Université Paris 1 Panthéon-Sorbonne, **UFR02**,
Licence de Sciences Economiques, **STATISTIQUE**, **Cours de Mme Pradel**

Corrigé Partiel 13 juin 2000

Exercice 1

$Y_1, ..., Y_{36} \sim i.i.d.N[m; 6.25] \Rightarrow \bar{Y} \sim N\left[m; \frac{6.25}{36}\right]$

1. $P[|N(0.1)| \le 1.960] = 0.95 \Rightarrow P\left[\bar{Y} - 1.96\frac{2.5}{6} \le m \le \bar{Y} + 1.96\frac{2.5}{6}\right] = 0.95$
   $P[\bar{Y} - 0.8167 \le m \le \bar{Y} + 0.8167] = 0.95$
2. Pour $\bar{Y} = 11.26$, l'intervalle de confiance à 95% pour m est $10.443 \le m \le 12.077$.
3. On suppose que $Y_{Fred} \sim N[m; 6.25]$ indépendant des 36 notes déjà observées. La meilleure prévision ponctuelle de Y_{Fred} serait m si on le connaissait, c'est donc $\bar{Y}$. L'erreur de prévision $Y_{Fred} - \bar{Y}$ a pour loi $N\left[0; 6.25\left(1 + \frac{1}{36}\right)\right]$.
   $P[|N(0.1)| \le 2.326] = 0.98 \Rightarrow P\left[\bar{Y} - 2.326 * 2.5\sqrt{1 + \frac{1}{36}} \le Y_{Fred} \le \bar{Y} + 2.326 * 2.5\sqrt{1 + \frac{1}{36}}\right] = 0.98$
   $P[\bar{Y} - 5.895 \le Y_{Fred} \le \bar{Y} + 5.8952] = 0.98$
   Numériquement, l'intervalle de prévision à 98% pour Y_{Fred} est $5.365 \le Y_{Fred} \le 17.155$.

Exercice 2

Comparaison de deux échantillons normaux indépendants : $X_1, ..., X_{16} \sim i.i.d.N[m_1; \sigma_1^2]$ et $X_{17}, ..., Y_{36} \sim i.i.d.N[m_2; \sigma_2^2]$.

1. Comparaison des variances = Test de $\{\sigma_1^2 = \sigma_2^2\}$ contre $\{\sigma_1^2 \ne \sigma_2^2\}$.(test bilatéral)
   - Refus de $\{\sigma_1^2 = \sigma_2^2\}$ si $\left\{\frac{S_1^2}{S_2^2} > A \text{ ou } \frac{S_2^2}{S_1^2} > B\right\}$ où S_1^2 et S_2^2 sont les variances empiriques des deux échantillons.
   - Sous Ho, $\frac{S_1^2}{S_2^2} \sim FISHER(15; 19)$ et $\frac{S_2^2}{S_1^2} \sim FISHER(19; 15)$. En mettant 5% de probabilité de chaque côté, on lit dans la table de FISHER à 5% les valeurs de A et B : $P[FISHER(15; 19) > 2.23] = 5\%$ et $P[FISHER(19; 15) > 2.34]$.
   - On observe $\frac{s_2^2}{s_1^2} = \frac{12.083}{5.811} = 2.079 < 2.34$, donc les observations sont en dehors de la région critique
   - Conclusion : au seuil de 10%, on ne peut refuser l'hypothèse d'égalité des variances des notes dans les deux sections.
2. Comparaison des moyennes : on suppose ici que les variances sont égales. La meilleure prévision de la variance commune est alors : $S^2 = \frac{15S_1^2 + 19S_2^2}{15+19}$. On teste $\{m_1 = m_2\}$ contre $\{m_1 < m_2\}$ (test unilatéral).
   - Refus de $\{m_1 = m_2\}$ si $Z = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{S^2(1/16 + 1/20)}} > A$.
   - Sous Ho, $Z \sim STUDENT(34)$. La table de Student nous fournit (colonne 0.20) $P[STUDENT(\infty) > 1.281] = 10\%$., et donc $A = 1.28$.
     Remarque : l'énoncé autorise l'utilisation de la loi Normale (student(∞)), et le calcul exact, non disponible sur la table fournie, aurait donné A = 1.307. On voit en fait que A = 1.3 serait tout à fait exact.
   - On observe $s^2 = \frac{15*5.811 + 19*12.083}{34} = 9.316$ et $z = \frac{11.247 - 9.222}{\sqrt{s^2(36/320)}} = 1.98$
   - Conclusion : au seuil de 10%, on décide que les notes dans la section S2 ont une espérance mathématique supérieure à celle des notes dans la section S1.

