

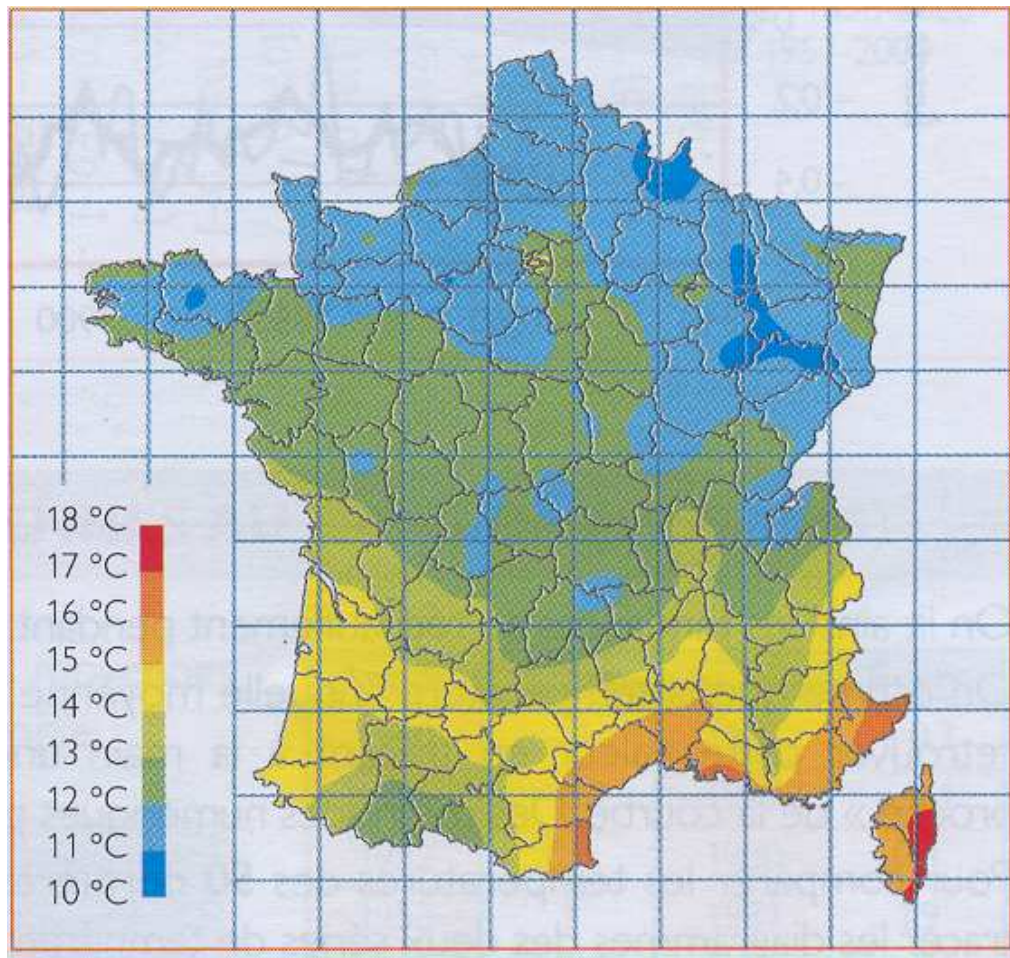
1 Thème 1 : Réchauffement de la planète ?

La question de l'élévation de la température du globe donne lieu à un débat scientifique car les variations de température :

- ne sont pas les mêmes en différents points du globe,
- sont très irrégulières et dépendent de multiples facteurs connus ou inconnus.

1.1 Recherche d'une température moyenne

Exemple : calcul de la température moyenne de la France.



Pour calculer la température moyenne de la France, on peut utiliser un quadrillage régulier : on repère la température à chaque nœud du quadrillage et on calcule la moyenne des températures obtenues.

Choisir deux sous-quadrillages différents et comparer les résultats obtenus.

Réponse :

Quadrillage le plus fin :

Températures : 11 – 11 – 12 – 11 – 11 – 10 – 11 – 11 – 11 – 11 – 11 – 11 – 12 – 11 – 10 – 11
 12 – 12 – 12 – 12 – 12 – 11 – 10 – 10 – 12 – 12 – 12 – 12 – 11 – 12 – 13 – 12 – 11 – 12 – 12
 12 – 13 – 14 – 13 – 13 – 12 – 13 – 13 – 14 – 14 – 14 – 14 – 14 – 15 – 14 – 15 – 13 – 12 – 13
 15 – 15 – 15 – 17

Moyenne calculée = 12,3 °C

Quadrillage moins fin : (4 carrés réunis)

Températures : 11 – 11 – 11 – 11 – 11 – 11 – 12 – 12 – 11 – 14 – 13 – 13 – 14 – 13 – 13 – 15

Moyenne calculée = 12,2 °C

Les moyennes calculées sont proches.

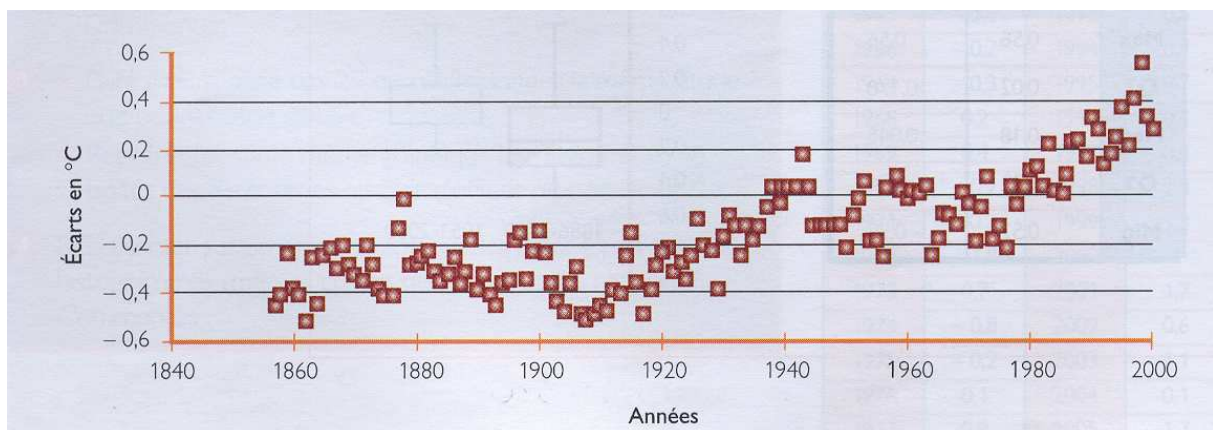
Cependant plus le maillage est fin et plus la moyenne calculée est fiable.

