

# Méthodes quantitatives d'analyse (POL 2809)

Séance 2, 11 septembre 2019

Enseignante: Florence Vallée-Dubois

Bureau: C-3114

Dispos: mercredis, 10h-11h30

[florence.vallee-dubois@umontreal.ca](mailto:florence.vallee-dubois@umontreal.ca)

# Aujourd'hui

Les types de variables.

Statistiques univariées et bivariées.

Différence entre substance et signification statistique.

# La recherche quantitative

# La recherche quantitative

**Donnée quantitative:** toute donnée pouvant être quantifiée; exprimée sous forme de chiffre.

# La recherche quantitative

**Donnée quantitative:** toute donnée pouvant être quantifiée; exprimée sous forme de chiffre.

**Statistiques:** méthodes permettant de décrire et de traiter des données quantitatives. On peut utiliser les statistiques pour expliquer des phénomènes en sciences sociales.

# Concepts

**Population:** l'ensemble des individus/objets qui nous intéressent et qui pourraient potentiellement être observés.

# Concepts

**Population:** l'ensemble des individus/objets qui nous intéressent et qui pourraient potentiellement être observés.

**Échantillon:** sous-groupe des individus qui composent la population.

# Concepts

**Stat. descriptives:** collecte, classement, synthèse, traitement, présentation de données numériques.



# Concepts

**Stat. descriptives:** collecte, classement, synthèse, traitement, présentation de données numériques.

**Stat. inférentielles:** induire les caractéristiques d'une population à partir des caractéristiques d'un échantillon de cette population.

# Types de variables

# Types de variables

Variables **nominales**: aucun ordre

# Types de variables

Variables **nominales**: aucun ordre

Le mode de scrutin dans un pays: (1) majoritaire, (2) proportionnel ou (3) mixte.

# Types de variables

Variables **nominales**: aucun ordre

Le mode de scrutin dans un pays: (1) majoritaire, (2) proportionnel ou (3) mixte.

Variables **nominales dichotomiques**: aucun ordre, 2 catégories

# Types de variables

Variables **nominales**: aucun ordre

Le mode de scrutin dans un pays: (1) majoritaire, (2) proportionnel ou (3) mixte.

Variables **nominales dichotomiques**: aucun ordre, 2 catégories

La participation électorale d'un individu: avoir participé (1) vs. ne pas avoir participé (0).

# Types de variables

Variables **ordinales**: ordonnées  
logiquement

# Types de variables

Variables **ordinales**: ordonnées logiquement

Niveau d'éducation: (1) primaire, (2) secondaire, (3) cégep, (4) bac, (5) maîtrise ou (5) doctorat.



# Types de variables

Variables **ordinales**: ordonnées logiquement

Niveau d'éducation: (1) primaire, (2) secondaire, (3) cégep, (4) bac, (5) maîtrise ou (5) doctorat.

Variables **continues (intervalles ou ratio)**: nombre infini de valeurs

# Types de variables

Variables **ordinales**: ordonnées logiquement

Niveau d'éducation: (1) primaire, (2) secondaire, (3) cégep, (4) bac, (5) maîtrise ou (5) doctorat.

Variables **continues (intervalles ou ratio)**: nombre infini de valeurs

PIB d'un pays; Taux de participation lors d'une élection, etc.

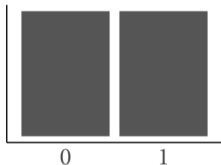
# Distributions

Distributions **discrètes**.

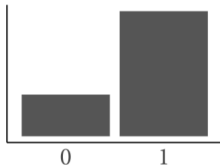
# Distributions

Distributions **discrètes**.

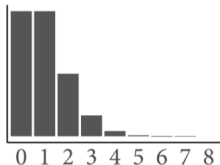
Bernoulli (0.50)



Bernoulli (0.75)

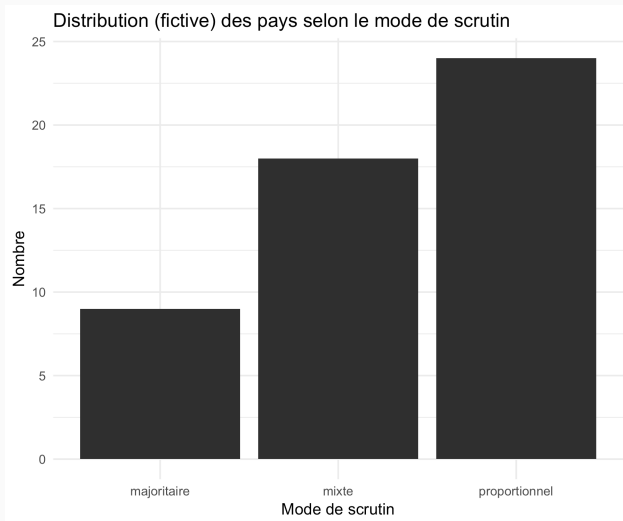


Poisson (1)



