Régression linéaire simple:

Test d'hypothèses et intervalle de confiance

(SW Chapitre 5)

<u>Plan de route</u>

1. Variance de $\hat{\beta}_1$ et son estimateur
2. Test d'hypothèses sur β_1
3. Intervalle de confiance pour β_1
4. Régression lorsque X est dichotomique (binaire)
5. Homoscédasticité et hétéroscédasticité
6. Efficacité des MCO et distribution t de Student

Vue d’ensemble

Nous souhaitons estimer la pente de la droite de régression à partir d’un échantillon aléatoire et ensuite quantifier l’erreur d’estimation. Pour cela, cinq étapes à suivre:

1. Identifier le paramètre d’intérêt.
2. Trouver un estimateur pour ce paramètre.
3. Établir la distribution de l’estimateur retenu sous les hypothèses du modèle.
4. Calculer l’écart-type de l’estimateur à partir de la variance de sa distribution.
5. Utiliser cet écart-type pour calculer le ratio *t* de Student qui servira à construire les tests d’hypothèses et les intervalles de confiance.

Objet d'intérêt: β_1 dans la régression suivante :

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

On rappelle que $\beta_1 = \Delta Y/\Delta X$ indique de combien Y varie lorsque X varie d'une unité.

Estimateur: L'estimateur MCO $\hat{\beta}_1$.

Distribution de $\hat{\beta}_1$: Nous établirons cette distribution sous les hypothèses suivantes en supposant que l'échantillon est assez grand :

Hypothèses du modèle linéaire:

1. L'erreur est de moyenne nulle et non corrélée au régresseur : $E(u|X = x) = 0$
2. Les observations (X_i,Y_i), $i = 1,\ldots,n$, sont i.i.d.
3. Le processus générateur des données (c-à-d, la distribution d'où proviennent les (X_i,Y_i)) ne produit pas beaucoup de valeurs extrêmes : $(E(X^4) < \infty$, $E(Y^4) < \infty$.

On peut enfin établir la distribution de $\hat{\beta}_1$:

Sous les hypothèses du modèle linéaire, l'estimateur $\hat{\beta}_1$ a une distribution approximativement normale (gaussienne) lorsque l'échantillon est suffisamment grand. On a :

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

Écart-type de $\hat{\beta}_1$ et tests d'hypothèses
(Section 5.1)

On souhaite tester une hypothèse de la forme $\beta_1 = 0$ en se basant sur les données. Ceci est l'hypothèse nulle (notée H_0), c'est-à-dire est l'hypothèse "par défaut".

Formulation des hypothèses :

Hypothèse nulle et alternative **bilatérale**:

$$H_0:\ \beta_1 = \beta_{1,0} \ \text{ contre } H_1:\ \beta_1 \neq \beta_{1,0}$$

où $\beta_{1,0}$ est la valeur à tester, choisie par l'économètre. En pratique, le choix de $\beta_{1,0}$ est guidé par l'objectif de l'étude.

Hypothèse nulle et alternative **unilatérale**:

$$H_0: \beta_1 = \beta_{1,0} \ \text{ contre } H_1:\ \beta_1 < \beta_{1,0} \text{ (ou } H_1:\ \beta_1 > \beta_{1,0}\text{)}$$

Approche générale:

(i) Calculer le *t de Student* ou la *p*-valeur.

(ii) Comparer le *t de Student* aux valeurs critiques de la loi *N*(0,1) ou comparer la p-valeur au seuil de significativité du test

- ***t de Student pour*** $\hat{\beta}_1$:

$$\boxed{t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}}$$

où $SE(\hat{\beta}_1)$ est l'écart-type de $\hat{\beta}_1$.

- ***t de Student pour un test sur la moyenne de Y***:

$$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y / \sqrt{n}}$$

Expression de SE($\hat{\beta}_1$)

Variance de $\hat{\beta}_1$ (lorsque *n* est grand):

$$\text{var}(\hat{\beta}_1) = \frac{\text{var}[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2} = \frac{\sigma_v^2}{n(\sigma_X^2)^2}$$, où $v_i = (X_i - \mu_X)u_i$.

Estimateur de la variance de $\hat{\beta}_1$: on remplace les variances en population σ_v^2 and σ_X^2 par des estimateurs construits à partir des données et les résidus des MCO:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \times \frac{\text{estimator of } \sigma_v^2}{(\text{estimator of } \sigma_X^2)^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{v}_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^2}$$

où $\hat{v}_i = (X_i - \bar{X})\hat{u}_i$.

Enfin,

$$\boldsymbol{SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}} \textbf{ = écart-type de } \hat{\beta}_1$$

- L'expression de $\hat{\sigma}^2_{\hat{\beta}_1}$ est moins compliqué qu'il en a l'air :
  Le numérateur est la variance empirique de $\hat{v}_i$ corrigé pour tenir compte du degré de liberté des résidus (on divise par *n-2* au lieu de *n*) et le dénominateur est le carré de la variance empirique de *X*.
- Le degré de liberté est $n - 2$ car on utilise déjà les mêmes données pour estimer deux coefficients (β_0 et β_1).
- Si on le souhaite, on peut mémoriser la formule de $SE(\hat{\beta}_1)$ mais ce n'est pas nécessaire. Sa valeur est calculée automatiquement par les logiciels statistiques.

En résumé: Pour tester H_0: $\beta_1 = \beta_{1,0}$ contre H_1: $\beta_1 \neq \beta_{1,0}$,

1. Calculer le t de Student :

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}}$$

2. Rejeter H_0 au seuil de significativité de 5% si $|t| > 1.96$.

2(bis). De manière équivalente, rejeter H_0 si la p-valeur est plus petite que 5%. La p-valeur est donnée par:

$$\begin{aligned} p\text{-valeur} &= Pr[|N(0,1)| > |t|] \\ &= 1 - Pr[-|t| < N(0,1) < |t|] \\ &= 2 - 2Pr[N(0,1) < |t|]. \end{aligned}$$

La procédure de test ci-dessus repose sur l'approximation normale de la loi de $\hat{\beta}_1$ lorsque *n* est grand (En pratique, on admet que *n* est assez grand s'il est supérieur ou égal à 50).