1.2 Etudes des températures moyennes sur la terre

Le graphique ci-dessous (dont le zéro est fixé arbitrairement), résulte de l'analyse de milliers d'observations.

Malgré une grande variabilité des résultats, une évolution ascendante se distingue.

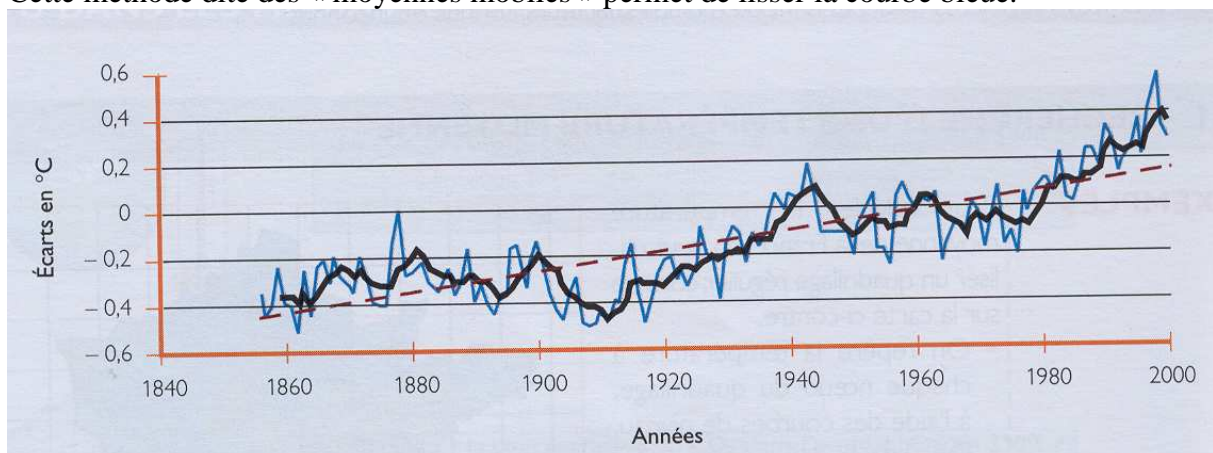
ECARTS DE LA TEMPERATURE MOYENNE DU GLOBE (1856-2000)



On peut rendre plus lisible l'évolution en reliant les points. Cependant, la courbe obtenue (bleue) est très irrégulière.

La courbe noire est tracée à partir des points obtenus en calculant la moyenne des températures des 5 années précédentes.

Cette méthode dite des « moyennes mobiles » permet de lisser la courbe bleue.

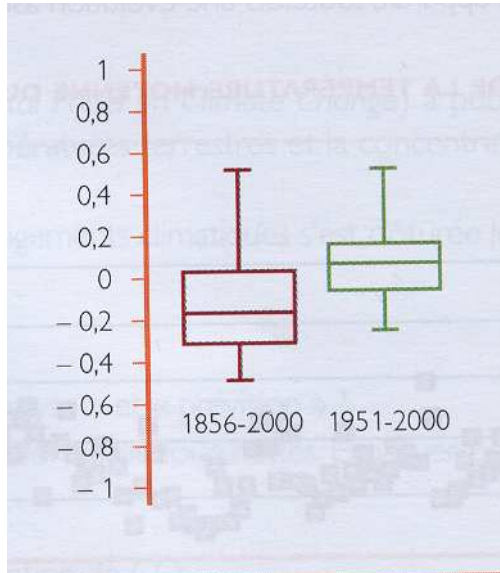


Enfin, on peut tracer la droite (en pointillés rouge) qui semble être « la plus proche » de la courbe. (Des méthodes numériques permettent d'obtenir une telle droite).

On peut retrouver ainsi à partir de cette droite que la température annuelle moyenne s'est élevée d'environ 0,6°C entre 1860 et 2000.

Pour comparer, les températures des **50 dernières années** à celles de l'**ensemble de la période**, on peut tracer les diagrammes en boîtes des deux séries.

La différence des deux séries est nettement marquée : les températures sont sensiblement plus élevées dans la dernière période.



1.3 Etude des températures moyennes au Canada (1950-2005)

Le tableau ci-dessous donne les écarts des températures moyennes annuelles par rapport à une température de référence choisie.

Année	Ecart	Année	Ecart
1950	-1	1978	-0,3
1951	-0,2	1979	-0,1
1952	0,9	1980	0,4
1953	1	1981	1,9
1954	0,2	1982	-0,9
1955	0,1	1983	0,2
1956	-0,5	1984	0,3
1957	0	1985	0
1958	0,8	1986	0,2
1959	-0,3	1987	1,4
1960	0,6	1988	0,8
1961	-0,1	1989	-0,2
1962	0,1	1990	0
1963	0,3	1991	0,5
1964	-0,5	1992	-0,1
1965	-0,6	1993	0,5
1966	-0,2	1994	0,4
1967	-0,3	1995	0,7
1968	0,2	1996	0,1
1969	0,4	1997	0,4
1970	-0,2	1998	2,4
1971	0	1999	1,8
1972	-1,8	2000	1
1973	0,7	2001	1,7
1974	-0,8	2002	0,6
1975	-0,2	2003	1,1
1976	0,1	2004	0,1

1977	0,9	2005	1,7
------	-----	------	-----

- 1) a) Représenter graphiquement le nuage de points.
Commenter
- b) Tracer une droite « proche des points ».
En déduire une estimation de la variation de température.

