

Méthodes quantitatives d'analyse (POL 2809)

Séance 4, 25 septembre 2019

Enseignante: Florence Vallée-Dubois

Bureau: C-3114

Dispos: mercredis, 10h-11h30

florence.vallee-dubois@umontreal.ca

Rappels

Remise du devoir 1: au plus tard vendredi le 27 (23h59).

Retards acceptés jusqu'à mardi prochain.

Correction en classe la semaine prochaine.

Retour sur la séance 3

Explications au tableau.

Plus d'infos aux pages 63 à 72 du manuel.

Aujourd'hui

Nos estimés sont-ils "trop" incertains?

Hypothèse nulle et calcul de la signification statistique.

Exercices et réalisation du devoir 1.

Signification statistique

La probabilité qu'on ait trouvé une relation dans l'échantillon alors qu'il n'y en a pas dans la population.

La probabilité que nos estimés soient le fruit du hasard.

Signification statistique

La probabilité qu'on ait trouvé une relation dans l'échantillon alors qu'il n'y en a pas dans la population.

La probabilité que nos estimés soient le fruit du hasard.

Si cette probabilité est trop élevée, les estimés ne sont pas statistiquement significatifs.

Signification statistique

La probabilité qu'on ait trouvé une relation dans l'échantillon alors qu'il n'y en a pas dans la population.

La probabilité que nos estimés soient le fruit du hasard.

Si cette probabilité est trop élevée, les estimés ne sont pas statistiquement significatifs.

Si cette probabilité n'est pas trop élevée, les estimés sont statistiquement significatifs.

Première étape: formulation d'une hypothèse nulle

Hypothèse nulle: Il n'y a pas de relation entre X (la VI) et Y (la VD).

$$H_0 : \beta = 0$$

Pour rejeter l'hypothèse nulle...

Pour pouvoir dire qu'une relation existe probablement dans la population ($\beta \neq 0$), la probabilité que nos estimés soient le fruit du hasard doit être faible.

Pour rejeter l'hypothèse nulle...

Pour pouvoir dire qu'une relation existe probablement dans la population ($\beta \neq 0$), la probabilité que nos estimés soient le fruit du hasard doit être faible.

Elle doit être inférieure à 5% ($< 0,05$).

Pour rejeter l'hypothèse nulle...

Si la probabilité $> 0,05$, on ne peut pas rejeter l'hypothèse nulle. Il y a plus de 5% de chances qu'il n'existe pas de relation dans la pop.

Pour rejeter l'hypothèse nulle...

Si la probabilité $> 0,05$, on ne peut pas rejeter l'hypothèse nulle. Il y a plus de 5% de chances qu'il n'existe pas de relation dans la pop.

Si la probabilité $< 0,05$, on peut rejeter l'hypothèse nulle. Il y a au moins 95% de chances qu'il y ait une relation dans la population.

Les limites du niveau de signification statistique

Ce n'est pas parce qu'un coefficient (β) est significatif qu'il est fort.

Les limites du niveau de signification statistique

Ce n'est pas parce qu'un coefficient (β) est significatif qu'il est fort.

Ce standard peut créer des incitatifs néfastes pour la recherche:

Les limites du niveau de signification statistique

Ce n'est pas parce qu'un coefficient (β) est significatif qu'il est fort.

Ce standard peut créer des incitatifs néfastes pour la recherche:

"p-hacking"

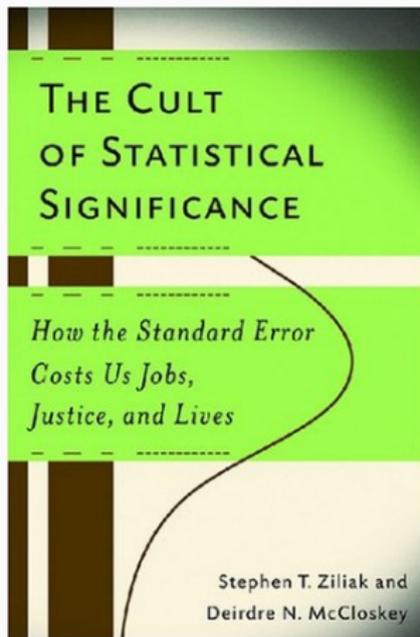
Les limites du niveau de signification statistique

Ce n'est pas parce qu'un coefficient (β) est significatif qu'il est fort.

