

Méthodes quantitatives d'analyse (POL 2809)

Séance 3, 18 septembre 2019

Enseignante: Florence Vallée-Dubois

Bureau: C-3114

Dispos: mercredis, 10h-11h30

florence.vallee-dubois@umontreal.ca

Retour sur la séance 2

Courbe normale, loi des grands nombres et théorème central limite.

Explications au tableau.

Plus d'infos aux pages 46 à 49 du manuel.

Exemple animé du théorème central limite: http://onlinestatbook.com/stat_sim/sampling_dist/index.html.

Aujourd'hui

Introduction à la régression linéaire.

Interprétation.

Biais et incertitude du coefficient.

Régression linéaire par les moindres carrés

Régression linéaire par les moindres carrés

Une méthode pour caractériser l'association entre deux variables.

Régression linéaire par les moindres carrés

Une méthode pour caractériser l'association entre deux variables.

En termes de sens (positive ou négative), de substance (force) et de signification statistique (généralisation à la population).

Régression linéaire par les moindres carrés

Une méthode pour caractériser l'association entre deux variables.

En termes de sens (positive ou négative), de substance (force) et de signification statistique (généralisation à la population).

Comme l'indique son nom: caractérise une relation **linéaire**.

Régression linéaire par les moindres carrés

Exemple au tableau.

Régression linéaire par les moindres carrés

Comme l'indique son nom: elle minimise
la somme des erreurs au carré.

Régression linéaire par les moindres carrés

Comme l'indique son nom: elle minimise
la somme des erreurs au carré.

Note: je vais utiliser le terme régression
linéaire ou régression **OLS**.

RL par les moindres carrés (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

RL par les moindres carrés (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

Une variable dépendante (pour l'instant, on s'en tient aux VD continues).

RL par les moindres carrés (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

Une variable dépendante.

Un intercept (ou constante).

RL par les moindres carrés (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

Une variable dépendante.

Un intercept (ou constante).

Un coefficient de régression (ou pente).

RL par les moindres carrés (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

Une variable dépendante.

Un intercept (ou constante).

Un coefficient de régression (ou pente).

Une variable indépendante.

RL par les moindres carrés (ou OLS)

$$Y = \hat{\alpha} + \hat{\beta} * X + \varepsilon$$

Une variable dépendante.

Un intercept (ou constante).

Un coefficient de régression (ou pente).

Une variable indépendante.

Un terme résiduel (ou terme d'erreur).

Exercices

$$Y = 5 + 0,03 * X$$

Quelle est la valeur de l'intercept?

Quelle est la valeur du coefficient de régression?

Réponses

$$Y = 5 + 0,03 * X$$

Quelle est la valeur de l'intercept? 5

Quelle est la valeur du coefficient de régression? 0,03

Exercices

$$Y = -0,7 + 300 * X$$

À quel valeur de Y la pente croise-t-elle $X = 0$?

Quand X augmente, Y augmente ou diminue?

Réponses

$$Y = -0,7 + 300 * X$$

À quel valeur de Y la pente croise-t-elle $X = 0$? -0,7

Quand X augmente, Y augmente ou diminue? Il augmente (β_1 est positif).

Exercices

$$Y = -0,03 - 2 * X$$

Si X est égal à 3, quelle est la valeur prédite de Y ?

Si X est égal à -2, quelle est la valeur prédite de Y ?

Si X est égal à 0, quelle est la valeur prédite de Y ?

Réponses

$$Y = -0,03 - 2 * X$$

Si ***X*** est égal à 3, quelle est la valeur prédite de ***Y***? -6,03

Si ***X*** est égal à -2, quelle est la valeur prédite de ***Y***? 3,97

Si ***X*** est égal à 0, quelle est la valeur prédite de ***Y***? -0,03 (la constante)

Questions?

Questions?

C'est la pause!

Interprétation de $\hat{\beta}$

Interprétation de $\hat{\beta}$

Si X est une variable...

... **continue**: Une augmentation de 1 unité dans la valeur de X est associée avec une augmentation/diminution de [...] unités dans la valeur de Y.

