Régression Linéaire Simple (une seule variable explicative)

(Stock/Watson Chapitre 4)

<u>Plan</u>

1. Le Modèle de Régression Linéaire dans la population
2. L'estimateur des Moindres Carrés Ordinaires (MCO) et la ligne de régression dans l'échantillon
3. Mesures de la qualité de l'ajustement
4. Les hypothèses des moindres carrés
5. La distribution d'échantillonnage de l'estimateur des MCO

La régression linéaire permet d’estimer la pente de la droite de régression de la population.

- La pente de la droite de régression de la population donne l’effet espéré sur Y d’un changement d’une unité de X.

- Notre but ultime est d’estimer l’effet causal sur Y d’un changement unitaire de X, mais pour l’instant, nous allons simplement voir comment ajuster une droite qui résume la relation entre deux variables, Y et X.

Le problème de l'inférence statistique pour la régression linéaire est très similaire à celui de l'estimation d'une moyenne ou d'une différence de moyennes. L'inférence statistique ou économétrique consiste en:

- L'estimation:
  - Comment tracer une ligne entre les points de l'échantillon, qui estime la pente de la régression dans la population?
    - Réponse: les moindres carrés ordinaires (MCO).
  - Quels sont les avantages et inconvénients des MCO?
- Les tests d'hypothèse:
  - Comment tester si la pente est nulle?
- Les intervalles de confiance:
  - Comment construire un intervalle de confiance pour la pente?

Le Modèle de Régression Linéaire
(SW Section 4.1)

La ***ligne de régression dans la population***:

$$\textit{Test Score} = \beta_0 + \beta_1 STR$$

β_1 = pente de la ligne de régression de la population

$$= \frac{\Delta \text{Test score}}{\Delta STR}$$

= changement dans le résultat des examens par changement unitaire de STR

Pourquoi β_0 et β_1 sont-ils des paramètres de la “population”?

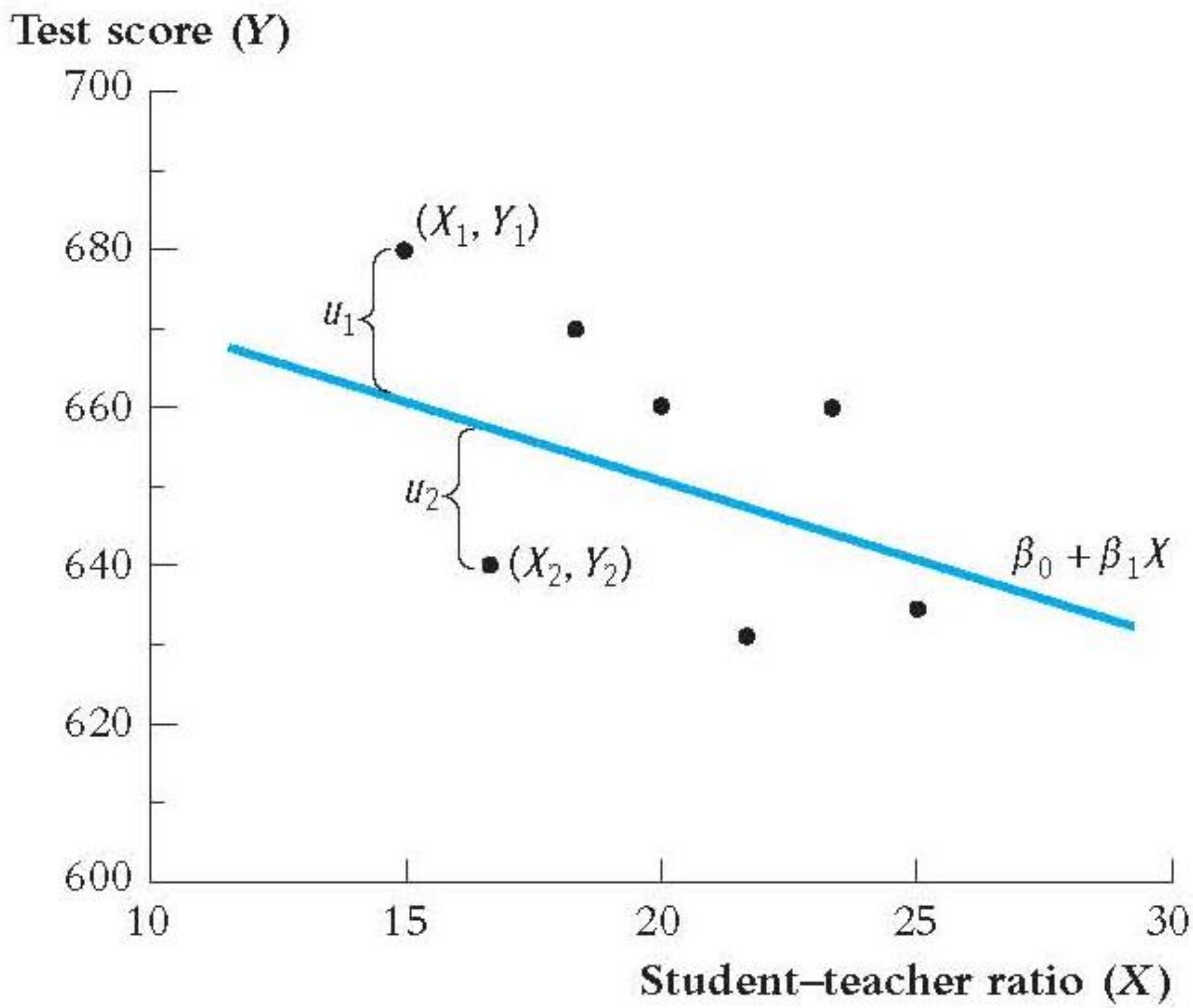
- On aimerait connaître la valeur dans la population, β_1.
- On ne connaît pas β_1, donc on doit l’estimer en utilisant des données.

Le Modèle de Régression Linéaire dans la Population

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

- On a n observations, (X_i, Y_i), $i = 1,.., n$.
- X est la ***variable indépendante*** ou le ***régresseur***
- Y est la ***variable dépendante***
- β_0 = ***constante***
- β_1 = ***pente***
- u_i = les ***erreurs*** de la régression
- Les erreurs de la régression captent l'effet de variables omises. En général ces variables omises sont des variables, autres que X, et qui affectent Y. Les erreurs de la régression contiennent aussi les erreurs de mesure de Y.

Le Modèle de Régression Linéaire dans la Population dans un graphe: Observations de Y et X (n = 7); droite de régression dans la population; et erreurs de régression ("terme d'erreur"):



L'Estimateur des Moindres Carrés Ordinaires (MCO) (SW Section 1.2)

Comment peut-on estimer β_0 et β_1 à partir des données?

