LES TESTS STATISTIQUES

1. PROBLEMATIQUE

Tests statistiques = Comparaison de 2 ou plusieurs groupes

Comparaison de fréquences d’une maladie

Région I = Fréquence f_1

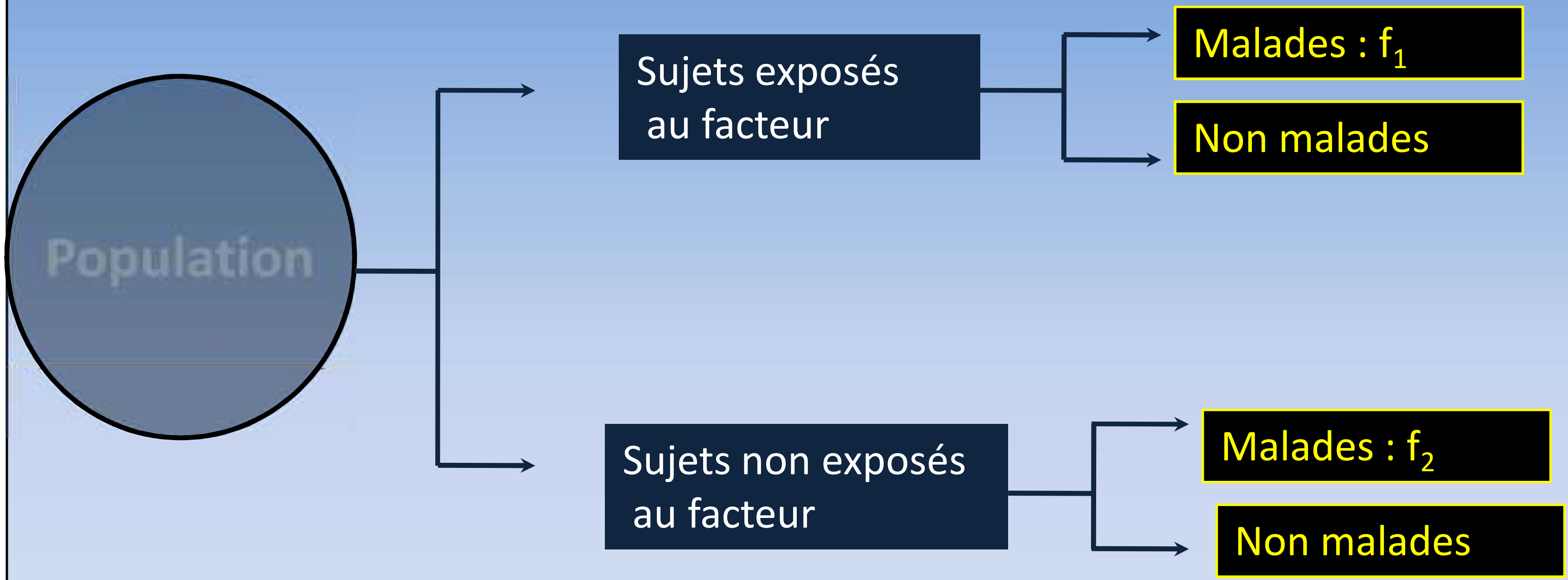
Région II = Fréquence f_2

Période I = Fréquence f_1

Période II = Fréquence f_2

Comparer les deux fréquences f_1 et f_2

Etude de la relation entre un facteur étiologique et une maladie



Comparer les deux fréquences f_1 et f_2

Comparer 2 ou plusieurs groupes

=

Etudier la relation (la liaison) entre 2 variables

**Répartition de 176 enfants
selon la qualité de l'eau de boisson
et la diarrhée**

Diarrhée / Eau	Mauvaise	Bonne	Total
Oui	54	17	**71**
Non	55	50	**105**
Total	**109**	**67**	**176**

2 Variables a tester :

- La diarrhée

- La qualité de l'eau

Comparer deux pourcentages :

- Pourcentage d'enfants diarrhéiques ayant une eau de mauvaise qualité : $P_1 = 54 / 109 = 49,5$ %