Exemple: *Performance des élèves et STR (ratio élève-enseignant)*, données californiennes

- Droite de régression estimée:

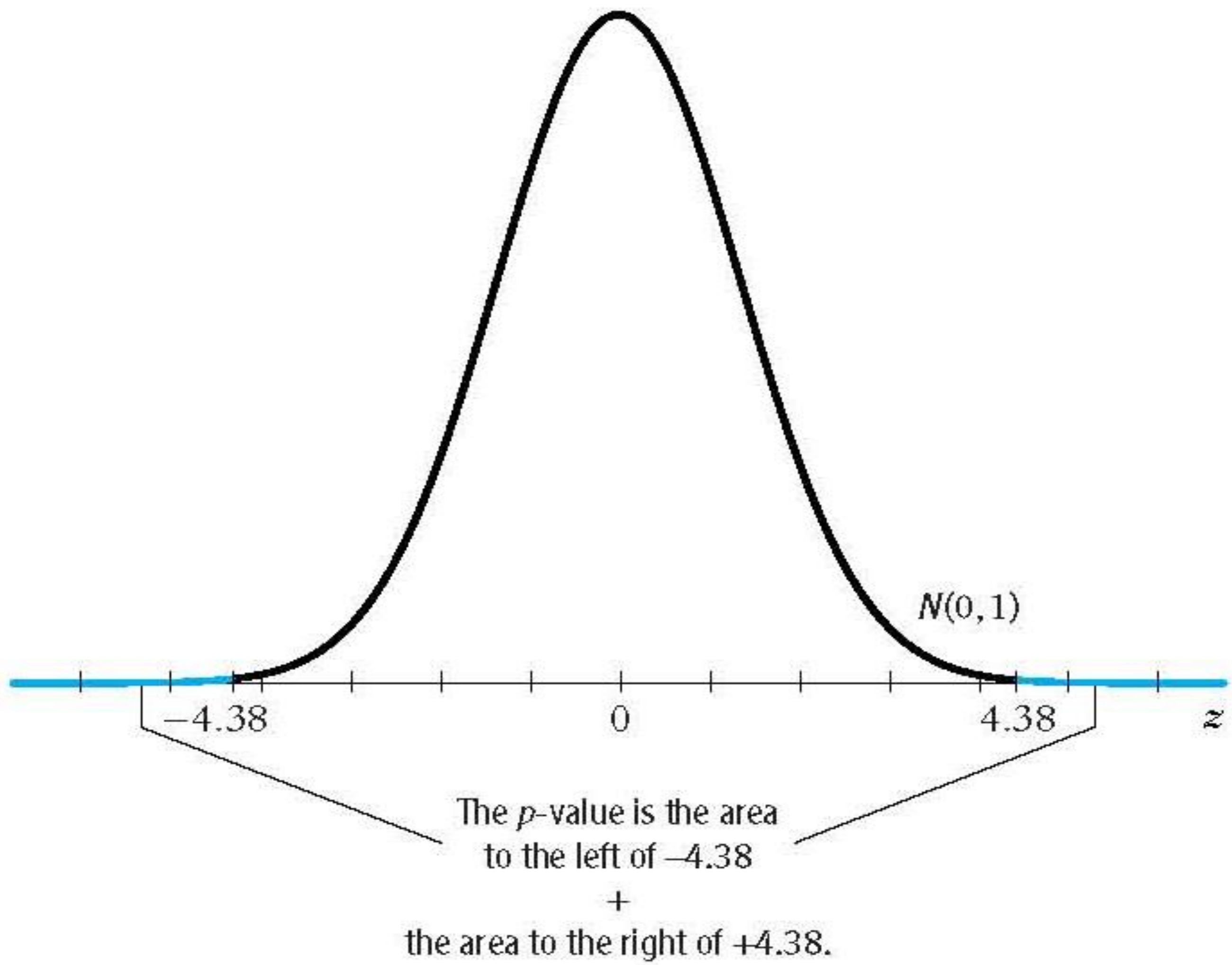
$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$

- Résultats fourni par le logiciel:

$$SE(\hat{\beta}_0) = 10.4 \qquad\qquad SE(\hat{\beta}_1) = 0.52$$

$$t \text{ de Student pour } \beta_{1,0} = 0: \qquad t = \frac{-2.28 - 0}{0.52} = -4.38$$

- À 1% de significativité, le seuil de rejet est égal à 2.58 pour un test bilatéral. Donc, on rejette l'hypothèse nulle au seuil de 1% car $|-4.38| = 4.38$ est plus grand que 2.58.
- Alternativement, on aurait pu calculer la *p*-valeur…



La p-valeur est égale à la surface de la région en bleu sur la figure (pas visible à l'œil nu car très petit)

$$p\text{-}valeur = 2\text{-}2Pr[N(0,1)<4.48]= 0.00001 = 10^{-5}.$$

Intervalle de Confiance pour β_1 (Section 5.2)

De manière équivalente, un intervalle de confiance à 95% est:

- Un ensemble de valeurs possibles pour β_1 qui ne peuvent être rejetées dans un test à 5% de significativité;
- Un intervalle dont les bornes sont déterminées à partir des données selon une certaine règle et qui contient la vraie valeur de β_1 avec 95% de probabilités.

