MAT 7381 - Solution # 4.20

a) Droite de regression : $y = -3,38 + 1,095803x$

b) À l'oeil, la relation semble plutôt faible, mais le coefficient de correlation de 0,8 dément cette impression. Le graphique des residus ne présente pas d'évidence de non-linéarité où d'hétéroscédasticité. Les statistiques de cook ne montre pas de données excessivement influentes.

c) L'intervalle de confiance est $17,65 \leq \beta_0 + \beta_1(20) \leq 19,47$: on peut affirmer avec 80% de confiance que la valeur moyenne de y pour un specimen evalué à 20 par le NBS se situe entre 17,65 et 19,47.

Les limites de prédiction sont $13,0 \leq \hat{y}_{20} \leq 24,1$. La teneur en plomb d'un specimen évalué à 20 par le NBS sera évaluée quelque part entre 13,0 et 24,1 par le laboratoire clinique.

d) $\hat{\beta}_1 = 1,095803$; $\hat{\sigma}_{\hat{\beta}_1} = 0,1173924$; $t_{n-2;\,.025} = 2,0129$. Intervalle de confiance : $0,8595 \leq \beta_1 \leq 1,3321$. On ne peut donc pas rejeter l'hypothèse que $\beta_1 = 0$.

e) Sous H_0, u_i est de moyenne 0 et $SCR_0 = \Sigma u_i^2 = 895,3625$, à n degrés de liberté. Dans le modèle de régression linéaire, $SCR = 818,8485$ à $n-2 = 46$ degrés de liberté, $MCR = 17,80105$. $F = \frac{(SCR_0 - SCR)/2}{MCR}$ $= \frac{38,25699}{17,80105} = 2,149$. À 2 et 46 degrés de liberté, $p = 0,128$. On ne peut pas rejeter l'hypothèse.

f) On utilise ici le fait que si $\begin{bmatrix} \underset{\sim}{x} \\ \underset{\sim}{y} \end{bmatrix}$ est de loi normale de moyenne $\begin{bmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{bmatrix}$ et de matrice de covariance $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ alors la loi conditionnelle de y étant donné x est normale de moyenne $\underset{\sim}{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu}_1)$ et de matrice de covariance $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

(i) Découle directement de l'énoncé ci-dessus

(ii) $E(y_i | x_i) = \mu_2 + \frac{\sigma_{12}}{\sigma_1^2}(x_i - \mu_1) = \left(\mu_2 - \frac{\sigma_{12}}{\sigma_1^2}\mu_1\right) + \frac{\sigma_{12}}{\sigma_1^2} x_i$: une fonction linéaire de x_i.

(iii) $Var(y_i | x_i) = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}$, effectivement indép. de x_i

La loi conditionnelle de y_i étant donné x_i est donc normale de moyenne $\beta_0 + \beta_1 x_i$ et de variance σ^2, où $\beta_0 = \mu_2 - \frac{\sigma_{12}}{\sigma_1^2}\mu_1$ et $\beta_1 = \frac{\sigma_{12}}{\sigma_1^2}$. La variance σ^2 est, en termes de la distribution conjointe, est $\sigma^2 = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}$.

g) $\beta_1 = 1 \Rightarrow \sigma_{12} = \sigma_1^2$, $(\beta_0 = 0 \text{ et } \beta_1 = 1) \Rightarrow \mu_1 = \mu_2$.