Rappelez-vous que $\bar{Y}$ est l'estimateur des moindres carrés de μ_Y: $\bar{Y}$ résout

$$\min_m \sum_{i=1}^{n}(Y_i - m)^2$$

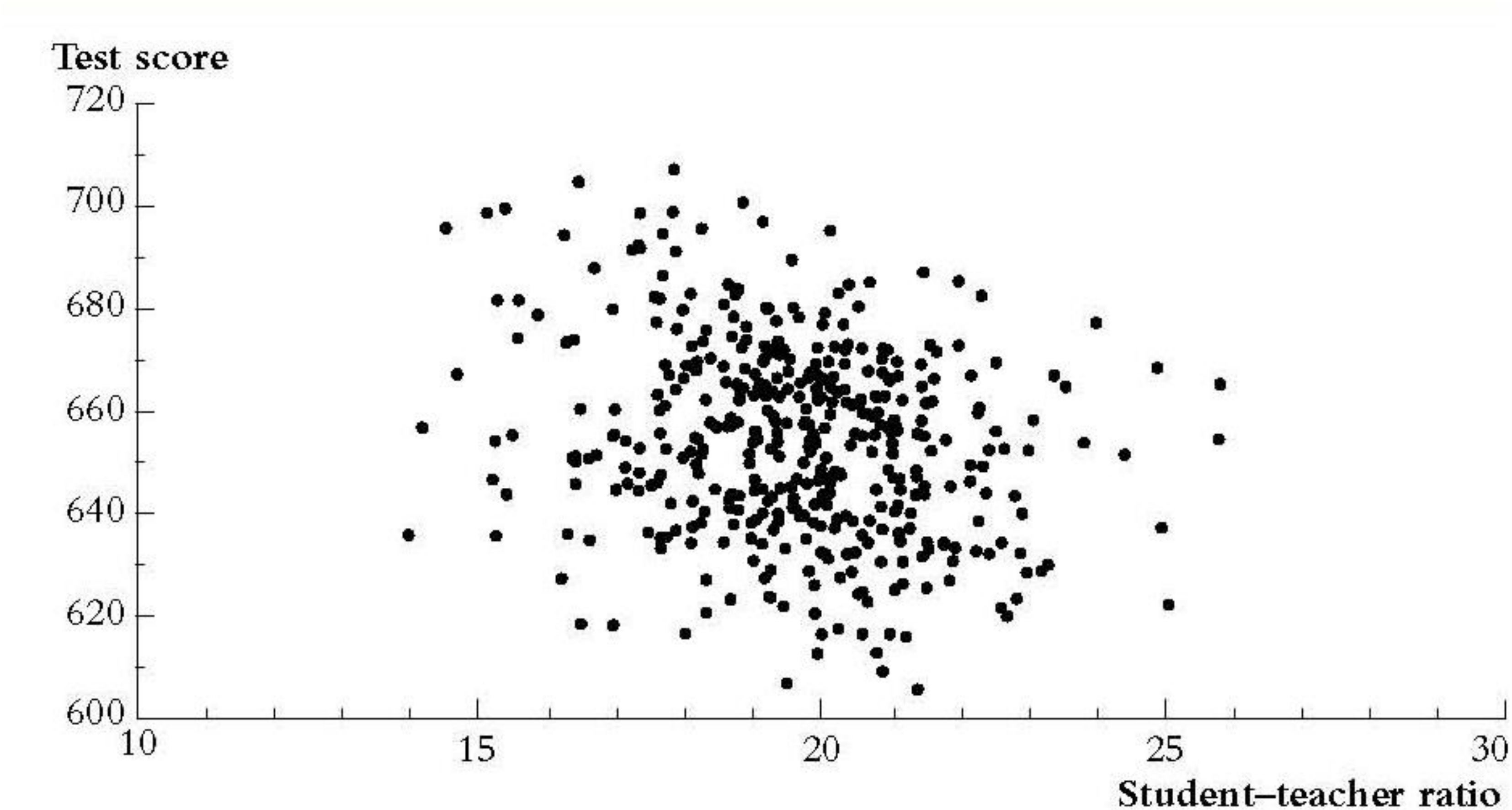
Par analogie, nous allons nous concentrer sur l'estimateur des moindres carrés ("moindres carrés ordinaires" ou "MCO") des **paramètres inconnus $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$**. L'estimateur MCO résout

$$\min_{b_0, b_1} \sum_{i=1}^{n}[Y_i - (b_0 + b_1 X_i)]^2$$

La **Mécanique des MCO**

La droite de régression dans la population: *Test Score* = $\beta_0 + \beta_1 STR$

$$\beta_1 = \frac{\Delta\text{Test score}}{\Delta STR} = ??$$



L’estimateur des MCO résout: $\min_{b_0, b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$

- L'estimateur des MCO minimise l'écart quadratique moyen entre les valeurs réelles de Yi et la prédiction ("valeur prédite") basée sur la droite de régression estimée.

- Ce problème de minimisation peut être résolu en utilisant de l’analyse (App. 1.2).

- **Il en résulte les estimateurs MCO de β_0 et β_1.**

The OLS Estimator, Predicted Values, and Residuals

The OLS estimators of the slope β_1 and the intercept β_0 are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.7}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.8}$$