Le *t* de Student est de loi N(0,1) en grande échantillon. Donc, l'IC à 95% pour β_1 se construit comme pour une moyenne:

Intervalle de confiance à 95% pour β_1:

$$[\hat{\beta}_1 - 1.96 \times SE(\hat{\beta}_1), \hat{\beta}_1 + 1.96 \times SE(\hat{\beta}_1)]$$

Par abus de notation:

$$[\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)] \text{ ou } \{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$$

Exemple: Performance des élèves et STR

Droite de régression:

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

$$SE(\hat{\beta}_0) = 10.4 \qquad\qquad SE(\hat{\beta}_1) = 0.52$$

Intervalle de confiance à 95% pour β_1:

$$\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\} = \{-2.28 \pm 1.96 \times 0.52\}$$
$$= (-3.30, -1.26)$$

Les affirmations suivantes sont équivalentes *(Pourquoi?)*

- L'intervalle de confiance à 95% ne contient pas 0;
- L'hypothèse $\beta_1 = 0$ est rejetée au seuil de 5%.

Une présentation concise (et conventionnelle) des résultats de régression:

On met les écarts-types estimés entre parenthèses en dessous du coefficient qui lui correspond.

$$\widehat{TestScore} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28}\times STR,\ R^2 = .05,\ SER = 18.6$$

<u>Lire de la manière suivante</u>

- La droite de régression estimée est:

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

- L'écart-type estimé de $\hat{\beta}_0$ est 10.4
- L'écart-type estimé de $\hat{\beta}_1$ est 0.52
- Le R^2 de la régression est .05;
- L'écart-type estimé du terme d'erreur est 18.6.

Commande de régression MCO dans STATA

```
regress testscr str, robust
```

Lecture de l'output

```
Regression 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     -4.38   0.000     -3.300945   -1.258671
  _cons |    698.933   10.36436     67.44   0.000      678.5602    719.3057
-------------------------------------------------------------------------------
```

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

$$(10.4)\quad(0.52)$$

Ratio t de student ($\beta_1 = 0$) = **–4.38**, p-valeur = **0.000**.

IC bilatéral à 95% pour β_1: [–3.30, –1.26]

Inférence sur β_0 et β_1: un résumé

Estimation:

- Estimateurs MCO : $\hat{\beta}_0$ et $\hat{\beta}_1$
- $\hat{\beta}_0$ et $\hat{\beta}_1$ ont des distributions approximativement normales en lorsque la taille d'échantillon est grande.

Test d'hypothèse:

- H_0: $\beta_1 = \beta_{1,0}$ contre $\beta_1 \neq \beta_{1,0}$ ($\beta_{1,0}$=valeur de β_1 sous H_0)
- $t = (\hat{\beta}_1 - \beta_{1,0})/SE(\hat{\beta}_1)$ et *p-valeur* $= 2\text{-}2Pr[N(0,1)<|t|]$

Intervalle de confiance (IC):

- IC à 95% pour β_1 est $\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$
- C'est l'ensemble des valeurs de β_1 qui ne sont pas rejetées dans un test de niveau 5%.
- C'est un intervalle qui contient la vraie valeur de β_1 avec 95% de chances.

Régression lorsque *X* est Binary
(Section 5.3)

Parfois, le régresseur *X* est binaire:

- $X = 1$ si “classe de petite taille” et $X = 0$ sinon
- $X = 1$ si “femme” et $X = 0$ si “homme”
- $X = 1$ si “fumeur” et $X = 0$ sinon

Les variables binaires sont aussi appelés variables « dichotomiques » ou encore variable « dummy ».

Jusqu’à présent, on appelait β_1 le “coefficient de pente”. Cette appellation doit être évitée lorsque X est une variable binaire.

Comment interprète-t-on β_1 dans ce cas ?

Interprétation d'une régression lorsque la variable explicative est binaire

$Y_i = \beta_0 + \beta_1 X_i + u_i$, où X_i prend deux valeurs possibles : {0,1}.

- Lorsque $X_i = 0$, on a $Y_i = \beta_0 + u_i$. La moyenne de Y_i est alors donnée par β_0 : $E(Y_i|X_i=0) = \beta_0$
- Lorsque $X_i = 1$, on a $Y_i = \beta_0 + \beta_1 + u_i$. Dans ce cas, la moyenne de Y_i est $\beta_0 + \beta_1$: $E(Y_i|X_i=1) = \beta_0 + \beta_1$

Donc, β_1 est la différence entre la moyenne du groupe désigné par X=1 et celle du groupe désigné par X=0.

$$\beta_1 = E(Y_i|X_i=1) – E(Y_i|X_i=0)$$

Exemple: Soit la variable $D_i = \begin{cases} 1 \text{ if } STR_i \le 20 \\ 0 \text{ if } STR_i > 20 \end{cases}$

Régression MCO: $\widehat{TestScore} = \mathbf{650.0} + \mathbf{7.4} \times D$

$(1.3) \quad (\mathbf{1.8})$

Tableau croisé des moyennes par groupe:

Taille de classe	Perf. moyenne ($\bar{Y}$)	Ecart-type (s_Y)	Effectifs
Petite ($STR > 20$)	657.4	19.4	238
Grande ($STR \ge 20$)	**650.0**	17.9	182

Diff. entre moyennes: $\bar{Y}_{\text{small}} - \bar{Y}_{\text{large}} = 657.4 - 650.0 = \mathbf{7.4}$

Écart-type: $SE = \sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} = \mathbf{1.8}$

Régression lorsque X_i est binaire : un résumé

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

où X est binaire.

- β_0 = moyenne de Y lorsque $X = 0$
- $\beta_0 + \beta_1$ = moyenne de Y lorsque $X = 1$
- β_1 = moyenne si X =1 moins moyenne si $X = 0$
- SE($\hat{\beta}_1$), le t de Student et les intervalles de confiance se calculent de manière usuelle.
- La régression MCO avec variable explicative binaire offre un cadre facile pour tester la différence entre deux moyennes, en particulier lorsqu'on doit contrôler l'effet d'autres variables.

