Exercice 1 : (6 points).

1. **Définition de la population dans une situation statistique :**

   L'ensemble des individus observables.

2. **Indiquer les différentes natures de variable :**

   Les variables nominales, ordinales, discrète et continues

3. **Que représente les notations ?**

   (a) p : le nombre de modalités de Y.

   (b) y_3 : la valeur de Y observée sur le troisième individu [de l'échantillon].

   (c) $E_{m'_2}$: le sous-échantillon induit par la seconde modalité de Y ;
   ou : les individus ayant la seconde modalité comme valeur pour Y.

   (d) $X_{m'_3}$: la distribution de X conditionnée par la troisième modalité de Y.

4. **Notation de l'effectif total de la seconde ligne :** $n_{2.}$

5. **Définition d'une distribution conditionnelle de X ; combien peut-on en construire ?**

   C'est la distribution de X observés sur les individus ayant la même valeur pour Y ; on peut en construire autant qu'il y a de modalité de Y, p.

6. **Formule de la fréquence de la seconde modalité de Y_{m_4} :**

   La distribution Y_{m_4} est décrite par la 4ème ligne ; c'est donc $\frac{n_{42}}{n_{4.}}$

7. **Notation et formule de l'effectif de la modalité conjointe (m_1, m'_4) sous hypothèse d'indépendance de X et Y.**

   $\tilde{n_{14}} = \frac{n1.*n.4}{n}$

8. **Donner une propriété qui permette d'assurer que X et Y sont indépendantes.**

   Les distributions conditionnelles de X en fréquence sont égales ;
   ou : les distributions conditionnelles de Y en fréquence sont égales ;
   ou : les effectifs observés sont égaux aux effectifs théoriques ;
   ou : le χ^2 est nul.

Exercice 2 : (12 points)

1. **Combien peut-on construire de distributions de X à partir du tableau ?**

   5 : 4 conditionnelles plus 1 marginale.

2. **Signification et valeur des expressions suivantes :**

   (a) n_{12} : c'est l'effectif de la modalite conjointe (A ; –) ; il vaut 14.

   (b) f_{24} : c'est la fréquence de la modalite conjointe (B ; ++) ; elle vaut $\frac{n_{24}}{n} = \frac{45}{330} = 13,6\%$.

   (c) $n_{3.}$: la somme de la troisième ligne ; elle vaut $330 - 104 - 90 = 136$

   (d) $f_{.2}$: la fréquence de la deuxième colonne ; elle vaut $100 - 23,6 - 18,5 - 39,4 = 18,5\%$

   (e) $\tilde{n}_{31}$: effectif sous hypothèse d'indépendance (ou effectif théorique) de la modalité conjointe $(C, 1)$; $\tilde{n}_{31} = \frac{n_{3.}*n_{.1}}{n} = \frac{78*90}{330} = 32,15$

3. **Compléter le tableau de contingence et les marges.**

X / Y	$--$	$-$	$+$	$++$	Marge	en %
A	35	14	11	30	90	27,3
B	23	12	24	45	104	31,5
C	20	35	26	55	136	41,2
Marge	78	61	61	130		
en %	23,6	18,5	18,5	39,4		

4. **Donner la valeur de $\sum_{i=1,2\ j=2,4} n_{ij}$.**

   $n_{12} + n_{13} + n_{14} + n_{22} + n_{23} + n_{24} = 14 + 11 + 30 + 12 + 24 + 45 = 136$

5. **Les individus dont la valeur pour X est B sont-ils relativement plus nombreux dans E_+ ou dans E_{++} ?**

   Dans E_+ : $\frac{24}{61} = 39,3\%$ contre $\frac{45}{130} = 34,6\%$ dans E_{++}.