- Pourcentage d'enfants diarrhéiques ayant une eau de bonne qualité : $P_2 = 17 / 67 = 25,4$ %

A priori : P_1 [P_2

Mais : La différence peut être due au hasard

Test statistique

Variable à plus de 2 modalités

Proportion de dysenterie selon la profession dans une épidémie survenue dans une localité de 8900 habitants

Profession	Population	Nbre de cas	%
Pécheurs	900	40	**4,4**
Agriculteurs	1800	15	**0,83**
Autres	6200	29	**0,47**

Profession

Pourcentage plus élevé chez les pécheurs

Test statistique pour confirmation

Taux d'uricémie et diabète

Uricémie	**Diabétiques**	**Non diabétiques**
Effectif	n_1	n_2
Moyenne	M_1	M_2
Variance	$s^2{}_1$	$s^2{}_2$

Variable qualitative : Diabète
Variable quantitative : Uricémie

Liaison entre deux variables = Liaison jamais constante

TEST STATISTIQUE

- Homogénéité entre deux ou plusieurs groupes : Comparaison
- Conformité de distributions : Ajustement

2. METHODE DU TEST D'HYPOTHESE

Liaison entre deux variables = Peut être due au hasard
(fluctuations d'échantillonnage)

Comparaison de deux traitements A et B chez 100 malades

Traitement	**A**		**B**	
Résultat	Effectif	%	Effectif	%
Succès	43	79,6	31	67,4
Echec	11	20,4	15	32,6
Total	54	**100**	46	100

% de succès avec A (P_A) > % de succès avec B (P_B)

Hasard ????

Test d'hypothèse :

- Proportions de succès avec A et B identiques ou différentes ?
- Existe –il une liaison entre le succès et le traitement

QUANTIFIER LE ROLE DU HASARD

Hypothèse nulle :

- Hypothèse d'absence de liaison entre les deux variables
- Hypothèse d'indépendance des deux variables

- ❖ **1ère étape du test**
- ❖ **Vérifiable après le test**

H_0 :

- Le traitement et le résultat produit sont indépendants
- Il n'existe pas de relation entre le traitement et le résultat produit
- Le résultat du traitement A n'est pas différent du résultat du traitement B
- Il n ya pas de différence entre les résultats des deux trainements

$$H_0 : P_A = P_B$$

Test d'hypothèse = Probabilité p

p = Probabilité que le hasard puisse expliquer les résultats

- **Si p á Seuil : Différence statistiquement significative**
  Hypothèse nulle rejetée
- **Si p > Seuil : Pas de Différence statistiquement significative**
  Hypothèse nulle retenue

Seuil de signification habituel = 5%

p = Probabilité de rejeter à tort une hypothèse nulle

= Probabilité de conclure une différence qui n'existe pas

(due au simple hasard)

3. ETAPES DU TEST D'HYPOTHESE

1. Position du problème et choix du test statistique
2. Formulation de l'hypothèse nulle
3. Vérification des conditions d'application
4. Application du test
5. Résultats
6. Conclusion

4. LIAISON ENTRE DEUX VARIABLES QUALITATIVES

4.1. Comparaison d'un pourcentage observé à un pourcentage théorique :

Soit un pourcentage théorique P
On observe sur un échantillon de taille n, un pourcentage P_0

Le pourcentage observé P_0 diffère-t-il du pourcentage théorique P ?

- e = z s
- e = | P – P_0 |
- p = 5%
- z = 1,96
- e = 1,96 s
- Plus p est élevé (z diminue) ; moins e est large

- z = e/s
- z = 1,96 ; IC à 95 %

- La valeur de z suit celle de e

- Calculer la différence absolue entre les deux pourcentages :

$$| P - P_0 |$$

- Calculer l'écart-type s du pourcentage théorique P :

$$s = \sqrt{P Q / n}$$

- Calculer l'écart-réduit z

$z = | P - P_0 | / s$

$$z = \frac{| P - P_0 |}{\sqrt{P Q / n}}$$

Si $z \geq 1{,}96$, Différence statistiquement significative

Si $z < 1{,}96$, Différence statistiquement non significative

Seuil de signification : Table de l'écart-réduit

- Conditions d'application :

$$nP \text{ et } nQ \geq 5$$

- Utilisation des fréquence absolues :

$$s = \sqrt{nPQ}$$

- Calculer l'écart-réduit z :

