Corrigé de l'interrogation n°2

Exercice [8points]

On dispose d'un échantillon $(X_1,\ X_2,\ ...,\ X_N)$ de taille N d'une variable suivant une loi Normale de paramètres m et σ^2 avec $\sigma > 0$, de densité :

$$f(x\ ; m,\ \sigma^2) = \frac{1}{\sqrt{2\ \Pi\ \sigma^2}}\ \exp\left(-\ \frac{1}{2\ \sigma^2}\,(x-m)^2\right)$$

1- La fonction de vraisemblance associée à la loi Normale de paramètre m et σ^2est donc:

$$\begin{aligned} L(x_1, x_2, ..., x_N; m,\ \sigma^2) &= \prod_{i=1}^{n} f(x\ ; m,\ \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\ \Pi\ \sigma^2}}\ \exp\left(-\ \frac{1}{2\ \sigma^2}\,(x-m)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{N}{2}} \prod_{i=1}^{N} exp\left(-\frac{(x_i-m)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-\frac{N}{2}} \quad exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-m)^2\right) \end{aligned}$$

La log-vraisemblance d'un échantillon $(x_1, x_2, ..., x_n)$ est donc égale à :

$$\begin{aligned} \ln L(x_1, x_2, ..., x_N; m,\ \sigma^2) &= \ln\left[(2\pi\sigma^2)^{-\frac{N}{2}}\right] + \ln\left[exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-m)^2\right)\right] \\ &= -\ \frac{N}{2}\ln\left[2\pi\sigma^2\right]\ -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-m)^2 \\ &= -\ \frac{N}{2}\ln\left[2\pi\right]\ -\ \frac{N}{2}\ln\left[\sigma^2\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-m)^2 \end{aligned}$$

2- En utilisant la condition du premier ordre portant sur m, déterminer l'estimateur du maximum de vraisemblance $\widehat{m}$ du paramètre m.

1.b. Afin de déterminer l'estimateur du maximum de vraisemblance $\widehat{m}$ du paramètre m, il faut maximiser la log-vraisemblance par rapport à la variable m à $(x_1, x_2, ..., x_N)$ donnés. La première étape consiste à résoudre l'équation correspondant à la condition du premier ordre :

$$\frac{\partial \ln L}{\partial m}(x_1, x_2, ..., x_N; m,\ \sigma^2) = 0$$

or $\frac{\partial \ln L}{\partial m}(x_1, x_2, ..., x_N; m,\ \sigma^2) = -\frac{1}{2\sigma^2}\sum\limits_{i=1}^{N}(-2)\ \ (x_i - m) = \frac{1}{\sigma^2}\sum\limits_{i=1}^{N}\ (x_i - m)$ donc : $\frac{\partial \ln L}{\partial m}(x_1, x_2, ..., x_N; m,\ \sigma^2) = 0 \Longleftrightarrow \frac{1}{\sigma^2}\sum\limits_{i=1}^{N}(x_i - \hat{m}) = 0 \Longleftrightarrow \sum\limits_{i=1}^{N}(x_i - \hat{m}) = 0$ d'où $\frac{\partial \ln L}{\partial m}(x_1, x_2, ..., x_N; m,\ \sigma^2) = 0 \Longleftrightarrow \sum\limits_{i=1}^{N}x_i - \sum\limits_{i=1}^{N}\hat{m} = 0 \Longleftrightarrow \sum\limits_{i=1}^{N}x_i = N\ \hat{m} \Longleftrightarrow$ $\hat{m} = \frac{1}{N}\sum\limits_{i=1}^{N}x_i = \bar{x}$

La fonction $m \to \ln L(x_1, x_2, ..., x_N; m,\ \sigma^2)$ atteint donc un extrémum (c'est-à-dire un maximum ou un minimum) au point $\hat{m} = \frac{1}{N}\sum\limits_{i=1}^{N} x_i$. Il s'agit bien d'un maximum car la fonction $m \to \ln L(x_1, x_2, ..., x_N; m,\ \sigma^2)$ est concave. En effet, pour tout $(x_1, x_2, ..., x_N; m,\ \sigma^2)$, on a :

$$\frac{\partial^2 \ln L}{\partial m^2}(x_1, x_2, ..., x_N; m,\ \sigma^2) = -\ \frac{N}{\sigma^2} < 0$$

En particulier, la condition du second ordre $\frac{\partial^2 \ln L}{\partial m^2}(x_1, x_2, ..., x_N; m,\ \sigma^2) < 0$ est satisfaite.