6. **Vérifier que les deux variables ne sont pas indépendantes.**

   l'effectif observé $n_{31} = 20$ est différent de l'effectif théorique correspondant $\tilde{n}_{31} = 32,15$

7. **Etudier l'indépendance statistique des deux variables par un test du χ^2 en utilisant le 95 ème centile de la distribution du χ^2 qui vaut 7,8 9,5 12,6 14,1 ou 21,0 selon que le d.d.l. est 3, 4, 6, 7 ou 12.**

   (a) d.d.l.$= (3-1) * (4-1) = 6$

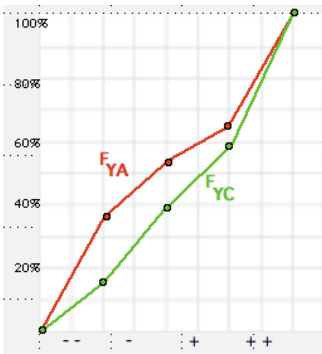
   (b) le seuil de rejet S est donc 12,6

   (c) calcul du χ^2
   $\tilde{n}_{11} = \frac{n_{78} * n_{90}}{330} = 21,27$ donc $\chi^2_{11} = \frac{(35-21,27)^2}{21,27} = 8,86$
   $\tilde{n}_{31} = 32,15$ donc $\chi^2_{31} = \frac{(20-32,15)^2}{32,15} = 4,59$
   conclusion $\chi^2 > \chi^2_{11} + \chi^2_{31} = 13,45 > S$: on rejette l'hypothèse d'indépendance statistique entre X et Y.

8. **Étude d'une liaison entre X et Y :**

   (a) **calculer les distributions conditionnelles cumulées de Y_A et Y_C et les représenter proprement sur la même grille.**

	$--$	$-$	$+$	$++$	Total
Y_A	35	14	11	30	90
cumul	35	49	60	90	
%	38,9	54,4	66,7	100,0	
Y_C	20	35	26	55	136
cumul	20	55	81	136	
%	14,7	40,4	59,6	100,0	



   (b) **tirer les conclusions de cette représentation concernant la liaison entre X et Y :**
   Comme $F_{Y_A} > F_{Y_A}$ on peut énoncer que $E_A <_Y E_C$: les individus ayant A pour X ont globalement des valeurs de Y inférieures aux individus ayant C pour X.

9. **Il s'agit de simuler 10 observations de la variable X en utilisant la série de chiffres au hasard suivante**

2 7 3 5 1 3 4 6 2 3 5 5 8 6 4 5 8 6 5 9 0 6 9 9 2 8 8 5 5 8 4 5 7 4 2 5 6 7 9 2 7 1 0 2 7

(a) **déterminer la population de simulation :** Elle compte 1000 individus ;

les 273 premiers notés de 1 à 273 ont l'étiquette A (le pourcentage est bien égal à $f_{1.}$)
les 315 suivants notés de 274 à 588 ont l'étiquette B (31,5%=$f_{2.}$) ;
les 412 derniers notés de 589 à 1000 ont l'étiquette C (41,2%=$f_{3.}$) ;

(b) **faire la simulation des 10 observations :**

273+1→B ; 513+1→B ; 462+1→B ; 355+1→B ; 864+1→C ;
586+1→B ; 590+1→C ; 699+1→C ; 288+1→B ; 558+1→B

(c) **construire la distribution de ces 10 observations :**

X	A	B	C	Total
effectif	0	7	3	10

Exercice 3 : (2 points) **Les rangs à deux épreuves passées par 12 individus sont donnés dans le tableau suivant :**

Calculer le τ de Kendall de cette série d'observations. On range les données par X croissant, et on calcul le bilan $cov_+ - cov_-$ pour chaque individu :

Individu	l	h	a	e	d	c	i	k	j	b	f	g	Total
Épreuve A	1	2	3	4	5	6	7	8	9	10	11	12	
Épreuve B	4	1	2	7	6	3	9	12	11	5	8	10	
Bilan	8-3	10-0	9-0	5-3	5-2	6-0	3-2	0-4	0-3	2-0	1-0		32

Et alors : $\tau = \frac{2*32}{12*11} = 0,48$