

Sommaire

REMERCIEMENTS.....	Page 2
DEDICACES.....	Page 3

<i>Introduction</i>	Page 6
---------------------------	--------

Chapitre I: Protocole d'étude.

I. Cancer du sein.....	Page 7
II. Objectifs de l'étude.....	Page 7
1. Objectif général	
2. Objectif spécifiques	
III. Matériels et méthodes	Page 8
1. Population de l'étude.....	Page 8
1.1 . sélection des cas.	
1.2 Sélection des témoins.	
2. La taille d'échantillonnage.	Page 8
3. La collecte des données.....	page 8
4. Considérations éthiques.....	Page 10
5. Plan d'analyse des données.....	page 11

Chapitre II: outils mathématiques.

⊙ L'effectif.....	Page 12
⊙ Fréquences relatives et cumulées.....	Page 12
⊙ La moyenne, l'écart type et la variance.....	Page 13
⊙ la médiane.....	Page 13
⊙ Le mode.....	Page 14
⊙ Les quartiles.....	Page 15
⊙ Le test de khi-deux d'indépendance.....	page 16
⊙ Le test d'hypothèse de comparaison de moyennes.....	page 18

Chapitre III: Traitement statistique des données

1. Analyse uni-variée

Tableau 1 : Répartition de la population d'étude selon l'âge	page 19
Tableau 2 : Répartition de la population d'étude selon la consommation passive du tabac.....	Page 20
Tableau 3 : Répartition de la population d'étude selon les antécédents familiaux.....	page 21
Tableau 4 : Répartition de la population d'étude selon l'âge à la première grossesse.....	page 22
Tableau 5 : Répartition de la population d'étude selon l'allaitement maternel.....	page 23
Tableau 6 : la fréquence de consommation de viandes rouges chez la population d'étude....	page 24
Tableau 7 : la fréquence de consommation de poissons chez la population d'étude.....	page 25
Tableau 8 : la fréquence de consommation de conserves chez la population d'étude.....	page 26
Tableau 9 : la fréquence de consommation de gras chez la population d'étude.....	Page 27
Tableau 10 : la fréquence de consommation de viandes de volaille chez la population d'étude.....	page 28
Tableau 11 : la fréquence de consommation de légumes secs, pomme de terre et les pâtes chez la population d'étude.....	page 29

2. Analyse bi-variée.

Tableau 12: Association entre le risque de cancer du sein et les viandes rouges.....	page30
Tableau 13 : Association entre le risque de cancer du sein et les conserves.....	page30
Tableau 14 : Association entre le risque de cancer du sein et le sucre.....	page31
Tableau 15: Association entre le risque de cancer du sein et le chocolat.....	page32
Tableau 16 : Association entre le risque de cancer du sein et le beurre.....	page32
Tableau 17 : Association entre le risque de cancer du sein et le lait.....	page33
Tableau 18: le test d'hypothèse de comparaison de moyennes d'âge des cas et des témoins.....	page 34

Chapitre IV : Discussion.

1. Quelques déductions suite aux résultats de l'analyse	page 35
2. Quelques conseils pour lutter contre le cancer du sein.....	page 36

<i>Conclusion</i>	page 37
Références bibliographiques.....	page 38

Introduction

Le stage est une période de formation très avantageuse pour les étudiants ; c'est le moment idéal pour mettre en pratique les connaissances théoriques acquises au cours des cursus universitaires; c'est aussi l'occasion d'acquérir de nouvelles compétences qui pourront être utilisées concrètement. Il permet à l'étudiant de profiter d'un point de vue parfois différent et d'apprendre énormément en observant des professionnels au travail.

Grace à l'importance du stage dans l'acquisition des connaissances, l'évaluation de degré de concentration, d'autonomie, de motivation et d'implication ; j'ai décidé de réaliser un stage au sein de la faculté de médecine et de pharmacie de Fès au laboratoire d'épidémiologie, Recherche Clinique et Santé Communautaire en vue d'être en contact direct avec des experts qui œuvrent dans le domaine de mon étude pour me perfectionner dans ma formation et améliorer mes points faibles durant ce stage de deux mois.

Le thème de mon étude est sous le titre "Etude statistique comparative sur le risque du cancer du sein chez les patientes de la région Fès-Maroc ", c'est un sujet qui m'a beaucoup attiré, il me permet de réaliser l'étude statistique de ce lourd fardeau socio sanitaire dans le monde.

Au cours de cette étude nous avons utilisé un logiciel statistique dénommé SPSS ; c'est un logiciel qui dispose de procédures avancées pour accélérer l'analyse des données complexes ou difficiles à manipuler.

Ce projet se compose de quatre parties, dans la première partie nous allons présenter le protocole d'étude qui détaille les objectifs de cette recherche, la manière et les critères de choix de la population ainsi que les questionnaires utilisés dans l'étude (questionnaire de base et questionnaire nutritionnel). La deuxième partie se penchera sur l'outil mathématique, quant à la troisième partie elle aborde un traitement statistique des données en se basant sur des graphiques et des interprétations des résultats obtenus. Enfin, une dernière partie où nous allons présenter une conclusion générale sur les facteurs de risque du cancer du sein.

Protocole d'étude

I. Cancer du sein

Le cancer du sein est une tumeur maligne de la glande mammaire, c'est un cancer qui naît dans les unités cellulaires dont la fonction est de sécréter le lait, les unités ducto-lobulaires du sein, essentiellement chez la femme ; le cancer du sein survient 200 fois moins souvent chez l'homme, qui possède lui aussi des seins, bien qu'atrophiés.

Dans le monde entier, le cancer du sein est le cancer le plus fréquent c'est la cause principale du décès des femmes. L'incidence de ce cancer varie considérablement selon les pays et les régions du monde. Il est trois fois plus fréquent dans les pays économiquement développés que dans les pays en développement. Les taux d'incidence les plus élevés sont observés aux États-Unis et en Europe du Nord, les taux les plus bas se situent en Afrique et en Asie.

Au Maroc le cancer du sein est le premier cancer chez la femme. Il représente 34,3 % de tous les cancers féminins.

II. Objectifs de l'étude :

1. Objectif général :

Déterminer les principaux facteurs de risque du cancer du sein chez les patientes de la région de Fès.

2. Objectifs spécifiques :

- Décrire les caractéristiques sociodémographiques des femmes atteintes du cancer du sein incluses dans notre étude.
- Déterminer les facteurs de risque du cancer du sein liés au mode de vie de la femme marocaine.

III . Méthodes et matériels :

Pour répondre à la question de recherche, ont a opté pour une étude cas-témoins avec appariement sur la variable âge, menée au CHU Hassan II de Fès.

1. Population de l'étude

1.1 Sélection des cas :

- Critères d'inclusion :

Les patientes admises à la consultation au centre d'oncologie de la région de Fès, pour un cancer du sein confirmé par un examen histologique.

- Critères d'exclusion :

- Diagnostic du cancer du sein non confirmé histologiquement.
- Avoir moins de 18 ans.
- Patiente refusant de participer.

1.2 Sélection des témoins :

- Critères d'inclusion :

- Être accompagnatrice d'un patient admis au centre des consultations externes du CHU de Fès.

Le recrutement des témoins sera fait sur la base d'un appariement sur l'âge.

- Critères d'exclusion :

- Accompagnatrices des patientes suivies pour un cancer du sein.
- Femme ayant des antécédents de cancer.
- Avoir moins de 18 ans.
- Femme refusant de participer.

2. la taille d'échantillonnage :

Notre échantillon se compose de 432 femmes admises pour dépistage du cancer au CHU. Il y a eu 216 cas atteints (les cas) et 216 cas non atteints (les témoins).

3. Collecte des données :

Pour les cas, la méthode de collecte des données a été faite par l'exploitation des dossiers médicaux des patientes ainsi qu'un questionnaire administré face à face de manière standardisée aux cas et aux témoins.

Les données seront recueillies à partir de deux questionnaires : un questionnaire de base (Annexe1), et un questionnaire nutritionnel (Annexe2).

