République démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Aboubeker Belkaide –Tlemcen
Faculté des sciences
Département de mathématiques
Mémoire de licence

Thème :

Le test du khi-deux

Présenter par :

Derkaoui Nadjia

Boumediene Ibtissam

Encadrer par :

Mme . Ben yelles

2009- 2010

Introdoction

Un test d'hypothèse est un procédé d'inférence permettant de contrôler (accepter ou rejeter) ou partir de l étude d'un ou plusieurs échantillons aléatoires , la validité d'hypothèses relatives à une ou plusieurs population.

En fonction de d'hypothèses testés plusieurs types de testes peuvent être réalisés:

*Les tests destinés à vérifier si un échantillon peut être considère comme extrait d'une population donnée , vis-à-vis d'un paramètre comme la moyenne ou la fréquence observée (test de conformité) ou par rapport à sa distribution observée (test d' ajustement).

*Les tests destinés à comparer plusieurs populations à l'aide d'un équivalent d'échantillon (test d'homogénéité)…etc. Mais pour faire ces tests, il faut utilise un parmi plusieurs des type comme le test γ, le test de R^2 ,et le test de X^2 .

Dans ce travail nous avons étudie le test khi-deux qui l'un des principaux tests appliqués pour le prise de décision les tests d'hypothèses.

Dans le chapitre I, nous définissions les tests d'hypothèses dans le cas générale pour pouvoir permettre une décision en évaluant les risques .

Dans le chapitre II, on s'intéresse à la loi du khi- deux : on définie T_i n variables aléatoires qui suit la loi normale centrée réduite,

On pose: $X = \sum_{k=1}^{n} T_i^2$ alors suit la loi de khi-deux à n degré de liberté puis faisons la preuve du théorème fondamentale du test du khi-deux .

Le chapitre III, est étudie d'une part de la conformité des paramètres d'un échantillon à l' aide du test de khi-deux .Et toute la théorie sera illustrée par un exemple sur la génétique .

D'autre part, on va s'intéresse à l'homogénéité de deux échantillons par le biais du test khi-deux et de la même manière nous explicitions le concept par un exemple, ce qui achèvera le travail.

Sweet

Chapitre I:

Généralités sur les tests

1-Les hypothèses:

Les statistiques développent des techniques et des méthodes qui permettent d'analyser les donnée issues des l'observation, afin de cerner les caractéristiques de la population concernée et d'identifier un modèle capable d'engendrer ces données.

Dans ce cadre, on est amené à faire des hypothèses, c'est-à-dire à émettre des assertions concernant ces caractéristiques de la population on ce modèle.

Le plus souvent, la situation se résume en une alternative constituée de deux hypothèses H_0 et H_1 qui s'excluent mutuellement et qui sont appelées respectivement l'hypothèse nulle, ou fondamentale, et l'hypothèse alternative, ou contraire.

En général, les hypothèses H_0 et H_1 ne jouent pas des rôles symétriques, et on choisit pour hypothèse nulle H_0 l'hypothèse à laquelle on croit ou on tient, ou encore celle qui permet de faire des calcules, ou encore celle dont le rejet est lourd de conséquences.

2-Règle de décision :

Un test d'hypothèses est une règle de décision qui permet, sur la base des données observées et avec des risques d'erreur déterminés ,d'accepte ou de refuser une hypothèse statistique. Elle est définie sous l'hypothèse « H_0 est vraie » et pour un seuil de signification α fixé.

-si la valeur de la statistique S calculée (S_{obs}) est supérieure à la valeur seuil(S_{seuil}) c'est-à-dire :S_{obs} > S_{seuil} alors l'hypothèse H_0 est rejetée au risque d'erreur α et l'hypothèse H_1 est acceptée.

-si la valeur de statistique S calculée (S_{obs}) est inférieure à la valeur seuil (S_{seuil}) c'est-à-dire :$S_{obs} \leq S_{seuil}$ alors l'hypothèse H_0 ne peut être rejetée.

3-les erreurs et les risques :

La règle de décision d'un test étant basé sur l'observation d'un échantillon, on n'est jamais sur de l'exactitude de la conclusion : il y a donc toujours un risque d'erreur.

L'erreur de première espèce consiste à rejeter H_0 à tort : le risque d'erreur de première espèce est noté , c'est le risque d'erreur que l'on prend en rejetant H_0 alors qu'elle est vraie. On l'appelle aussi le niveau du test.