Conclusion: L'estimateur du maximum de vraisemblance du paramètre m est $\hat{m} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$.

3- On pose $s = \sigma^2$.

$$\begin{aligned}\ln L(x_1, x_2, ..., x_N; m,\ \sigma^2) &= -\ \frac{N}{2}\ln[2\pi]\ -\ \frac{N}{2}\ln\left[\sigma^2\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - m)^2 \\ &= -\ \frac{N}{2}\ln[2\pi]\ -\ \frac{N}{2}\ln[s] - \frac{1}{2\ s}\sum_{i=1}^{N}(x_i - m)^2\end{aligned}$$

La condition du premier ordre sur s: $\frac{\partial \ln L}{\partial s}(x_1, x_2, ..., x_N; m,\ \sigma^2) = 0$or

$$\frac{\partial \ln L}{\partial s}(x_1, x_2, ..., x_N; m,\ \sigma^2) = -\ \frac{N}{2}\ \frac{1}{s} + \frac{1}{s^2}\frac{1}{2}\sum_{i=1}^{N}(x_i - m)^2 = \frac{1}{2s}\left[-\ N\ + \frac{1}{s}\sum_{i=1}^{N}(x_i - m)^2\right]$$

donc : $\frac{\partial \ln L}{\partial s}(x_1, x_2, ..., x_N; m,\ \sigma^2) = 0$ et pour $m = \hat{m} = \bar{x} \Longleftrightarrow -\ N\ + \frac{1}{s}\sum\limits_{i=1}^{N}(x_i - \bar{x})^2 = 0 \Longleftrightarrow \frac{1}{s}\sum\limits_{i=1}^{N}(x_i - \bar{x})^2 = N$ d'où $\frac{\partial \ln L}{\partial m}(x_1, x_2, ..., x_N; m,\ \sigma^2) = 0$ $\Longleftrightarrow \sum\limits_{i=1}^{N}(x_i - \bar{x})^2 = N\ \hat{s} \Longleftrightarrow \hat{s}\ = \frac{\sum\limits_{i=1}^{N}(x_i - \bar{x})^2}{N}$

La fonction $s \to \ln L(x_1, x_2, ..., x_N; m,\ \sigma^2)$ atteint donc un extrémum (c'est-à-dire un maximum ou un minimum) au point $\hat{s}\ = \frac{\sum\limits_{i=1}^{N}(x_i - \bar{x})^2}{N}$. Il s'agit bien d'un maximum car, pour tout $(x_1, x_2, ..., x_N; m,\ \sigma^2)$, on a :

$$\frac{\partial^2 \ln L}{\partial s^2}(x_1, x_2, ..., x_N; m,\ \sigma^2) = \frac{N}{2}\ \frac{1}{s^2} - \frac{2\ s}{s^4}\frac{1}{2}\sum_{i=1}^{N}(x_i - m)^2 = \frac{N}{2}\ \frac{1}{s^2} - \frac{1}{s^3}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Pour $s = \hat{s} = \frac{\sum\limits_{i=1}^{N}(x_i - \bar{x})^2}{N}$, $\frac{\partial^2 \ln L}{\partial s^2}(x_1, x_2, ..., x_N; m,\ \sigma^2) = \frac{N}{2}\ \frac{1}{\hat{s}^2} - \frac{1}{\hat{s}^3}N\quad \hat{s} = -\frac{N}{2}\frac{1}{\hat{s}^2} < 0$

D'où, la condition du second ordre $\frac{\partial^2 \ln L}{\partial s^2}(x_1, x_2, ..., x_N; m,\ \sigma^2) < 0$ est satisfaite.