$z = |x - x_0| / s$

$$z = \frac{|x - x_0|}{\sqrt{n\,P\,Q}}$$

Exemple :

Une variété de souris présente un cancer avec une fréquence de 20 %
Après l'administration d'une certaine substance chez un lot de 100 souris, on observe un pourcentage de cancers de 14 %
Cette substance a-t-elle un effet sur le cancer ?

$P = 0,2$
$Q = 1 – P = 0,8$
$n = 100$
$P_0 = 0,14$

1. Position du problème et choix du test statistique :

Il s'agit de la comparaison d'un pourcentage observé à un pourcentage théorique

2. Hypothèse nulle :

La substance en question n'a pas d'effets sur le cancer

$$H_0 : P = P_0$$

$$H_0 : \; z = \frac{| P - P_0 |}{\sqrt{P Q / n}} < 1{,}96$$

3. Vérification des conditions d'application :

n p = 100 x 0,2 = 20

n q = 100 x 0,8 = 80

4. Application :

$$z = \frac{| P - P_0 |}{\sqrt{P\,Q / n}}$$

$$z = \frac{| 0{,}2 - 0{,}14 |}{\sqrt{0{,}2 \times 0{,}8 / 100}} = 1{,}5$$

5. Résultat :

$z = 1{,}5 < 1{,}96$

Il n y a pas de différence statistiquement significative

L'hypothèse nulle est retenue

6. Conclusion :

La substance en question n'a pas d'effets sur le cancer

- Utilisation des fréquence absolues :

  $x = 20$

  $x_0 = 14$

$$z = \frac{|20 - 14|}{\sqrt{0{,}2 \times 0{,}8 \times 100}} = 1{,}5$$

4.2. Comparaison de deux pourcentages observés

Soit un pourcentage P_A observé sur un échantillon n_A
et
un pourcentage P_B observé sur un échantillon n_B

Le pourcentage observé P_A diffère-t-il du pourcentage observé P_B ?

- Calculer la différence absolue entre les deux pourcentages :

$$| P_A – P_B |$$

- Calculer l'écart-type s des deux pourcentages :

$$s = \sqrt{} P Q / n_A + P Q / n_B$$

P et Q : Proportions estimés sur l'ensemble de deux échantillons

$P = x_A + x_B / n_A + n_B$

- Calculer l'écart-réduit z :

$z = | P_A – P_B | / s$

$$z = \frac{| P_A – P_B |}{\sqrt{} P Q / n_A + P Q / n_B}$$

Si $z \geq 1{,}96$, Différence statistiquement significative

Si $z < 1{,}96$, Différence statistiquement non significative

Seuil de signification : Table de l'écart-réduit

- Conditions d'application :

$$n_A P\ ,\ n_A Q,\ n_B P \text{ et } n_B Q \geq 5$$

Exemple :

On sélectionne deux groupes de 200 et de 400 sujets qui reçoivent respectivement deux traitements A et B. Les pourcentages de succès respectifs sont de 30 % et de 45 %.
L'action des deux traitements est-elle différente?

$n_A = 200$
$n_B = 400$
$P_A = 0{,}30$
$Q_A = 1 - P_A = 0{,}70$
$P_B = 0{,}45$
$Q_B = 1 - P_B = 0{,}55$

Estimation de P et de Q :

$x_A = 0{,}30 \times 200 = 60$
$x_B = 0{,}45 \times 400 = 180$
$P = x_A + x_B / n_A + n_B = 60 + 180 / 200 + 400 = 0{,}4$