L'erreur de deuxième espèce consiste à rejeter H_1 à tort : le risque d'erreur de deuxième espèce est noté β , c'est le risque d'erreur que l'on prend en rejetant H_1 alors qu'elle est vraie.

Les risques lies aux tests d'hypothèses peuvent se résumer ainsi :

	SITUATION VRAIE			
	H_0 est vraie		H_1 est vraie	
	La Décision est	Probabilité de prendre cette décision avant expérience	La Décision est	Probabilité de prendre cette décision avant expérience

Conclusion du test	*Accepter* H_0	*bonne*	*1-α*	*fausse*	*β(risque de deuxième espèce)*
	Rejeter H0	*fausse*	*α(risque de première espèce)*	*bonne*	*1-β*

Remarque :

La probabilité complémentaire du risque de deuxième espèce (1-β) définit la puissance du test à l'égard de la valeur du paramètre dans l'hypothèse alternative H_1 *.*

La puissance du test représente la probabilité de rejeter l'hypothèse nulle H_0 *lorsque l'hypothèse vraie est* H_1 *plus* β *est petit, plus le test est puissant.*

Sweet

Chapitre II:

La loi du khi-deux

I. Loi du khi deux à un degré de liberté:

i. Définition :

Soit X une variable aléatoire suit la loi normale centrée réduite N (0,1)

, la variable aléatoire définie par : $Y = X^2$ suit la loi du khi-deux à un degré de liberté.

ii. Densité de probabilité :

La densité de probabilité de X est la fonction définie sur R par :

$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$ *Donc, la densité de probabilité de Y est définie*

Par: $g(x) = \frac{1}{\sqrt{2\pi}} x^{\frac{-1}{2}} \exp\left(\frac{-x}{2}\right)$ *si x>0*

Pour calculer $g(x)$, on fait:

Soit F et G deux fonctions de répartition des variables aléatoires X et Y respectivement pour tout réel x, $G(x) = P(Y \leq x)$, $donc$ $G(x) = P(X^2 \leq x)$

.si x<0, alors l'évènement$(X^2 \leq x)$ est impossible donc

$G(x) = 0$

.si x≥0, alors $(X^2 \leq x)$ équivaut à $-\sqrt{x} \leq X \leq \sqrt{x}$ donc $G(x) = P(-\sqrt{x} \leq X \leq \sqrt{x})$ c'est-à-dire $G(x) = F(\sqrt{x}) - F(-\sqrt{x})$

On conclue:

.si $x\leq 0$, $G(x) = 0$

.si $x>0$, $G(x) = F(\sqrt{x}) - F(-\sqrt{x})$

On passer à la fonction de densité:

.si $x\leq 0$, $g(x)=0$

.si $x>0$, $g(x) = \frac{1}{2\sqrt{x}}|f(\sqrt{x}) + f(-\sqrt{x})|$

C'est –à-dire $$g(x) = \frac{1}{\sqrt{2\pi}}\frac{1}{2\sqrt{x}}[e^{\frac{-(\sqrt{x})^2}{2}} + e^{\frac{-(-\sqrt{x})^2}{2}}]$$

Donc: $g(x) = \frac{1}{\sqrt{2\pi}} x^{\frac{-1}{2}} \exp\left(\frac{-x}{2}\right)$ si $x>0$.

iii. Espérance mathématique:

Par définition $E(Y) = E(X^2)$ d'après le théorème de könig

$V(X) = E(X) - E(X^2)$ Donc $E(Y) = V(X) + (E(X))^2$

C'est-à-dire $E(Y)=1$.

iv. Variance mathématique:

On a $V(Y) = E(Y) - E(Y^2) = E(Y^2) - 1$, il reste de calculer $E(Y^2)$:

Alors $$E(Y^2) = \int_{-\infty}^{+\infty} x^2\, g(x)dx = \frac{1}{\sqrt{2\pi}}\int_0^{+\infty} x^{\frac{3}{2}}\ e^{\frac{-x}{2}}\, dx$$

En intégrant par parties:

On utilise se changement de variable:

$$U = x^{\frac{3}{2}} \qquad , \qquad dU = \frac{3}{2}x^{\frac{1}{2}}dx$$

$$dV = e^{\frac{-x}{2}}dx \qquad , \qquad V = -2e^{\frac{-x}{2}}$$

Alors:

$$E(Y^2) = \frac{1}{\sqrt{2\pi}}\left[\left(-2x^{\frac{3}{2}}e^{\frac{-x}{2}}\right) + 3\int_0^{+\infty} x^{\frac{1}{2}}\ e^{\frac{-x}{2}}\, dx\right|$$

$$= \frac{3}{\sqrt{2\pi}} \int_0^{+\infty} x^{\frac{1}{2}} e^{\frac{-x}{2}} dx$$

$$= 3E(Y)$$

Donc: $E(Y^2) = 3$ et finalement: $V(Y) = 2.$

II. Loi du khi deux à n degré de liberté:

1. Définition:

$T_1, T_2, \ldots, T_n$ Désignant n variables aléatoires indépendantes qui suivent toutes la loi normale centrée réduite, soit X la variable aléatoire définie par: $X = \sum_{k=1}^{n} {T_k}^2$,On dit que la variable aléatoire X suit la loi du khi-deux à n degrés de liberté

2. Densité de probabilité:

D'après la densité de la loi de khi-deux à n degré de liberté ;

On a $g(x) = \frac{1}{\sqrt{2\pi}} x^{\frac{-1}{2}} e^{\frac{-x}{2}}$ *. En remarque que: $g(x)$ la densité d'une variable aléatoire suit la loi Gamma $(\frac{1}{2}, \frac{1}{2})$ alors la densité de variable aléatoire X suit la loi de Gamma $(\frac{1}{2}, \frac{n}{2})$ est :*

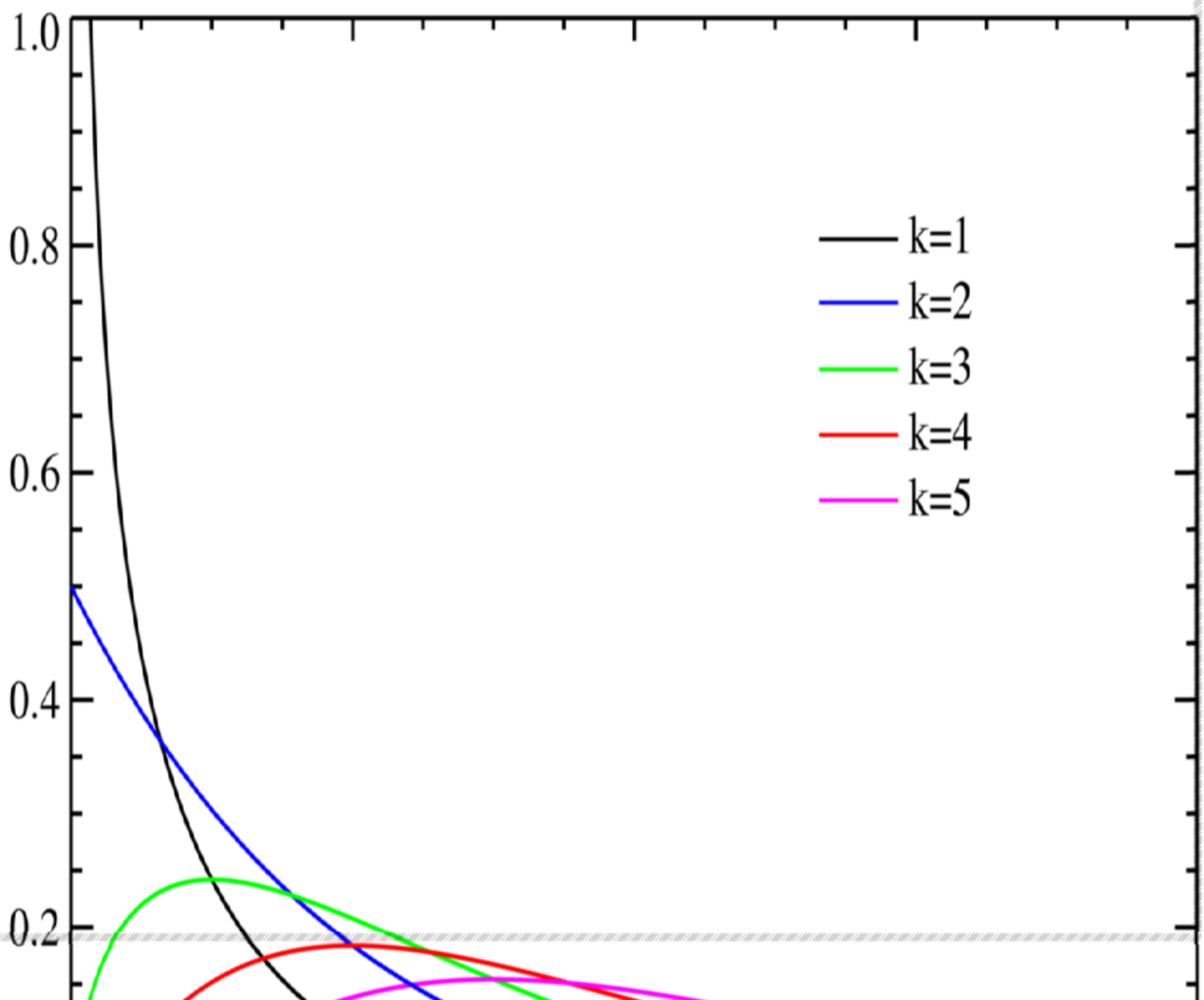
$g(x) = \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} x^{n-2/2} e^{\frac{-x}{2}}$ *alors, cette densité est le même que la densité d'une variable X suit la loi de khi-deux à n degré de liberté.*

