de son effectif par la taille de la population :

$$n_{ij} = \mathrm{card}\{u \in \text{échantillon } | X(u) = x_i \text{ et } Y(u) = y_j\} \text{ et } f_{ij} = \frac{n_{ij}}{n}$$

Exemple

n_{ij}	A	B	C	D
Oui	8	8	8	0
Non	0	0	0	8

f_{ij}	A	B	C	D
Oui	$8/32$	$8/32$	$8/32$	$0/32$
Non	$0/32$	$0/32$	$0/32$	$8/32$

Effectifs marginaux et fréquences marginales

Liaison de deux variables qualitatives

Soient $i \in \{1, \ldots, p\}$ et $j \in \{1, \ldots, q\}$

Definition (Effectifs marginaux)

L'effectif marginal de x_i (respectivement y_j), noté $n_{i\bullet}\cdot$ (respectivement $n_{\bullet j}$), est l'effectif de x_i (respectivement y_j) pour la variable X (respectivement Y) :

$$n_{i\bullet} = \sum_{j=1}^{q} n_{ij} \text{ et } n_{\bullet j} = \sum_{i=1}^{p} n_{ij}$$

Definition (Fréquences marginales)

La fréquence marginale de x_i (respectivement y_j), notée $f_{i\bullet}\cdot$ (respectivement $f_{\bullet j}$), est l'effectif de x_i (respectivement y_j) pour la variable X (respectivement Y) :

$$f_{i\bullet} = \frac{n_{i\bullet}}{n} = \sum_{j=1}^{q} f_{ij} \text{ et } f_{\bullet j} = \frac{n_{\bullet j}}{n} = \sum_{i=1}^{p} f_{ij}$$

Effectifs marginaux et fréquences marginales

Liaison de deux variables qualitatives

Soient $i \in \{1, \ldots, p\}$ et $j \in \{1, \ldots, q\}$

Propriétés

$$\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij} = n, \sum_{i,j} f_{ij} = 1$$

$$\sum_{i=1}^{p} n_{i\bullet} = \sum_{j=1}^{q} n_{\bullet j} = n \text{ et } \sum_{i=1}^{p} f_{i\bullet} = \sum_{j=1}^{q} f_{\bullet j} = 1$$

Exemple : somme sur les lignes ou les colonnes

n_{ij}	A	B	C	D	$n_{i,.}$
Oui	8	8	8	0	24
Non	0	0	0	8	8
$n_{.,j}$	8	8	8	8	**32**

f_{ij}	A	B	C	D	$f_{i,.}$
Oui	8/32	8/32	8/32	0/32	0,75
Non	0/32	0/32	0/32	8/32	0,25
$f_{.,j}$	0,25	0,25	0,25	0,25	**1**

Tables de contingence des effectifs et des fréquences

Liaison de deux variables qualitatives

Effectifs

X \ Y	y_1	$\dots$	y_j	$\dots$	y_q	**Total**
x_1	n_{11}	$\dots$	n_{1j}	$\dots$	n_{1q}	$n_{1\bullet}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	$\dots$	n_{ij}	$\dots$	n_{iq}	$n_{i\bullet}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_p	n_{p1}	$\dots$	n_{pj}	$\dots$	n_{pq}	$n_{i\bullet}$
Total	$n_{\bullet 1}$	$\dots$	$n_{\bullet j}$	$\dots$	$n_{\bullet q}$	**n**

Fréquences

X \ Y	y_1	$\dots$	y_j	$\dots$	y_q	**Total**
x_1	f_{11}	$\dots$	f_{1j}	$\dots$	f_{1q}	$f_{1\bullet}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	f_{i1}	$\dots$	f_{ij}	$\dots$	f_{iq}	$f_{i\bullet}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_p	f_{p1}	$\dots$	f_{pj}	$\dots$	f_{pq}	$f_{i\bullet}$
Total	$f_{\bullet 1}$	$\dots$	$f_{\bullet j}$	$\dots$	$f_{\bullet q}$	**1**

En extrayant une colonne (respectivement une ligne) du tableau de contingence, on obtient une distribution conditionnelle de X (respectivement de Y). Précisément, l'extraction de la colonne j (respectivement de la ligne i) permet de répondre à la question «Sachant que Y vaut y_j (respectivement X vaut x_i), comment se comporte X (respectivement Y) ? ».