Ce standard peut créer des incitatifs néfastes pour la recherche:

- "p-hacking"
- problème du tiroir

Les limites du niveau de signification statistique



Les limites du niveau de signification statistique

Crise de la reproductibilité? Bonne vulgarisation ici:

`https:`

`//www.wnycstudios.org/podcasts/
radiolab/articles/stereothreat
(Stereothreat, de Radiolab)`

Les limites du niveau de signification statistique



Comment savoir si β est significatif?

Trois manières

Calculer la statistique t

Se référer à la valeur p

Calculer l'intervalle de confiance de 95% autour de notre coefficient.

1ère option: Calculer la statistique t

Elle compare le coefficient et son erreur type.

1ère option: Calculer la statistique t

Elle compare le coefficient et son erreur type.

$$t = \frac{\hat{\beta} - H_0}{\hat{\sigma}_{\hat{\beta}}}$$

1ère option: Calculer la statistique t

Elle compare le coefficient et son erreur type.

$$t = \frac{\hat{\beta} - H_0}{\hat{\sigma}_{\hat{\beta}}}$$

Vu que notre H_0 est qu'il n'y a pas de relation, ça revient à:

$$t = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$$

1ère option: Calculer la statistique t

Elle compare le coefficient et son erreur type.

$$t = \frac{\hat{\beta} - H_0}{\hat{\sigma}_{\hat{\beta}}}$$

Vu que notre H_0 est qu'il n'y a pas de relation, ça revient à:

$$t = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$$

Le coefficient divisé par son erreur type.

Interpréter la statistique t

Prendre la valeur absolue de t ($|t|$).

Interpréter la statistique t

Prendre la valeur absolue de t ($|t|$).

Plus $|t|$ est grande, plus le coefficient est gros par rapport à son erreur type.

Interpréter la statistique t

Règle approximative: pour obtenir un niveau de signification de $< 0,05$, la valeur absolue de t doit être plus grande ou égale à **2**.

Interpréter la statistique t

Règle approximative: pour obtenir un niveau de signification de $< 0,05$, la valeur absolue de t doit être plus grande ou égale à **2**.

En d'autres mots, le coefficient doit être au moins deux fois plus grand que son erreur type (en valeur abs.)!

Statistique t : exemples et exercices

VD: Nb de pots de vin acceptés

	estimés	err. type	stat. t
Constante	30	2,4	12,5
Isolationnisme	2,5	0,56	?

Interprétez le coefficient (en substance).

Calculez la statistique t . Est-ce significatif (à $< 0,05$)?

Statistique t : exemples et exercices

VD: Nb heures de bénévolat

	estimés	err. type	stat. t
Constante	200	10	20
Cours citoyen.	20	45	?

Interprétez le coefficient (en substance).

Calculez la statistique t . Est-ce significatif ($\alpha < 0,05$)?

Questions?

2e option: Consulter la valeur p

Explications au tableau.

2e option: Consulter la valeur p

Explications au tableau.

Valeur p bilatérale: probabilité d'obtenir une statistique t aussi extrême.

2e option: Consulter la valeur p

Explications au tableau.

Valeur p bilatérale: probabilité d'obtenir une statistique t aussi extrême.

En d'autres mots, elle vous dit s'il y a des fortes chances que votre statistique t (et donc, le coefficient) soit le fruit du hasard.

Interpréter la valeur p

La valeur p bilatérale doit être $< 0,05$.