Annexe 1 : Questionnaire de base

Nom d'enquêteur :	_____	
Date de l'enquête :		/_/_/_/_/_/_/_
Enquêtée :	_____	
Paire N° :		/_/_/_/_
Statut participante :		
Cas = 1	Témoin =2	/_

1. Age :

2. Etes-vous exposée au tabac dans votre entourage familial :

Oui=1

Non=2

3. Dans votre famille, y a-t-il quelqu'un qui a développé une tumeur maligne?

Oui=1

Non=2

/_/_

4. Âge à la première grossesse (avortement ou naissance) /_/_/_/

5. Allaitement maternel

0 = Pas d'allaitement maternel pour des raisons personnelles ou médicales.

Nombre de mois pour chaque naissance = Durée de l'allaitement maternel en mois .

I	II	III	IV	V	VI	VII	IIX	
/_/_	/_/_	/_/_	/_/_	/_/_	/_/_	/_/_	/_/_	/_/_
IX	X	Dernière						
/_/_	/_/_	/_/_						

Annexe 2 : Questionnaire nutritionnel

29- حليب (البقر)/ حليب صوجا (قياس كأس د العصور او دالعا)	نادرا / أبدا	1- 3مرات في الشهر	مرة في الاسبوع	2-4 مرات في الاسبوع	5-6 مرات في الاسبوع	مرة في اليوم	2-3مرات في اليوم	4 مرات في اليوم أو أكثر
q29p1 - حليب كامل	1	2	3	4	5	6	7	8
q29p2 - لبن (بوحدو ولا مع اللواكه)	1	2	3	4	5	6	7	8
q29p3 - حليب كامل الدسم	1	2	3	4	5	6	7	8
q29p4 - حليب دبال الرجين (نصف لثندة، Demi écrémé)	1	2	3	4	5	6	7	8
q29p5 - حليب دبال الرجين (خال من الدسم)	1	2	3	4	5	6	7	8
q29p6 - الرائب	1	2	3	4	5	6	7	8
q29p7 - حليب الصوجا	1	2	3	4	5	6	7	8
q29p8 - مسكو دالبن (سكوك)	1	2	3	4	5	6	7	8
q29p9 - يوغورت (أي نوع)	1	2	3	4	5	6	7	8
q29p10 - اليوغورت باليو بيوتيكس activia	1	2	3	4	5	6	7	8
q29p11 - اليوغورت المنسوب بالصوجا	1	2	3	4	5	6	7	8

30- الفورماج قياس طرف واحد ولا منتهون أطرف (دلخيز)	نادرا / أبدا	1- 3مرات في الشهر	مرة في الاسبوع	2-4 مرات في الاسبوع	5-6 مرات في الاسبوع	مرة في اليوم	2-3مرات في اليوم	4 مرات في اليوم أو أكثر
q30p1 - أي فورماج	1	2	3	4	5	6	7	8
q30p2 - فورماج قاصح (الندار cheddar، بارميزان Parmesan)	1	2	3	4	5	6	7	8
q30p3 - الفورماج الرطب ، مثال : (الكومبير، البري، فياتيليا)	1	2	3	4	5	6	7	8
q30p4 - فورماج نصف قاصح مثل: Gouda، Emmentaler/Edam (الفورماج لحم)	1	2	3	4	5	6	7	8
q30p5 - الجبن (طبيعي أو منعم)	1	2	3	4	5	6	7	8
q30p6 - اروفماج طري مثال: فينا/موزاريلا	1	2	3	4	5	6	7	8
q30p7 - لافش كوري، كوري، لقب لحليب، جوليور...	1	2	3	4	5	6	7	8

4. Considérations éthiques :

Il s'agira d'une étude sur des femmes adultes, avec ou sans cancer du sein, capables à donner un consentement éclairé. La participation à l'étude est volontaire, chaque femme participante à l'étude, et après explication des objectifs de l'étude, les modalités de participation, ainsi que les risques et les bénéfices de cette étude, une déclaration de consentement sera demandée et signée. Toute participante a le droit de quitter l'étude à tout moment sans donner des justifications.

5. Plan d'analyse des données :

Pour les questionnaires, les données ont été saisies dans une base de données à l'aide du logiciel SPSS. Après, nous avons procédé à une description de la population enquêtée, calcul des fréquences absolues, des fréquences relatives et des moyennes des réponses pour structurer l'analyse statistique des données, des tableaux d'analyse et des figures ont été élaborés afin de faciliter leur exploitation et mieux visualiser la répartition des effectifs et pourcentages trouvés.

La technique du Khi-carré nous a permis d'établir s'il y a indépendance ou lien entre les aliments consommés et la maladie, de même que les tests d'hypothèses incluant les tests de signification et les tests de comparaisons nous ont permis comparer entre des moyennes observées et des moyennes théoriques, ainsi que la comparaison entre deux moyennes au risque de 5%.

Outils mathématiques

En statistique, on manipule parfois de très grandes quantités d'informations. Pour en simplifier l'analyse, on effectuera quelques calculs ou diagrammes : moyennes, fréquences, diagrammes en bâtons ou circulaires.

La statistique étudie certaines caractéristiques : caractères ou variables d'un ensemble fini appelé **population**. Les éléments de cette population étudiée sont appelés **individus**

Type de variables :

Une variable peut être :

- **Quantitative** : numérique et fait l'objet de calcul (âge, taille, poids, notes, nombres d'heures etc. ...)
- **Qualitative** : c'est le contraire de quantitative, mais la variable peut très bien être numérique.
- **Discrète** : si la variable ne prend qu'un nombre fini de valeurs (ces valeurs sont appelées modalités et notées x_i).
- **Continue** : si la variable prend ses valeurs dans un intervalle (classe)

Effectif :

On appelle effectif d'une valeur donnée x_i le nombre de fois ou cette valeur apparaît dans la population statistique étudiée ce nombre est noté n_i .

On appelle effectif total de la population étudiée noté N la somme des p effectifs particulières n_i correspondant à chacune des valeurs soit $N = n_1 + n_2 + \dots + n_p = \sum_{i=1}^p n_i$

Fréquences :

On appelle fréquence de la valeur (ou modalité) x_i noté f_i le rapport de l'effectif n_i correspondant à la valeur x_i et de l'effectif n de la population observée $f_i = \frac{n_i}{N}$

Ce rapport est égal au pourcentage d'individus présentant la valeur (ou modalité) x_i par rapport à l'ensemble de la population observée $\sum_{i=1}^p f_i = 1$

Fréquences cumulées :

La fréquence cumulée d'une modalité est la somme des fréquences des modalités qui lui sont inférieures ou égales.

Moyenne :

La moyenne d'une série statistique est un élément important qui permet de caractériser les séries

statistiques. Elle est définie par : $\bar{x} = \frac{\sum_{i=1}^p n_i x_i}{\sum_{i=1}^p n_i}$

Avec $n_1, n_2, n_3, \dots, n_p$ sont les **effectifs** correspondants aux modalités $x_1, x_2, x_3, \dots, x_p$, si la série est discrète, ou les centres de chaque classe si la série est continue

L'écart-type :

L'écart-type d'une série statistique est un nombre positif qui mesure la dispersion autour de la moyenne. Plus les valeurs sont concentrées autour de la moyenne plus l'écart-type est petit

On a :

$$\sigma = \sqrt{\frac{\sum_{i=1}^p n_i (x_i - \bar{x})^2}{\sum_{i=1}^p n_i}}$$

La variance :

On appelle la variance d'une série, noté V, le réel : $V = \sigma^2$

La médiane :

La médiane est un paramètre de position, qui permet de couper la population étudiée en deux groupes contenant le même nombre d'individus.

Ce paramètre est utile pour donner la répartition du caractère étudié, car 50 % environ de la population étudiée a une modalité inférieure à la médiane et 50 % une modalité supérieure à la médiane.

✓ Cas d'une variable discrète :

Pour la déterminer, on utilise les effectifs cumulés croissants

- lorsque l'effectif total est un nombre impair (de type $N = 2n+1$), la médiane est la valeur prise par le $n+1$ ème effectif Statistiques.

- lorsque l'effectif total est un nombre pair (de type $N = 2n$), la médiane est la valeur prise par la moyenne entre le n ème effectif et le $n+1$ ème effectif.