Fréquences conditionnelles

Liaison de deux variables qualitatives

Ainsi, une distribution conditionnelle est l'étude d'un critère lorsque l'autre reste fixe. On note $X|_{Y=y_j}$ (respectivement $Y|_{X=x_i}$) la distribution conditionnelle de X sachant $Y = y_j$ (respectivement de Y sachant $X = x_i$).

Definition (Fréquence conditionnelle)

Soient $i \in \{1, \ldots, p\}$ et $j \in \{1, \ldots, q\}$. La **fréquence conditionnelle** de x_i sachant y_j (respectivement y_j sachant x_i), notée $f_{i|Y=y_j}$ ou $f_{i|j}$ (respectivement $f_{j|X=x_i}$ ou $f_{j|i}$), est la fréquence d'occurrence de x_i dans la colonne j (respectivement y_j dans la ligne i) du tableau de contingence :

$$f_{i|Y=y_j} = \frac{n_{ij}}{n_{\bullet j}} \text{ et } f_{j|X=x_i} = \frac{n_{ij}}{n_{i\bullet}}$$

Propriétés

$$\forall j \in \{1, \ldots, q\}, \sum_{i=1}^{p} f_{i|Y=y_j} = 1 \text{ et } \forall j \in \{1, \ldots, q\}, \sum_{j=1}^{q} f_{j|X=x_i} = 1$$

Table des profils-colonnes et profils-lignes

Liaison de deux variables qualitatives

Definition (Profils lignes et profils colonnes)

La table des **profils-colonne** représente les distributions conditionnelles de $X|_Y$.
Le tableau des **profils-ligne** représente les distributions conditionnelles $Y|_X$.

$X\|_Y$ \ Y	y_1	$\cdots$	y_q
x_1	$f_{1\|Y=y_1}$	$\cdots$	$f_{1\|Y=y_q}$
$\vdots$	$\vdots$		$\vdots$
x_p	$f_{p\|Y=y_1}$	$\cdots$	$f_{p\|Y=y_q}$
Total	1	$\cdots$	1

Tableau des profils-colonne

X \ $Y\|_X$	y_1	$\cdots$	y_q	Total
x_1	$f_{1\|X=x_1}$	$\cdots$	$f_{q\|X=x_1}$	1
$\vdots$	$\vdots$			$\vdots$
x_p	$f_{1\|X=x_p}$	$\cdots$	$f_{q\|X=x_p}$	1

Tableau des profils-ligne

Propriétés

$$\forall j \in \{1, \ldots, q\} \text{ et } j \in \{1, \ldots, q\},$$
$$f_{ij} = f_{i|Y=y_j} f_{\bullet j} \text{ et } f_{ij} = f_{j|X=x_i} f_{i\bullet}$$

Remarque sur les fréquences

Liaison de deux variables qualitatives

Toutes les fréquences sont des estimations de probabilité d'une population à partir d'un échantillon...Lesquelles ?

- fréquence $f_{ij} \longrightarrow \mathrm{p}\left(X = x_i \text{ et } Y = y_j\right)$
- fréquence marginale $f_{i\bullet} \longrightarrow \mathrm{p}\left(X = x_i\right)$ (resp $f_{\bullet j}$)
- fréquence conditionnelle $f_{i|Y=y_j} \longrightarrow \mathrm{p}\left(X = x_i | Y = y_j\right)$ (resp $f_{j|X=x_i}$)

Indépendance de deux variables qualitatives

Liaison de deux variables qualitatives

Definition (indépendance)

Les variables X et Y sont dites indépendantes si

$$\forall i \in \{1, \dots, p\} \text{ et } j \in \{1, \dots, q\}, f_{ij} = f_{i\bullet} f_{\bullet j}$$

Propriétés

Les assertions suivantes sont équivalentes

(i) X et Y sont indépendantes.