$P = 0{,}4$
$Q = 0{,}6$

1. Position du problème et choix du test statistique :

Il s'agit de la comparaison de deux pourcentages observés

2. Hypothèse nulle :

L'action des deux traitements est identique

$$H_0 : P_A = P_B$$

$$H_0 :$$

$$z = \frac{| P_A - P_B |}{\sqrt{PQ / n_A + P Q / n_B}} < 1,96$$

3. Vérification des conditions d'application :

$n_A\ p = 200 \text{ x } 0{,}4 = 80$

$n_A\ q = 200 \text{ x } 0{,}6 = 120$

$n_B\ p = 400 \text{ x } 0{,}4 = 160$

$n_B\ q = 400 \text{ x } 0{,}6 = 240$

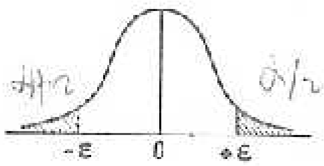
4. Application :

$$z = \frac{|\,P_A - P_B\,|}{\sqrt{P\,Q / n_A + P\,Q / n_B}}$$

$$z = \frac{|\,0{,}35 - 0{,}40\,|}{\sqrt{0{,}4 \text{ x } 0{,}6 / 200 + 0{,}4 \text{ x } 0{,}6 / 400}} = 3{,}5$$

Table de l'écart-réduit (loi normale) (*)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour $\varepsilon = 1{,}960$ la probabilité est $\alpha = 0{,}00 \quad 0{,}05 = 0{,}05$.

Table pour les petites valeurs de la probabilité

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
ε	3,29053	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

(*) D'après Fisher et Yates, Statistical tables for biological, agricultural, and medical research (Oliver and Boyd, Edinburgh) avec l'aimable autorisation des auteurs et des éditeurs.

5. Résultat :

$z = 3{,}5 > 1{,}96$

L'hypothèse nulle est rejetée

La différence est statistiquement significative

Le seuil de signification < 1 p. 1000

6. Conclusion :

L'action des deux traitement est différente (p < 1 p. 1000)

5. LIAISON ENTRE UNE VARIABLE QUALITATIVE ET UNE VARIABLE QUANTITATIVE

5.1. Comparaison d'une moyenne observée à une valeur théorique :

On observe sur un échantillon de taille n,
une moyenne m_0 avec un écart-type s;
qu'on doit comparer à une valeur théorique m

La moyenne observée m_0 diffère-t-elle de la valeur théorique m ?

- Calculer l'écart-réduit z

$$z = \frac{|m - m_0|}{s / \sqrt{n}}$$

Si $z \geq 1,96$, Différence statistiquement significative
Si $z < 1,96$, Différence statistiquement non significative

Seuil de signification : Table de l'écart-réduit

- Conditions d'application :

  - $n \geq 30$

Exemple :

Sur un échantillon de 250 sujets, on observe une glycémie moyenne de 1,17 g/l avec un écart-type de 0,32 g/l.
Cette mesure doit être considérée comme normale en adoptant la valeur théorique de 1,26 g/l ?

$n = 250$
$m_0 = 1,17$
$s = 0,32$
$m = 1,26$

1. Position du problème et choix du test statistique :

Il s'agit de la comparaison d'une moyenne observée à une valeur théorique

2. Hypothèse nulle :

La mesure observée est considérée comme normale en adoptant une valeur théorique de 1,26 g/l

$$\mathbf{H_0 : m_0 = m}$$

$\mathbf{H_0}$ **:**

$$z = \frac{| m - m_0 |}{s / \sqrt{n}} < 1{,}96$$

3. Vérification des conditions d'application :

n = 250 > 30

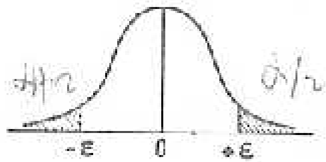
4. Application :

$$z = \frac{| m - m_0 |}{s / \sqrt{n}}$$

$$z = \frac{| 1{,}26 - 1{,}17 |}{0{,}32 / \sqrt{250}} = 4{,}5$$

Table de l'écart-réduit (loi normale) (*)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour $\varepsilon = 1{,}960$ la probabilité est $\alpha = 0{,}00 \quad 0{,}05 = 0{,}05$.