✓ Cas d'une variable continue :

Les fréquences cumulées (ou les effectifs cumulés) permettent de déterminer dans quelle classe

$[b_i ; b_{i+1}[$ se situe la médiane. La classe ainsi obtenue est appelée classe médiane. Une fois qu'on a la classe médiane, on trouve la médiane par interpolation linéaire.

Soit $[b_{i-1} ; b_i [$ la classe médiane, a_i l'amplitude de la classe médiane, N_i l'effectif cumulé croissant de la classe médiane, N_{i-1} l'effectif cumulé croissant de la classe avant la classe médiane et N l'effectif total.

L'expression de la médiane est donnée par : $Mé = b_{i-1} + a_i \left(\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}} \right)$

La même démarche pourrait être utilisée en remplaçant les fréquences absolues par les fréquences

relatives : $Mé = b_{i-1} + a_i \left(\frac{0.5 - F_{i-1}}{F_i - F_{i-1}} \right)$

Où F_i désigne la fréquence cumulée croissante de la classe médiane, F_{i-1} la fréquence cumulée croissante de la classe qui précède la classe médiane

Mode :

On appelle mode ou valeur dominante d'une série statistique la valeur observée de la variable ayant le plus grand effectif (ou la fréquence la plus élevée). On note généralement le mode M_0

✓ Pour une variable discrète :

Le mode est la modalité qui représente le plus grand effectif élevée

✓ Pour une variable continue :

Dans le cas d'une variable continue groupée en classes, on parle plutôt de classe modale. La classe modale est la base du rectangle ayant la hauteur la plus élevée cependant on distingue deux cas selon que les amplitudes des classes sont égales ou inégales.

▪ **Cas d'amplitude identiques :**

Dans ce cas, la classe modale est la classe d'effectif ni le plus élevé, soit $[b_{i-1} ; b_i [$. L'effectif de la classe qui précède la classe modale est n_{i-1} et celui de la classe modale est n_i . La détermination du mode à partir de la classe modale se fait de la façon générale suivante : $M_0 = b_{i-1} + a_i \left(\frac{m_1}{m_1 + m_2} \right)$

Avec :

b_{i-1} : borne inférieure de la classe modale

b_i : borne supérieure de la classe modale

a_i : amplitude de la classe modale

$$m_1 = n_i - n_{i-1}$$

$$m_2 = n_i - n_{i+1}$$

▪ **Cas d'amplitudes inégales :**

Dans le cas où les amplitudes sont différentes, la classe modale est la classe de densité (ou de fréquence corrigée) la plus élevée, ou encore d'effectif corrigé le plus élevé.

Le mode est donnée par $M_0 = b_{i-1} + a_i \left(\frac{m_1}{m_1 + m_2} \right)$ avec

b_{i-1} : borne inférieure de la classe modale

b_i : borne supérieure de la classe modale

a_i : amplitude de la classe modale

$$m_1 = h_i - h_{i-1}$$

$$m_2 = h_i - h_{i+1}$$

h_i, h_{i-1}, h_{i+1} sont les effectifs corrigés

Les quartiles :

Les quartiles sont des valeurs qui séparent un ensemble de données placées en ordre croissant en quatre parties égales, c'est-à-dire comprenant le même nombre de données.

Chaque ensemble de données possède trois quartiles :

Le premier quartile Q1 d'une série statistique ordonnée est la valeur qui sépare cette série en deux groupes :

- Le premier groupe contient un quart des effectifs (25 %)
- Le deuxième groupe contient trois quarts des effectifs (75 %)

L'expression du premier quartile est donnée par :

$$Q_1 = b_i + a_i \left(\frac{\frac{N}{4} - N_{i-1}}{N_i - N_{i-1}} \right) = b_i + a_i \left(\frac{0.25 - F_{i-1}}{F_i - F_{i-1}} \right)$$

Le troisième quartile Q3 d'une série statistique ordonnée est la valeur qui sépare cette série en deux groupes :

- Le premier groupe contient trois quarts des effectifs (75 %)
- Le deuxième groupe contient un quart des effectifs (25 %)

L'expression du troisième quartile est donnée par :

$$Q_3 = b_i + a_i \left(\frac{\frac{3N}{4} - N_{i-1}}{N_i - N_{i-1}} \right) = b_i + a_i \left(\frac{0.75 - F_{i-1}}{F_i - F_{i-1}} \right)$$

La médiane d'une série statistique correspond au deuxième quartile, les deux premiers quarts étant égaux à la moitié (50%)

$$Q_2 = b_i + a_i \left(\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}} \right) = b_i + a_i \left(\frac{0.5 - F_{i-1}}{F_i - F_{i-1}} \right) = M_e$$

Test du χ^2 d'indépendance :

Principe du test :

Le test d'indépendance du khi-deux vise à déterminer si deux variables observées sur un échantillon sont indépendantes ou non. Les variables étudiées sont des variables qualitatives catégorielles présentant plusieurs modalités.

- Les données sont structurées sous forme d'un tableau des effectifs observés pour les deux caractères comparés ou table de contingence

		Caractère A					
		<i>modalité 1</i>		<i>Modalité i</i>		<i>modalité p</i>	total
Caractère B	<i>modalité 1</i>	n_{11}		n_{i1}		n_{p1}	$n_{.1}$
	<i>modalité j</i>	n_{1j}		n_{ij}		n_{pj}	$n_{.j}$
	<i>modalité q</i>	n_{1q}		n_{iq}		n_{pq}	$n_{.q}$
Total	$n_{1.}$		$n_{i.}$		$n_{p.}$	$n_{..} = N$	

Avec l'effectif n_{ij} correspond au nombre d'individus ayant la modalité i du caractère A et la modalité j du caractère B avec $1 \leq i \leq p$ et $1 \leq j \leq q$

L'effectif $n_{i.}$ est la somme des effectifs de la colonne i

L'effectif $n_{.j}$ est la somme des effectifs de la ligne j

L'effectif $n_{..}$ est l'effectif total de la table de contingence

- Le tableau des effectifs attendus sous l'hypothèse H_0 : indépendance entre le caractère A et le caractère B.

		Caractère A			Total
		modalité 1	modalité i	modalité p	
Caractère B	modalité 1	$\frac{n_{1.} \times n_{.1}}{N}$	$\frac{n_{i.} \times n_{.1}}{N}$	$\frac{n_{p.} \times n_{.1}}{N}$	$n_{.1}$
	modalité j	$\frac{n_{1.} \times n_{.j}}{N}$	$\frac{n_{i.} \times n_{.j}}{N}$	$\frac{n_{p.} \times n_{.j}}{N}$	$n_{.j}$
	modalité q	$\frac{n_{1.} \times n_{.q}}{N}$	$\frac{n_{i.} \times n_{.q}}{N}$	$\frac{n_{p.} \times n_{.q}}{N}$	$n_{.q}$
	Total	$n_{1.}$	$n_{i.}$	$n_{p.}$	$n_{..} = N$

Sous H_0 , l'effectif attendu t_{ij} correspondant à la modalité i du caractère A (A_i) et à la modalité j du caractère B (B_j) peut être obtenu de la façon suivante :

$$P(A_i \cap B_j) = p_{ij} = P(A_i) \times P(B_j) \text{ sous } H_0 : \text{indépendance entre les deux caractères}$$

$$\text{D'où } p_{ij} = \frac{n_{i.}}{N} \times \frac{n_{.j}}{N} \text{ avec } t_{ij} \text{ effectif attendu}$$

$$\text{D'où } t_{ij} = N \times p_{ij} \text{ ainsi } t_{ij} = \frac{n_{.j} \times n_{i.}}{N}$$

Hypothèses testées :

H_0 : les deux caractères sont indépendants.

H_1 : les deux caractères ne sont pas indépendants.

Conditions d'application du test :

Tous les effectifs attendus doivent être supérieurs ou égaux à 5. Dans le cas contraire, il faut regrouper certaines classes de la variable.