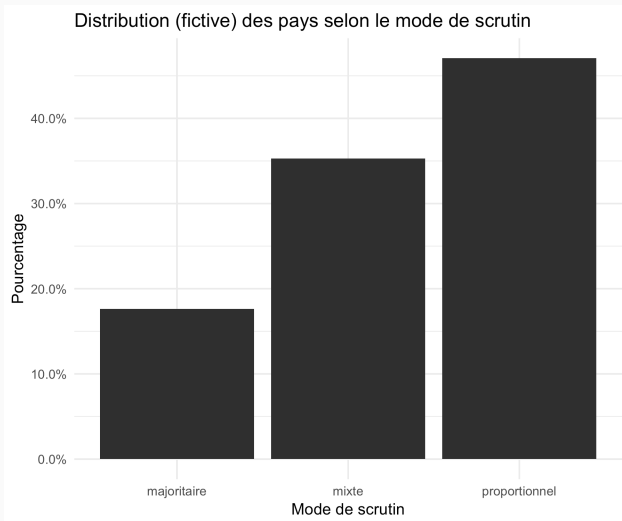
# Distributions

## Distributions discrètes.



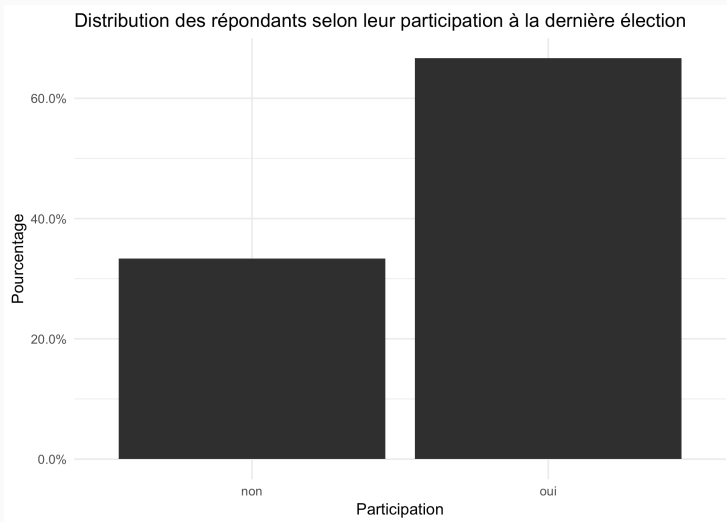
# Distributions

## Distributions discrètes.



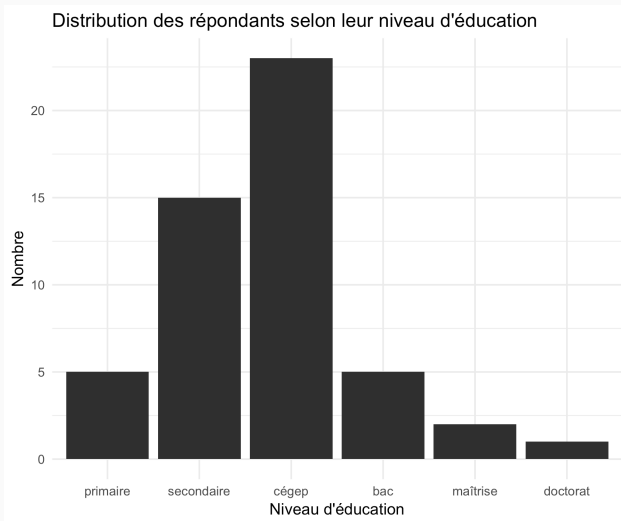
# Distributions

## Distributions discrètes.



# Distributions

## Distributions discrètes.





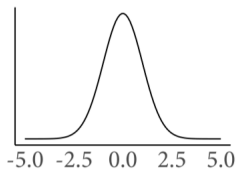
# Distributions

Distributions **continues**.

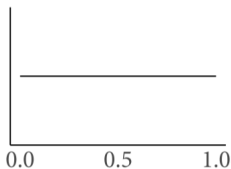
# Distributions

Distributions continues.