Conclusion: L'estimateur du maximum de vraisemblance du paramètre s est $\hat{s} = \frac{\sum\limits_{i=1}^{N}(x_i-\bar{x})^2}{N}$

4- $E(\hat{m}) = \frac{1}{N}\sum\limits_{i=1}^{N} E(X_i) = \frac{N\ m}{N} = m$ (l'estimateur $\hat{m}$ est sans biais).

$V(\hat{m}) = V\left(\frac{1}{N}\sum\limits_{i=1}^{N} X_i\right) = \frac{1}{N^2}V\left(\sum\limits_{i=1}^{N} X_i\right)$ or $V\left(\sum\limits_{i=1}^{N} X_i\right) = \sum\limits_{i=1}^{N} V(X_i)$ car les X_i sont supposés indépendants, donc

$V(\hat{m}) = \frac{1}{N^2}\sum\limits_{i=1}^{N} V(X_i) = \frac{1}{N^2} N\ \sigma^2 = \frac{\sigma^2}{N}$

L'expression de la quantité d'information de Fisher est :

$$I_n(m) = E\left(\left[\frac{\partial \ln\ln L(X_1, ..., X_N\ ; m,\ \sigma^2)}{\partial m}\right]^2\right) = -E\left(\frac{\partial^2 \ln\ln L(X_1, ..., X_N\ ; m,\ \sigma^2)}{\partial m^2}\right)$$

D'après ce qui précède, on a :$\frac{\partial^2 \ln L}{\partial m^2}(x_1, x_2, ..., x_N\ ; m,\ \sigma^2) = -\ \frac{N}{\sigma^2}$ donc $E\left(\frac{\partial^2 \ln L(X_1, ..., X_N\ ; m,\ \sigma^2)}{\partial m^2}\right) = -\ \frac{N}{\sigma^2}$ et $I_n(m) = \frac{N}{\sigma^2}$

L'estimateur $\hat{m}$ est un estimateur sans biais et sa variance vérifie $V(\hat{m}) = \frac{1}{I_n(p)}$. Il s'agit donc d'un estimateur efficace.

5***(*Bonus* 2 *points*) Etudier le biais de $\widehat{s}$. Comment peut-on le modifier pour qu'il devienne sans biais ?

$E(\hat{s}) = E\left[\frac{\sum\limits_{i=1}^{N}(x_i-\bar{x})^2}{N}\right] = \frac{1}{N}\ E\left[\sum\limits_{i=1}^{N}(x_i-\bar{x})^2\right] = \frac{1}{N}\ \sum\limits_{i=1}^{N} E\left[(x_i-\bar{x})^2\right]$ Or $E\left[(x_i-\bar{x})^2\right] = V\left[(x_i-\bar{x})\right] + \left[E(x_i-\bar{x})\right]^2$

$E(x_i - \bar{x}) = E(x_i) - E(\bar{x}) = m - m = 0$

$V\left[(x_i-\bar{x})\right] = V(x_i) - 2Cov(x_i, \bar{x}) + V(\bar{x})$

$Cov(x_i, \bar{x}) = Cov\left(x_i, \frac{1}{N}\sum\limits_{i=1}^{N} x_i\right) = \frac{1}{N}Cov(x_i, x_i) + \frac{1}{N}Cov\left(x_i, \sum\limits_{j\neq i}^{N} x_i\right) = \frac{1}{N}Cov(x_i, x_i) + 0 = \frac{1}{N}V(x_i) = \frac{\sigma^2}{N}$

D'où, $V\left[(x_i-\bar{x})\right] = V(x_i) - 2Cov(x_i, \bar{x}) + V(\bar{x}) = \sigma^2 - 2\frac{\sigma^2}{N} + \frac{\sigma^2}{N} = \sigma^2 - \frac{\sigma^2}{N} = \frac{N-1}{N}\ \sigma^2$ et donc $E\left[(x_i-\bar{x})^2\right] = V\left[(x_i-\bar{x})\right] + \left[E(x_i-\bar{x})\right]^2 = V\left[(x_i-\bar{x})\right]$ $E(\hat{s}) = \frac{1}{N}\ \sum\limits_{i=1}^{N} E\left[(x_i-\bar{x})^2\right]$ $E(\hat{s}) = \frac{1}{N}\ \sum\limits_{i=1}^{N}\frac{N-1}{N}\ \sigma^2 = \frac{N-1}{N}\ \sigma^2 \neq \sigma^2$ donc l'estimateur des MVS $\hat{s}$ est biaisé. L'estimateur sans biais de σ^2 (s) est $\frac{1}{N-1}\sum\limits_{i=1}^{N}(x_i-\bar{x})^2$ soit $\frac{N}{N-1}\hat{s}$.