L'écart à l'indépendance :

$$\chi^2_{obs.} = \sum_{i=1}^p \sum_{j=1}^q \frac{(n_{ij} - t_{ij})^2}{t_{ij}}$$

q = nombre de lignes, c'est-à-dire le nombre de modalités de la variable présentée en lignes

p = nombre de colonnes, c'est-à-dire le nombre de modalités de la variable présentée en colonnes

n_{ij} l'effectif observé

t_{ij} l'effectif théorique attendu sous H_0

$\chi^2_{obs.}$ est comparée avec la valeur seuil, χ^2_{seuil} lue sur la table du khi-deux pour $(p-1)(q-1)$ ddl (degrés de liberté) et pour un risque d'erreur α fixé.

La règle de décision :

- si $\chi^2_{obs.} > \chi^2_{seuil}$ l'hypothèse H_0 est rejetée au risque d'erreur α : il n'y a pas indépendance statistique entre les deux caractères étudiés dans la population.
- si $\chi^2_{obs.} \leq \chi^2_{seuil}$ l'hypothèse H_0 est acceptée: les deux caractères étudiés dans la population sont statistiquement indépendants.

Test d'hypothèse de comparaison de moyennes :

Soit α le seuil de signification du test

1) L'hypothèse statistique :

$$H_0 = m_1 = m_2$$

2) La contre hypothèse :

$$H_1 = m_1 < m_2 \text{ ou}$$

$$H_1 = m_1 > m_2 \text{ ou}$$

$$H_1 = m_1 \neq m_2$$

3) Conditions d'application du test :

Les deux échantillons sont prélevés au hasard et indépendamment de deux populations normales ou de taille n_1 et $n_2 > 30$

4) Choix de la statistique :

La statistique qui convient pour effectuer le test de l'hypothèse est $H_0 = m_1 = m_2$ est $\bar{X}_1 - \bar{X}_2$

5) sa loi de probabilité :

Sous H_0 et sous les conditions d'application du test

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ obéit à une loi normale } Z(0,1)$$

$$\text{On calcul } Z_{exp} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

6) Règle de décision :

a. $H_1 = m_1 < m_2$ rejeter H_0 si $Z_{\text{exp}} < -Z_{\alpha}$

b. $H_1 = m_1 > m_2$ rejeter H_0 si $Z_{\text{exp}} > Z_{\alpha}$

c. $H_1 = m_1 \neq m_2$ rejeter H_0 si Z_{exp} n'est pas dans l'intervalle $[-Z^{\alpha/2}; +Z^{\alpha/2}]$

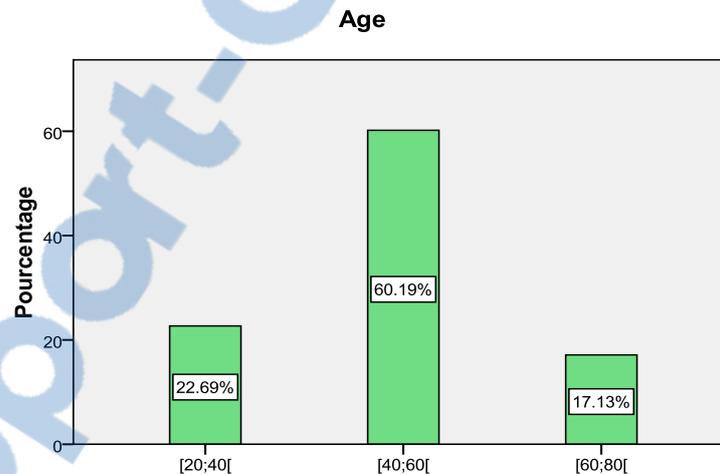
Traitement statistique des données

Analyse des données :

1- ANALYSE UNI VARIEE:

Tableau 1 : Répartition de la population d'étude selon l'âge

AGE				
	Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
[20;40[49	22.7	22.7	22.7
[40;60[130	60.2	60.2	82.9
[60;80[37	17.1	17.1	100.0
Total	216	100.0	100.0	



D'après la représentation graphique de l'âge des femmes ayant un cancer du sein, on constate que celles ayant un âge compris entre 40 et 60 ans représentent un risque maximal de 60,19% ; tandis que le risque du cancer du sein est moins élevé dans la classe d'âge entre 20 et 40 et ne représente que 22.69%, donc le cancer du sein se développe le plus souvent chez les femmes qui ont plus de 40 ans.

	Valide	216
	Manquante	0
Moyenne		48,44
Médiane		49.07
Mode		50
Ecart-type		11,12
Variance		123,64
Quartiles	25	40,75
	50	49 ,07
	75	55,25

✚ **Moyenne** : dans cette étude la moyenne d'âge des femmes qui ont un cancer est de 48,44 ans.

✚ **Mode** : 50ans est la valeur la plus fréquente dans cette étude d'âge des femmes qui ont un cancer du sein.

✚ **Médiane** : Au moins la moitié des femmes contractant un cancer du sein ont un âge supérieur ou égal à 49 ans.

Au moins la moitié des femmes contractant un cancer du sein ont un âge inférieur à 49ans.

✚ **Écart-type** : la distribution de l'âge des femmes ayant un cancer du sein s'étend à plus ou moins 11.12 ans, l'écart type dans ce cas est moins élevé on peut dire que les valeurs sont étroitement distribuées.

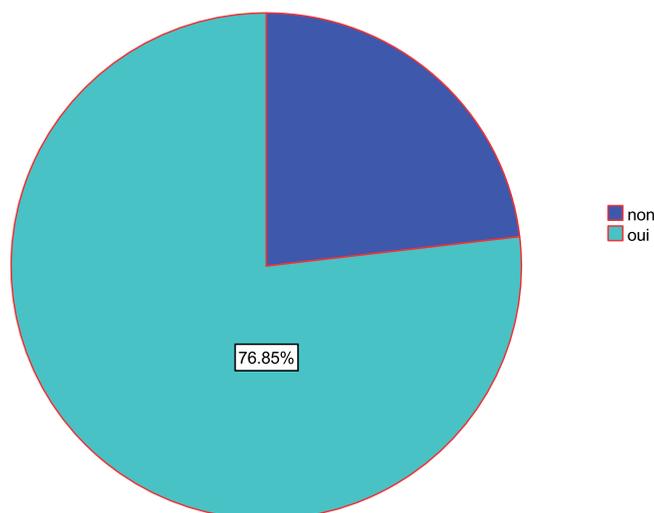
✚ **Quartiles** : Au moins un quart (25%) des femmes contractant un cancer du sein ont un âge inférieures ou égales à 40,75 ans

Au moins trois quart (75%) des femmes contractant un cancer du sein ont un âge inférieures ou égales à 55.25 ans

Tableau 2 : Répartition de la population d'étude selon la consommation passive du tabac

		Tabagisme			
		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Non	50	23.1	23.1	23.1
	Oui	166	76.9	76.9	100.0
Total		216	100.0	100.0	

Tabagisme



Les femmes ont un risque plus élevé de développer un cancer du sein si elles ont été fumeuses de façon passive, la valeur y représente plus de 76% de la population.