Homoscédasticité et hétéroscédasticité
(Section 5.4)

1. Que signifient homoscédasticité et hétéroscédasticité?
2. Conséquences de l'homoscédasticité.
3. Calcul de l'écart-type de $\hat{\beta}_1$ sous l'hypothèse d'homoscédasticité.

Définitions

Si la variance du terme d'erreur *u* conditionnellement à *X* (notée $Var(u|X=x)$) est constante (c'est-à-dire, ne dépends pas de la valeur *x* prise par *X*) alors, on dit que *u* est ***homoscédastique***. Sinon, *u* est ***hétéroscédastique***.

Exemple: Régresseur binaire.

- $\hat{\beta}_1$ = différence entre les moyennes des deux groupes
- Écart-type de $\hat{\beta}_1$ la lorsque les variances diffèrent:

$$SE = \sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}$$

- Écart-type de $\hat{\beta}_1$ lorsque les variances sont égales:

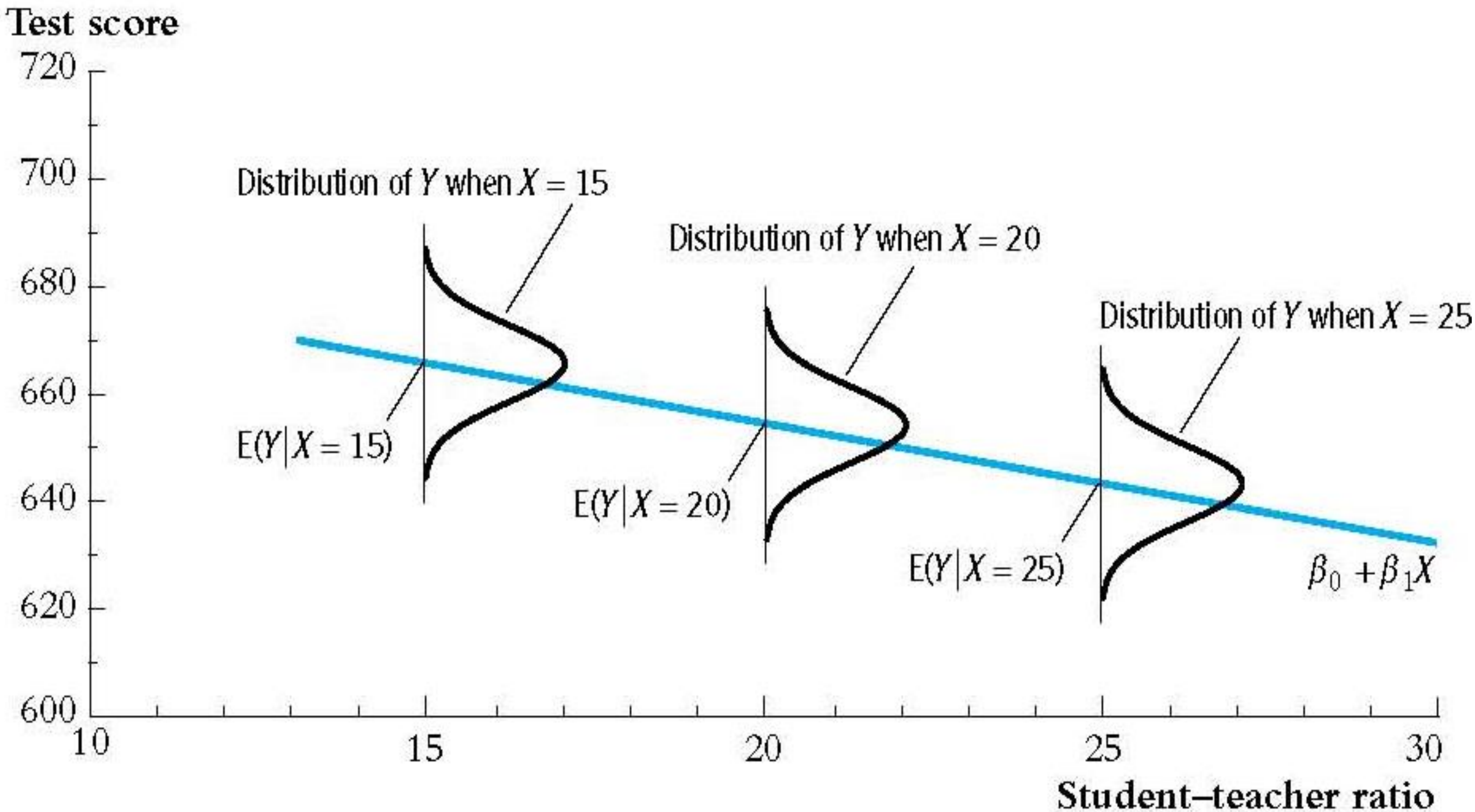
$$SE = s_p \sqrt{\frac{1}{n_s} + \frac{1}{n_l}}$$

où $s_p^2 = \dfrac{(n_s - 1)s_s^2 + (n_l - 1)s_l^2}{n_s + n_l - 2}$ (SW, Section 3.6)