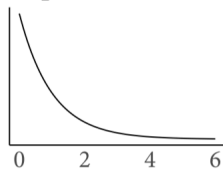
Normale (0, 1)



Uniforme (0, 1)

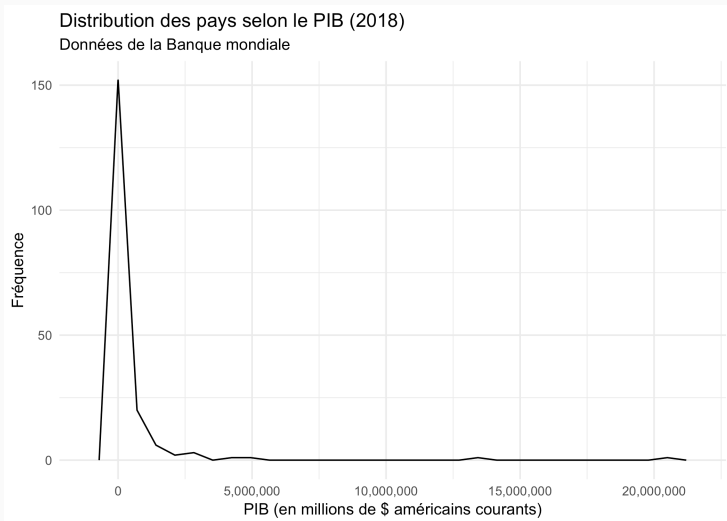


Exponentielle (1)



# Distributions

## Distributions continues.



# Distributions continues

Somme de l'aire sous la courbe = 1  
(100%)

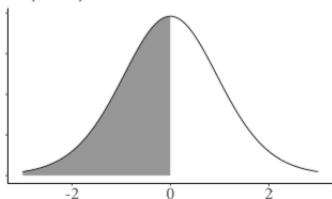
# Distributions continues

Somme de l'aire sous la courbe = 1  
(100%)

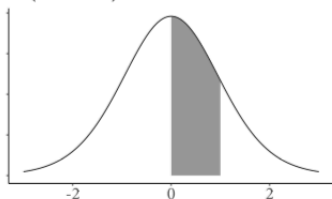
FIGURE 2.3. —

L'aire sous la courbe d'une distribution continue mesure la probabilité de piger un nombre situé dans cet intervalle.

$\Pr(x < 0)$



$\Pr(0 < x < 1)$



Questions?

Questions?

C'est la pause!

# Analyse univariée

Mesures de **tendance centrale**:  
mode, moyenne, médiane

Mesures de **position**: quartiles,  
déciles, centiles

Mesures de **variation**: variance,  
écart-type, écart interquartile



# Mesures de tendance centrale

Moyenne

# Mesures de tendance centrale

Moyenne

$\{1, 2, 2, 4, 5, 10, 20\}$

$$\bar{x} = 6,29$$

# Mesures de tendance centrale

Moyenne

$\{1, 2, 2, 4, 5, 10, 20\}$

$$\bar{x} = 6,29$$

Médiane

# Mesures de tendance centrale

Moyenne

$\{1, 2, 2, 4, 5, 10, 20\}$

$$\bar{x} = 6,29$$

Médiane

$\{1, 2, 2, \boxed{4}, 5, 10, 20\}$

$$\text{médiane} = 4$$

# Mesures de tendance centrale

Moyenne

$\{1, 2, 2, 4, 5, 10, 20\}$

$$\bar{x} = 6,29$$

Médiane

$\{1, 2, 2, \boxed{4}, 5, 10, 20\}$

$$\text{médiane} = 4$$

Mode

# Mesures de tendance centrale

Moyenne

$\{1, 2, 2, 4, 5, 10, 20\}$

$$\bar{x} = 6,29$$

Médiane

$\{1, 2, 2, \boxed{4}, 5, 10, 20\}$

$$\text{médiane} = 4$$

Mode

$\{1, \boxed{2, 2}, 4, 5, 10, 20\}$

$$\text{mode} = 2$$

# Mesures de position

Quartiles



# Mesures de position

Quartiles



Déciles





# Mesures de position

Quartiles



Déciles



Centiles



# Mesures de dispersion

Variance

# Mesures de dispersion

Variance

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

pour un échantillon

# Mesures de dispersion

Variance

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

pour un échantillon

Écart-type

# Mesures de dispersion

Variance

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

pour un échantillon

Écart-type

$$s_x = \sqrt{\textit{variance}_x}$$

# Mesures de dispersion

Variance

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

pour un échantillon

Écart-type

$$s_x = \sqrt{\textit{variance}_x}$$

Écart interquartile

# Mesures de dispersion

Variance

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

pour un échantillon

Écart-type

$$s_x = \sqrt{\textit{variance}_x}$$

Écart interquartile



Questions?