(ii) $\forall (i,j) \in \{1, \dots, p\} \times \{1, \dots, q\}, n_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$

(iii) $\forall (i,j) \in \{1, \dots, p\} \times \{1, \dots, q\}, f_{i|Y=y_j} = f_{i\bullet}$

(iv) $\forall (i,j) \in \{1, \dots, p\} \times \{1, \dots, q\}, f_{j|X=x_i} = f_{\bullet j}$

Ecart à l'indépendance de deux variables qualitatives

Liaison de deux variables qualitatives

En pratique, on va quantifier l'écart à l'indépendance. Pour cela, on peut calculer la quantité suivante :

Definition (χ^2 observé)

$$\chi^2 = n \sum_{i,j} \frac{(f_{ij} - f_{i\bullet} f_{\bullet j})^2}{f_{i\bullet} f_{\bullet j}} = \sum_{i,j} \frac{(n_{ij} - n_{i\bullet} n_{\bullet j})^2}{n_{i\bullet} n_{\bullet j}}$$

Propriétés

- Le χ^2 est toujours positif ou nul
- $\chi^2 = 0 \Leftrightarrow$ X et Y sont indépendantes
- Plus le χ^2 est grand, plus les variables sont indépendantes

Ecart à l'indépendance de deux variables qualitatives

Liaison de deux variables qualitatives

Il se trouve que le χ^2 observé croît avec la taille de la population. Pour cette raison, il est utile de le renormaliser de manière adéquate.

Definition (Mesures dérivées du χ^2 observé)

- Φ^2 : $\Phi^2 = \frac{\chi^2}{n}$
- Tschuprow : $T = \frac{\Phi^2}{\sqrt{(p-1)(q-1)}}$
- V de Cramer : $V = \sqrt{\frac{\Phi^2}{\min(p,q)-1}} \in \{0,1\}$

On verra plus tard une méthodologie permettant d'apporter une réponse à cette question (test non paramétrique du χ^2 d'indépendance)

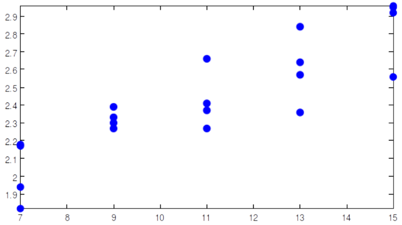
Liaison de deux variables quantitatives

Variables et données concernées

Les outils présentés dans cette partie sont valables uniquement pour deux variables quantitatives discrètes ou continues regroupées par intervalles ou non. Nous pourrons donc utiliser les données brutes (comme dans l'exemple ci-dessous) ou les données regroupées dans une table statistique des effectifs ou des fréquences (comme dans la partie précédente)

Nous avons à disposition la statistique double $\{(X(1), Y(1)), \ldots, (X(n), Y(n))\}$, il est donc possible de la représenter par un nuage de n points dans $\mathbb{R}^2$

Lancer	H_a	Temps
1	7	1,94
2	9	2,3
3	11	2,66
4	13	2,36
5	15	2,96
6	7	2,18
7	9	2,33
8	11	2,37
9	13	2,64
10	15	2,56
11	7	1,82
12	9	2,39
13	11	2,41
14	13	2,57
15	15	2,92
16	7	2,17
17	9	2,27
18	11	2,27
19	13	2,84
20	15	2,95



Décrire les deux variables individuellement

Moyenne marginale, variance marginale, écart-type marginal

Marginale signifie que l'on considère les variables X et Y indépendamment l'une de l'autre. Pour les données brutes, se reporter aux formules définies pour l'analyse univariées.

Definition (Moyenne marginale à partir d'une table des ~~effectifs~~)

Attention manque nij

- Moyennes marginales $\bar{x}$ et $\bar{y}$,

$$\bar{x} = \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} x_i \text{ et } \bar{y} = \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} y_j$$

- variance marginale s_X^2 et s_Y^2

$$s_X^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} \left(x_i - \bar{x}\right)^2 \text{ et } s_Y^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} f_{ij} \left(y_j - \bar{y}\right)^2$$

- écart-types marginaux s_X et s_Y^2 : $s_X = \sqrt{s_X^2}$ et $s_Y = \sqrt{s_Y^2}$

Quantifier le lien entre les deux variables quantitatives

Covariance

Soit X et Y deux variables quantitatives,

Definition (Covariance à partir des données brutes)

$$\mathrm{Cov}\left\{X, Y\right\} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\bar{y}$$