s_p^2 = estimateur de σ^2 lorsque $\sigma_l^2 = \sigma_s^2$

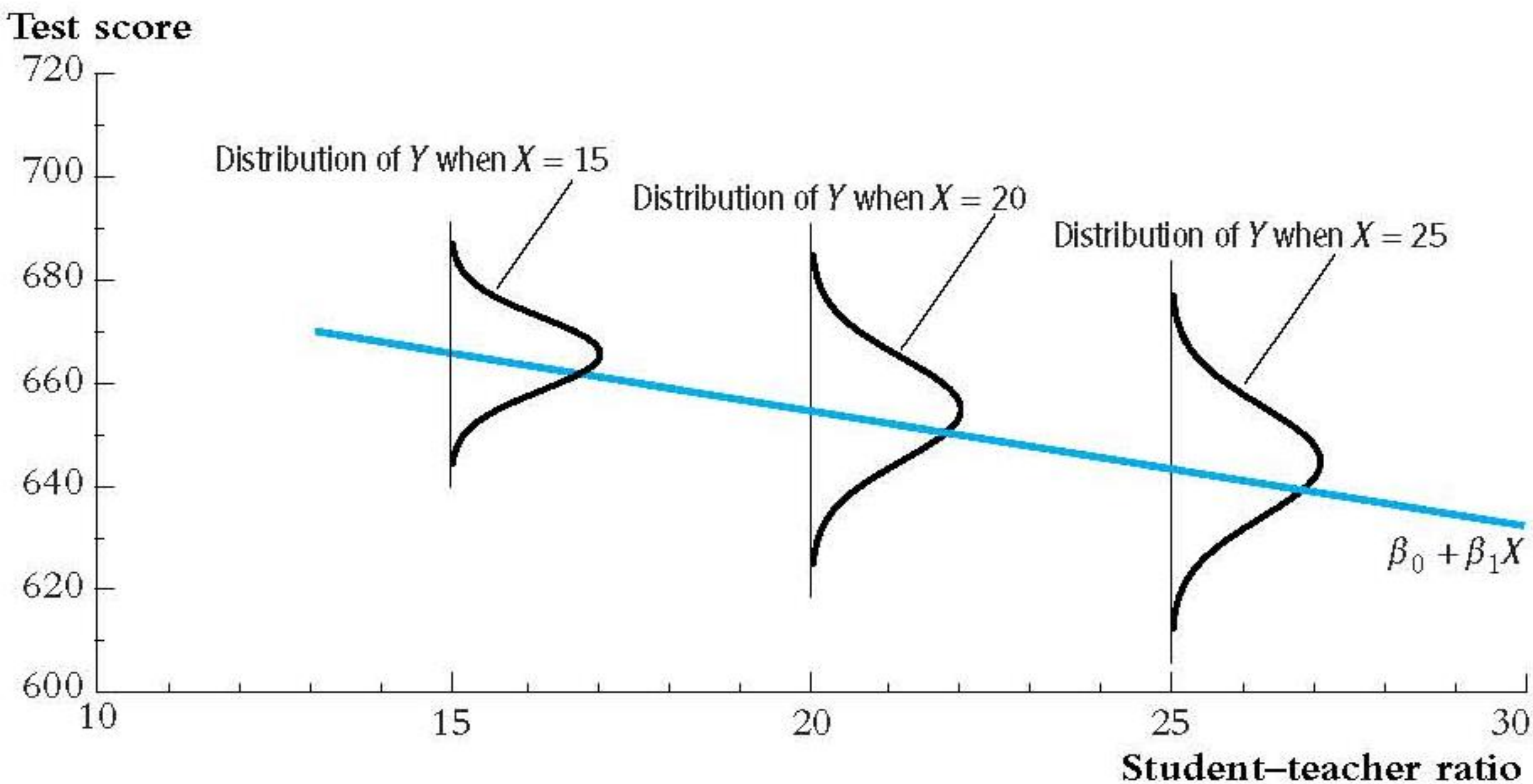
- Variances constantes = **homo**scédasticité
- Variances inégales = **hetero**scédasticité

Visualisation de l'homoscédasticité



- $E(u|X=x) = 0$ (u satisfait l'hypothèse de modèle linéaire #1)
- La variance de *u* ***ne dépend pas*** de *x*.