Problème [15*points*] **:**

1.Le modèle : $y_i = \beta_1 + \beta_2\ x_i + \varepsilon_i \qquad i = 1, ...N$ peut s'écrire sous la forme :

$y = X\beta + \varepsilon$ avec $y = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$, et $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_N \end{bmatrix}$

2. L'estimation par la méthode des moindres carrés ordinaires se fait à partir de la minimisation de la somme des carrés des résidus $\sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i)^2$. Les équations normales s'obtiennent à partir des conditions du premier ordre :

$$\frac{\partial \left(\sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i)^2 \right)}{\partial \beta_2} = 0$$

$$\frac{\partial \left(\sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i)^2 \right)}{\partial \beta_1} = 0$$

ces deux conditions étant vérifiées pour $\beta_2 = \widehat{\beta_2}$ et $\beta_1 = \widehat{\beta_1}$.

Or $\frac{\partial \left(\sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i)^2 \right)}{\partial \beta_2} = -2 \sum_{i=1}^{N} x_i (y_i - \beta_1 - \beta_2 x_i) = -2 \sum_{i=1}^{N} \left(x_i y_i - \beta_1 x_i - \beta_2 a x_i^2 \right) = -2 \left[\sum_{i=1}^{N} x_i y_i - \beta_2 \sum_{i=1}^{N} x_i^2 - \beta_1 \sum_{i=1}^{N} x_i \right]$ et $\frac{\partial \left(\sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i)^2 \right)}{\partial \beta_1} = -2 \sum_{i=1}^{N} (y_i - \beta_1 - \beta_2 x_i) = -2 \left[\sum_{i=1}^{N} y_i - \beta_1 N - \beta_{2i} \sum_{i=1}^{N} x_i \right]$ donc les estimateurs $\widehat{\beta_1}$ et $\widehat{\beta_2}$ vérifient les deux équations suivantes (dites équations normales) :

$$\sum_{i=1}^{N} x_i y_i - \widehat{\beta_2} \sum_{i=1}^{N} x_i^2 - \widehat{\beta_1} \sum_{i=1}^{N} x_i = 0 \tag{1}$$

$$\sum_{i=1}^{N} y_i - \widehat{\beta_2} \sum_{i=1}^{N} x_i - \widehat{\beta_1} N = 0 \tag{2}$$

3.En divisant par N les deux membres de l'équation (2) on obtient l'expression de $\widehat{\beta_1}$ en fonction de $\widehat{\beta_2}$:

$$\widehat{\beta_1} = \bar{y} - \widehat{\beta_2} \bar{x}$$

et en remplaçant $\widehat{\beta_1}$dans l'équation (1) par son expression, on obtient :

$$\sum_{i=1}^{N} x_i y_i - \widehat{\beta_2} \sum_{i=1}^{N} x_i^2 - \left(\bar{y} - \widehat{\beta_2} \bar{x} \right) \sum_{i=1}^{N} x_i = 0$$

or $\sum_{i=1}^{N} x_i = N\bar{x}$ donc

$$\sum_{i=1}^{N} x_i y_i - \widehat{\beta_2} \sum_{i=1}^{N} x_i^2 - N\bar{x}\bar{y} + \widehat{\beta_2} N \bar{x}^2 = 0$$

d'où

$$\widehat{\beta_2}\sum_{i=1}^{N}x_i^2 - \widehat{\beta_2}N\bar{x}^2 = \sum_{i=1}^{N}x_iy_i - N\bar{x}\bar{y}$$

et par conséquent :

$$\widehat{\beta_2} = \frac{\sum_{i=1}^{N}x_iy_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N}x_i^2 - N\bar{x}^2} = \frac{S_{xy}}{S_{xx}}$$