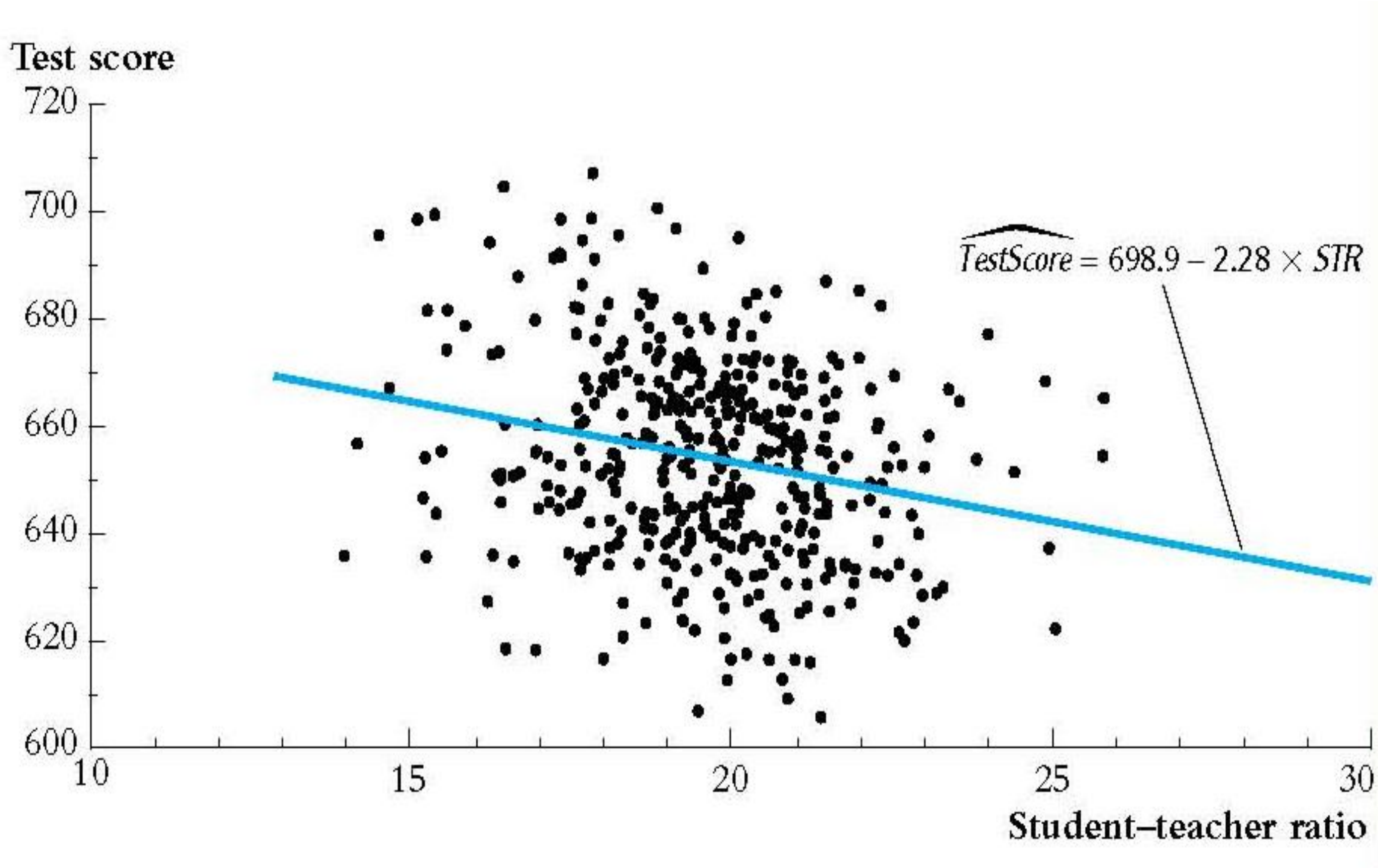
The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,\ i = 1, \ldots, n \tag{4.9}$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \tag{4.10}$$

The estimated intercept $(\hat{\beta}_0)$, slope $(\hat{\beta}_1)$, and residual $(\hat{u}_i)$ are computed from a sample of n observations of X_i and Y_i, $i = 1, \ldots, n$. These are estimates of the unknown true population intercept (β_0), slope (β_1), and error term (u_i).

Application aux données sur les résultats d’examen *Test Score – Taille des Classes Size*



Pente estimée = $\hat{\beta}_1$ = – 2.28

Constante = $\hat{\beta}_0$ = 698.9

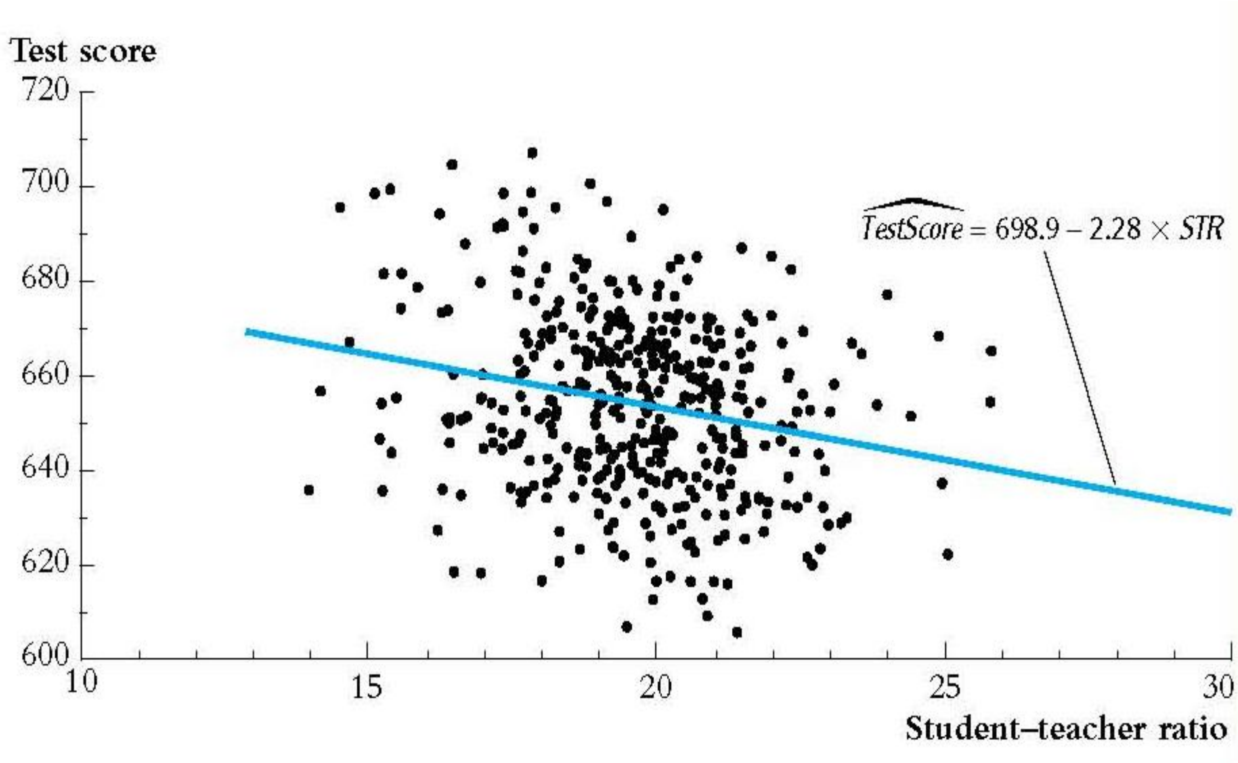
Droite de régression estimée: $\widehat{TestScore}$ = 698.9 – 2.28×*STR*

Interprétation de la pente et de la constante estimées

$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$

- Les académies (districts) avec un étudiant supplémentaire par enseignant ont des résultats moyens d'examen inférieurs de 2.28 points.
- C'est-à-dire, $\dfrac{\Delta \text{Test score}}{\Delta STR} = -2.28$
- La constante (prise littéralement) signifie que, selon cette droite estimée, les académies avec zéro élève par enseignant auraient un résultat (prédit) des tests de 698,9. Mais cette interprétation de la constante n'a pas de sens - il extrapole la droite en dehors de l'intervalle des données - ici, la constante n'a pas d'interprétation économique.