Definition (Covariance à partir des données d'une table d'effectifs ou de fréquences)

$$\mathrm{Cov}\left\{X, Y\right\} = \sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}(x_i - \bar{x})(y_j - \bar{y})$$

$$= \left(\sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij} x_i y_j\right) - \bar{x}\bar{y} = \left(\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij} x_i y_j\right) - \bar{x}\bar{y}$$

Quantifier le lien entre les deux variables quantitatives

Coefficient de corrélation linéaire

Definition (Coefficient de corrélation linéaire)

Le coefficient de corrélation ρ de X et Y est défini par

$$\rho_{X,Y} = \frac{\mathrm{Cov}\left\{X, Y\right\}}{s_X s_Y} \in [-1, +1]$$

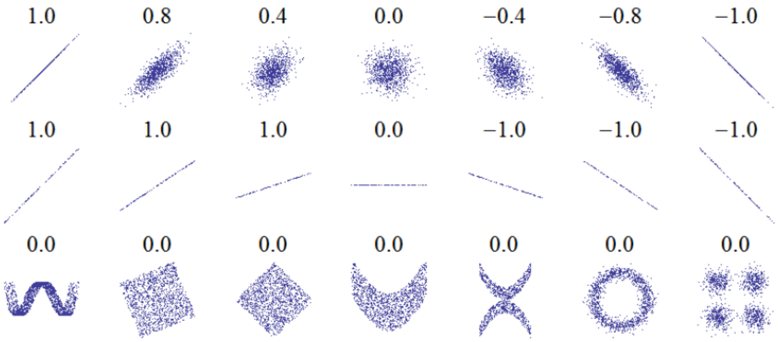
Si $\rho_{X,Y} =$

- 1, il existe une relation linéaire parfaite entre X et Y. De plus, quand X augmente, Y augmente.
- ~~1, il existe une relation linéaire parfaite entre X et Y~~
- -1, il existe une relation linéaire parfaite entre X et Y. De plus, quand X augmente, Y diminue.

Quantifier le lien entre les deux variables quantitatives

Coefficient de corrélation linéaire

Lien entre la valeur de $\rho_{X,Y}$ et la forme du nuage de points :



Quantifier le lien entre les deux variables quantitatives

Régression linéaire / ajustement affine / prédiction

On suppose ici qu'il y a une relation linéraire importante entre X et Y . Le principe d'un ajustement affine, aussi appelé régression linéaire, est de déterminer la « meilleure » droite pour exprimer la liaison affine entre X et Y =, c'est-à-dire la droite qui approxime au mieux le nuage de points.

Méthodes des moindres carrés

Si X n'est pas constante, la droite de régression par la méthode des moindres carrés de Y en fonction de X est la droite d'équation $y = \hat{a}x + \hat{b}$ dans le repère du nuage de points de (X, Y) avec :

$$\hat{a} = \rho_{X,Y}\frac{s_Y}{s_X} \text{ et } \hat{b} = \bar{y} - \rho_{X,Y}\frac{s_Y}{s_X}\bar{x}$$

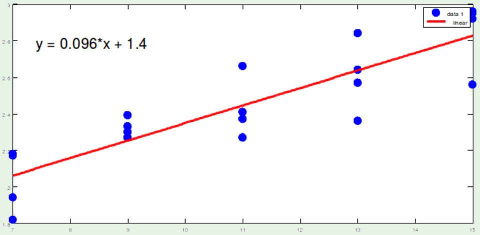
Quantifier le lien entre les deux variables quantitatives

Exemple

Temps de vol des hélico en fonction de H_a

Lancer	H_a	Temps
1	7	1,94
2	9	2,3
3	11	2,66
4	13	2,36
5	15	2,96
6	7	2,18
7	9	2,33
8	11	2,37
9	13	2,64
10	15	2,56
11	7	1,82
12	9	2,39
13	11	2,41
14	13	2,57
15	15	2,92
16	7	2,17
17	9	2,27
18	11	2,27
19	13	2,84
20	15	**2,95**

$\rho = 0.88$



Liaison entre deux variables : qualitative vs quantitative

Variables et données concernées

Les outils présentés dans cette partie sont valables uniquement pour une variable qualitative et une variable quantitative discrète ou continue regroupée par intervalle. Nous pourrons allons donc utiliser les données brutes.