4. $\hat{\beta} = (X'X)^{-1}\ X'y = (X'X)^{-1}\ [X'X\beta + \varepsilon] = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon = \beta + (X'X)^{-1}X'\varepsilon$

$\widehat{\beta_2} = \frac{\sum_{i=1}^{N}x_iy_i - N\bar{x}\bar{y}}{\sum_{i=1}^{N}x_i^2 - N\bar{x}^2} = \beta_2 + \frac{\sum_{i=1}^{N}x_i\varepsilon_i - N\bar{x}\bar{\varepsilon}}{S_{xx}}$

$\widehat{\beta_1} = \bar{y} - \widehat{\beta_2}\bar{x} = \bar{y} = \beta_1 + \beta_2\bar{x} + \bar{\varepsilon} - \left[\beta_2 + \frac{1}{S_{xx}}\sum_{i=1}^{N}x_i\varepsilon_i - N\bar{x}\bar{\varepsilon}\right]\bar{x} = \beta_1 + \frac{\bar{\varepsilon}\sum_{i=1}^{N}x_i^2 - \bar{x}\sum_{i=1}^{N}x_i\varepsilon_i}{S_{xx}}$

5. Les estimateurs des MCO $\widehat{\beta_1}$ et $\widehat{\beta_2}$ sont sans biais, linéaires par rapport aux variables dépendantes y_i et de variance minimale dans la classe des estimateurs linéaires sans biais. Cette dernière propriété porte sur l'efficacité *relative* des estimateurs des MCO par rapport aux autres estimateurs linéaires sans biais (Théorème de Gauss-Markov). En aucun cas, ceci n'implique que les estimateurs $\widehat{\beta_1}$ et $\widehat{\beta_2}$ sont efficaces au sens où leur variance atteint la borne FDCR.

$E\left(\widehat{\beta}\right) = E\left(\beta + (X'X)^{-1}\varepsilon\right) = E(\beta) + E\left((X'X)^{-1}X'\varepsilon\right) = \beta + (X'X)^{-1}X'E(\varepsilon)$ or $E(\varepsilon) = 0$

$E\left(\widehat{\beta}\right) = \begin{bmatrix} E\left(\widehat{\beta_1}\right) \\ E\left(\widehat{\beta_2}\right) \end{bmatrix} = \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

6.1 L'estimateur $\widehat{\beta_2}$ est une combinaison linéaire d'une variable aléatoire normale donc $\widehat{\beta_2}$ suit une loi Normale.

$E\left(\widehat{\beta_2}\right) = \beta_2$ et $V\left(\widehat{\beta_2}\right) = \frac{\sigma^2}{S_{xx}} = \sigma_2^2$ D'où $\widehat{\beta_2} \sim N\left(\beta_2, \sigma_2^2\right)$. de même pour l'estimateur $\widehat{\beta_1} \sim N\left(\beta_1, \sigma_1^2\right)$ avec $\sigma_1^2 = \frac{\sigma^2}{N}\frac{\sum_{i=1}^{N}x_i^2}{S_{xx}}$

6.2 On a montré plus haut que $\widehat{\beta_2} \sim N\left(\beta_2, \sigma_2^2\right)$, et donc $\frac{\widehat{\beta_2} - \beta_2}{\sigma_2} \sim N(0,1)$

Posons $U = \frac{\widehat{\beta_2} - \beta_2}{\sigma_2}$, $P(-u \leq U \leq u) = 0.98 = 2F(u) - 1$ où F est la fonction de répartition de la loi $N(0,1)$, d'où $F(u) = 0.99$. Sur la table statistique de $N(0,1)$, $u = 2.33$

$$P(-2.33 \leq U \leq 2.33) = P\left(-2.33 \leq \frac{\widehat{\beta_2} - \beta_2}{\sigma_2} \leq 2.33\right) = 0.98$$

donc :

$$P\left(\widehat{\beta_2} - 2.33\ \sigma_2 \leq \beta_2 \leq \widehat{\beta_2} + 2.33\ \sigma_2\right) = 0.98$$