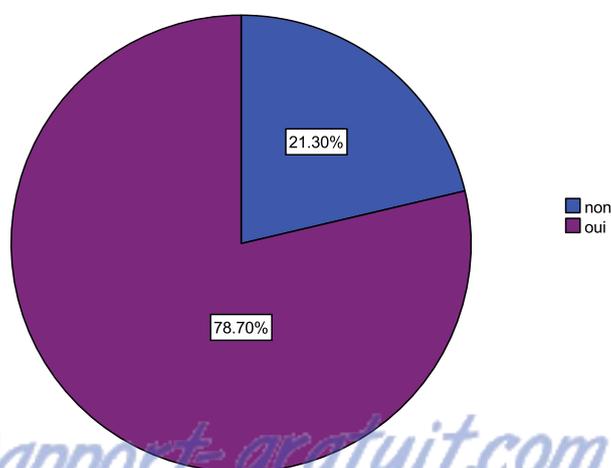
Cette représentation révèle qu'il pourrait y avoir un lien entre le tabagisme passif et le cancer du sein

Tableau 3 : Répartition de la population d'étude selon les antécédents familiaux

Histoire du cancer dans la famille

		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	Non	46	21.3	21.3	21.3
	Oui	170	78.7	78.7	100.0
Total		216	100.0	100.0	

Histoire du cancer dans la famille

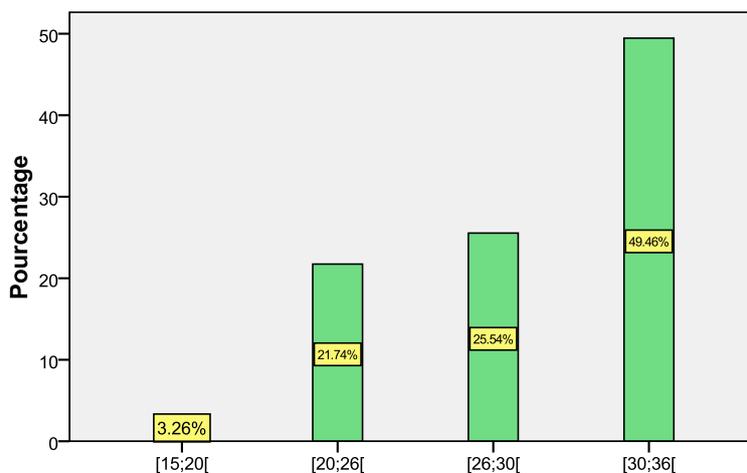


La valeur la plus élevée dans ce graphe représente les femmes atteintes d'un cancer et qui ont des antécédents familiaux de ce cancer ou d'un autre type de cancer.

Tableau 4 : Répartition de la population d'étude selon l'âge à la première grossesse

		L'âge à la première grossesse			
		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	[15;20[6	2.8	3.3	3.3
	[20;26[40	18.5	21.7	25.0
	[26;30[47	21.8	25.5	50.5
	[30;36[91	42.1	49.5	100.0
	Total	184	85.2	100.0	
Manquante	Système manquant	32	14.8		
Total		216	100.0		

L'âge à la première grossesse



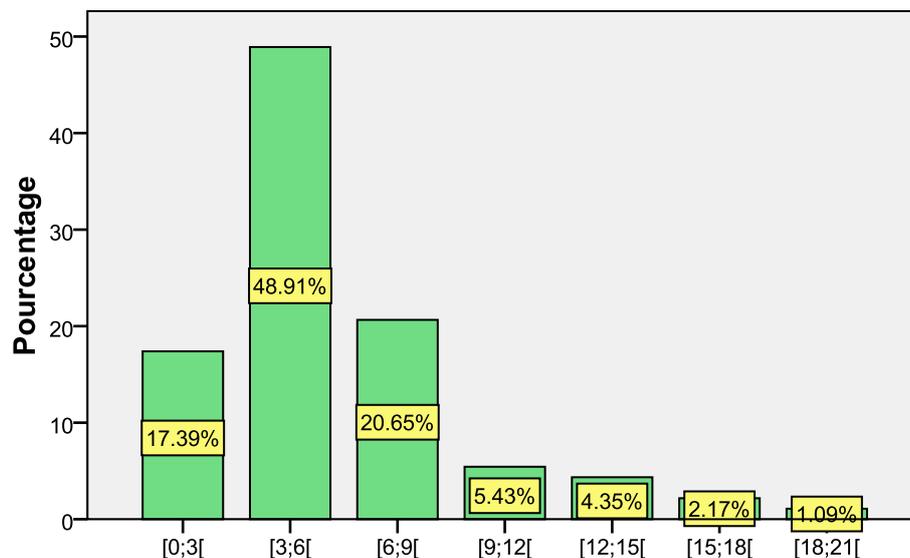
A partir de cet histogramme on peut conclure que les femmes qui ont eu leur première grossesse après 30 ans sont à risque légèrement plus élevé de développer un cancer du sein que les femmes qui ont eu leur première grossesse à un âge plus précoce.

L'absence ou l'âge tardif de la première grossesse peuvent augmenter les chances d'être touchée par cette tumeur. Avoir son premier enfant après 30 ans double le risque d'avoir un cancer du sein par rapport à une femme devenue mère avant 20 ans

Tableau 5 : Répartition de la population d'étude selon l'allaitement maternel

		Allaitement			
		Effectifs	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	[0;3[31	14.4	16.8	16.8
	[3;6[90	41.7	48.9	65.8
	[6;9[38	17.6	20.7	86.4
	[9;12[10	4.6	5.4	91.8
	[12;15[8	3.7	4.3	96.2
	[15;18[4	1.9	2.2	98.4
	[18;21[2	.9	1.1	99.5
	[21;24[1	.5	.5	100.0
	Total	184	85.2	100.0	
Manquante	Système manquant	32	14.8		
	Total	216	100.0		

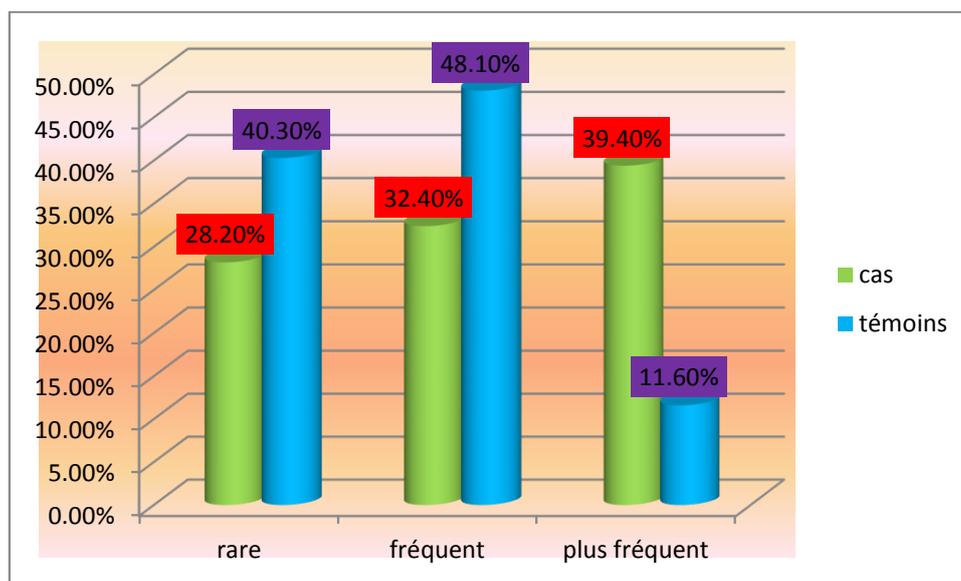
Allaitement



Le graphique au dessus illustre le pourcentage des femmes ayant un cancer de sein selon la durée d'allaitement d'un enfant, on peut constater que les femmes qui allaitent pendant une durée de 21 à 24 mois ont un risque plus faible de développer un cancer du sein par rapport aux femmes qui n'allaitent que 3 à 6 mois ; donc on peut conclure que l'allaitement maternel pendant une période de temps prolongée protégerait les femmes contre le cancer du sein.

Tableau 6 : la fréquence de consommation de viandes rouges chez la population d'étude

		Effectifs	Pourcentages
Rare	Cas	61	28.2%
	Témoins	87	40.3%
fréquent	Cas	70	32.4%
	Témoins	104	48.1%
Plus fréquent	Cas	85	39.4%
	Témoins	25	11.6%



L'histogramme retrace la consommation de la viande rouge chez les cas et les témoins.

