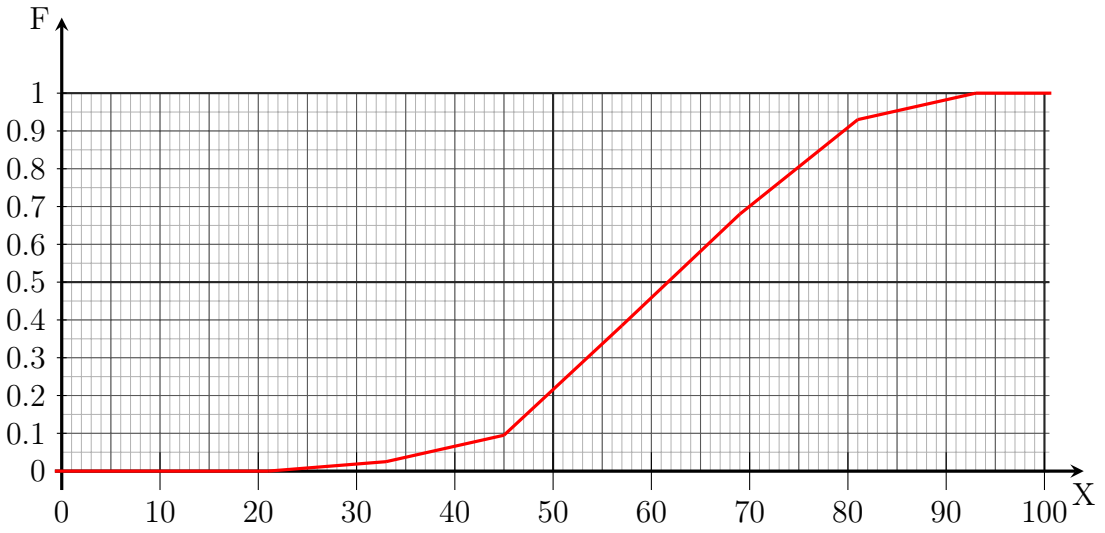
Corrigé de l'examen de Statistiques – 21 Avril 2015

1 Estime de soi : groupe témoin

1. $n = 5 + 14 + 58 + 59 + 50 + 14 = 200$
2. moyenne : $m(X) = \frac{\sum_i c_i\, n_i}{N} = \frac{27\times5+39\times14+51\times58+63\times59+75\times50+87\times14}{200} = \frac{12324}{200} = 61,62$
   $m(X^2) = \frac{\sum_i c_i^2\, n_i}{N} = \frac{27^2\times5+39^2\times14+51^2\times58+63^2\times59+75^2\times50+87^2\times14}{200} = \frac{797184}{200}$
   $Var(X) = m(X^2) - m(X)^2 = \frac{797184}{200} - \left(\frac{12324}{200}\right)^2 \simeq 188,9$
   Écart-type : $s(X) = \sqrt{Var(X)} \simeq 13,74$
3. On obtient les fréquences et fréquences cumulées suivantes :

Classes	[21 ; 33[	[33 ; 45[	[45 ; 57[	[57 ; 69[	[69 ; 81[	[81 ; 93[
Effectifs	5	14	58	59	50	14
Fréquences	0.025	0.07	0.29	0.295	0.25	0.07
Fréquences cumulées	0.025	0.095	0.385	0.68	0.93	1

d'où



4. Classe du premier quartile : [45 ; 57[
   $Q_1 \simeq x_i + \frac{x_{i+1}-x_i}{F(x_{i+1})-F(x_i)}(0,25 - F(x_i)) \simeq 45 + \frac{57-45}{0,385-0,095}(0,25 - 0,095) \simeq 51,41$ Il y a donc 25% des lycéens de l'échantillon dont le score est inférieur à 51,41 .
5. Comme n= 200 >30, on cherche z_α tel que $F(z_\alpha) = \frac{c+1}{2}$
   On lit sur la table inverse que pour la confiance 0.95 on a $z_\alpha \simeq 1.96$ d'où $a_\alpha = z_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 1.909$
   On estime donc que μ est dans l'intervalle $[61.62 - 1.909\,;\, 61.62 + 1.909] \simeq [59.7\,;\, 63.5]$, avec la confiance $c = 0.95$.
6. En supposant que $X \sim \mathcal{N}(60\,;\,15)$, on a :
   (a) $\mathbb{P}[45 \leq X < 57] = \mathbb{P}[\frac{45-60}{15} < \frac{X-60}{15} < \frac{57-60}{15}] = \mathbb{P}[-1,0 < Z < -0,2]$
   $= F(-0,2) - F(-1) \simeq (1 - 0,5793) - (1 - 0,8413)$
   $\simeq 0,4207 - 0,1587 \simeq 0,262$
   d'où l'effectif théorique $0,262 \times 200 \simeq 52,4$.

(b) $\mathbb{P}[57 \leq X < 69] = \mathbb{P}[\frac{57-60}{15} < \frac{X-60}{15} < \frac{69-60}{15}] = \mathbb{P}[-0,2 < Z < 0,6]$
$= F(0,6) - F(-0,2) \simeq 0,7257 - (1 - 0,5793)$
$\simeq 0,7257 - 0,4207 \simeq 0,305$
d'où l'effectif théorique $0,305 \times 200 \simeq 61$.