Interpréter la valeur p

La valeur p bilatérale doit être $< 0,05$.

Si elle est $< 0,05$, il y a moins de 5% de chances que nous ayons obtenu un tel coefficient s'il n'existe aucune relation dans la population.

Interpréter la valeur p

La valeur p bilatérale doit être $< 0,05$.

Si elle est $< 0,05$, il y a moins de 5% de chances que nous ayons obtenu un tel coefficient s'il n'existe aucune relation dans la population.

Manière plus rapide et directe de savoir si c'est significatif. Les outils informatiques nous la fourniront.

Interpréter la valeur p

La valeur p bilatérale doit être $< 0,05$.

Si elle est $< 0,05$, il y a moins de 5% de chances que nous ayons obtenu un tel coefficient s'il n'existe aucune relation dans la population.

Manière plus rapide et directe de savoir si c'est significatif. Les outils informatiques nous la fourniront.

Exemples et exercices à venir dans un instant.

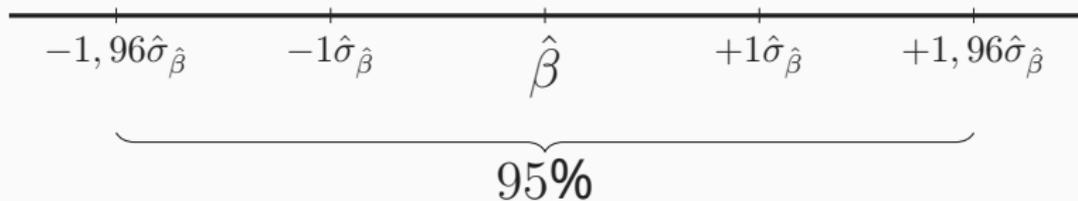
Questions?

3e option: Calculer l'intervalle de confiance de 95%

Explications au tableau.

3e option: Calculer l'intervalle de confiance de 95%

Explications au tableau.



Interpréter l'intervalle de confiance de 95%

0 ne doit pas se retrouver dans l'intervalle.

Interpréter l'intervalle de confiance de 95%

0 ne doit pas se retrouver dans l'intervalle.

Sinon, on ne peut pas rejeter la possibilité que $\beta = 0$.

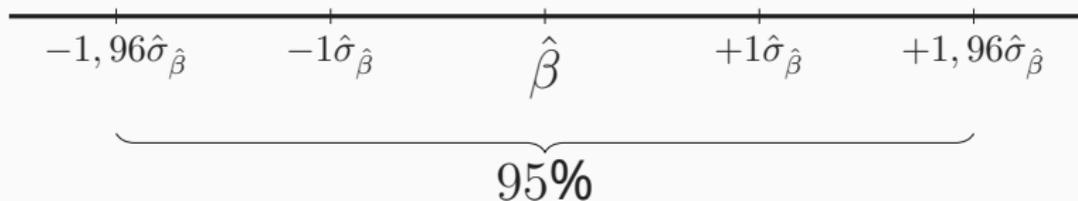
Calcul de l'intervalle de confiance de 95%: exemples et exercices

VD: Nb de pots de vin acceptés

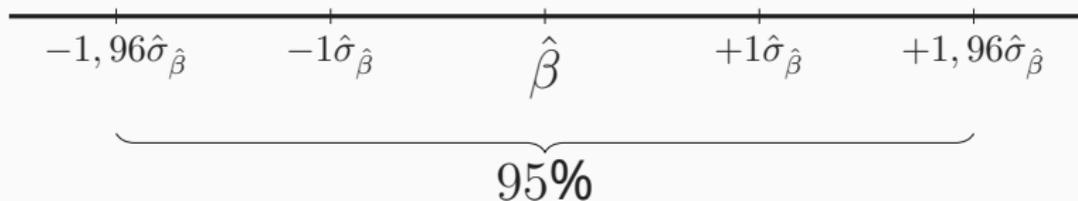
	estimés	err. type	stat. t
Constante	30	2,4	12,5
Isolationnisme	2,5	0,56	4,46