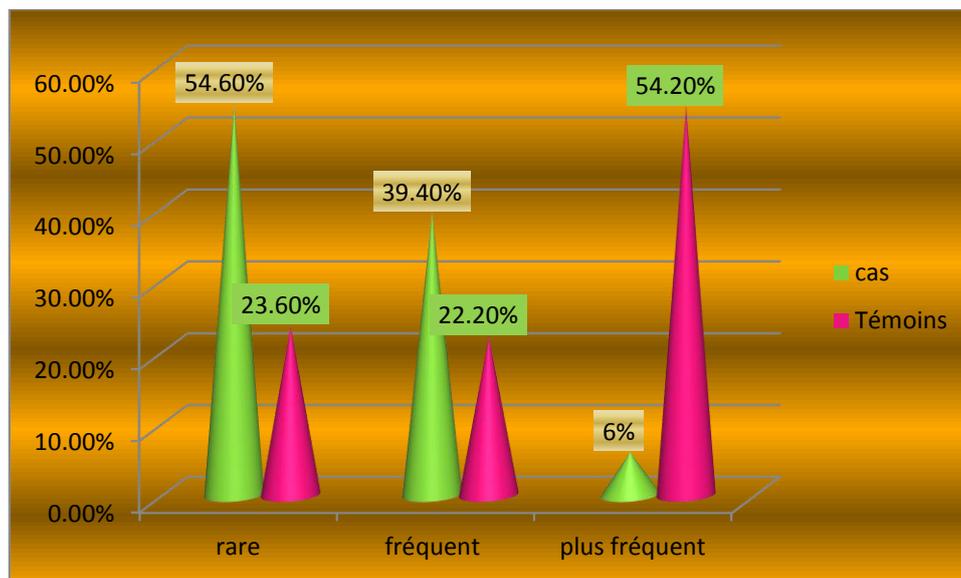
Les témoins qui consomment rarement les viandes représentent un pourcentage de 40.30%, tandis que ceux qui mangent fréquemment les viandes représentent une faible moyenne de 11, 60%.

Pour les cas on a obtenu une grande valeur de 39,40% des patientes qui consomment les viandes avec une énorme quantité, ces dernières souffrent d'un cancer du sein.

A partir de cette représentation on peut déduire que la grande consommation des viandes rouges pourrait augmenter les risques de contracter un cancer du sein.

Tableau 7 : la fréquence de consommation de poissons chez la population d'étude

		Effectifs	Pourcentages
Rare	Cas	118	54.6%
	Témoins	51	23.6%
fréquent	Cas	85	39.4%
	Témoins	48	22.2%
Plus fréquent	Cas	13	6%
	Témoins	117	54.2%

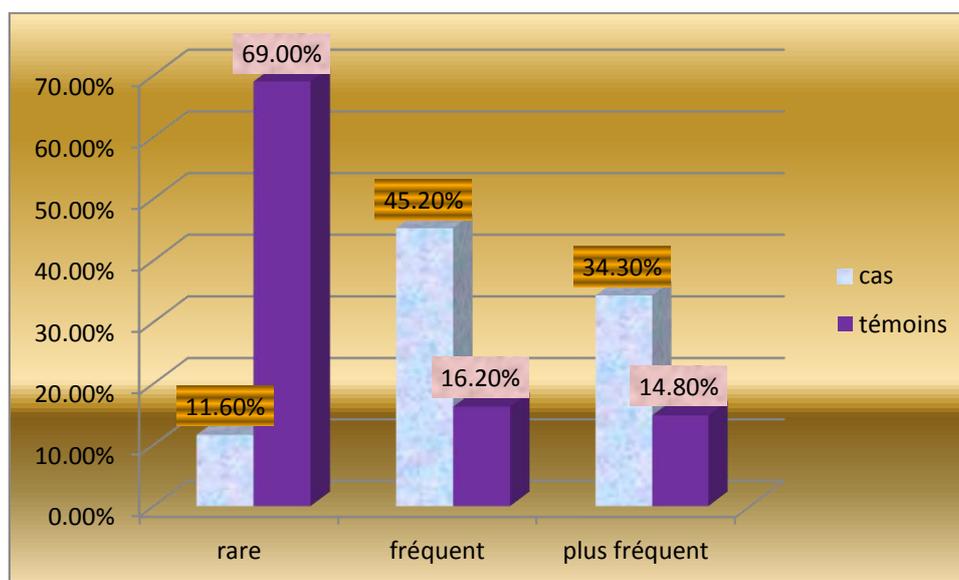


L'histogramme représente le taux de consommation des poissons chez les cas et témoins. Pour les témoins on remarque que la valeur des femmes qui consomment les poissons fréquemment représente 54,20%, ces témoins ne souffrent plus de cette maladie. Pour les cas on observe que la valeur de consommation de poisson pour celles qui ont un cancer du sein représente 6%.

Donc on constate que la consommation des poissons deux fois par semaine réduirait le taux de développer un cancer du sein.

Tableau 8 : la fréquence de consommation de conserves chez la population d'étude

		Effectifs	Pourcentages
Rare	Cas	25	11.6%
	Témoins	149	69.0%
fréquent	Cas	117	45.2%
	Témoins	35	16.2%
Plus fréquent	Cas	74	34.3%
	Témoins	32	14.8%

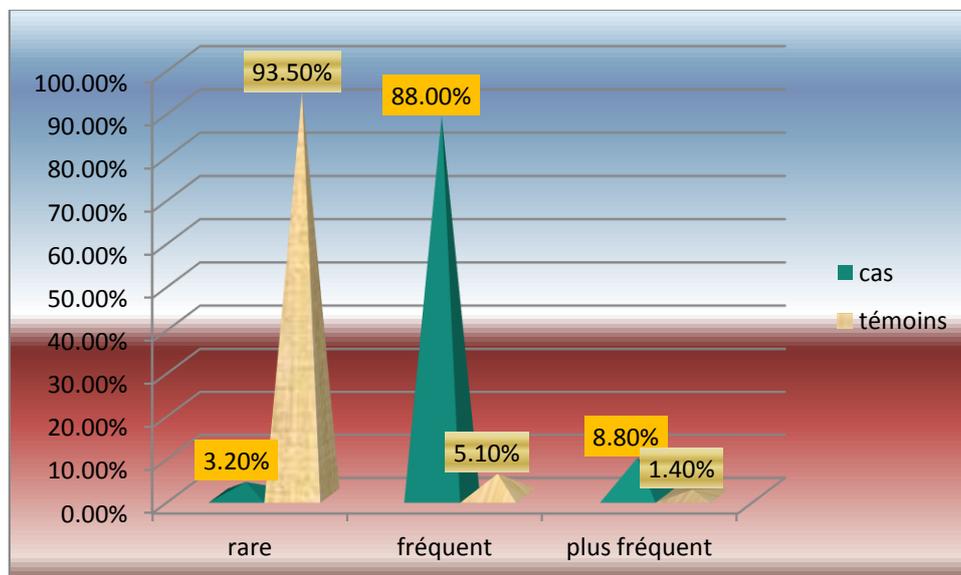


Il est figuré dans ce graphique un taux très élevé de 80% des cas qui se nourrissent des aliments conservés, ces femmes les consomment presque tous les jours. Comme il apparaît clairement une valeur de 69% des témoins qui prennent rarement ces produits.

Donc on peut déduire qu'il existe une matière dans les boîtes de conserves engendrant cette grave tumeur.

Tableau 9 : la fréquence de consommation de gras chez la population d'étude

		Effectifs	Pourcentages
Rare	Cas	7	3.2%
	Témoins	202	93.5%
fréquent	Cas	190	88.0%
	Témoins	11	5.1%
Plus fréquent	Cas	19	8.8%
	Témoins	3	1.4%

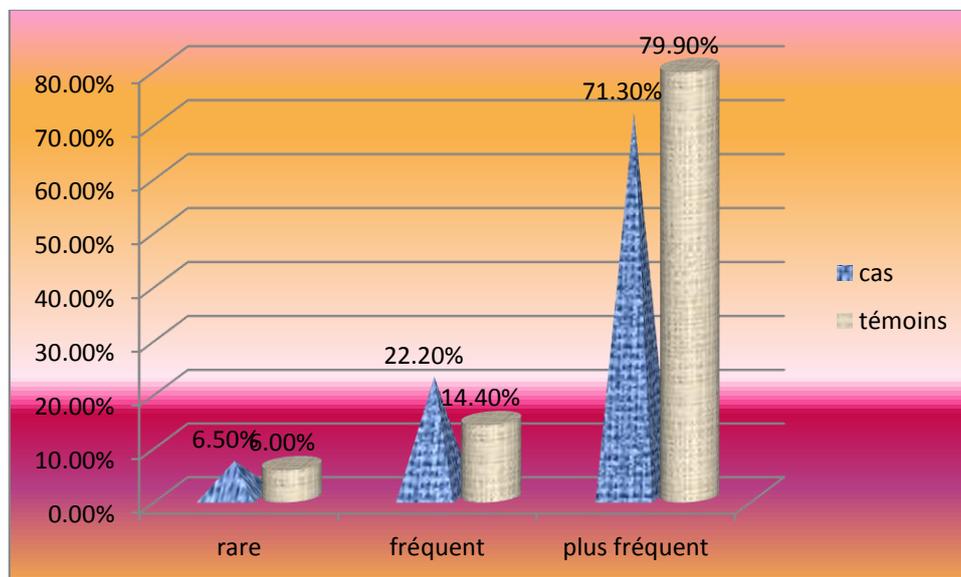


Dans cette représentation graphique on remarque qu'un grand nombre des témoins ne mangent plus les gras, le taux de ces derniers augmente avec la diminution de la consommation des gras en rétrogradant d'une valeur de 93,50% à 1,40%. Quant à les cas il est bien mentionné une valeur de 88% à cause de la consommation fréquenté des gras.