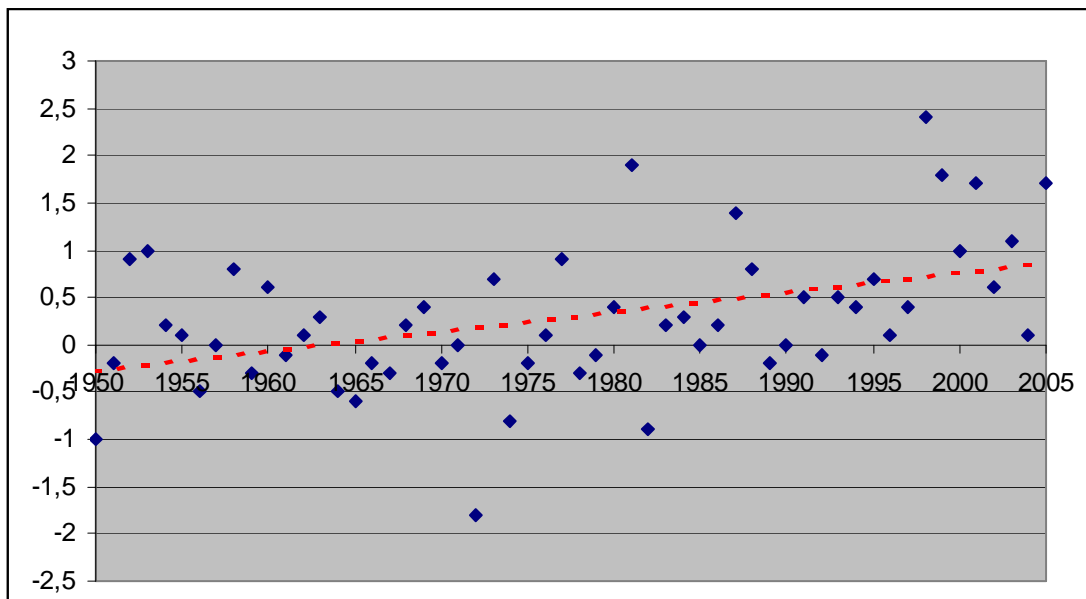
- 2) En utilisant les fonctions statistiques de la calculatrice ou du tableur :
 - a) Déterminer la médiane, la moyenne et les quartiles de la série.
Comparer médiane et moyenne. Commenter.
 - b) Représenter la série par un diagramme en boîte.
Est-il symétrique ? Proposer une explication.
 - c) Déterminer les déciles D1 et D9. Représenter la série par un diagramme en boîte limité par ces déciles. Est-il symétrique ? Commenter.

- 3) a) Faire avec la série des 20 dernières années la même étude que pour la série globale.
b) Représenter sur le même graphique les diagrammes en boîtes des deux séries (choisir d'utiliser ou non les déciles).

- 4) Représenter sur un même graphique les deux séries par des histogrammes avec des classes de températures d'amplitude 0,1°C. Commenter.

Réponse :

1) a) b)



Equation de la droite d'ajustement affine : $y = 0,020588x - 40,4232$

Estimation de la température en 1950 : $0,020588 \times 1950 - 40,4232 = -0,28 \text{ } ^\circ\text{C}$

Estimation de la température en 2005 : $0,020588 \times 2005 - 40,4232 = 0,855 \text{ } ^\circ\text{C}$

Ecart de température entre 1985 et 2005 : $0,855 + 0,28 = 1,135 \text{ } ^\circ\text{C}$

Soit une augmentation annuelle moyenne de : $1,135/55 \approx 0,021 \text{ } ^\circ\text{C/an}$

2) a) Médiane = 0,2

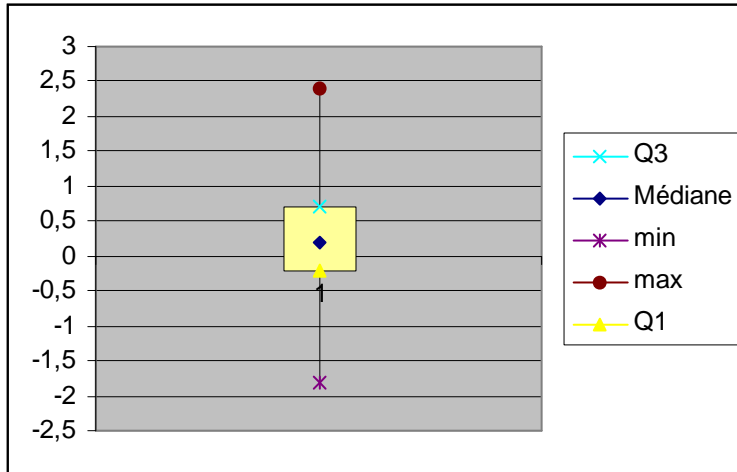
Moyenne = 0,28

Q1 = -0,2

$Q3 = 0,7$

La moyenne est sensiblement supérieure à la médiane (+ 40%) car des températures plus élevées sont concentrées sur les 30 dernières années.

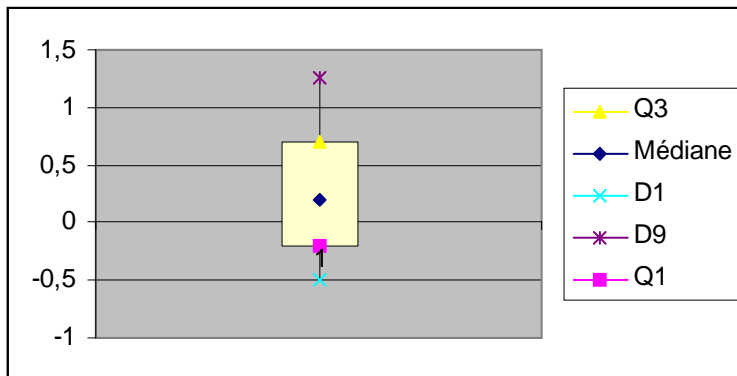
b)



Le diagramme est symétrique.

Les valeurs extrêmes minimum et maximum sont symétriques par rapport à 0.

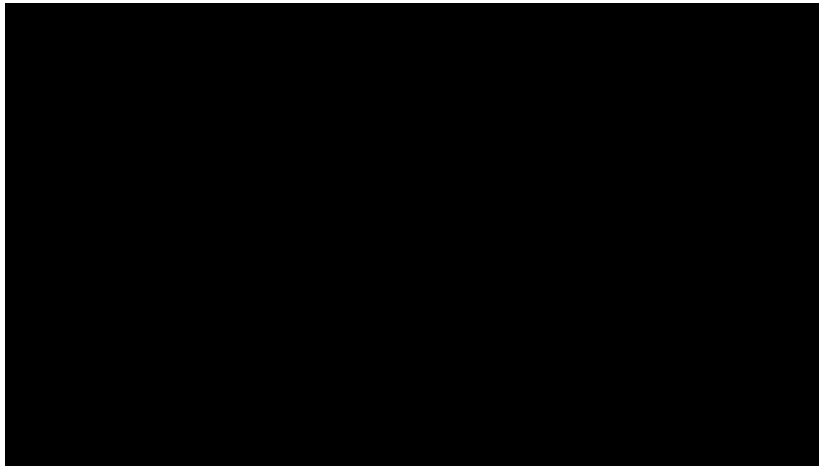
c) $D1 = -0,5$ et $D9 = 1,25$



Le diagramme n'est pas symétrique.