Variable X continue: exemple

Déterminants du vote par circonscription, 2007-2012

	PLQ	PQ	ADQ-CAQ
Anglophone (%)	0,45 (0,04)***	-0,44 (0,05)***	-0,08 (0,05)***
Minorité (%)	0,30 (0,05)***	-0,25 (0,06)***	-0,17 (0,07)***
Montréal-Laval	-1,98 (1,37)	-2,09 (1,54)	-2,05 (1,65)
Couronne	-2,48 (1,07)***	-2,15 (1,20)***	5,09 (1,29)***
Saguenay	3,03 (1,65)***	0,30 (1,86)	-2,34 (2,00)
Outaouais	2,63 (1,76)	-5,65 (1,98)***	-1,70 (0,43)
Sud-Est	3,91 (1,29)***	-16,94 (1,45)***	12,60 (1,56)***
Mauricie	4,67 (1,69)***	-6,95 (1,90)***	2,05 (2,04)
Estrie	-1,56 (1,96)	-4,59 (2,21)***	4,02 (2,37)*
Région de Québec	2,67 (1,25)***	-13,90 (1,41)***	11,05 (1,51)***
Sortant PLQ	12,61 (1,16)***	2,81 (1,30)***	-12,00 (1,39)***
Sortant PQ	1,32 (1,17)	10,49 (1,31)***	-11,99 (1,41)***
Candidats effectifs (t-1)	-10,06 (1,00)***	1,83 (1,12)*	2,79 (1,20)**
Nombre de candidats (t)	-0,51 (0,30)***	0,12 (0,34)	-0,30 (1,14)
2008	16,35 (0,95)***	6,83 (1,06)***	-18,82 (1,14)***
2012	-1,89 (0,94)***	4,79 (1,06)***	-2,99 (1,14)***
Constante	50,47 (2,97)***	26,43 (3,34)***	34,15 (3,58)***
R ² ajusté	0,86	0,73	0,71
N	362	362	362

Figure: Godbout, Jean-François. "Les élections au Québec de 1973 à 2012," dans Les Québécois aux urnes (2013), p.37

Variable X continue: exercice

VD: Le nombre de morts durant la 2e Guerre mondiale, dans un pays donné.

VI: L'ampleur du commerce international entre ce pays et l'Allemagne, mesuré en dollars américains.

Équation: $\text{morts} = 10000 - 20 * \text{dollars}$

Interprétez le coefficient.

Variable X continue: réponse

VD: Le nombre de morts durant la 2e Guerre mondiale, dans un pays donné.

VI: L'ampleur du commerce international entre ce pays et l'Allemagne, mesuré en dollars américains.

Équation: $\text{morts} = 10000 - 20 * \text{dollars}$

Interprétation: Chaque dollar supplémentaire de commerce entre un pays donné et l'Allemagne est associé à 20 morts de moins dans ce pays lors de la DGM.

Interprétation de $\hat{\beta}$

Si X est une variable...

... **dichotomique**: Une augmentation de 0 à 1 dans la valeur de X est associée avec une augmentation/diminution de [...] unités dans la valeur de Y.

Variable X dichotomique: exemple fictif

VD: Revenu annuel d'une personne, en milliers de dollars.

VI: Appartenir à un syndicat (= 1), ou pas (= 0).

VD: Revenu (en millier de \$)	
Constante	35
Appartenance à un syndicat	2,04

Variable X dichotomique: exemple fictif

VD: Revenu (en millier de \$)	
Constante	35
Appartenance à un syndicat	2,04

Appartenir à un syndicat augmente le revenu annuel de 2,04 milliers de dollars.

Variable X dichotomique: exercice

VD: La probabilité d'aller voter aux prochaines élections fédérales, mesurée en pourcentage.

VI: Le fait d'être étudiant.e à l'université (= 1), ou pas (= 0).

Équation:

probabilité de voter = $60 + 0,03 * \text{étudiant}$

Interprétez le coefficient.

Variable X dichotomique: réponse

VD: La probabilité d'aller voter aux prochaines élections fédérales, mesurée en pourcentage.

VI: Le fait d'être étudiant.e à l'université (= 1), ou pas (= 0).

Équation:

probabilité de voter = $60 + 0,03 * \text{étudiant}$

Interprétation: Être étudiant.e à l'université augmente la probabilité d'aller voter aux prochaines élections de 0,03 points de pourcentage.

Variable X dichotomique: exercice

VD: Le nombre de lois pour la protection de l'environnement dans un pays donné.