Valeurs Prédites et Résidus:



Un des districts dans les données est Antelope, CA, pour lequel STR = 19.33 et $Test\ Score$ = 657.8

valeur prédite: $\hat{Y}_{Antelope} = 698.9 - 2.28 \times 19.33 = 654.8$

résidu: $\hat{u}_{Antelope} = 657.8 - 654.8 = 3.0$

Régression MCO: output de STATA

```
regress testscr str, robust

Régression with robust standard errors               Number of obs =     420
                                                     F(  1,    418) =   19.26
                                                     Prob > F      =  0.0000
                                                     R-squared     =  0.0512
                                                     Root MSE      =  18.581

-------------------------------------------------------------------------
        |               Robust
testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+----------------------------------------------------------------
    str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
  _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
-------------------------------------------------------------------------
```

$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$

(On expliquera le reste de cet output de STAT plus tard)

Mesure de la Qualité de l'Ajustement (Section 4.3)

Il existe deux statistiques complémentaires calculées à partir de la régression qui mesurent la façon dont la ligne de régression explique les données:

- Le ***R^2* de la *régression*** mesure la fraction de la variance de *Y* expliquée par *X*; cette mesure n'a pas d'unité et elle est entre zéro (pas d'ajustement) et un (ajustement parfait)

- L'***erreur-type de l'estimation* (*ETE*)** mesure la taille moyenne de l'écart entre les points et la droite, en unités de Y.

Le *R^2* de la *régression* est la fraction de la variance de Y_i dans l'échantillon, "expliquée" par la régression.

$Y_i = \hat{Y}_i + \hat{u}_i$ = prédiction MCO + résidu MCO

$\Rightarrow$ var (Y) = var($\hat{Y}_i$) + var($\hat{u}_i$) (pourquoi?)

$\Rightarrow$ Somme de carrés totale = SC "expliquée" + SC "résidus"

Définition du R^2:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- $R^2 = 0$ implique $SCR = 0$
- $R^2 = 1$ implique $SCE = SCT$
- $0 \leq R^2 \leq 1$
- Pour une régression avec un seul X, R^2 = le carré du coefficient de corrélation entre X et Y

L'erreur-type de la régression (*SER*)

La SER mesure la variabilité de la distribution de *u*. La SER est (presque) l'écart-type des résidus MCO :

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}})^2}$$

$$= \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

La deuxième égalité est vérifiée parce que $\bar{\hat{u}} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$.

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

SER:

- Est mesurée en unités de u, qui sont les unités de Y
- Mesure la “taille” moyenne des résidus (l’“erreur” moyenne de la droite de régression des MCO)
- L’***erreur quadratique moyenne*** (EQM) est très proche de l’SER:

$$EQM = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2}$$

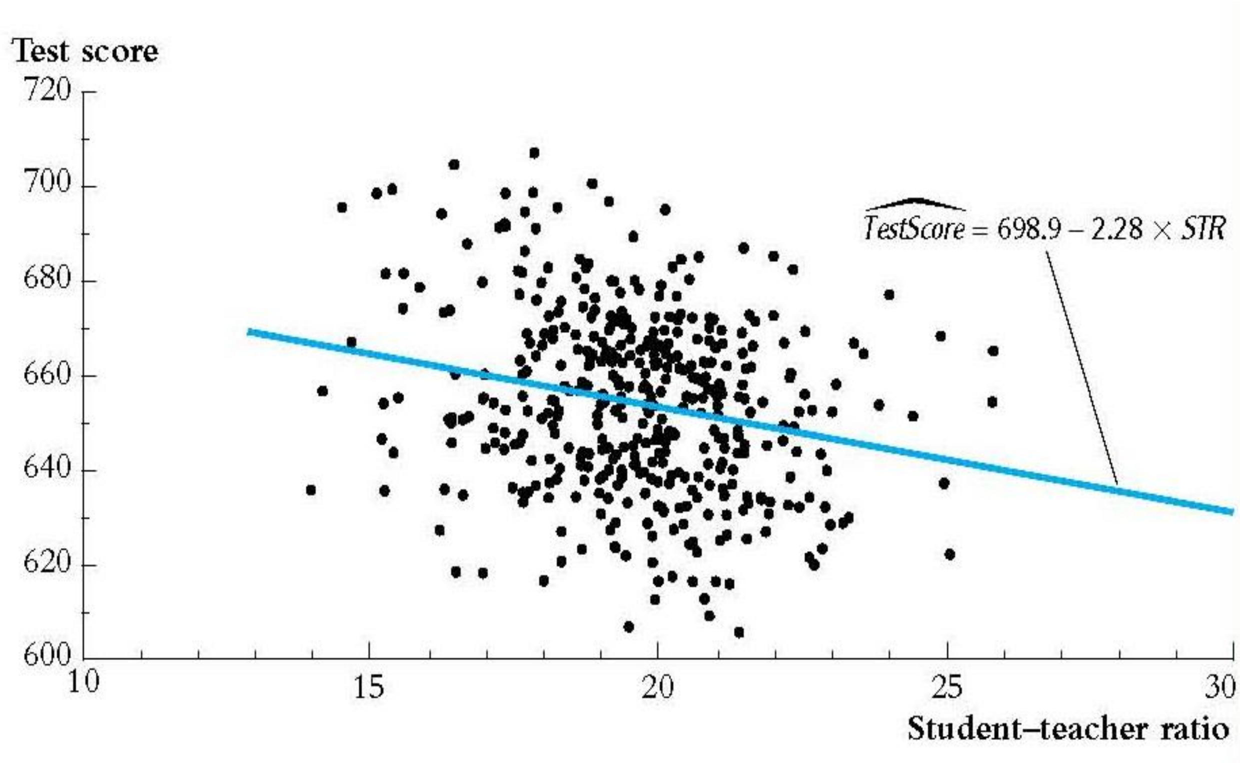
La seule différence entre les deux mesures est la division par n plutôt que par (n–2).

Détail technique: pourquoi diviser par n–2 plutôt que par n–1?

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

- La division par n–2 est une correction pour les “degrés de liberté”, de la même façon diviser par n–1 dans s_Y^2 avec la différence que pour la SER, deux paramètres ont été estimés (β_0 et β_1, par $\hat{\beta}_0$ par $\hat{\beta}_1$), alors que pour s_Y^2 on n’a estimé qu’un seul paramètre (μ_Y, par $\bar{Y}$).
- Lorsque n est grand, cela ne fait aucune différence de diviser par n, n–1, or n–2 – même si la formule conventionnelle utilise n–2, avec un seul régresseur.
- Voir les détails dans l’annexe web.