Les valeurs proches du neuvième décile sont plus proches d'une augmentation de $+1,3^{\circ}\text{C}$ Alors que les valeurs proches du premier décile sont proches d'une diminution de $-0,5^{\circ}\text{C}$. Ceci confirme une augmentation perceptible des températures.

3)



Equation de la droite d'ajustement affine : $y = 0,05312x - 105,249$

Estimation de la température en 1985 : $0,05312 \times 1985 - 105,249 = 0,188 \text{ } ^\circ\text{C}$

Estimation de la température en 2005 : $0,05312 \times 2005 - 105,249 = 1,250 \text{ } ^\circ\text{C}$

Ecart de température entre 1985 et 2005 : $1,25 - 0,188 = 1,062$

Soit une augmentation annuelle moyenne de : $1,062/20 \approx 0,053 \text{ } ^\circ\text{C/an}$

L'augmentation moyenne annuelle entre 1985 et 2005 est 2,5 fois plus importante que celle entre les années 1950 et 2005.

Médiane : 0,5

Moyenne : 0,72

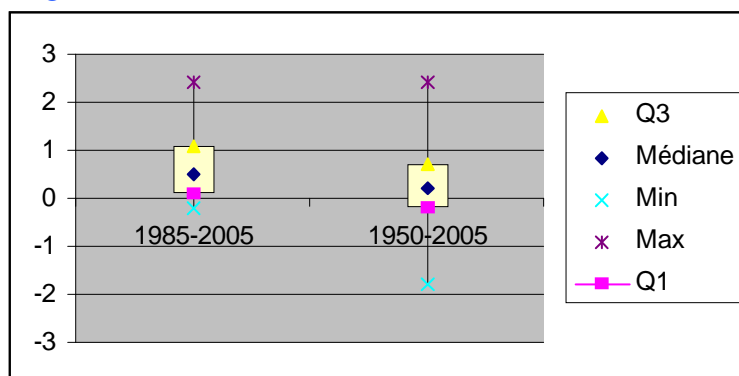
Q1 : 0,1

Q3 : 1,1

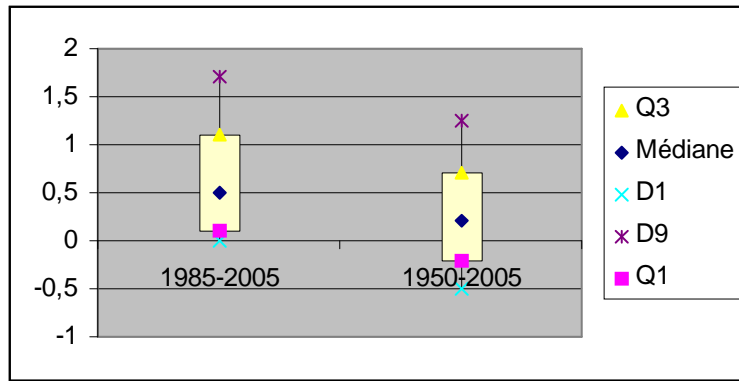
D1 : 0

D9 : 1,7

Diagrammes en boîtes avec le min-max comme bornes :



Diagrammes en boîtes avec les déciles comme bornes :



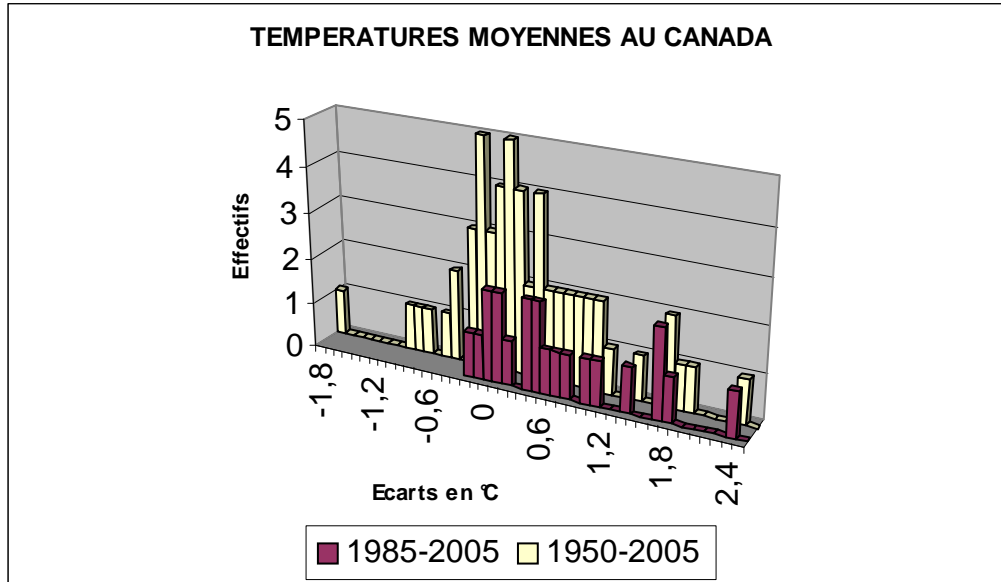
On constate encore une hausse des températures concentrée sur les 20 dernières années.

4) On construit un tableau des effectifs pour les deux séries 1950-2005 et 1985-2005.

Ecart	1985-2005	1950-2005
-1,8		1,00
-1,7		0
-1,6		0
-1,5		0
-1,4		0
-1,3		0
-1,2		0
-1,1		0
-1		1
-0,9		1
-0,8		1
-0,7		0
-0,6		1
-0,5		2
-0,4		0
-0,3		3
-0,2	1	5
-0,1	1,00	3
0	2	4
0,1	2	5
0,2	1,00	4
0,3	0,00	2
0,4	2	4
0,5	2	2
0,6	1	2
0,7	1,00	2
0,8	1	2
0,9	0	2
1	1	2
1,1	1	1
1,2	0	0
1,3	0	0
1,4	1	1
1,5	0	0
1,6	0	0
1,7	2	2
1,8	1	1
1,9	0	1
2	0	0
2,1	0	0
2,2	0	0