VI: Le fait que ce pays soit producteur de pétrole (= 1), ou pas (= 0).

Équation: nombre lois = $2 - 1 * \text{prod. pétrole}$

Interprétez le coefficient.

Variable X dichotomique: réponse

VD: Le nombre de lois pour la protection de l'environnement dans un pays donné.

VI: Le fait que ce pays soit producteur de pétrole (= 1), ou pas (= 0).

Équation: nombre lois = $2 - 1 * \text{prod. pétrole}$

Interprétation: Les pays producteurs de pétrole ont en moyenne 1 loi de moins pour la protection de l'environnement que les pays qui ne sont pas producteurs de pétrole.

Interprétation de $\hat{\beta}$

Si X est une variable...

... **nominale**: En comparaison avec la catégorie de référence, la catégorie [...] est associée avec une augmentation/diminution de [...] unités dans la valeur de Y.

Variable X nominale: exemple fictif

VD: Revenu (en millier de \$)	
Constante	35
Secteur d'emploi (référence = industriel)	
Services publics	-0,5
Arts et culture	-4,2
Administration	0,006
Construction	0,9

Variable X nominale: exercice

VD: L'appréciation du gouvernement sortant, mesuré en pourcentage.

VI: Le fait d'être un donateur d'un des partis politiques suivants: Libéral, Conservateur, NPD, Parti vert, BQ, aucun.

Intercept	50
(référence = Libéral)	
Conservateur	-5
NPD	-0,3
Part vert	-0,0005
BQ	-3,5
aucun	-0,04

Variable X nominale: réponse

Les donateurs du parti Conservateur ont une appréciation du gouvernement sortant qui est de 5 points inférieure à celle des donateurs du parti Libéral.

Les donateurs du NPD ont une appréciation du gouvernement sortant qui est de 0,3 points inférieure à celle des donateurs du parti Libéral.

etc.

Interprétation de $\hat{\beta}$

Interprétation de $\hat{\beta}$

Si X est une variable...

... **ordinaire**: parfois interprété comme une variable continue, parfois comme une variable nominale

Questions?

Questions?

C'est la pause!

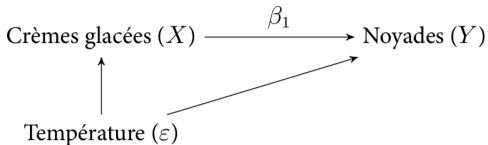
Coefficient de régression non biaisé

Coefficient de régression non biaisé

Condition la plus importante: que X soit indépendant de toutes les variables ignorées par le modèle (en d'autres mots, tout de qui se trouve dans ε).

Coefficient de régression non biaisé

Condition la plus importante: que X soit indépendant de toutes les variables ignorées par le modèle (en d'autres mots, tout de qui se trouve dans ε).



Arel-Bundock, Vincent. Analyse causale et méthodes quantitatives (à venir).

Uncertainty of β

Incertitude de β

Variance échantillonnale: mesure comment nos estimés (de β_1 , de α) changent lorsqu'on les calcule à partir de différents échantillons.

$$\sigma_{\hat{\beta}_1}^2 = \frac{\sigma_\varepsilon^2}{n \cdot \sigma_X^2}$$

Incertitude de β

La mesure de la variance n'est pas à la même échelle que nos mesures. On préfère prendre la racine carré de la variance, qui est:

Incertitude de β

La mesure de la variance n'est pas à la même échelle que nos mesures. On préfère prendre la racine carré de la variance, qui est:

L'erreur-type.

En termes mathématiques: $\sigma_{\hat{\beta}_1} = \frac{\sigma_{\varepsilon}}{\sqrt{n} \sigma_X}$

Incertitude de β

L'incertitude des estimés dépend:

Incertitude de β

L'incertitude des estimés dépend:
Du nombre d'observations (n).

Incertitude de β

L'incertitude des estimés dépend:

- Du nombre d'observations (n).

- De la taille des erreurs de prédictions (ε).

Incertitude de β

L'incertitude des estimés dépend:

- Du nombre d'observations (n).

- De la taille des erreurs de prédictions (ε).

- De la variation dans X .

Analogie:



Questions?

Attention!

Peu d'incertitude \neq Association statistique forte

Association statistique \neq Causalité

**À la semaine
prochaine!**