Calculez l'intervalle de confiance de 95%

Calcul de l'intervalle de confiance de 95%: exemples et exercices

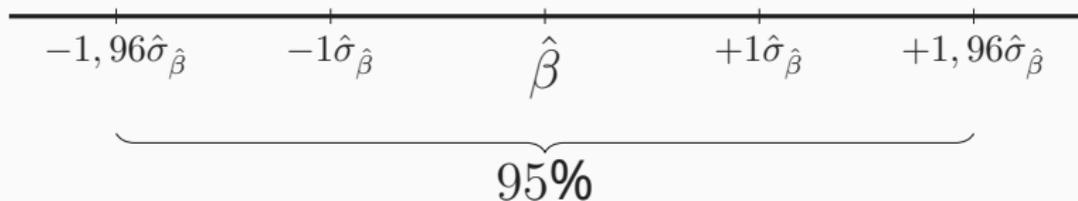


Calcul de l'intervalle de confiance de 95%: exemples et exercices

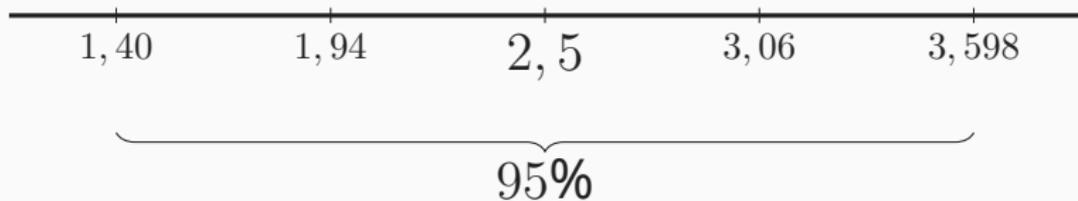


Si $\hat{\beta} = 2,5$ et erreur type = $0,56$

Calcul de l'intervalle de confiance de 95%: exemples et exercices



Si $\hat{\beta} = 2,5$ et erreur type = $0,56$



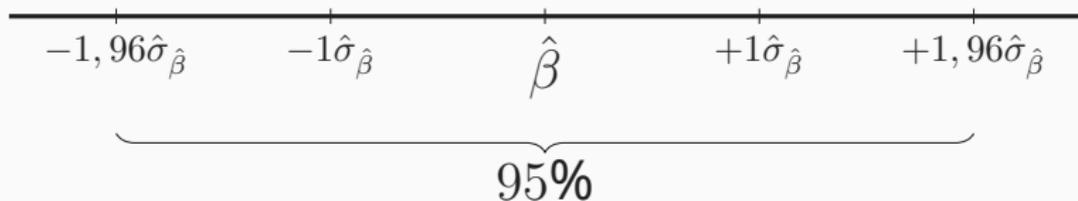
Calcul de l'intervalle de confiance de 95%: exemples et exercices

VD: Nb heures de bénévolat

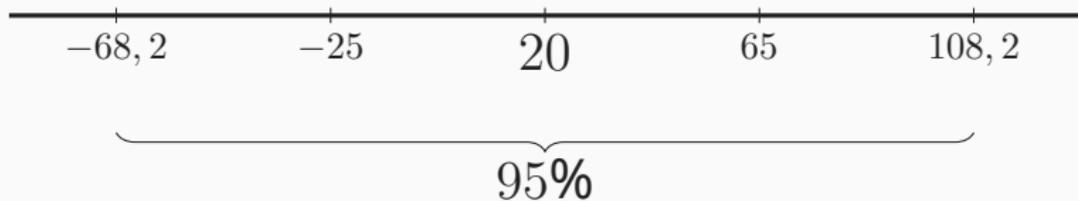
	estimés	err. type	stat. t
Constante	200	10	20
Cours citoyen.	20	45	0,44

Calculez l'intervalle de confiance de 95%

Calcul de l'intervalle de confiance de 95%: exemples et exercices



Si $\hat{\beta} = 20$ et erreur type = 45



Questions?