2,3	0	0
2,4	1	1
2,5	0	0

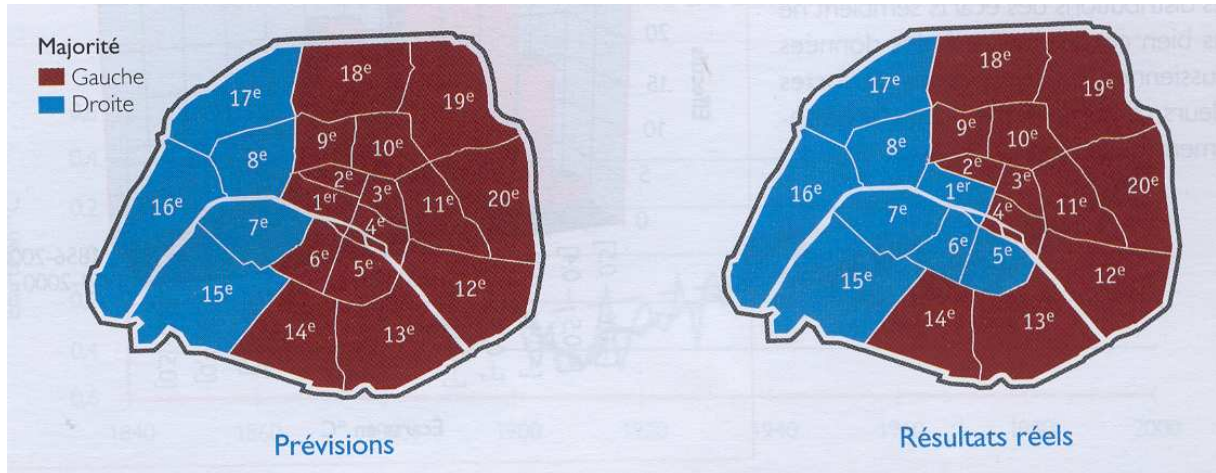
On trace l'histogramme correspondant pour les deux séries.



Commentaire : On constate une concentration des écarts de températures positifs sur la période 1950-2005 témoignant d'un réchauffement du climat.

2 Thème 2 : Sondages d'opinion

ELECTIONS MUNICIPALES A PARIS



Les cartes ci-dessus représentent une prévision des résultats des élections municipales de 2001 à Paris, comparée aux résultats réels. Dans chaque arrondissement, l'institut de sondage avait interrogé de 500 à 1000 personnes soit au total moins de 2% des électeurs. On constate que la prévision était **assez bonne**.

Principe des sondages d'opinion

Supposons que 43 % des électeurs souhaitent voter pour le candidat A. Si plusieurs sondages différents sont réalisés auprès de plusieurs échantillons de n personnes choisis au hasard, les pourcentages d'intentions de vote pour A obtenus sont des données gaussiennes, dont la loi normale associée a pour moyenne 43% et un écart-type inférieur à $\frac{50}{\sqrt{n}}$.

On peut donc dire que dans 95% des cas , le résultat du sondage ne s'éloignera pas de 43% de plus de deux écarts-types.

Si on interroge 601 personnes, comme dans le sondage ci-dessous, la différence avec les intentions de vote ne devrait pas dépasser $2 \times \frac{50}{\sqrt{601}} \approx 4,1\%$

Exemple : Le cas des municipales à Toulouse en 2001.

- 1) Calculer pour les trois listes principales l'intervalle $p - \frac{100}{\sqrt{601}}$; $p + \frac{100}{\sqrt{601}}$ (p désigne la prévision)
- 2) Quelle aurait dû être la taille de l'échantillon pour que ces intervalles aient une largeur de 2% ?
- 3) Les résultats appartiennent-ils aux fourchettes de prévision ? Commenter.

Listes	Sondage (2/02)	Résultat (18/03)
100% à gauche	4	2,44
Lutte ouvrière	1	1,67
Motivé-e-s (Zebda)	14	12,38
Gauche plurielle	31	27,83
Les verts	7	6,16
Droite unie	36	41,56

100% Toulouse	2	0,9
Front National	3	4,03
MNR	2	2,65
Total	100	100
Abstention	36	40,04

Réponse :

1) Liste Droite unie : $p = 36$

$$\left[p - \frac{100}{\sqrt{601}} ; p + \frac{100}{\sqrt{601}} \right] \approx [31,9 ; 40,1]$$

Liste Gauche plurielle : $p = 31$

$$\left[p - \frac{100}{\sqrt{601}} ; p + \frac{100}{\sqrt{601}} \right] \approx [26,9 ; 35,1]$$

Liste Motivé-e-s : $p = 14$

$$\left[p - \frac{100}{\sqrt{601}} ; p + \frac{100}{\sqrt{601}} \right] \approx [9,9 ; 18,1]$$

2) Soit n la taille de l'échantillon cherché.

$$\text{On veut } \frac{200}{\sqrt{n}} < 2$$

$$\text{Soit } \sqrt{n} > \frac{200}{2} = 100$$

Soit $n > 10\,000$

Pour que les intervalles aient une largeur de 2%, la taille de l'échantillon aurait dû être de 10000 personnes.

3) Oui sauf pour la liste Droite unie pour laquelle le sondage est inférieur au résultat.

3 Thème 3 : Statistiques médicales

3.1 Enquête sur un marqueur

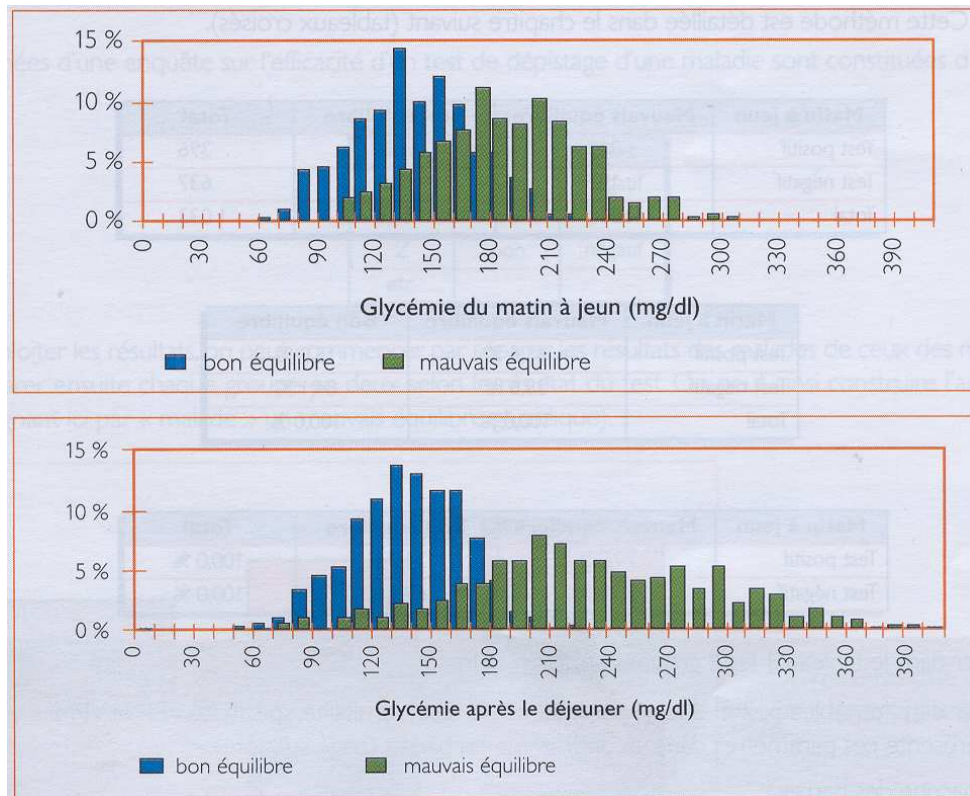
Pour vérifier l'équilibre glucidique, la référence est le taux d'hémoglobine glyquée (HbA1c), mais sa mesure est encore difficile à établir. On utilise donc comme marqueur (ou test) la **mesure de la glycémie** (poids de sucre par dl de sang).