Cette étude indique que le risque de développer un cancer du sein est presque doublé chez les femmes dont le sang renferme une grande quantité des gras.

Tableau 10 : la fréquence de consommation de viandes de volaille chez la population d'étude

		Effectifs	Pourcentages
Rare	Cas	14	6.5%
	Témoins	13	6.0%
fréquent	Cas	48	22.2%
	Témoins	31	14.4%
Plus fréquent	Cas	154	71.3%
	Témoins	172	79.6%

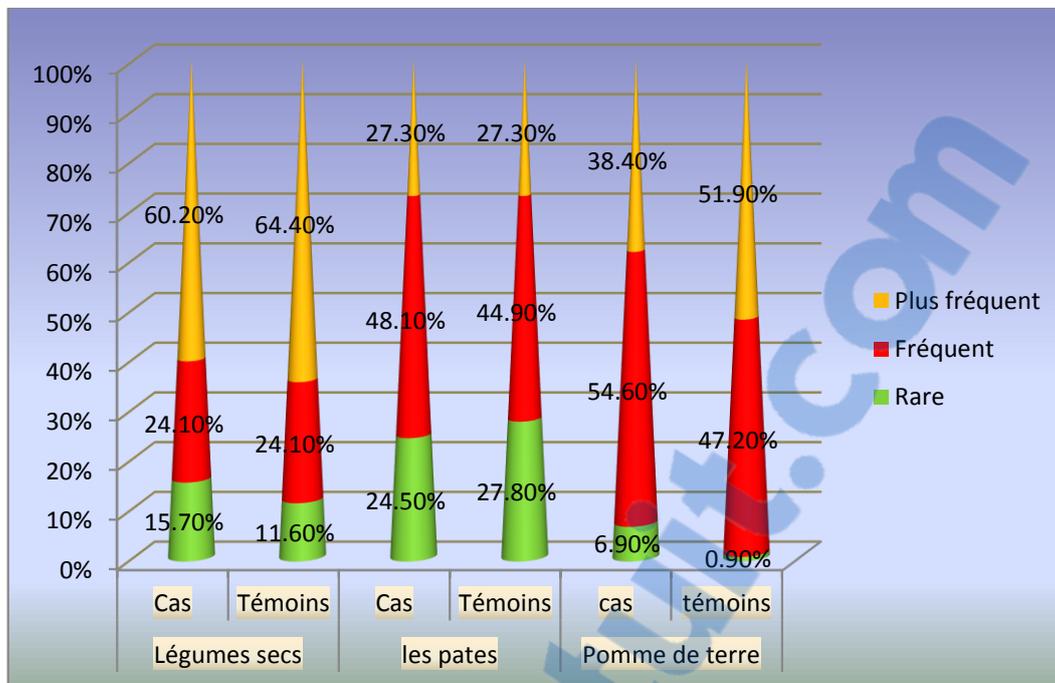


La représentation graphique montre que le taux des patients qui consomment fréquemment les viandes de volaille est presque le même chez les cas et les témoins, (les cas : 71,30% / les témoins : 79,90%).

En somme les viandes de volaille n'ont aucun effet sur l'apparition du cancer du sein

Tableau 11 : la fréquence de consommation de légumes secs, pomme de terre et les pâtes chez la population d'étude

		Rare	Fréquent	Plus fréquent
Légumes secs	Cas	15,70%	24,10%	60,20%
	Témoins	11,60%	24,10%	64,40%
les pâtes	Cas	24,50%	48,10%	27,30%
	Témoins	27,80%	44,90%	27,30%
Pomme de terre	cas	6,90%	54,60%	38,40%
	témoins	0,90%	47,20%	51,90%



On ce qui concerne les trois types de consommation des légumes secs, on constate que tous les pourcentages sont les mêmes chez les cas et témoins, ainsi la même chose pour les pâtes et les pommes de terres.

On peut conclure que les légumes, les pâtes et les pommes de terre n'ont aucune influence sur l'engendrement du cancer du sein.

2-ANALYSE BI VARIEE :

Tableau 12: Association entre le risque de cancer du sein et les viandes rouges

Tableau croisé

			cancer		Total
			cas	témoins	
Viande rouge	Rare	Effectif	61	87	148
		Effectif théorique	74.0	74.0	148.0
	Fréquent	Effectif	70	104	174
		Effectif théorique	87.0	87.0	174.0
	plus fréquent	Effectif	85	25	110
		Effectif théorique	55.0	55.0	110.0
Total		Effectif	216	216	432
		Effectif théorique	216.0	216.0	432.0

Khi-deux de Pearson	Khi-deux théorique	Degré de liberté
43.939	5.991464547	2

La valeur de χ^2 observée est à la fois élevée et supérieure à la valeur critique correspondant au seuil de signification statistique de 5% et avec un degré de liberté qui égal à 2. ce résultat nous permet de rejeter l'hypothèse nulle, autrement dit le cancer et la consommation de viande sont indépendants, dont on peut conclure qu'il existe bien une relation entre le cancer et la consommation de la viande dans la population observée.

Tableau 13 : Association entre le risque de cancer du sein et les conserves

Tableau croisé

			CANCER		Total
			cas	témoins	
CONSERVES	Rare	Effectif	25	149	174
		Effectif théorique	87.0	87.0	174.0
	Fréquent	Effectif	117	35	152
		Effectif théorique	76.0	76.0	152.0
	pus fréquent	Effectif	74	32	106
		Effectif théorique	53.0	53.0	106.0
Total	Effectif	216	216	432	
	Effectif théorique	216.0	216.0	432.0	

Khi-deux de Pearson	Khi-deux théorique	Degré de liberté
149.246	5.991464547	2

La statistique de test χ^2 calculé valant 149.246 est largement supérieure à 6 (χ^2 théorique) ; nous pouvons donc très franchement rejeter l'hypothèse nulle que la variable est distribuée uniformément avec un risque d'erreur de 5%. On conclut qu'il y a un lien entre la consommation des conserves et le cancer du sein.

Tableau 14: Association entre le risque de cancer du sein et le sucre

Tableau croisé

			cancer		Total
			cas	témoins	
Sucre	Rare	Effectif	20	25	45
		Effectif théorique	22.5	22.5	45.0
	Fréquent	Effectif	25	33	58
		Effectif théorique	29.0	29.0	58.0
	plus fréquent	Effectif	171	158	329
		Effectif théorique	164.5	164.5	329.0
Total	Effectif	216	216	432	
	Effectif théorique	216.0	216.0	432.0	

Khi-deux de Pearson	Khi-deux théorique	Degré de liberté
2.173	5.991464547	2

La statistique de test χ^2 calculé valant 2.173, elle est inférieure à χ^2 théorique=5.99; nous ne pouvons donc pas rejeter l'hypothèse que les deux variables «le cancer» et « le sucre » sont indépendantes. Autrement dit, avec une marge d'erreur de 5%, nous pouvons dire que les deux variables «le cancer» et « sucre » sont indépendantes.