7. $\mathbb{P}[X \leq 51.41] = \mathbb{P}[\frac{X-60}{15} < \frac{51.41-60}{15}] \simeq \mathbb{P}[Z < -0.5727]$
$\simeq F(-0.5727) \simeq 1 - 0.7166 \simeq 0.2834$
Où l'on a utilisé une interpolation linéaire :
   - $F(0.5727) \simeq 0.7157 + \frac{0.719-0.7157}{0.58-0.57}(0.5727 - 0.57) \simeq 0.7166$.

2 Estime de soi et échec scolaire

1. moyenne : $m(X) = \frac{\sum x_i}{N} = \frac{46.7+54+\cdots+31.8}{12} = \frac{572.7}{12} \simeq 47.72$
$m(X^2) = \frac{\sum x_i^2}{N} = \frac{46.7^2+54^2+\cdots+31.8^2}{12} = \frac{29171.55}{12}$
$Var(X) = m(X^2) - m(X)^2 = \frac{29171.55}{12} - \left(\frac{572.7}{12}\right)^2 \simeq 153.29$
Écart-type : $s(X) = \sqrt{Var(X)} \simeq 12.38$
2. Comme $n = 12 \leq 30$, on cherche t_α à partir de la table inverse de Student avec $p = \frac{\alpha}{2} = 0.025$ et $n - 1 = 11$ degrés de liberté (ddl)
On lit $t_\alpha \simeq 2.201$ d'où $a_\alpha = t_\alpha \frac{s_e}{\sqrt{n-1}} \simeq 8.2163$.
On estime donc que μ est dans l'intervalle $[39.51; 55.94]$ avec la confiance $c = 0.95$
3. À partir de ces intervalles, on peut conclure, avec la confiance $c = 0.95$, que l'estime de soi moyenne est plus faible pour les lycéens en échec scolaire.

3 Traitement contre la douleur

1. On calcule les rangs :

Sujet	1	2	3	4	5	6	7	8	9	10
X	8	6	9	8	9	7	8	7	7	5
Y	5	6	7	5	9	5	7	10	5	3
X''	7	2	9,5	7	9,5	4	7	4	4	1
Y''	3,5	6	7,5	3,5	9	3,5	7,5	10	3,5	1
$(X'' - Y'')^2$	12,25	16	4	12,25	0,25	0,25	0,25	36	0,25	0

le coefficient de correlation des rangs de spearman est donc $1 - \left(6\,\frac{12,25+16+4+12,25+0,25+\cdots+0}{10\left(10^2-1\right)}\right) \simeq 0,506$
2. On calcule d'abord les moyennes, écart-types et covariance :
moyenne : $m(X) = \frac{\sum x_i}{N} = \frac{8+6+\cdots+5}{10} = \frac{74}{10} = 7,4 \quad m(X^2) = \frac{\sum x_i^2}{N} = \frac{8^2+6^2+\cdots+5^2}{10} = \frac{562}{10}$
$Var(X) = m(X^2) - m(X)^2 = \frac{562}{10} - \left(\frac{74}{10}\right)^2 = 1,44$
$m(Y) = \frac{\sum y_i}{N} = \frac{62}{10} = 6,2 \qquad m(XY) = \frac{\sum x_i\, y_i}{N} = \frac{471}{10} = 47,1$
$Cov(X,Y) = m(XY) - m(X)m(Y) = \frac{471}{10} - \frac{74}{10}\frac{62}{10} = 1,22$
on pose $a = \frac{Cov(X,Y)}{Var(X)} = \frac{1,22}{1,44} \simeq 0,847$ et $b = m(Y) - a\,m(X) \simeq 6,2 - 0,847 \times 7,4 \simeq -0,068$
D'où l'équation de la droite $D_{Y|X} : Y = 0,847\,X - 0,068$

4 Attention des écoliers

1. (a) Cas avec remise : $X \sim \mathcal{B}(23; 0,1)$.

(b) Cas sans remise :
Le nombre total d'écoliers dissipés est $4\,000\,000 \times \frac{10}{100} = 400\,000$
donc $X \sim \mathcal{H}(4\,000\,000; 400\,000; 23)$

2. Le coefficient d'exhaustivité est $\sqrt{\frac{4000000-23}{4000000-1}} \simeq 1$, donc $\mathcal{B}(23; 0, 1) \simeq \mathcal{H}(4\,000\,000; 400\,000; 23)$.

3. On ne peut pas utiliser d'approximation pour cette loi binomiale. On a
$$\begin{aligned}\mathbb{P}[X < 3] &= \mathbb{P}[X = 0] + \mathbb{P}[X = 1] + \mathbb{P}[X = 2] \\ &= \binom{23}{0}(1-0,1)^{23} + \binom{23}{1}(0,1)(1-0,1)^{23-1} + \cdots + \binom{23}{2}(0,1)^2(1-0,1)^{23-2} \\ &\simeq 0,089 + 0,226 \simeq 0,59\end{aligned}$$

4. La proportion expérimentale est $p_e = \frac{14}{250} = 0,056$. On a $n = 250 > 30$, et $n\,p_e = 14 > 5$, et $n(1-p_e) = 236 > 5$ donc on peut utiliser la pocédure du formulaire pour estimer la proportion p.
On lit sur la table inverse que pour la confiance 0,96 on a $z_\alpha \simeq 2,054$
d'où $a_\alpha = z_\alpha\sqrt{\frac{p_e(1-p_e)}{n}} \simeq 0,03$.
On estime donc que p est dans l'intervalle $[0,026; 0,086]$ avec la confiance $c = 0,96$.