Pour déterminer le moment de la journée qui permet la meilleure discrimination entre les patients dont l'équilibre glucidique est bon, et les autres, on a réalisé une enquête auprès de 1 033 diabétiques (un groupe de 523 dont l'équilibre glucidique est bon, d'après le taux HbA1c, et un groupe de 510 dont l'équilibre glucidique est mauvais).

On relève le taux de glycémie :

- le matin à jeun ;
- après le déjeuner de midi

On obtient les répartitions suivantes :



On observe :

- il y a chevauchement des histogrammes
- le taux de glycémie varie peu entre les deux analyses chez les patients bien équilibrés ; mais augmente nettement dans l'autre groupe.
- Ce phénomène va permettre une meilleure discrimination entre les deux groupes.

On définit un seuil de déséquilibre pour chaque heure de test :

- le matin, le test sera positif si le taux dépasse 165 mg/dl
- l'après-midi s'il dépasse 180 mg / dl

3.2 Evaluation des résultats

L'efficacité du test s'évalue à partir des tableaux suivants :

A jeun	Mauvais équilibre	Bon équilibre
Test positif	313	83
Test négatif	197	440

Après-midi	Mauvais équilibre	Bon équilibre
Test positif	408	38
Test négatif	102	485

Dans chaque cas, il y a 4 groupes dénommé de façon claire :

- les vrais positifs (vp)
- les faux positifs (fp)
- les faux négatifs (fn)
- les vrais négatifs (vn)

Il y a 4 paramètres d'évaluation d'un test de dépistage d'une maladie :

- Le pourcentage de malades parmi les positifs ou VPP (valeur prédictive positive)
- Le pourcentage de non-malades parmi les négatifs ou VPN (valeur prédictive négative)
- Le pourcentage de positifs parmi les malades ou sensibilité.
- Le pourcentage de négatifs parmi les non malades ou spécificité

Pour préparer le calcul des ces paramètres, on peut indiquer la somme des effectifs des lignes et des colonnes.

On divise ensuite chaque effectif par le total de sa colonne ou de sa ligne.

tableau 1 :

A jeun	Mauvais équilibre	Bon équilibre	Total
Test positif	313	83	396
Test négatif	197	440	637
Total	510	523	1033

Tableau 2 :

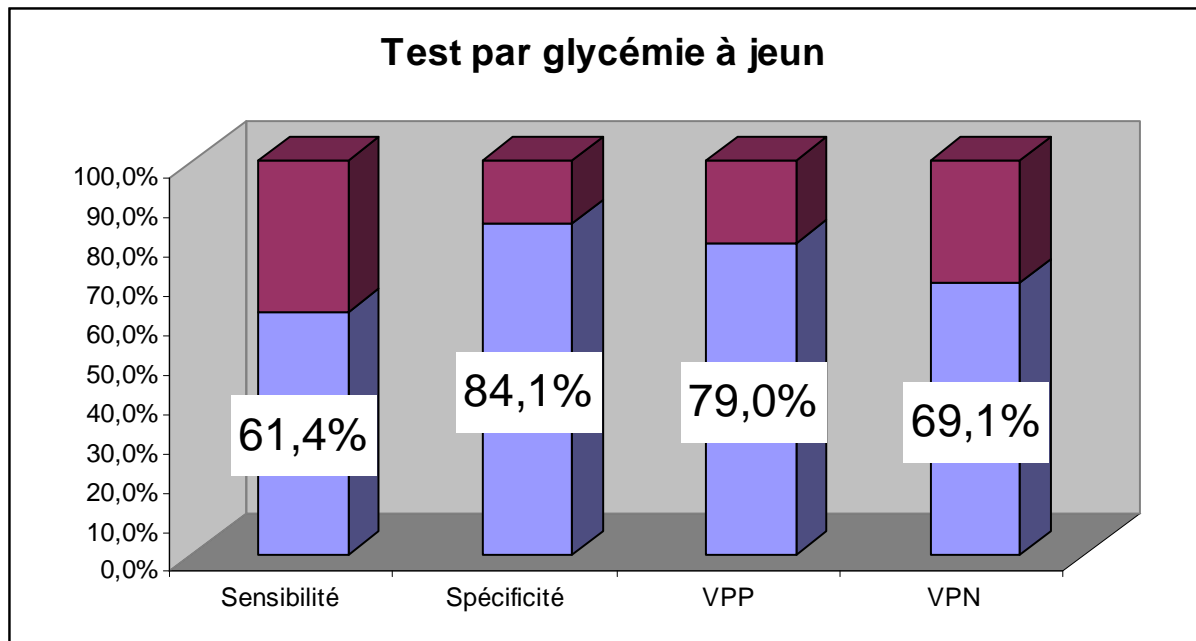
A jeun	Mauvais équilibre	Bon équilibre
Test positif	61,4%	15,9%
Test négatif	38,6%	84,1%
Total	100%	100%

Tableau 3 :

A jeun	Mauvais équilibre	Bon équilibre	Total
Test positif	79,0%	21,0%	100%
Test négatif	30,9%	69,1%	100%

- 1) Identifier dans le tableau 1 les 4 groupes vp, fp, vn et fn.
- 2) Identifier dans les tableaux 2 et 3 les 4 paramètres du test : sensibilité, spécificité, VPP et VPN.