Questions?

C'est la pause!

Récapitulons

Pour que le coefficient trouvé soit significatif...

$$t \geq 2 \text{ ou}$$

$$p < 0,05 \text{ ou}$$

0 ne se trouve pas dans l'intervalle de confiance de 95%.

Pourquoi apprendre les 3?

Les textes ne rapportent pas toujours toutes les informations (erreurs type seulement, *p* seulement, etc.)

Pourquoi apprendre les 3?

Les textes ne rapportent pas toujours toutes les informations (erreurs type seulement, *p* seulement, etc.)

Le tout peut être simplifié grâce aux ***

Exemples et exercices: valeur p et étoiles

Political Psychology, Vol. 20, No. 2, 1999

Forming Impressions of Political Leaders: A Cross-National Comparison

S. Mark Pancer

*Department of Psychology
Wilfrid Laurier University, Waterloo, Ontario*

Steven D. Brown and Cathy Widdis Barr

*Department of Political Science
Wilfrid Laurier University, Waterloo, Ontario*

Exemples et exercices: valeur p et étoiles

Table IV. Ordinary Least Squares Regression of Thermometer Scores on the Three Summary Image Factors

Target figure	Constant	Charisma	Factor Competence	Integrity	R^2 for equation	N
Clinton (U.S.)	-13.10**	18.95***	48.81***	32.01***	.52***	261
Bush (U.S.)	-20.27***	25.85***	36.60***	47.19***	.55***	255
Quayle (U.S.)	-3.69	12.73*	32.47***	27.07***	.37***	243

Note. Cell entries for factor coefficients are unstandardized b -weights. Because factor metrics range from 0 to 1 and thermometer scores range from 0 to 100, b -weights represent the change in thermometer rating associated with a shift from the lowest to the highest rating on each factor. The Clinton (U.S.) and Thatcher (U.K.) rows each appear twice as an aid to comparison among groups.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Exemples et exercices: valeur p et étoiles

American Political Science Review

Vol. 88, No. 3 September 1994

ISSUES, CANDIDATE IMAGE, AND PRIMING: THE USE OF PRIVATE POLLS IN KENNEDY'S 1960 PRESIDENTIAL CAMPAIGN

LAWRENCE R. JACOBS *University of Minnesota*

ROBERT Y. SHAPIRO *Columbia University*

Exemples et exercices: valeur p et étoiles

TABLE 2

Effect of Kennedy's Poll Results on the Content of His Public Statements during the 1960 Election

WEEK OF CAMPAIGN	BIVARIATE REGRESSION (b)	CORRELATION COEFFICIENTS (r)
Opening ^a		
September		
Week 3	.37*	(.33*)
Week 4	.08	(.15)
October		
Week 5	.40*	(.30*)
Home stretch ^b		
Week 7	.55*	(.54*)
Week 9	.78*	(.54*)
November		
Week 10	.32*	(.41*)
Average	.42*	(.38*)

Note: The independent variable is the average percentage of respondents in Kennedy's polls ranking an issue and policy direction as important. The dependent variable is the sum of the content scores for Kennedy's statements about the issues, where each statement was coded on a five-point scale (-2 to +2) depending on the degree of support for the policy. N = 55. No analysis is shown for weeks 1, 2, 6, and 8 because there were no polls in the appropriate preceding period.

^aTwo-week lag between polls and statements.

^bOne-week lag between polls and statements.

* $p \leq .05$.

Attention!

Signification stat. \neq Association stat. forte

*** \neq Association stat. forte

Pour le reste de la séance: devoir en classe (je peux répondre à vos questions!)

**À la semaine
prochaine!**