3. Espérance et variance mathématique:

La variable aléatoire X est la somme de n variables indépendantes suivant toutes une loi de khi-deux à un degré de liberté d'où :$E(X)=n$ et $V(X)=2n$.

III. Les courbes de khi-deux :

A chaque valeur de n correspond, un type de courbe pour la densité de la loi du khi-deux



Graphe n 01

Remarque:

Si la variable aléatoire X suit une loi du khi-deux à n degrés de liberté, la table donne, pour un risque α choisi, le nombre X_α^2 tel que: $P(X \geq X_\alpha^2) = \alpha$

IV. Théorème fondamentale de test du khi-deux :

Théorème 1:

Si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires normales indépendantes, la variable aléatoire $S=X_1^2+X_2^2+\ldots+X_n^2$ suit une loi khi-deux à n degrés de liberté.

Théorème 2:

Sous l'hypothèse H_0, $(X_1, X_2, \ldots, X_n)$ est un échantillon d'une loi entièrement spécifié alors la statistique : $X_{\theta_0}^2 = \sum_{j=1}^{k} \frac{(N_j - nP_j(\theta_0))^2}{nP_j(\theta_0)}$ *à pour loi asymptotique la loi* X_{k-1}^2.

Preuve:

A/ Montrons tout d'abord que les variables aléatoires $N_1, N_2, \ldots, N_k$ obéissent à la loi multinomiale à s'avoir :

$$P(N_1 = n_1, \dots, N_k = n_k) = n! \prod_{j=1}^{k} \frac{{P_j}^{n_j}}{n_j!} \qquad ou \quad P_j = P_j(\theta_0)$$

Soient X la variable aléatoire étudiée ($X_1, X_2, \dots, X_n$) un n échantillon de X ,et μ la mesure de probabilité de X lorsque H_0 est vraie.

On partage l'ensemble des valeurs X (Ω) en k classes (Cj) Avec $1 \leq j \leq k$

Si X est une variable aléatoire discret, les C_j sont en générale des points.

Si X est une variable aléatoire continue, les C_j sont en générale des intervalles ou des produits des intervalles.

Pour tout indice j de 1 à k, on note P_j la probabilité théorique de la classe C_j donnée la loi de μ avec $P_j=P(X \in C_j)$.

On note N_j le nombre de variable X_j prenant leur valeur dans C_j ; si H_0 est vraie.

On suppose que :

$$Y_{jl} = \begin{cases} 1 & si \quad X_j \in C_j \\ O & sinon \end{cases}$$

Alors $N_j = \sum_{l=1}^{n} Y_{jl}$ *et on a* $P_{\theta_0}(Y_{jl} = 1) = P_j$ *l étant fixé , la fonction caractéristique de ($Y_{1l}, Y_{2l}, \dots, Y_{kl}$) est :*

$E(\exp(i \sum_{j=1}^{k} t_j Y_{jl}))$

Si $Y_{jl}=1$ i.e $X_l \in C_j$ .donc tous les autres Y_{ml} sont nuls est cet événement a pour probabilité P_j et pour conséquent la fonction caractéristique pour l fixé est : $\sum_{j=1}^{k} P_j \exp(it_j)$