# Analyse bivariable

# Analyse bivariable

Covariance

# Analyse bivariable

Covariance

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

# Analyse bivariable

Covariance

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

x	$\bar{x}$	$x_i - \bar{x}$	y	$\bar{y}$	$y_i - \bar{y}$
1	3	-2	2	6	-4
2	3	-1	4	6	-2
3	3	0	6	6	0
4	3	1	8	6	2
5	3	2	10	6	4

# Analyse bivariable

Covariance

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

x	$\bar{x}$	$x_i - \bar{x}$	y	$\bar{y}$	$y_i - \bar{y}$
1	3	-2	2	6	-4
2	3	-1	4	6	-2
3	3	0	6	6	0
4	3	1	8	6	2
5	3	2	10	6	4

# Analyse bivariée

Covariance

$$\text{Cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

x	$\bar{x}$	$x_i - \bar{x}$	y	$\bar{y}$	$y_i - \bar{y}$
1	3	-2	2	6	-4
2	3	-1	4	6	-2
3	3	0	6	6	0
4	3	1	8	6	2
5	3	2	10	6	4

$$(8 + 2 + 2 + 8) \div 4 = 5$$

# Analyse bivariable

## Corrélation de Pearson

# Analyse bivariable

## Corrélation de Pearson

$$r_{xy} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$$



# Analyse bivariable

## Corrélation de Pearson

$$r_{xy} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$$

Exemple précédent:  $\sigma_x = 1,58$  et  $\sigma_y = 3,16$

$$1,58 * 3,16 = 5$$

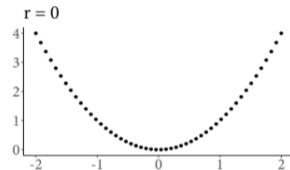
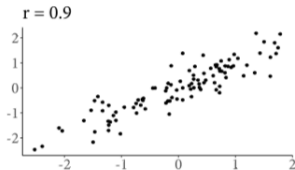
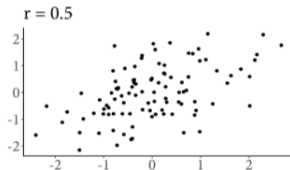
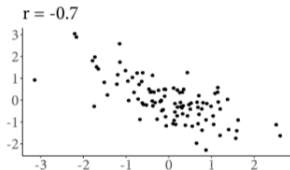
$$\text{Donc } r_{xy} = 5 \div 5 = 1$$

# Analyse bivariée

## Corrélation de Pearson

FIGURE 4.3. –

Association linéaire (corrélation de Pearson) entre deux variables.



Questions?

Questions?

C'est la pause!

# Inférence

# Inférence

Loi des grands nombres: lorsque la taille des échantillons augmente, les moyennes de ces échantillons convergent vers la vraie moyenne.

# Inférence

Loi des grands nombres: lorsque la taille des échantillons augmente, les moyennes de ces échantillons convergent vers la vraie moyenne.

Un petit exemple.

# La loi des grands nombres: un petit exemple

Dés roses: 5 lancés.

Dés bleus: 10 lancés.

Dés oranges: 15 lancés.

Dés verts: 20 lancés.

Lancez, puis faites la moyenne.



# Inférence d'une moyenne

Erreur type (écart-type des moyennes):

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

# Inférence d'une moyenne

Erreur type (écart-type des moyennes):

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

À partir d'une seule moyenne, on peut inférer la moyenne d'une population (avec intervalles et niveau de confiance).

**Exemple:** la taille des étudiant.e.s en science politique.

# Inférence d'une relation

Est-il probable que la relation trouvée entre 2 variables existe réellement dans la population?

# Inférence d'une relation

Est-il probable que la relation trouvée entre 2 variables existe réellement dans la population?

**Niveau de signification statistique:** la probabilité de trouver une association dans l'échantillon quand il n'y en a pas dans la population.

# Inférence d'une relation

Formulation d'une hypothèse nulle ( $H_0$ ):  
Il n'existe pas de relation entre  $X$  et  $Y$ .

# Inférence d'une relation

Formulation d'une hypothèse nulle ( $H_0$ ):  
Il n'existe pas de relation entre  $X$  et  $Y$ .

Si la probabilité est  $< 0,05$  (5%), on peut rejeter l'hypothèse nulle.

Questions?

# Attention!

Signification statistique  $\neq$  Association statistique

Association statistique  $\neq$  Lien causal



**À la semaine  
prochaine!**