Exercice 3

1. Le modèle est linéaire, car $E(Y_i) = a_1x1_i + b_1e1_i + a_2x2_i + b_2e2_i$ est combinaison linéaire de (a_1, b_1, a_2, b_2) , en posant de façon usuelle $e1_i = 1$ si $i \in S1$ et 0 sinon, $e2_i = 1$ si $i \in S2$ et 0 sinon, $x1_i = x_ie1_i$ et $x2_i = x_ie2_i$.
   Pour que le modèle soit de plus standard, il faut de plus que $\sigma_1^2 = \sigma_2^2$, car l'absence d'autocorrélations est déjà acquise (les individus sont indépendants).
2. Comparaison de variances : on teste $\{\sigma_1^2 = \sigma_2^2\}$ contre $\{\sigma_1^2 \ne \sigma_2^2\}$.
   - Refus de $\{\sigma_1^2 = \sigma_2^2\}$ si $\left\{\frac{S_1^2}{S_2^2} > A \text{ ou } \frac{S_2^2}{S_1^2} > B\right\}$ où $S_1^2 = \frac{SCR1}{14}$ et $S_2^2 = \frac{SCR2}{18}$ sont les variances résiduelles des deux régressions.

- Sous Ho, $\frac{S_1^2}{S_2^2}\sim FISHER(14;18)$ et $\frac{S_2^2}{S_1^2}\sim FISHER(18;14)$. En mettant 5% de probabilité de chaque côté, on lit dans la table de FISHER à 5% les valeurs de A et B : $P[FISHER(14;18) > 2.29] = 5\%$ et $P[FISHER(18;14) > 2.41]$.
- On observe $s_1^2 = \frac{0.4394}{14} = 0.0314$ et $s_2^2 = \frac{0.7008}{18} = 0.0389$
  $\frac{s_2^2}{s_1^2} = \frac{0.7008/18}{0.4394/14} = 1.24 < 2.41$, donc les observations sont en dehors de la région critique
- Conclusion : au seuil de 10%, on peut accepter l'hypothèse d'égalité des variances des perturbations dans les deux sections.

3. modèle contraint :
   a. égalité des paramètres entre les deux sections :$\{a_1 = a_2, b_1 = b_2\}$.
   b. Test de sous hypothèse linéaire = test simultané des deux contraintes. La SCR non contrainte est la somme des SCR des deux régressions partielles : $SCR1 + SCR2$ avec $14 + 18 = 32$ degrés de liberté, la régression sous contrainte étant celle effectuée sur toutes les données, donnant SCR.
      - refus de Ho si $F = \frac{(SCR-SCR1-SCR2)/2}{(SCR1+SCR2)/32} > A$.
      - Sous Ho, $F\sim FISHER(2;32)$. On lit dans la table de FISHER à 5% la valeur de A : $P[FISHER(2;32) > 3.30] = 5\%$
      - On observe $F = \frac{(1.1707-0.4394-0.7008)/2}{(0.4394+0.7008)/32} = 0.428 < 3.3$
      - Conclusion : au seuil de 5%, les observations ne permettent pas de déceler un changement des paramètres entre les deux sections.
4. Régression (II) $y = 0.788x + 3.105e + \hat{u}$,
   a. les variances covariances du couple $\left(\hat{a},\hat{b}\right)$ sont données par $\sigma^2(X'X)^{-1}$, où $X = \begin{pmatrix} x & e \end{pmatrix}$:
      $Var(\hat{a}) = 0.331\sigma^2$, $Var\left(\hat{b}\right) = 0.00283\sigma^2$, $Cov\left(\hat{a},\hat{b}\right) = -0.0293\sigma^2$
      On acceptera également la version $Var(\hat{a}) = 0.00283\sigma^2$, $Var\left(\hat{b}\right) = 0.331\sigma^2$, $Cov\left(\hat{a},\hat{b}\right) = -0.0293\sigma^2$ qui correspond à l'interprétation $X = \begin{pmatrix} e & x \end{pmatrix}$, puisque les résultats numériques sont données en mettant la constante d'abord.
   b. $\hat{\sigma}^2 = \frac{SCR}{36-2} = \frac{1.1707}{34} = 0.0344$
   c. On en déduit : $\widehat{Var(\hat{a})} = 0.011397$, $\widehat{Var\left(\hat{b}\right)} = 9.74 * 10^{-5}$, $\widehat{Cov\left(\hat{a},\hat{b}\right)} = -0.00101$
      Variante acceptée : $\widehat{Var\left(\hat{b}\right)} = 0.011397$, $\widehat{Var(\hat{a})} = 9.74 * 10^{-5}$, $\widehat{Cov\left(\hat{a},\hat{b}\right)} = -0.00101$
5. Test de $\{a = 0\}$ contre $\{a \neq 0\}$: test de student bilatéral.
   - refus de Ho si $\left|\frac{\hat{a}}{\widehat{\sigma_a}}\right| > A$, où $\widehat{\sigma_a} = \sqrt{\widehat{Var(\hat{a})}}$.
   - Sous Ho, $\frac{\hat{a}}{\widehat{\sigma_a}}\sim STUDENT(34)$. On choisit de faire un test de seuil 5%. On lit dans la table de Student (colonne 0.05) : $P[STUDENT(34) > 2.0] = 5\%$.
   - On observe $\left|\frac{\hat{a}}{\widehat{\sigma_a}}\right| = \frac{3.105}{\sqrt{0.0114}} = 29.1 > 2$... Variante acceptée : $\left|\frac{\hat{a}}{\widehat{\sigma_a}}\right| = \frac{3.105}{\sqrt{0.0000974}} = 314.6 > 2$.
   - Conclusion : au seuil de 5%, la note moyenne annuelle explique les variations de la note du bac.
6. Prévision de Y_{Fred} : la meilleure prévision sachant que $x_{Fred} = 11$ est $\widehat{Y} = 11\hat{a} + \hat{b}$, indépendante de Y_{Fred}.
   L'erreur de prévision $Y_{Fred} - \widehat{Y} \sim N[0,V]$ où
   $V = \left(\sigma^2 + (11)^2 Var(\hat{a}) + 2 * 11 Cov\left(\hat{a},\hat{b}\right) + Var\left(\hat{b}\right)\right) = \sigma^2(1 + 40.409)$ est estimée par $41.409\hat{\sigma}^2$.
   Variante acceptée :
   $(11)^2 Var(\hat{a}) + 2 * 11 Cov\left(\hat{a},\hat{b}\right) + Var\left(\hat{b}\right) = \sigma^2(121 * 0.00283 - 11 * 0.0293 + 0.331) = 0.3511\sigma^2$
   $\frac{Y_{Fred}-\widehat{Y}}{\sqrt{\hat{\sigma}^2 41.41}}\sim STUDENT(34)$ et $P[|STUDENT(34)| < 2.33] = 98\%$
   L'intervalle de prévision à 98% pour Y_{Fred} est donc :
   $P\left[11\hat{a} + \hat{b} - 2.33 * \sqrt{\hat{\sigma}^2 41.41} \le Y_{Fred} \le 11\hat{a} + \hat{b} + 2.33 * \sqrt{\hat{\sigma}^2 41.41}\right] = 0.98$
   Sa réalisation est ici : $8.991 \le Y_{Fred} \le 14.555$
   Variante acceptée : $11.517 \le Y_{Fred} \le 12.029$