Exemple de R^2 et de SER



$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \ \mathbf{R^2 = .05,\ SER = 18.6}$$

La SER explique seulement une petite fraction de la variabilité des résultats d'examen. Est-ce raisonnable ? Est-ce que cela signifie que la taille des classes STR n'est pas importante ?

Les hypothèses des Moindres Carrés (SW Section 1.4)

Quelles sont, précisément les propriétés de la distribution d'échantillonnage de l'estimateur des MCO? Quand $\hat{\beta}_1$ sera-t-il sans biais? Quelle est sa variance?

Pour répondre à toutes ces questions, nous devons faire des hypothèses sur la façon dont Y et X sont reliés, et sur la manière dont ces variables ont été collectées (la façon dont l’'échantillonnage a été réalisé)

Ces hypothèses - il y en a trois - sont connues comme les hypothèses des moindres carrés.

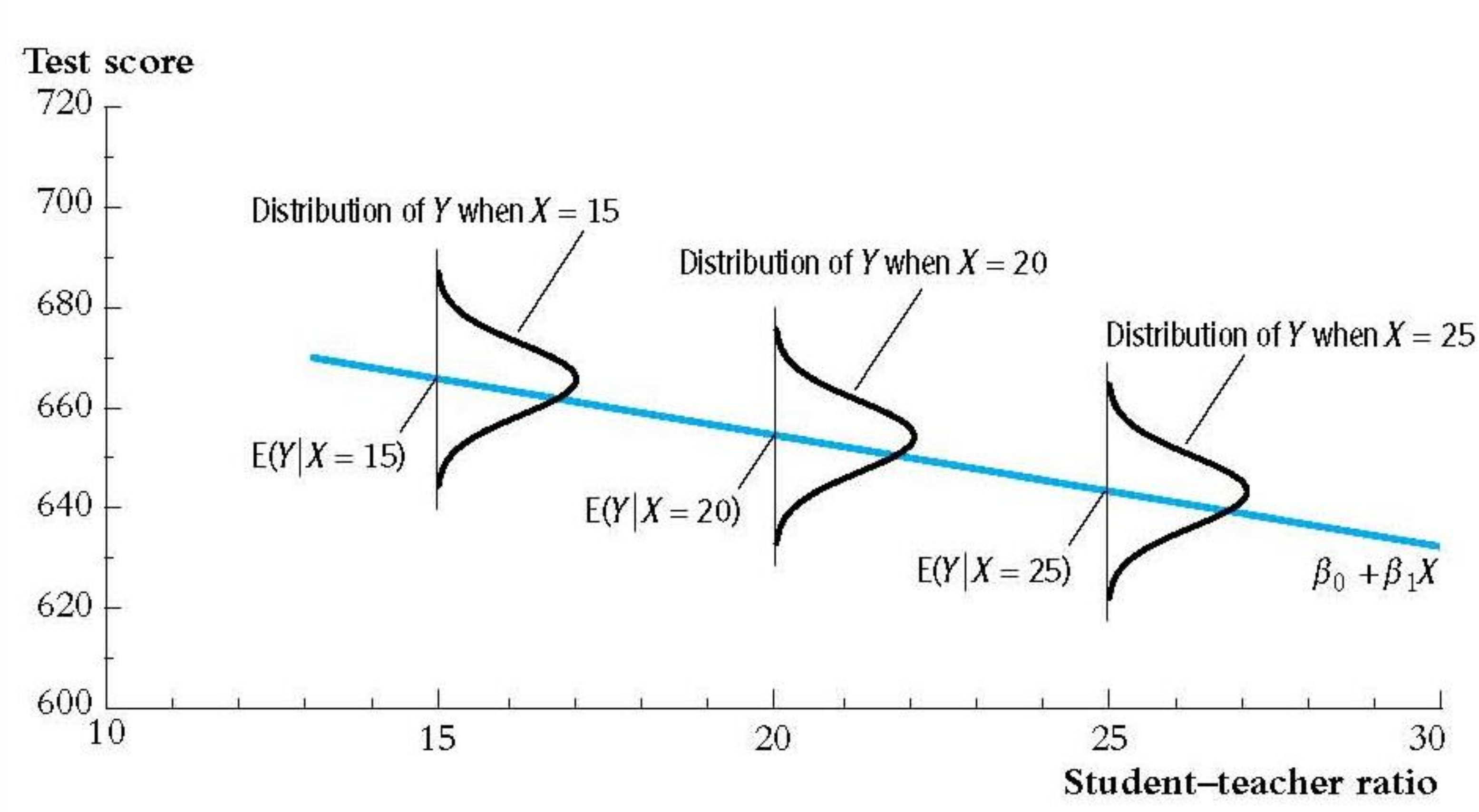
Les hypothèses des Moindres Carrés

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

1. L'espérance conditionnelle de u étant donné X a une moyenne nulle, c'est-à-dire $E(u|X = x) = 0$.
   - *Ceci implique que* $\hat{\beta}_1$ *est sans biais*
2. (X_i,Y_i), $i =1,\ldots,n$, sont i.i.d.
   - *C'est le cas si* (X, Y) *sont collectés par échantillonnage simple*
   - *Cela donne la distribution d'échantillonnage de* $\hat{\beta}_0$ *et* $\hat{\beta}_1$
3. Les valeurs aberrantes (outliers) de X et/ou Y sont rares.
   - *Techniquement,* X *et* Y *ont un moment d'ordre 4 fini*
   - *Les outliers peuvent invalider les valeurs de* $\hat{\beta}_1$

Hypothèse des moindres carrés no 1: $E(u|X = x) = 0$.

Pour toute valeur de X, la moyenne de u est nulle:



Exemple: $Test\ Score_i = \beta_0 + \beta_1 STR_i + u_i$, u_i = autres facteurs

- Quels sont ces “autres facteurs”? Est-ce que l’hypothèse $E(u|X=x) = 0$ est plausible pour ces autres facteurs?

Hypothèse des moindres carrés no. 1, *suite.*

Le cas de référence : l'expérience aléatoire idéale dans des conditions contrôlées (type laboratoire):

• X est attribué au hasard aux sujets (étudiants mis au hasard dans des classes de tailles différentes, patients qui reçoivent ou non des traitements médicaux au hasard). Les groupes sont générés par ordinateur – sans utiliser d'information sur les individus.