$\bar{u} = -1,160625$; $\hat{\sigma}_{\bar{u}} = s_u/\sqrt{n} = 4,204112/\sqrt{48}$

$T = \frac{\bar{u}}{\hat{\sigma}_{\bar{u}}} = -1,912662$. Sous H_0, $T \sim t_{n-1}$. La valeur p est 0,06189577.

h) $w = L'\begin{bmatrix} x_i \\ y_i \end{bmatrix}$ suit une loi normale de moyenne $L'\mu$ et de matrice de covariance $L'\Sigma L$, où $L' = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$. w est de moyenne $\begin{bmatrix} \mu_1 \\ \mu_2 - \mu_1 \end{bmatrix}$ et de matrice de covariance $\begin{bmatrix} \sigma_1^2 & \sigma_{12} - \sigma_1^2 \\ \sigma_{12} - \sigma_1^2 & \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} \end{bmatrix}$

h) (Suite) Sous H_0, $\mu_2-\mu_1=0$, $\sigma_{12}=\sigma_1^2$, $\sigma_1^2+\sigma_2^2-2\sigma_{12}=(1-\rho^2)\sigma_2^2$

Donc $\underset{\sim}{\psi}_i$ est normale, de moyenne $\begin{bmatrix}\mu_1\\0\end{bmatrix}$ et de matrice de covariance $\sigma_2^2\begin{bmatrix}\rho^2 & 0\\ 0 & 1-\rho^2\end{bmatrix}$. Les estimateurs du maximum de vraisemblance sont $\hat{\mu}_1=\bar{\omega}_1=\bar{x}$; $\widehat{\rho^2\sigma_2^2}=\hat{\psi}_1^2 = \Sigma(\omega_{1i}-\bar{\omega}_1)^2/n$; $\widehat{(1-\rho^2)\sigma_2^2}=\Sigma\omega_{2i}^2/n=\hat{\psi}_2^2$

$\bar{\omega}_1=22{,}95229$; $\widehat{\psi_1^2}=26{,}91069$; $\widehat{\psi_2^2}=447{,}6812$.

i) Sans les restrictions H_{00}, la moyenne de $\underset{\sim}{\omega}_i$ est estimée par $[\bar{\omega}_1,\bar{\omega}_2]'=[22{,}95229 ; -1{,}160625]$ et la matrice de covariance par $\Sigma(\underset{\sim}{\omega}_i-\bar{\underset{\sim}{\omega}})(\underset{\sim}{\omega}_i-\bar{\underset{\sim}{\omega}})'/n=\begin{bmatrix}26{,}9107 & 2{,}57812\\ 2{,}57812 & 17{,}3063\end{bmatrix}$

Alors $\ln L=-191{,}2104$ (maximum) et $\max \ln L_0=-269{,}6276$

$-2\ln\frac{L_0}{L}=-2[-269{,}6276+191{,}2104]=156{,}83$. Sous H_0 $-2\ln(L_0/L)$ est à peu près de loi χ_2^2 (une restriction sur le vecteur moyenne et une restriction sur la matrice de covariance. Donc on rejette fortement H_{00}.

j) Sous ces hypothèses, la distribution conditionnelle de $y_i | x_i$ est de moyenne $\mu(1-\frac{\sigma_{12}}{\sigma_1^2})$ et de matrice de covariance $\sigma^2\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$. Alors

$\beta_1=\frac{\sigma_{12}}{\sigma_1^2}=\rho$

b) La matrice de covariance de v_i est $L'\left[\sigma_z^2 ee' + \sigma^2 I\right]L$, où $L' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Elle est donc egale à $\begin{bmatrix} 2\sigma^2 & 0 \\ 0 & 4\sigma_z^2 + 2\sigma^2 \end{bmatrix}$

Cette matrice est estimée sans biais par la matrice de covariance échantillonnale $\hat{\Sigma}_v = \begin{bmatrix} 17{,}67455 & -22{,}9405 \\ -22{,}9405 & 138{,}1395 \end{bmatrix}$

Donc σ^2 peut être estimée sans biais par $\hat{\Sigma}_v[1,1]/2 : \hat{\sigma}^2 = 8{,}837277$. De même une estimation de $4\sigma_z^2 + 2\sigma^2$ est $\hat{\Sigma}_v[2,2] = 138{,}1395$, ce qui permet d'estimer σ_z^2 par $\left[\hat{\Sigma}_v[2,2] - 2\hat{\sigma}^2\right]/4 = 30{,}11623$