Visualisation de l'hétéroscédasticité

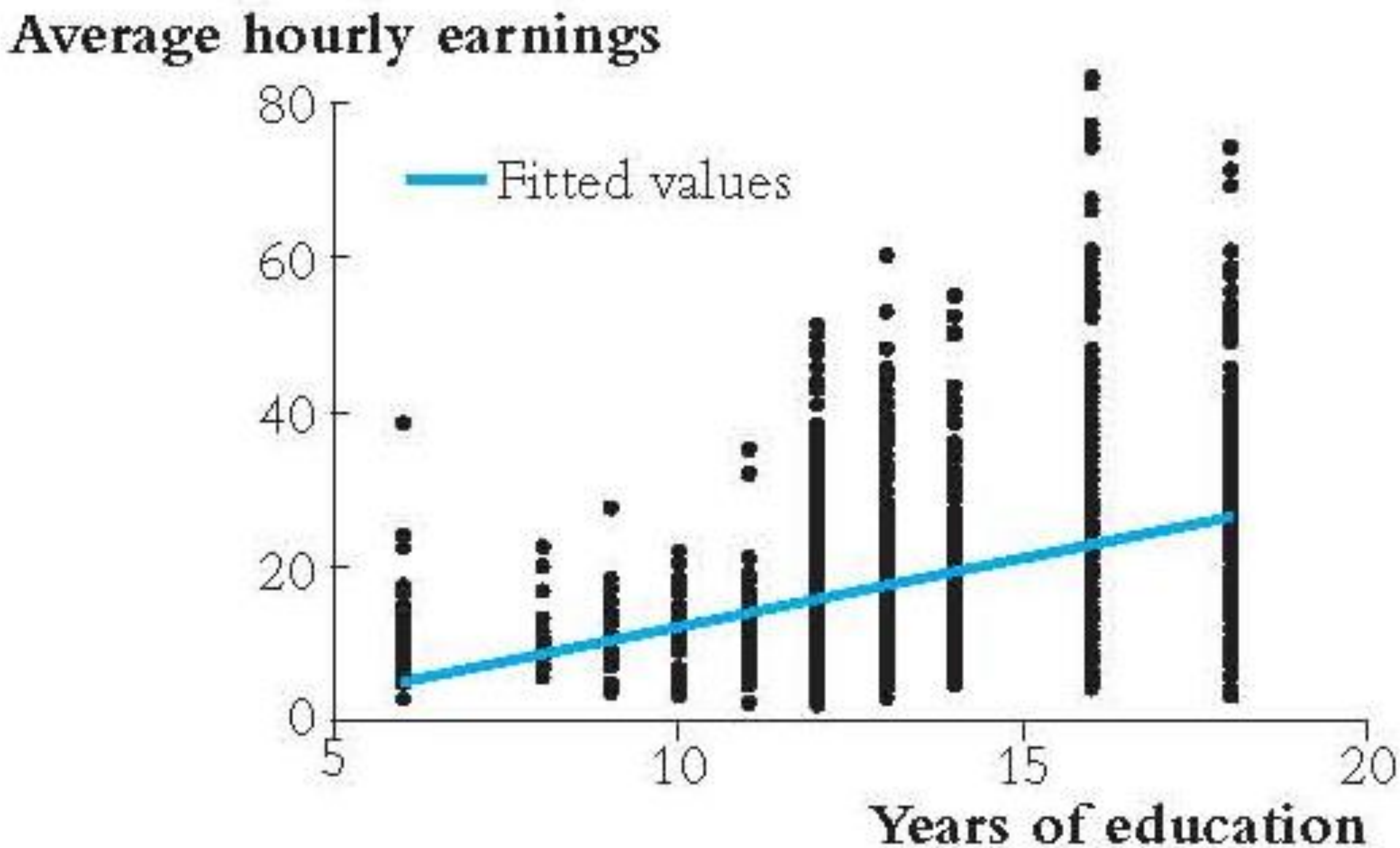


- $E(u|X=x) = 0$ (u satisfait l'hypothèse de modèle linéaire #1)
- La variance de *u* ***depend*** de *x*.

Autre exemple en "Économie du travail"

« Salaire horaire moyen » et « Nombre d'années d'études »

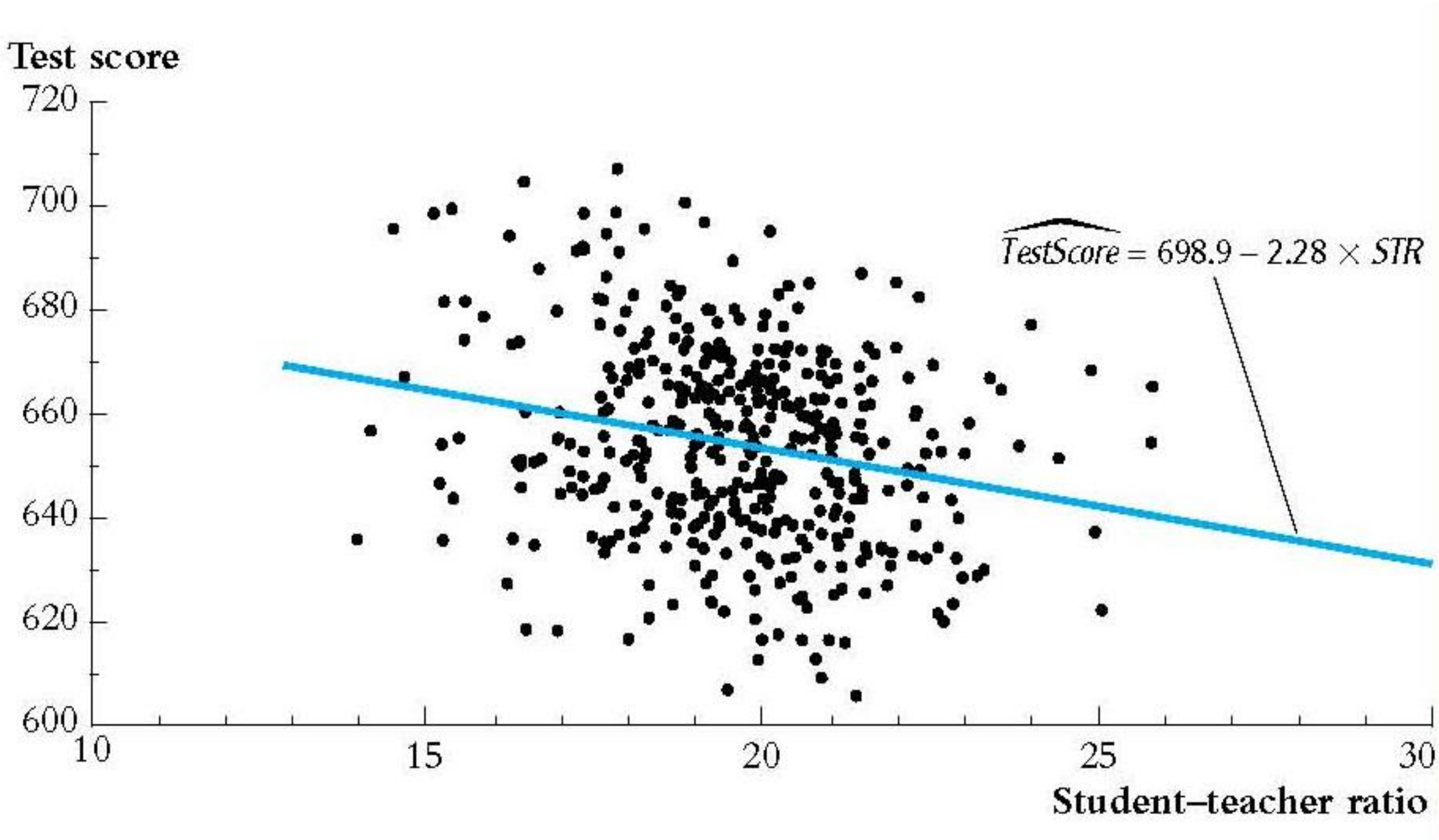
Source: *Current Population Survey*.