• Etant donné que X est attribué au hasard, toutes les autres caractéristiques individuelles – ce que l'on récupère dans les u - sont distribuées indépendamment de X, donc u et X sont indépendantes.

• Ainsi, dans une expérience aléatoire idéale dans des conditions contrôlées, $\boldsymbol{E(u|X = x) = 0}$ (càd, l'HMC no. 1 est vérifiée)

- Dans des expériences réelles, ou avec des données d'observation, il faudra très sérieusement se poser la question de savoir si $E(u|X = x) = 0$ est vérifiée.

Hypothèse des moindres carrés no 2: (X_i,Y_i), $i = 1,\ldots,n$ sont i.i.d.

C'est automatiquement le cas si l'unité d'observation (individu, district) est obtenue par échantillonnage aléatoire simple:

- Les unités sont choisies à partir de la même population, donc (X_i, Y_i) sont *identiquement distribués* pour tout $i = 1,\ldots, n$.
- Les unités sont choisies au hasard, donc les (X, Y) pour des unités différentes sont *distribués indépendamment*.

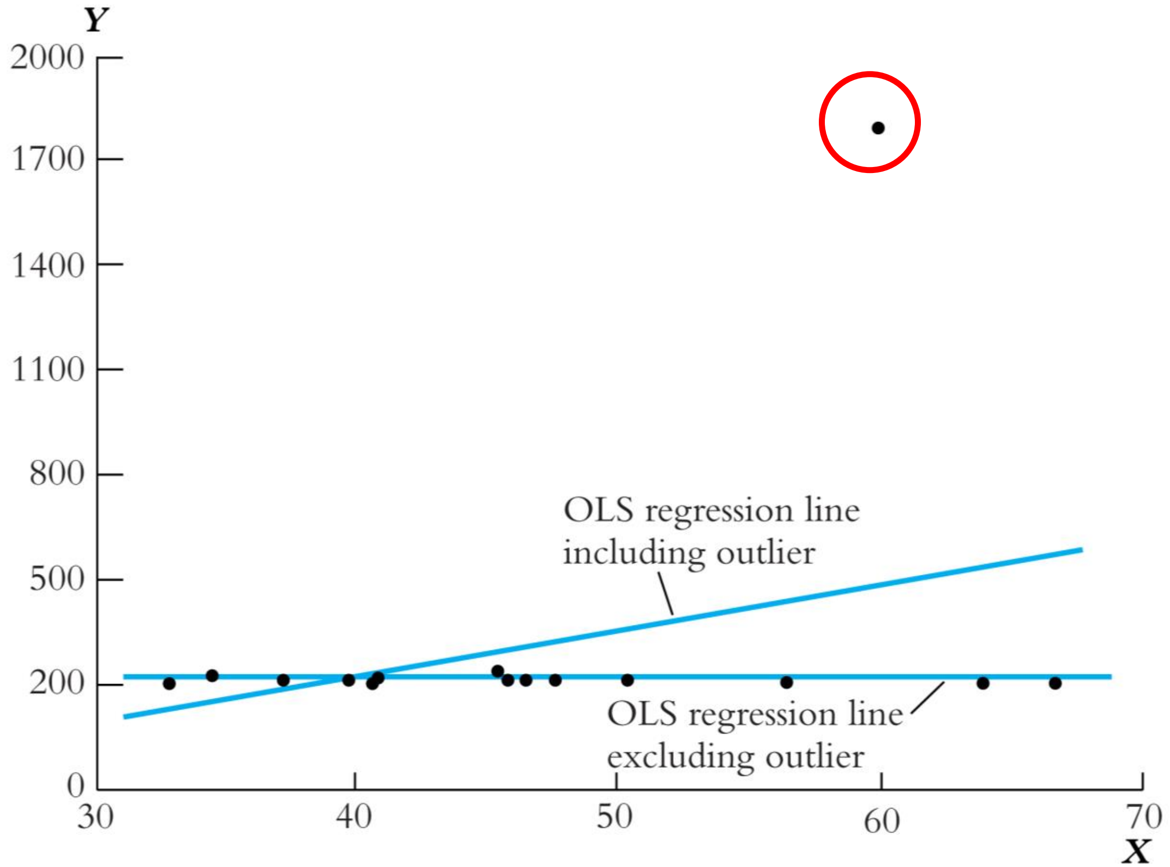
L'échantillonnage non-i.i.d. est pertinent lorsque les observations sont recueillies dans l'ordre chronologique (cela concerne les séries temporelles et le panel).

Hypothèse des moindres carrés no 3: ***Les grands outliers sont rares***

Techniquement: $E(X^4) < \infty$ **et** $E(Y^4) < \infty$

- Un grand outlier est une valeur extrême de X ou Y
- Du point de vue technique, si X et Y sont bornées, alors elles ont des moments d'ordre 4 finis. (Les notes de tests standards sont d'office bornées; *STR*, le revenu familial, etc. aussi.)
- En substance, on fait l'hypothèse qu'il n'y a pas de grand outlier, parce qu'un outlier peut très fortement influer sur les résultats.
- Regardez vos données ! Si vous avez un grand outlier, s'agit-il d'une erreur de frappe ? Est-ce que cela fait partie des données ? Pourquoi cette valeur est-elle là?

Les MCO peuvent être très sensibles aux outliers:



- *Ce point isolé est-il un outlier de X ou de Y?*
- En pratique, les outliers sont souvent des erreurs dans les données (des problèmes d'encodage). Parfois ce sont des observations qui ne devraient pas faire partie des données. Faites un graphique des données!

La d'échantillonnage de l'Estimateur des MCO
(SW Section 1.5)

L'estimateur MCO est calculé à partir d'un échantillon de données. Un échantillon différent donne une valeur différente de $\hat{\beta}_1$. C'est là la source de l'"incertitude liée à l'échantillonnage" de $\hat{\beta}_1$. On veut:

- quantifier l'incertitude associée à l'échantillonnage de $\hat{\beta}_1$
- utiliser $\hat{\beta}_1$ pour tester des hypothèses telles que $\beta_1 = 0$
- construire un intervalle de confiance pour β_1
- Pour tout cela, on a besoin de connaître la distribution d'échantillonnage de l'estimateur des MCO. Deux étapes pour y parvenir…
    - Cadre Probabiliste pour la régression linéaire
    - Distribution de l'estimateur des MCO

Cadre Probabiliste pour la régression linéaire

Le Cadre Probabiliste pour la régression linéaire se résume à trois hypothèses.

Population

- Le groupe qui nous intesse (ex: tous les districts scolaires possibles)

Les variables aléatoires*: *Y, X

- Ex: (*Test Score*, *STR*)

***Distribution jointe de* (*Y*, *X*). On fera les hypothèse suivantes:**

- La fonction de régression est linéaire
- $E(u|X) = 0$ (1^{re} Hypothèse des Moindres Carrés)
- X, Y ont des moments d'ordre 4 finis (3^{e} H. des MC.)