Exemple : Hélico Vs Temps de vol

- y'a-t-il un effet de la variable qualitative sur la quantitative ?
- Y'a-t-il un lien entre les caractéristiques statistiques des temps de vol pour chacune des valeurs de Hélico ?
- Les stats dépendent-elles du type d'hélico ?
- On parlera d'**analyse de la variance** (dispersion des données)

Numéro du lancer	Helico	Temps
1	A	1,97
2	A	2,03
3	A	2,1
4	A	1,86
5	A	2
6	A	2,14
7	A	2,07
8	A	2,19
9	B	2,52
10	B	3,01
11	B	2,69
12	B	2,71
13	B	2,93
14	B	2,83
15	B	2,7
16	B	3,11
17	C	2,52
18	C	2,25
19	C	2,47
20	C	2,44
21	C	2,44
22	C	2,45
23	C	2,36
24	C	2,27
25	D	1,87
26	D	2,12
27	D	1,84
28	D	2,08
29	D	1,98
30	D	1,98
31	D	1,78
32	D	1,88

Liaison entre deux variables : qualitative vs quantitative

L'analyse de la liaison consiste à découper l'échantillon en plusieurs sous-échantillons (un pour chaque valeur de la variable qualitative)

Exemple : Hélico Vs Temps de vol

Numéro du lancer	Helico	Temps
1	A	1,97
2	A	2,03
3	A	2,1
4	A	1,86
5	A	2
6	A	2,14
7	A	2,07
8	A	2,19
9	B	2,52
10	B	3,01
11	B	2,69
12	B	2,71
13	B	2,93
14	B	2,83
15	B	2,7
16	B	3,11
17	C	2,52
18	C	2,25
19	C	2,47
20	C	2,44
21	C	2,44
22	C	2,45
23	C	2,36
24	C	2,27
25	D	1,87
26	D	2,12
27	D	1,84
28	D	2,08
29	D	1,98
30	D	1,98
31	D	1,78
32	D	1,88

Numéro du lancer	Helico	Temps
1	A	1,97
2	A	2,03
3	A	2,1
4	A	1,86
5	A	2
6	A	2,14
7	A	2,07
8	A	2,19

	Helico	Temps
9	B	2,52
10	B	3,01
11	B	2,69
12	B	2,71
13	B	2,93
14	B	2,83
15	B	2,7
16	B	3,11

	Helico	Temps
17	C	2,52
18	C	2,25
19	C	2,47
20	C	2,44
21	C	2,44
22	C	2,45
23	C	2,36
24	[illegible]	[illegible]

Numéro du lancer	Helico	Temps
25	D	1,87
26	D	2,12
27	D	1,84
28	D	2,08
29	D	1,98
30	D	1,98
31	D	1,78
32	D	1,88

Définitions et notations

Soient

- X la variable qualitative prenant p valeurs $\{x_1, \ldots, x_p\}$ d'effectifs $\{n_1, \ldots, n_p\}$
- $\sum_{i=1}^{p} n_i = n$
- Y la variable quantitative de moyenne $\bar{y}$ et de variance s_Y^2
- chaque valeur x_i de X définit un sous échantillon Y_i d'effectifs $\{n_1, \ldots, n_p\}$ possédant une moyenne $\bar{y}_i$ et une variance associées $s_{Y_i}^2$:

Formule de décomposition de la variance

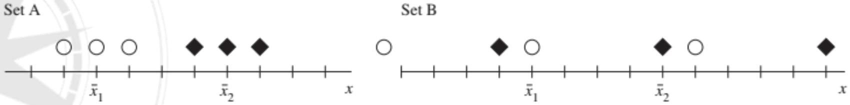
Propriété

La variance totale se décompose en somme de la variance interclasse et intraclasse :

$$
\begin{aligned}
s_Y^2 &= \frac{1}{n}\sum_{i=1}^{p} n_i.(\bar{y}_i - \bar{y})^2 && + \frac{1}{n}\sum_{i=1}^{p} n_i s_{Y_i}^2 \\
&= s_B^2 && + s_W^2 \\
&= \text{variance inter-classe} && + \text{variance intra-classe}
\end{aligned}
$$

attention
multiplication!