On représente ces paramètres dans un diagramme en barres.



- 3) Pour chacune des barres :
 - a) Quels sont les effectifs de référence ?
 - b) Quels sont les effectifs représentés par leur fréquence ?

- 4) En appliquant la même méthode :
 - a) Calculer les paramètres du test de glycémie pratiqué après le déjeuner.
 - b) Représenter dans le même diagramme les paramètres des deux tests.

- 5) Quel est le meilleur moment pour pratiquer un test par la glycémie dans le but de vérifier l'équilibre glucidique ?

Réponse :

1) $vp = 313$ $fp = 83$ $vn = 440$ $fn = 197$

- 2) sensibilité = 61,4 % (test positif parmi les malades : se lit dans le tableau 2)
 spécificité = 84,1 % (test négatif parmi les non-malades : se lit dans le tableau 2)
 VPP = 79 % (malades parmi les positifs : se lit dans le tableau 3)
 VPN = 69,1 % (non-malades parmi les négatifs : se lit dans le tableau 3)

- 3) a) Sensibilité : effectifs de référence = malades (effectif total = 510)
 Spécificité : effectifs de référence = non-malades (effectif total = 523)
 VPP : effectifs de référence = test positif (effectif total = 396)
 VPN : effectifs de référence = test négatif (effectif total = 637)

4) a)

tableau 1' :

Après-midi	Mauvais équilibre	Bon équilibre	Total
Test positif	408	38	446
Test négatif	102	485	587
Total	510	523	1033

Tableau 2' :

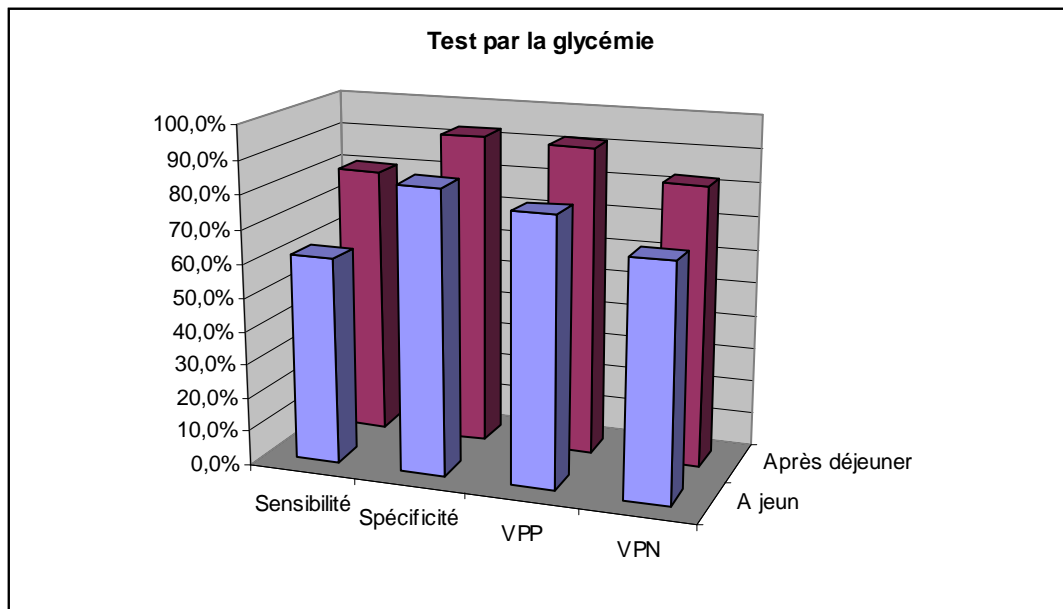
Après-midi	Mauvais équilibre	Bon équilibre
Test positif	80%	7,3%
Test négatif	20%	92,7%
Total	100%	100%

Tableau 3 :

Après-midi	Mauvais équilibre	Bon équilibre	Total
Test positif	91,5 %	8,5%	100%
Test négatif	17,4 %	82,6%	100%

sensibilité = 80 %
 spécificité = 92,7 %
 VPP = 91,5 %
 VPN = 82,6 %

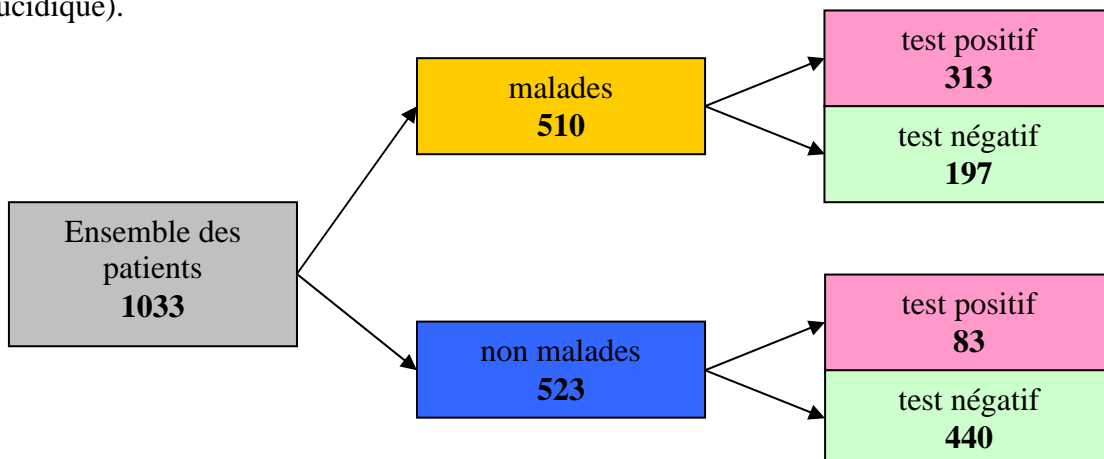
b)



4) Les 4 paramètres du test sont plus élevés (donc plus discriminants) l'après-midi. Il est donc préférable d'effectuer le test de glycémie l'après-midi.

3.3 Dénombrement des résultats à l'aide d'un arbre

On peut construire l'arbre suivant (en désignant par « malade » un mauvais équilibre glucidique).



