Exercice 1 : (7 points).

1. **Dans un tableau de contingence, que désignent les notations ?** n_{23}, f_{54}, $n_{.2}$ **et** $\tilde{n}_{32}$ **?**

   n_{23} : effectif [observé] de la modalité (m_2 ; m'_3) (situé au croisement de la 2ème ligne et de la 3ème colonne)
   f_{54} : proportion [observée] de la modalité (m_5 ; m'_4)
   $n_{.2}$: effectif de la 2ème modalité de Y ou la somme de la 2ème colonne
   $\tilde{n}_{32}$: effectif théorique de la modalité (m_3 ; m'_2)
   Calcul de f_{54} à partir des effectifs : $f_{54} = n_{54}/n$

2. **Qu'appelle-t-on moyenne conditionnelle de X, dans le contexte d'une variable conjointe XxY ?**

   La moyenne de X pour les individus ayant même modalité pour Y ; ou autre ...

3. **Définition de la variance inter de Y, et formule dans le cas où X a 4 modalités.**

   C'est la variance des moyennes conditionnelles de Y.
   var inter $= \frac{\sum_{i=1,4} n_{i.} * (\overline{Y_i} - \overline{Y})^2}{n}$ ou $\frac{n_{1.} * \overline{Y_1}^2 + n_{2.} * \overline{Y_2}^2 + n_{3.} * \overline{Y_3}^2 + n_{4.} * \overline{Y_4}^2}{n} - \overline{Y}^2$

4. **Construire un exemple de tableau de contingence avec 3 modalités pour X et 4 pour Y, où la covariance a un sens. Donner l'expression de cette covariance (le calcul n'est pas obligatoire).**

   Les variables doivent être quantitatives, discrète ou continue.
   On note c_i et c'_j les centres des classes de X et Y.
   $Cov(X,Y) = \frac{\sum_{i=1,3}^{j=1,4} n_{ij} * (c_i - \overline{X}) * (c'_j - \overline{Y})}{n}$ ou
   moyenne des produits - produit des moyennes $= \frac{\sum_{i=1,3}^{j=1,4} n_{ij} * c_i * c'_j}{n} - \overline{X} * \overline{Y}$

5. $F'(0.3)$ **vaut 0.382 pour la variable gaussienne** ($P(Z > 0.3) = 0.382$)**, et** $F'(0.6)$ **vaut 0.274. Quelle est la proportion des individus dont la valeur appartient aux intervalles** $]-\infty\,;\,0.3[$, $]-0.6\,;\,0.3[$ **et** $]-0.6\,;\,-0.3[$ **?**

   $P(Z < 0.3) =_{cplt} 1 - P(Z > 0.3) = 1 - 0.382 = 0.618$
   $P(-0.6 < Z < 0.3) = P(Z < 0.3) - P(Z < -0.6) = P(Z < 0.3) - P(Z > 0.6) = 0.618 - 0.274 = 0.344$
   $P(-0.6 < Z < -0.3) =_{sym} P(0.3 < Z < 0.6) = P(Z > 0.3) - P(Z > 0.6) = 0.382 - 0.274 = 0.108$

Exercice 2 : (3 points).

Dans cet exercice, les mots sont formés avec l'alphabet A = {a,b,c,d,e,f,g}.

1. **Combien y a-t-il de mots à 5 lettres toutes différentes ?**

   Le nombre de 5-énumération de 7 $= A_7^5 = 2520$

2. **Combien y a-t-il de mots à 5 lettres ayant au moins une lettre avec plusieurs occurrences ?**

   Le nombre de mots à 5 lettres diminué du nombre de mots à 5 lettres toutes différentes : $7^5 - A_7^5 = 16807 - 2520 = 14287$

3. **Combien y a-t-il de mots à 6 lettres ayant exactement 3 a ?**

   On place d'abord les 3 a dans une des parties de 3 parmi 6 : il y a $\binom{6}{3} = 20$ possibilités ; puis on remplit les 3 places restantes avec un mot sans a : $(7-1)^3 = 216$ possibilités ; au total on trouve 20*216=4320 mots.