***Collecte de données par échantillonnage aléatoire simple, ce qui implique*:**

- $\{(X_i, Y_i)\}$, $i = 1,\ldots, n$, sont i.i.d. (2me H. des MC.)

La Distribution d'échantillonnage de $\hat{\beta}_1$

Comme $\bar{Y}$, $\hat{\beta}_1$ a une distribution d'échantillonnage.

- Qu'est-ce que $E(\hat{\beta}_1)$?
  - Si $E(\hat{\beta}_1) = \beta_1$, alors les MCO sont sans biais– une bonne chose!
- Que vaut var ($\hat{\beta}_1$)? (mesure de l'incertitude liée à l'échantillonnage)
  - Il nous faudra calculer l'écart-type de $\hat{\beta}_1$.
- Quelle est la distribution of $\hat{\beta}_1$ dans des échantillons petits?
  - Elle est très compliquée, en général
- Quelle est la distribution of $\hat{\beta}_1$ dans des échantillons grands?
  - Dans de grands échantillons, $\hat{\beta}_1$ est normal.

La moyenne et la variance de la distribution d'échantillonnage de $\hat{\beta}_1$

Un peu d'algèbre préliminaire:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u}$$

donc $\quad Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + (u_i - \bar{u})$

Et,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})[\beta_1(X_i - \bar{X}) + (u_i - \bar{u})]}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \beta_1 \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Donc $$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Et $$\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[\sum_{i=1}^{n}(X_i - \bar{X})\right]\bar{u}$$

$$= \sum_{i=1}^{n}(X_i - \bar{X})u_i - \left[\left(\sum_{i=1}^{n} X_i\right) - n\bar{X}\right]\bar{u}$$

$$= \sum_{i=1}^{n}(X_i - \bar{X})u_i$$

On Substitue $\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u}) = \sum_{i=1}^{n}(X_i - \bar{X})u_i$ dans l'expression de $\hat{\beta}_1 - \beta_1$:

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

donc

$$\boxed{\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

On peut alors cal*culer* $E(\hat{\beta}_1)$ *et var*$(\hat{\beta}_1)$*:*

$$E(\hat{\beta}_1) - \beta_1 = E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right]$$

$$= E\left\{ E\left[\frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right] \middle| X_1, \ldots, X_n \right\}$$

$= 0$ **parce que $E(u_i|X_i=x) = 0$ par HMC1**

- Donc HMC1 Implique que $E(\hat{\beta}_1) = \beta_1$
- Càd que, $\hat{\beta}_1$ **est un estimateur sans bias de β_1.**
- Pour les détails voir App. 1.3

Ensuite, calculons var($\hat{\beta}_1$):

On peut écrire

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})u_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\frac{1}{n}\sum_{i=1}^{n}v_i}{\left(\frac{n-1}{n}\right)s_X^2}$$

où $v_i = (X_i - \bar{X})u_i$. Si n est grand, $s_X^2 \approx \sigma_X^2$ et $\frac{n-1}{n} \approx 1$, donc

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n}\sum_{i=1}^{n}v_i}{\sigma_X^2},$$

avec $v_i = (X_i - \bar{X})u_i$ (voir App. 1.3). Donc,

$$\hat{\beta}_1 - \beta_1 \approx \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\sigma_X^2}$$

donc $\quad \mathrm{var}(\hat{\beta}_1 - \beta_1) = \mathrm{var}(\hat{\beta}_1)$

$$= \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} v_i\right)\Big/(\sigma_X^2)^2 = \frac{\mathrm{var}(v_i)/n}{(\sigma_X^2)^2}$$

Et la dernière égalité découle de l’hypothèse 2. Donc,

$$\boxed{\mathrm{var}(\hat{\beta}_1) = \frac{1}{n}\times\frac{\mathrm{var}[(X_i - \mu_x)u_i]}{(\sigma_X^2)^2}} \quad .$$

Résumé jusqu’ici

1. $\hat{\beta}_1$ est sans bais: $E(\hat{\beta}_1) = \beta_1$ – comme $\bar{Y}$!
2. $\mathrm{var}(\hat{\beta}_1)$ est inversement proportionnelle à n – comme $\bar{Y}$!

Quelle est la distribution d'échantillonnage de $\hat{\beta}_1$?

La distribution d'échantillonnage exacte est compliquée – elle dépend de la distribution dans la population de (Y, X) – mais quand n est grand on a des approximations simples (et précises):

(1) Comme var($\hat{\beta}_1$) $\propto 1/n$ et $E(\hat{\beta}_1) = \beta_1$, $\hat{\beta}_1 \xrightarrow{p} \beta_1$

(2) Lorsque n est grand, la distribution d'échantillonnage de $\hat{\beta}_1$ est bien approximée par une distribution normale (CLT)

Souvenez-vous du ***CLT***: supposons que $\{v_i\}$, $i = 1,\ldots, n$ est i.i.d. avec $E(v) = 0$ et var(v) = σ^2. Alors quand n est grand, $\frac{1}{n}\sum_{i=1}^{n} v_i$ est approximativement distribué $N(0,\sigma_v^2/n)$.

Approximation pour *n* grand, de la distribution de $\hat{\beta}_1$:

$$\hat{\beta}_1 - \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\left(\frac{n-1}{n}\right) s_X^2} \approx \frac{\frac{1}{n}\sum_{i=1}^{n} v_i}{\sigma_X^2}, \text{ avec } v_i = (X_i - \bar{X})u_i$$

- Quand n est grand, $v_i = (X_i - \bar{X})u_i \approx (X_i - \mu_X)u_i$, qui est i.i.d. (*pourquoi*?) et var(v_i) < ∞ (*pourquoi*?). Donc, par le CLT, $\frac{1}{n}\sum_{i=1}^{n} v_i$ est approximativement distribué $N(0, \sigma_v^2 / n)$.
- Donc, pour n grand, $\hat{\beta}_1$ est approximativement distribué

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_v^2}{n(\sigma_X^2)^2}\right), \text{ where } v_i = (X_i - \mu_X)u_i$$

Plus la variance de *X* est grande, plus la variance de $\hat{\beta}_1$ est faible