Exercice 4

1. Test de $\{\lambda = 1\}$ contre $\{\lambda = 2\}$, avec $X_1, ..., X_4 \sim i.i.d.POISSON(\lambda)$
   a. Le test le plus puissant parmi ceux de seuil 10% exactement est celui de région critique $\sum_{i=1}^{4} X_i > A$, $\left(\text{car } 2 > 1\right)$, où A est déterminé par :$P\left[\sum_{i=1}^{4} X_i > A \mid \lambda = 1\right] = 0.10$

Or sous Ho, $\sum_{i=1}^{4} X_i \sim POISSON(4)$. Il n'existe pas de réel A qui corresponde exactement au seuil de 10% : on passe directement de 0.1106 à 0.0511, puisque la variable de Poisson prend des valeurs entières.

b. Le test le plus puissant sera celui de seuil maximal, parmi ceux de seuil inférieur à 10% On prendra donc la règle de refuser $\{\lambda = 1\}$ si $\sum_{i=1}^{4} X_i \geq 8$.

c. Le seuil exact de ce test est $P\left[\sum_{i=1}^{4} X_i \geq 8 \mid \lambda = 1\right] = 0.0511$

d. La puissance de ce test est $P\left[\sum_{i=1}^{4} X_i \geq 8 \mid \lambda = 2\right] = 0.5471$, lu dans loi de Poisson de paramètre 2*4 = 8.

2. Le test ainsi défini ne dépend pas de la valeur de λ dans l'alternative, tant que cette valeur reste supérieure à 1 : il est donc uniformément le plus puissant pour $\{\lambda = 1\}$ contre $\{\lambda > 1\}$.
3. On regarde dans la table entre les colonnes $\lambda = 4$ et $\lambda = 8$:
   $P[Z \geq 7 \mid \lambda = 1] = 0.1106$ $P[Z \geq 7 \mid \lambda = 2] = 0.6867 < 1 - 0.1106 = 0.8894$
   $P[Z \geq 6 \mid \lambda = 1] = 0.2148$ $P[Z \geq 7 \mid \lambda = 2] = 0.8088 > 1 - 0.2148 = 0.7852$
   Il faut donc prendre $\alpha = 21.5\%$ pour que la puissance en $\lambda = 2$ soit supérieure à $1 - \alpha$.