On obtient alors l'intervalle de confiance pour β_2 au seuil de 98% suivant : $\widehat{\beta_2} - 2.33\ \sigma_2 \leq \beta_2 \leq \widehat{\beta_2} + 2.33\ \sigma_2$

De la même manière l'intervalle de confiance pour β_1 au seuil de 98% suivant : $\widehat{\beta_1} - 2.33\ \sigma_1 \leq \beta_1 \leq \widehat{\beta_1} + 2.33\ \sigma_1$

6.3 Sachant, $N = 5$, $\sum\limits_{i=1}^{5} y_i = 75$, $\sum\limits_{i=1}^{5} x_i = 55$, $\sum\limits_{i=1}^{5} x_i y_i = 880$, $\sum\limits_{i=1}^{5} x_i^2 = 633$ et $\sigma^2 = 0,25$

$\overline{y} = \frac{1}{N}\sum\limits_{i=1}^{N} y_i = 15$

$\overline{x} = \frac{1}{N}\sum\limits_{i=1}^{N} x_i = 11$

$S_{xy} = \sum\limits_{i=1}^{N} x_i y_i - N\overline{xy} = 55$

$S_{xx} = \sum\limits_{i=1}^{N} x_i^2 - N\left(\overline{x}\right)^2 = 28$

D'où $\widehat{\beta_2} = \frac{S_{xy}}{S_{xx}} = \frac{55}{28} = 1.96 \simeq 2$ et $\widehat{\beta_1} = \bar{y} - \widehat{\beta_2}\bar{x} = 15 - 2 \times 11 = -7$

$\sigma_1^2 = \frac{\sigma^2}{N}\frac{\sum_{i=1}^{N} x_i^2}{S_{xx}} = (1.6)^2$ et $\sigma_2^2 = \frac{\sigma^2}{S_{xx}} = (0.09)^2$

Pour ces valeurs, on obtient l'intervalle de confiance pour β_2 au seuil de 98% suivant : $[1.79\ ,\quad 2.21]$.

Ainsi que l'intervalle de confiance pour β_1 au seuil de 98% suivant : $[-9.47\ ,\quad -4.53]$

7.1 $z_i = y_i - x_i = \beta_1 + \beta_2 x_i + \varepsilon_i - x_i = \beta_1 + (\beta_2 - 1)x_i + \varepsilon_i = a + b\ x_i + v_i$

d'où $a = \beta_1$, $b = \beta_2 - 1$ et $v_i = \varepsilon_i$.

7-2 On considère le modèle : $z_i = a + b\ x_i + v_i$

$$\begin{aligned} \widehat{a} &= \bar{z} - \widehat{b}\bar{x} \\ \widehat{\beta_2} &= \frac{S_{xz}}{S_{xx}} \end{aligned}$$

7-3 $S_{xz} = \sum\limits_{i=1}^{N} x_i z_i - N\overline{xz} = \sum\limits_{i=1}^{N} x_i(y_i - x_i) - N\overline{x}(\bar{y} -\ \bar{x}) = \sum\limits_{i=1}^{N}\left(x_i y_i - x_i^2\right) - N\overline{x}\overline{y} -\ N\bar{x}^2$

$S_{xz} = \sum\limits_{i=1}^{N} x_i y_i - N\overline{x}\overline{y} - \left(\sum\limits_{i=1}^{N} x_i^2 - N x_i^2\right) = S_{xy} - S_{xx}$

7-4

$$\begin{aligned} \widehat{a} &= \bar{z} - \widehat{b}\bar{x} = (\bar{y} - \bar{x}) - \left(\widehat{\beta_2} - 1\right)\bar{x} = \bar{y} - \widehat{\beta_2}\bar{x} = \widehat{\beta_1} \\ \widehat{b} &= \frac{S_{xz}}{S_{xx}} = \frac{S_{xy} - S_{xx}}{S_{xx}} = \frac{S_{xy}}{S_{xx}} - 1 = \widehat{\beta_2} - 1 \end{aligned}$$