Les maths

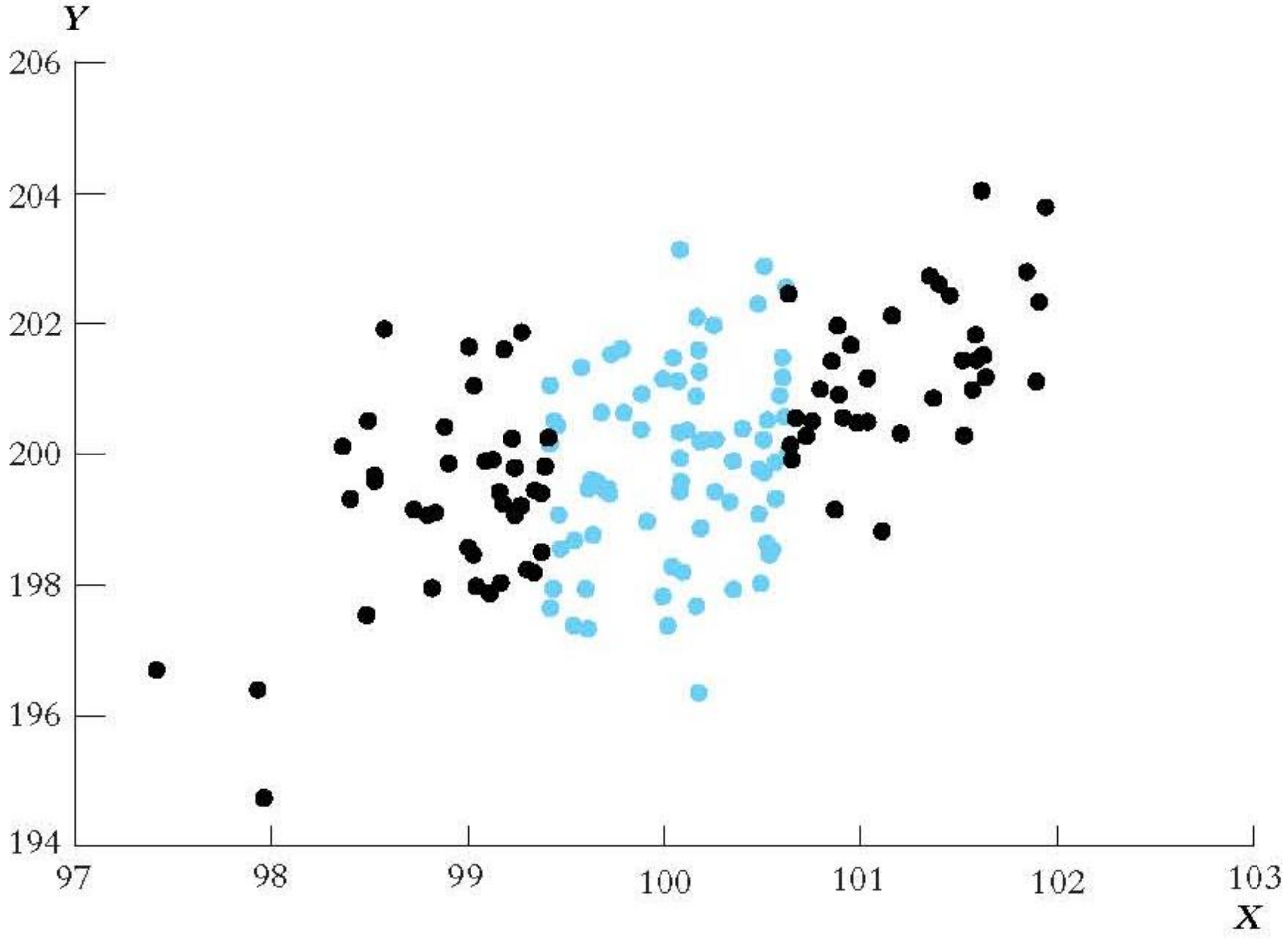
$$\text{var}(\hat{\beta}_1 - \beta_1) = \frac{1}{n} \times \frac{\text{var}[(X_i - \mu_x)u_i]}{(\sigma_X^2)^2}$$

avec σ_X^2 = var(X_i). La variance de X apparaît (au carré) au dénominateur – donc lorsque l'on augmente la variabilité de X on décroît la variance de β_1.

L'intuition

Plus il y a de variation dans X, plus on a d'information dans les données que l'on peut utiliser pour ajuster la droite de régression. Cela se voit facilement dans un graphique…

Plus la variance de X est grande, plus la variance de $\hat{\beta}_1$ est faible



On a le même nombre de points noirs et bleus. Lesquels donneraient une droite de régression plus précise?

Résumé de la distribution d'échantillonnage de $\hat{\beta}_1$:

Si les trois hypothèses des Moindres Carrés sont vérifiées

- La distribution d'échantillonnage exacte (en échantillon fini) de $\hat{\beta}_1$ est telle que:
  - $E(\hat{\beta}_1) = \beta_1$ (càd que, $\hat{\beta}_1$ est sans biais)
  - $\text{var}(\hat{\beta}_1) = \dfrac{1}{n} \times \dfrac{\text{var}[(X_i - \mu_x)u_i]}{\sigma_X^4} \propto \dfrac{1}{n}$.
- Outre sa moyenne et sa variance, la distribution exacte de $\hat{\beta}_1$ est très compliquée et dépend de la distribution de (X, u)
- $\hat{\beta}_1 \xrightarrow{p} \beta_1$ (càd que, $\hat{\beta}_1$ est convergent)
- Quand n est grand, $\dfrac{\hat{\beta}_1 - E(\hat{\beta}_1)}{\sqrt{\text{var}(\hat{\beta}_1)}} \sim N(0,1)$ (CLT)
- *Ce résultat présente un parallélisme parfait avec la distribution d'échantillonnage* de $\bar{Y}$.

Large-Sample Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$

If the least squares assumptions in Key Concept 4.3 hold, then in large samples $\hat{\beta}_0$ and $\hat{\beta}_1$ have a jointly normal sampling distribution. The large-sample normal distribution of $\hat{\beta}_1$ is $N(\beta_1, \sigma^2_{\hat{\beta}_1})$, where the variance of this distribution, $\sigma^2_{\hat{\beta}_1}$, is

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{\mathrm{var}\left[(X_i - \mu_X)u_i\right]}{\left[\mathrm{var}(X_i)\right]^2}. \tag{4.21}$$

The large-sample normal distribution of $\hat{\beta}_0$ is $N(\beta_0, \sigma^2_{\hat{\beta}_0})$, where

$$\sigma^2_{\hat{\beta}_0} = \frac{1}{n}\frac{\mathrm{var}(H_i u_i)}{\left[E(H_i^2)\right]^2}, \text{ where } H_i = 1 - \left[\frac{\mu_X}{E(X_i^2)}\right]X_i. \tag{4.22}$$

Nous sommes dès lors prêts à nous intéresser aux tests d'hypothèse et aux intervalles de confiances…