Hétéroscédastique ou homoscédastique ?

Autre exemple en “Économie de l’éducation”

“Performance moyenne des élèves” et “STR”



Hétéroscédastique ou homoscédastique ?

Jusqu'à présent, on a supposé implicitement que u pourrait être hétéroscédastique.

Rappelons les trois hypothèses du modèle linéaire:

1. $E(u|X = x) = 0$
2. (X_i,Y_i), $i = 1,\ldots,n$, sont i.i.d.
3. Rareté des observations aberrantes.

Ajoutons l'hypothèse d'homoscédasticité :

4. $\text{var}(u|X=x) = \sigma_u^2$ pour tout x.

Conséquences de l'homoscédasticité

- Avec l'ajout de l'hypothèse #4, on peut prouver que l'estimateur MCO est le meilleur parmi les estimateurs sans biais qui s'écrivent comme une fonction linéaire des Y_i. Voir: Théorème de Gauss-Markov.
- La variance de $\hat{\beta}_1$ a une expression plus simple. En effet, si var($u_i|X_i$=x) = σ_u^2, alors :

$$var(\hat{\beta}_1) = \frac{\text{var}[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2} \quad \textit{(formule générale)}$$

$$= \frac{\sigma_u^2}{n\sigma_X^2} \quad \textit{(simplification due à l'homoscédasticité)}$$

- La formule suivante de l'écart-type estimé de $\hat{\beta}_1$ en découle :

$$SE(\hat{\beta}_1) = \sqrt{\frac{1}{n} \times \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}.$$

En somme, deux façons d'estimer la variance de $\hat{\beta}_1$:

- ***Formule sous l'hypothèse d'homoscédasticité:*** n'est valide que si les erreurs sont réellement homoscédastique. Elle sous-estime l'écart-type de $\hat{\beta}_1$ si l'hypothèse d'homoscédasticité est fausse.
- ***Formule robuste à l'hétéroscédasticité:*** conçue pour le cas où l'erreur est hétéroscédastique, mais également valide en cas d'homoscédasticité.

Implications pratiques

- Les deux formules précédentes donnent rarement les mêmes valeurs : les écart-types estimés sous l’hypothèse d’homoscédasticité sont plus petits que ceux que l’on obtient sous l’hypothèse d’hétéroscédasticité.
- Les logiciels font l’hypothèse d’homoscédasticité par défaut. En générale, on peut spécifier des options pour avoir des résultats robustes à l'hétéroscédasticité.
- Si les données sont hétéroscédastiques et qu’on oublie de le spécifier au logiciel, le t-stat, la p-valeur et les intervalles de confiance affichés par le logiciel seront erronés.

Écart-types robustes sous STATA

```
regress testscr str, robust

Regression 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
-------------------------------------------------------------------------
```

- L'ajout de l'option "**, robust**" à la commande indique à STATA de faire les calculs sous l'hypothèse d'hétéroscédasticité.
- Sinon STATA fera les calculs sous l'hypothèse d'homoscédasticité (l'option par défaut).

Autres fondements théoriques des MCO (Section 5.5)

- Nous en savons maintenant beaucoup sur les estimateurs MCO: ils sont sans biais et convergents.
- Nous savons comment calculer leurs écarts-types, avec ou sans l'hypothèse d'homoscédasticité.
- Nous savons comment utiliser les résultats de l'estimation par MCO pour faire des tests d'hypothèses et construire des intervalles de confiance.
- L'estimation par MCO est une des méthodes les plus utilisées. C'est une bonne raison pour vous de l'utiliser également.
- MCO est facile à expliquer : l'utilisation d'un autre estimateur nécessite un plus grand effort de justification.

Quelques questions peuvent quand même subsister:

- N'y a-t-il pas des estimateurs de variance plus faible que celle de l'estimateur MCO ?
- Où est passé la distribution de Student, celle qu'on aurait dû attribuer à la statistique t?

Pour répondre à ces questions, nous allons postuler des hypothèses plus fortes que précédemment.

Hypothèses du modèle linéaire (suite et fin)

1. $E(u|X = x) = 0$.
2. (X_i,Y_i), $i = 1,\ldots,n$, sont i.i.d.
3. Données aberrantes rares : $E(Y^4) < \infty$ et $E(X^4) < \infty$.
4. u est homoscédastique
5. u suit la loi normale $N(0,\sigma^2)$

- Les hypothèses #4 et #5 sont assez fort : elles ont tendance à ne pas être vérifiées en pratique. Cependant, leur ajout permet d'établir des résultats plus forts au sujet des MCO.
- Ci-après, nous commençons par une discussion sur l'efficacité des MCO.

Efficacité des MCO, partie I

Théoreme de Gauss-Markov : Sous les hypothèses du modèle linéaire #1 à #4, l'estimateur des MCO $\hat{\beta}_1$ a la plus petite variance parmi tous les estimateurs sans biais qui s'écrivent comme une fonction linéaire de (Y_1,…, Y_n).

- De ce fait, ont dit que $\hat{\beta}_1$ est le meilleur estimateur linéaire et sans biais (en englais, BLUE qui signifie « Best Linear Unbiased Estimator »).
- Pour la preuve de ce théorème, voir SW Appendice 5.2

- $\hat{\beta}_1$ est un “estimateur linéaire” signifie qu’il peut s’écrire comme une fonction linéaire des $Y_1,\ldots, Y_n$:

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

où $w_i = \dfrac{(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$.