Remplacer x par y dans la figure ci-dessous, placer $\bar{y}$, à quoi correspondent les deux variances ?



Quantifier la liaison entre deux variables : quantitative vs qualitative

Rapport de corrélation

Definition

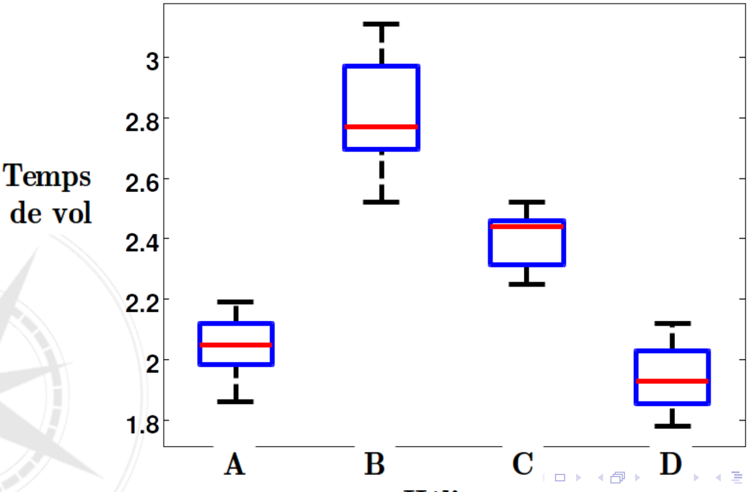
La rapport de corrélation entre X variable qualitative et Y variable quantitative est défini par :

$$S_{Y/X} = \sqrt{\frac{s_B^2}{s_Y^2}} \in \{0, 1\}$$

- $S_{Y/X}$ exprime l'impact de la partition en p valeurs de X sur la dispersion des mesures de Y
- En pratique, si $S_{Y/X}$
  - est de proche de 0 $\Rightarrow s_B^2 \ll s_W^2 \Rightarrow$ faible séparabilité des sous-échantillons $\Rightarrow$ la partition de X n'a aucun impact sur la variable Y
  - est de proche de 1 $\Rightarrow s_B^2 \gg s_W^2 \Rightarrow$ bonne séparabilité des sous-échantillons $\Rightarrow$ la partition de X a un impact sur la variable Y

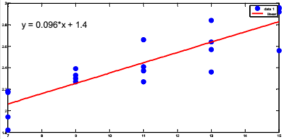
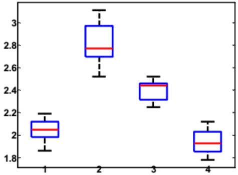
Représenter la liaison entre une variable quantitative et qualitative

Exemple des boîtes à moustache



Analyse des statistiques bivariées

Résumé des trois cas possibles

<table>
<tr><th>Quantitatif Vs Quantitatif</th><th>Quantitatif Vs Qualitatif</th><th>Qualitatif Vs Qualitatif</th></tr>
<tr>
<td>Nuage de points

</td>
<td>Boites à moustache parallèles

</td>
<td>Table de contingence
<table>
<tr><th>n_{ij}</th><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td>Oui</td><td>8</td><td>8</td><td>8</td><td>0</td></tr>
<tr><td>Non</td><td>0</td><td>0</td><td>0</td><td>8</td></tr>
</table></td>
</tr>
<tr>
<td>Coefficient de corrélation linéaire
$$\rho = \frac{\operatorname{Cov}\left(X,Y\right)}{S_N^X S_N^Y} \in \left[-1,1\right]$$</td>
<td>Rapport de corrélation
$$S_{Y/X} = \sqrt{\frac{S_B^2}{S_{N,Y}^2}} \in \left[0,1\right]$$</td>
<td>Distance du χ^2
$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i,\bullet} n_{\bullet,j}}{N}\right)^2}{\frac{n_{i,\bullet} n_{\bullet,j}}{N}}$$</td>
</tr>
</table>

Figure