Tableau 15 : Association entre le risque de cancer du sein et le chocolat

Tableau croisé

			cancer		Total
			cas	témoins	
chocolat	Rare	Effectif	179	172	351
		Effectif théorique	175.5	175.5	351.0
	fréquent	Effectif	27	36	63
		Effectif théorique	31.5	31.5	63.0
	plus fréquent	Effectif	10	8	18
		Effectif théorique	9.0	9.0	18.0
Total	Effectif	216	216	432	
	Effectif théorique	216.0	216.0	432.0	

Khi-deux de Pearson	Khi-deux théorique	Degré de liberté
1.648	5.991464547	2

χ^2 calculé (= 1.648) < à χ^2 théorique (= 5,991) donc nous pouvons affirmer avec un risque d'erreur maximal de 5% qu'il y'a une indépendance entre le cancer et le chocolat

Tableau 16 : Association entre le risque de cancer du sein et le beurre

Tableau croisé

			cancer		Total
			cas	témoins	
Beurre	Rare	Effectif	125	125	250
		Effectif théorique	125.0	125.0	250.0
	fréquent	Effectif	62	59	121
		Effectif théorique	60.5	60.5	121.0
	plus fréquent	Effectif	29	32	61
		Effectif théorique	30.5	30.5	61.0
Total	Effectif	216	216	432	
	Effectif théorique	216.0	216.0	432.0	

Khi-deux de Pearson	Khi-deux théorique	Degré de liberté
0.222	5.991464547	2

χ^2 calculé=0.222 et χ^2 théorique=5.99 donc χ^2 calculé < χ^2 théorique ; on ne rejette pas l'hypothèse nulle au seuil de signification de 5% donc on peut conclure qu'il ya une indépendance entre le cancer du sein et le beurre

Tableau 17 : Association entre le risque de cancer du sein et le lait

Tableau croisé

			Cancer		Total
			Cas	témoins	
Lait	Rare	Effectif	26	27	53
		Effectif théorique	26.5	26.5	53.0
	Fréquent	Effectif	76	70	146
		Effectif théorique	73.0	73.0	146.0
	plus fréquent	Effectif	114	119	233
		Effectif théorique	116.5	116.5	233.0
Total	Effectif	216	216	432	
	Effectif théorique	216.0	216.0	432.0	

Khi-deux de Pearson	Khi-deux théorique	Degré de liberté
0.373	5.991464547	2

La statistique $\chi^2_{obs}=0.373$ est inférieure à la valeur critique $\chi^2_{0.05;2}=5.99$, par conséquent on accepte l'hypothèse nulle au seuil $\alpha=0.05$ qui correspond à l'indépendance entre le lait et la tumeur.

Donc on peut conclure que le lait n'a pas d'effet sur le cancer du sein.

Tableau 18: le test d'hypothèse de comparaison de moyennes d'âge des cas et des témoins

	<i>Témoins</i>	<i>Cas</i>
Moyenne	48.2638889	48.4444444
Variances	120.67	123.64
Observations	216	216

La statistique Z	-0.16977241
Valeur critique de z (unilatéral)	1.64485363

La statistique $Z = -0.16977241 > -Z_{0.05} = -1.64485363$, par conséquent on accepte l'hypothèse nulle au seuil $\alpha=0.05$.

On conclut à partir de ce test que la moyenne d'âge des patientes qui ont un cancer du sein est la même que celle des femmes qui n'ont pas le cancer du sein.

Discussion

1. Quelques déductions suite aux résultats de l'analyse :

Dans cette étude cas-témoins, nous avons trouvé plusieurs facteurs de risque associés à la survenue du cancer du sein chez les femmes, En effet, les résultats ont montré que chez les femmes âgées de plus de 40 ans avaient plus de risque de développer le cancer du sein que celles âgées de 20 à 39 ans.

On a constaté également que les femmes allaitantes leurs enfants une durée très courte avaient plus de risque d'avoir un cancer du sein que celles allaitantes deux ans.

Notre étude a montré aussi que le tabac n'avait aucun effet significatif sur le risque de cancer du sein. Néanmoins l'exposition au tabagisme indirect (tabagisme passif) augmente légèrement le risque de cancer du sein.

Les femmes de la population générale ayant eu leur première grossesse après l'âge de 30 ans ont un risque plus élevé de contracter un cancer du sein comparativement aux femmes ayant eu leur première grossesse avant l'âge de 26 ans

Le risque de développer un cancer du sein augmente aussi en fonction des antécédents familiaux. Ainsi, le risque de développer un cancer du sein est plus grand si votre mère ou une de vos sœurs ont eu un cancer du sein.

Parmi les facteurs qui influencent le risque de développer un cancer, l'alimentation semble jouer un rôle non négligeable

Les résultats de notre enquête ont montré que la consommation d'excès de conserves dans les aliments est très significativement liée à la maladie avec un risque très élevé.

Une consommation régulière de poissons réduirait le risque de développer un cancer du sein

Notre étude montre que manger quelques morceaux de chocolat noir par jour peut diminuer les risques de cancer.

La consommation de divers laits et produits laitiers, dont le yogourt, le fromage et le beurre, n'est pas associée à un risque accru de cancer du sein et pourrait en fait être associée à une réduction du risque de cancer du sein

2. Quelques conseils pour lutter contre le cancer du sein:

Les causes du cancer du sein sont multiples et parfaitement identifiées. Elles se surajoutent et leur importance varie d'une femme à l'autre selon la durée d'exposition à telle ou telle cause.

Voici donc les principales recommandations que l'on peut retenir à cet égard pour réduire le risque de cancer du sein.

- **Consommer autant que possible, au moins 5 fruits et légumes (400g) par jour :**
Les fruits et les légumes jouent un rôle protecteur grâce à leur richesse en fibres, vitamines et minéraux antioxydants.
- **Limiter la consommation de viande rouge :** la consommation excessive de viande rouge favorise le cancer du sein, de plus une alimentation riche en graisses animales a souvent une teneur élevée en calorie, ce qui augmente le risque de prise de poids. C'est pourquoi il est conseillé de limiter sa consommation de viande rouge (moins de 500 g par semaine et de privilégier plutôt les volailles et certains poissons.
- **Surveiller son poids :** l'excès de poids et l'obésité augmentent le risque de développer le cancer du sein chez les femmes ménopausées. le maintien d'un poids optimal tout au long de la vie permettrait donc de réduire la probabilité de développer le cancer du sein.
- **Faire du sport :** une activité physique régulière réduit de façon significative le risque de développer un cancer du sein. 4 heures de marche ou de vélo par semaine seraient suffisantes pour faire baisser de façon significative le risque de cancer.
- **Allaiter vos enfants :** les risques du cancer du sein diminuent de 3% après un allaitement d'une durée de 6 mois et de 4,5% lorsque la durée de l'allaitement d'une ou plusieurs enfants est 12 mois. en d'autre mot, le nombre total de mois d'allaitement qui compte et non le nombre d'enfants allaités. Pour bénéficier d'une protection minimale allaiter vos enfants au moins 6 mois.
- **Ne pas prendre les contraceptifs oraux pendant une durée longue :** la prise de contraceptifs oraux accroît très légèrement le risque de cancer du sein chez les femmes qui les utilisent durant plus de 5 ans.

conclusion

Cette expérience de stage constitue pour moi une phase importante dans ma formation, elle me permet l'exploration et l'ouverture dans le monde professionnel.

Ce stage a été pour moi très enrichissant, J'ai appris à appliquer et exploiter mes compétences acquises en probabilités et statistiques à des problèmes réels.

J'ai appris aussi à manipuler un nouveau logiciel de statistique

J'ai eu l'occasion durant ce stage de mettre en avant les compétences d'analyses et de traitements des données acquises durant ma formation.

Pour conclure cette étude a révélé un certain nombre de facteurs de risque associés à la survenue du cancer du sein chez les femmes, à savoir la grossesse tardive, l'absence d'allaitement maternel, le tabagisme, l'âge, les antécédents familiaux, et la nutrition.....

Références bibliographiques

- Polycopie de probabilité et statistiques Pr MME Ezzaki, FST de Fès
- Analyse des données et statistique, G.saporta
- le cancer du sein, wikipédia