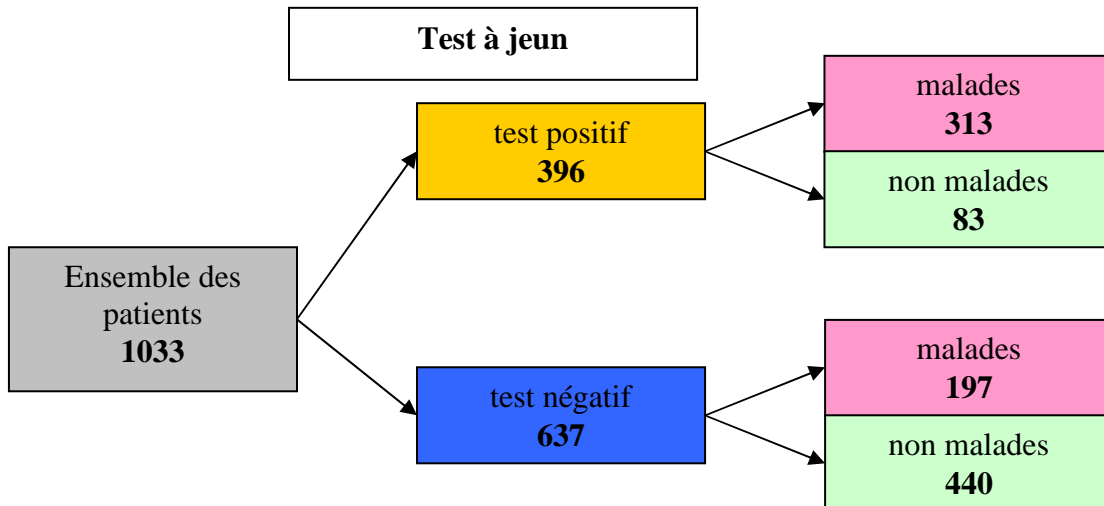
- 1) Comment calculer à l'aide de cet arbre la sensibilité et la spécificité du test du matin ?
- 2) Calculer de même, en construisant un autre arbre, la VPP et la VPN de ce test.
- 3) Etudier de la même manière le test de l'après-midi.

Réponse :

1) sensibilité = % de positifs parmi les malades = $\frac{313}{510} \approx 61,4 \%$.

spécificité = % de négatifs parmi les non-malades = $\frac{440}{523} \approx 84,1 \%$.

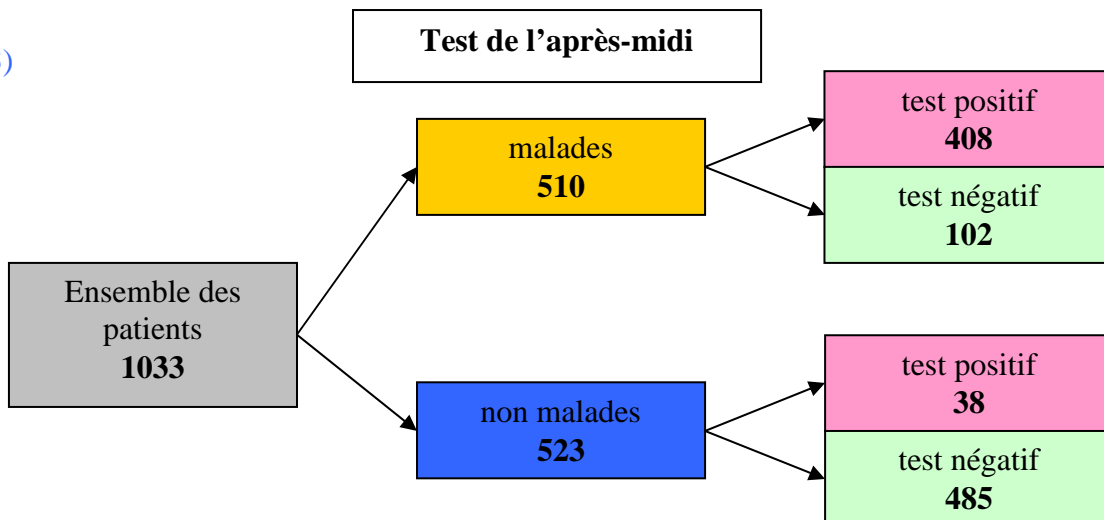
2)



VPP = % de malades parmi les positifs = $\frac{313}{396} \approx 79,0 \%$

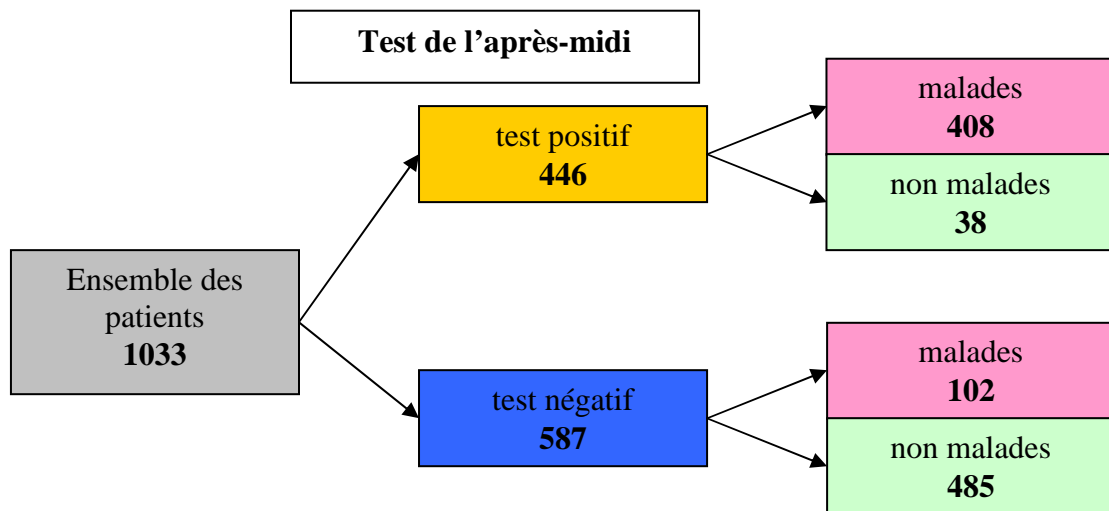
VPN = % de non-malades parmi les négatifs = $\frac{440}{637} \approx 69,1 \%$

3)



sensibilité = % de positifs parmi les malades = $\frac{408}{510} = 80 \%$

spécificité = % de négatifs parmi les non-malades = $\frac{485}{523} \approx 92,3 \%$.



$$\text{VPP} = \% \text{ de malades parmi les positifs} = \frac{408}{446} \approx 91,5 \%$$

$$\text{VPN} = \% \text{ de non-malades parmi les négatifs} = \frac{485}{587} \approx 82,6 \%$$