- Le théorème de Gauss-Markov stipule que parmi tous les choix de pondération $\{w_i\}$ possibles, les MCO affectent les poids qui font en sorte que $\hat{\beta}_1$ a la plus petite variance.

Efficacité des MCO, partie II

- Sous les hypothèses du modèle linéaire #1 à #5, $\hat{\beta}_1$ est de variance minimale parmi tous les estimateurs convergents de β_1 (linéaire ou non).
- Ce résultat est très intéressant. Il stipule que sous les hypothèses #1 à #5 (c'est-à-dire hypothèses #1 à #3 + l'homoscédasticité et normalité des erreurs), l'estimateur des MCO est le meilleur parmi tous les estimateurs convergents. Comme un estimateur non convergent est un estimateur plutôt médiocre, cela signifie que l'estimateur des MCO est « le meilleur parmi les meilleurs », à condition que les hypothèses #1 à #5 soient satisfaites
- La preuve de ce résultat dépasse le cadre de ce cours.

Limitations des MCO

Malgré les résultats impressionnants qui précèdent, l'estimateur des MCO présente quelques limites.

1. Le théorème de Gauss-Markov n'est valide qu'en cas d'homoscédasticité, une condition rarement remplie en pratique. De plus, il se focalise sur les estimateurs linéaires qui ne représentent qu'un petit sous-ensemble des estimateurs possibles.
2. Le résultat d'optimalité (Efficacité, partie II) nécessite la normalité de l'erreur en plus de l'homoscédasticité, une condition peu plausible en pratique.

3. Les MCO sont assez sensibles aux valeurs extrêmes. Par exemple, dans le cadre de l'estimation de la moyenne en population, on doit préférer la médiane empirique à la moyenne empirique si les valeurs aberrantes sont trop fréquentes. En effet, la médiane empirique est plus robuste aux valeurs extrêmes et de ce fait, présente un risque plus faible que la moyenne empirique en présence de valeurs extrêmes.

4. L'estimateur de Moindres Écarts Absolus (défini ci-après) est plus robuste que les MCO à la présence de valeurs aberrantes dans l'échantillon.

$$\min_{b_0, b_1} \sum_{i=1}^{n} \left| Y_i - (b_0 + b_1 X_i) \right|$$

Mais cet estimateur est plus difficile à analyser que les MCO.

Inférence lorsque *u* est homoscédastique et de distribution normale (Section 5.6)

Reprenons les cinq hypothèses du modèle linéaire:

1. $E(u|X = x) = 0$.
2. (X_i,Y_i), $i = 1,\ldots,n$, sont i.i.d.
3. Rareté des valeurs extrêmes : $E(Y^4) < \infty$, $E(X^4) < \infty$.
4. u est homoscédastique.
5. u est de loi $N(0,\sigma^2)$.

Si ces cinq hypothèses sont vérifiées, alors:

- $\hat{\beta}_0$ et $\hat{\beta}_1$ suivent de lois normales pour tout n.
- Le t de Student suit une loi de Student à $n - 2$ degrés de libertés. Ceci est également vrai pour tout n.

Preuve de la normalité de $\hat{\beta}_1$ sous les hypothèses #1à #5:

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

$$= \frac{1}{n}\sum_{i=1}^{n} w_i u_i \text{, où } w_i = \frac{(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Une combinaison linéaire de variables suivant des lois normales suit également une loi normale. Nous avons:

$$\hat{\beta}_1 - \beta_1 \sim N\left(0, \frac{1}{n^2}\left(\sum_{i=1}^{n} w_i^2\right)\sigma_u^2\right) \qquad (*)$$

où w_i est donnée ci-dessus. Lorsqu'on substitue w_i dans (*), on obtient la variance de $\hat{\beta}_1$ dans le cas homoscédastique.

De plus, sous les hypothèses #1 à #5, la statistique t suit une loi de Student à n – 2 degrés de libertés (la loi t_{n-2}).

- Pourquoi n – 2? Parce qu'on estime 2 paramètres (β_0 et β_1) à partir des données.
- Lorsque n est inférieure à 30, les valeurs critiques de la loi de Student sont plus grands que celles de la loi $N(0,1)$.
- Lorsque n est plus grand que 50, la différence entre les deux lois devient négligeable.

Loi de Student

degrés de liberté	Valeurs critiques à 5% bilatéral
10	2.23
20	2.09
30	2.04
60	2.00
∞	1.96

Implications pratiques

- Si $n < 50$ et vous êtes convaincus que u est homoscédastique et de loi normale, alors vous pouvez faire vos inférences en supposant que t suit la loi t_{n-2} plutôt que sur la loi $N(0,1)$.
- En pratique, u n'est ni homoscédastique ni de loi normale dans la plupart des cas.
- Heureusement, on a assez de données de nos jours dans la plupart des applications (c'est-à-dire, $n > 50$). Ceci nous permet de faire l'approximation normale sur la base du théorème central limite et de faire les inférences comme si u suivait une loi normale.

Retour à la question de départ (Section 5.7)

- Supposons qu'on recrute plus d'enseignants de sorte à pouvoir diminuer le ratio élève-enseignant d'une unité. Quel serait l'effet attendu d'une telle politique ?
- Notre régression linéaire simple basée sur les données californiennes nous permet-il de répondre de manière convaincante à cette question ?

  Pas vraiment. Les districts ayant les STR plus faibles ont aussi tendance à être les plus riches. Les élèves de ces districts bénéficient donc d'un environnement qui favorise l'apprentissage extra-scolaire. Il a donc violation de l'hypothèse #1car $corr(u_i, STR_i) > 0$.
- L'ajout d'autres régresseurs peut remédier à ce problème.