Table pour les petites valeurs de la probabilité

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
ε	3,29053	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

(*) D'après Fisher et Yates, Statistical tables for biological, agricultural, and medical research (Oliver and Boyd, Edinburgh) avec l'aimable autorisation des auteurs et des éditeurs.

5. Résultat :

$z = 4,5 > 1,96$

L'hypothèse nulle est rejetée

La différence est statistiquement significative

Le seuil de signification < 1 p. 100000

6. Conclusion :

La mesure observée n'est pas considérée comme normale
en adoptant une valeur théorique de 1,26 g/l ($p < 1$ p. 1 00000)

5.2. Comparaison de deux moyennes observées

Soit une moyenne m_A observée sur un échantillon $n_{A,}$ avec une variance $s^2{}_A$
et
une moyenne m_B observée sur un échantillon $n_{B,}$ avec une variance $s^2{}_B$

La moyenne observée m_A diffère-t-elle de la moyenne observée m_B ?

- Calculer la différence absolue entre les deux moyennes :

$$| m_A - m_B |$$

- Calculer l'écart-type s des deux moyennes :

$$s = \sqrt{s^2_A / n_A + s^2_B / n_B}$$

- Calculer l'écart-réduit z

$z = | m_A - m_B | \ / s$

$$z = \frac{| m_A - m_B |}{\sqrt{s^2_A / n_A + s^2_B / n_B}}$$

Si $z \geq 1{,}96$, Différence statistiquement significative

Si $z < 1{,}96$, Différence statistiquement non significative

Seuil de signification : Table de l'écart-réduit

- Conditions d'application :

$$n_A \text{ et } n_B \geq 30$$

Exemple :

Le taux moyen de cholestérol sanguin chez un groupe de 150 diabétiques est de 1,46 g/l avec un écart-type de 0,28 g/l. Celui d’un groupe de 300 sujets non diabétiques est de 1,45 g/L avec un écart-type de 0,25 g/l.

Existe-il une relation entre l’hypercholestérolémie et le diabète?

$n_A = 150$

$n_B = 300$

$m_A = 1,46$

$s_A = 0,28$

$m_B = 1,45$

$s_B = 0,25$

1. Position du problème et choix du test statistique :

Il s'git de la comparaison de deux moyennes observées

2. Hypothèse nulle :

Il n y a pas de relation entre le diabète et l'hypercholestérolémie

$$H_0 : m_A = m_B$$

H_0 :

$$z = \frac{| m_A - m_B |}{\sqrt{s^2_A / n_A + s^2_B / n_B}} < 1,96$$

3. Vérification des conditions d’application :

$n_A = 150 > 30$

$n_B = 300 > 30$

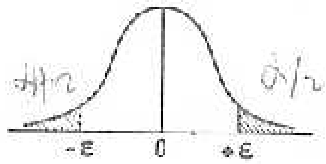
4. Application :

$$z = \frac{|m_A - m_B|}{\sqrt{s^2_A / n_A + s^2_B / n_B}}$$

$$z = \frac{|1{,}46 - 1{,}45|}{\sqrt{(0{,}28)^2 / 150 + (0{,}25)^2 / 300}} = 3{,}8$$

Table de l'écart-réduit (loi normale) (*)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour $\varepsilon = 1{,}960$ la probabilité est $\alpha = 0{,}00 + 0{,}05 = 0{,}05$.

Table pour les petites valeurs de la probabilité

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
ε	3,29053	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

(*) D'après Fisher et Yates, Statistical tables for biological, agricultural, and medical research (Oliver and Boyd, Edinburgh) avec l'aimable autorisation des auteurs et des éditeurs.

5. Résultat :

$z = 3{,}8 > 1{,}96$

L'hypothèse nulle est rejetée

La différence est statistiquement significative

Le seuil de signification < 1 p. 1000

6. Conclusion :

Il y a une relation entre le diabète et l'hypercholestérolémie ($p <$ 1 p. 1000)