4. **Combien y a-t-il de mots à 6 lettres ayant exactement 3 a et au moins 1 b ?**

   Nombre de mots ayant 3a et 1 b : $\binom{6}{3} * \binom{3}{1} * (7-2)^2 = 20 * 3 * 25 = 1500$
   Nombre de mots ayant 3a et 2 b : $\binom{6}{3} * \binom{3}{2} * (7-2) = 20 * 3 * 5 = 300$
   Nombre de mots ayant 3a et 3 b : $\binom{6}{3} * \binom{3}{3} = 20 * 1$
   Au total on trouve 1500+ 300+20=1820 mots

Exercice 3 : (5 points).

Il s'agit d'effectuer un test du χ^2 du tableau de contingence suivant, portant sur le vote par groupe d'âges de 1280 individus d'un échantillon ; on prendra comme seuil de décision le 95ème centile du χ^2 égal à 12,6 pour ce tableau.

X=Candidat - Y=Âge	18-35	35-50	50-80
A	127	113	140
B	98	84	98
C	101	113	241
D	57	41	67

1. **Quel degré de liberté a-t-on choisi pour calculer le seuil ?**

   (4-1)*(3-1)=6

2. **Formule et valeur de l'effectif de la modalité (C ; 18-35) si X et Y sont indépendantes.**

   $\tilde{n}_{31} = \frac{n_{3.} * n_{.1}}{n} = \frac{455*383}{1280} = 136.14$

3. **Formule et valeur du taux de liaison t_{13}.**

   $t_{13} = \frac{(n_{13} - \tilde{n}_{13})}{\sqrt{\tilde{n}_{13}}}$; comme $\tilde{n}_{13} = \frac{380*546}{1280} = 162.09$, on a
   $t_{13} = \frac{(140-162.09)}{\sqrt{162.09}} = -1,73$

4. **Faire le test en indiquant précisément les conclusions auxquelles vous aboutissez.**

   $\chi^2_{11} = \frac{(127-113.70)^2}{113.70} = 1.55$ car $\tilde{n}_{11} = \frac{383*380}{1280} = 113.70$
   $\chi^2_{13} = t^2_{13} = 3.01$
   $\chi^2_{31} = \frac{(101-136.14)^2}{136.14} = 9.07$
   Si bien que :
   $\chi^2 \geq \chi^2_{11} + \chi^2_{13} + \chi^2_{31} = 1.55 + 3.01 + 9.07 \geq 12,6$
   On décide donc de rejeter l'hypothèse d'indépendance de X et Y [avec un risque d'erreur de 5%

Exercice 4 : (5 points).

On a mesuré une variable continue X sur 4 groupes disjoints d'individus, et on a obtenu les résultats suivants :

	Taille du groupe	Moyenne	Variance
Groupe 1	380	46.4	300.9
Groupe 2	280	45.6	298.4
Groupe 3	455	52.2	294.0
Groupe 4	165	47.1	320.8

ainsi que la variance des moyennes égale à 8.2.

1. **Calculez la moyenne globale de X.**

   D'après la formule de décomposition de la moyenne, la moyenne globale est la moyenne des moyennes conditionnelles pondérée par la distribution de Y :
   $\overline{X} = \frac{\sum_{j=1,4} n_{.j} * \overline{X_j}}{n} = \frac{380*46.4+280*45.6+455*52.2+165*47.1}{1280} = 48.38$

2. **Calculez la variance globale de X.**

   Variance globale = variance inter + variance intra

   Variance inter = variance des moyennes conditionnelles = 8.2

   Variance intra = moyenne des variances conditionnelles pondérée par la distribution de Y :

   var inter = $\frac{\sum_{j=1,4} n_{.j} * var(X_j}{n} = \frac{380*300.9+280*298.4+455*294.0+165*320.8}{1280} = 300.47$

   Au total :

   Variance globale = 8.2 + 300.5 = 308.7

3. **Faire l'analyse de la variance de X, avec le dernier centile de la distribution du χ^2 égal à 11,35. Vous préciserez le degré de liberté utilisé, et vous indiquerez précisément les conclusions auxquelles vous aboutissez.**

   On teste l'égalité des moyennes conditionnelles : on sait que si c'est le cas, le nombre (1280-4)*var-inter/var-intra suit une loi du χ^2 à 3 degrés de liberté.

   Pour cet échantillon, ce nombre est égal à 1276*8.2/300,47=34.8 ; comme il est supérieur à 11.35 on rejette l'hypothèse d'égalité des moyennes conditionnelles en considérant que l'appartenance à un groupe a un effet sur la mesure de X.