Les Y_{jl} pour les valeurs différentes de l sont indépendants d'où la fonction caractéristique de l'ensemble de Y_{jl} es :

$\prod_{l=1}^{n} (\sum_{j=1}^{k} P_j \exp(it_j))$ *donc, la fonction caractéristique de* $(N_1, \dots, N_k)$ *est:*

$$E\left(\exp\left(i \sum_{j=1}^{k} N_j t_j \right) \right) = E\left(\exp\left(i \sum_{j,l} Y_{jl} t_j \right) \right)$$

$$= \prod_{l=1}^{n} (\sum_{j=1}^{k} P_j \exp(it_j))$$

$$= (\sum_{j=1}^{k} P_j \exp(it_j))^n$$

Qui est la fonction caractéristique de la loi multinomiale

Conclusion :

$$P(\, N_1 = n_1, \dots, N_k = n_k \,) = n! \prod_{j=1}^{k} \frac{P_j^{n_j}}{n_j!}$$

B/Montrons que lorsque n tend vers l'infini, la loi de probabilité des variables $U_j = \frac{N_j - nP_j}{\sqrt{nP_j}}$ $\quad j = 1, \dots, k,$ *tend vers une loi normale N(0,1).*

En effet, la fonction caractéristique des U_j *est :*

$$E\left(\exp\left(i \sum_{j=1}^{k} t_j U_j\right)\right) = E\left(\exp\left(i \sum_{j=1}^{k} t_j \frac{N_j - nP_j}{\sqrt{nP_j}}\right)\right)$$

$$= \exp\left(-i \sum_{j=1}^{k} \sqrt{nP_j}\, t_j\right) E\left(\exp\left(i \sum_{j=1}^{k} \frac{t_j}{\sqrt{nP_j}} N_j\right)\right)$$

$$= \exp\left(-i \sum_{j=1}^{k} \sqrt{nP_j}\, t_j\right) \left[\sum_{j=1}^{k} P_j \exp\left(i \frac{t_j}{\sqrt{nP_j}}\right)\right]^n$$

Ecrivons le développement limité à l'ordre 2 en $\frac{1}{\sqrt{n}}$:

On a :

$$\exp\left(i \frac{t_j}{\sqrt{nP_j}}\right) = 1 + i \frac{t_j}{\sqrt{nP_j}} + i^2 \frac{t_j^2}{2nP_j} + \frac{1}{n} \varepsilon\left(\frac{1}{n}\right) \quad \textit{avec} \quad \varepsilon\left(\frac{1}{n}\right) \text{----} \to 0$$

$$= 1 + i \frac{t_j}{\sqrt{nP_j}} - \frac{t_j^2}{2nP_j} + \frac{1}{n} \varepsilon\left(\frac{1}{n}\right)$$

D'où $\sum_{j=1}^{k} P_j \exp\left(i \frac{t_j}{\sqrt{nP_j}}\right) = \sum_{j=1}^{k} P_j \left[1 + i \frac{t_j}{\sqrt{nP_j}} - \frac{t_j^2}{2nP_j} + \frac{1}{n} \varepsilon\left(\frac{1}{n}\right)\right]$

$$= 1 + i \sum_{j=1}^{k} t_j \sqrt{\frac{P_j}{n}} - \frac{1}{2n} \sum_{j=1}^{k} t_j^2 + \frac{1}{n} \varepsilon\left(\frac{1}{n}\right)$$

De même:

$$\exp\left(-i\sum_{j=1}^{k} t_j\sqrt{\frac{P_j}{n}}\right) = 1 - i\sum_{j=1}^{k} t_j\sqrt{\frac{P_j}{n}} - \frac{1}{2n}\left(\sum_{j=1}^{k} t_j\sqrt{P_j}\right)^2 + \frac{1}{n}\varepsilon\left(\frac{1}{n}\right)$$

; à l'ordre 2 le produit vaut:

$$\exp\left(-i\sum_{j=1}^{k} t_j\sqrt{\frac{P_j}{n}}\right)\sum_{j=1}^{k} P_j \exp\left(i\frac{t_j}{\sqrt{nP_j}}\right) = \quad \ldots\ldots\ldots\ldots..(1)$$

$$(1) = 1 - \frac{1}{2n}\left(\sum_{j=1}^{k} t_j\sqrt{P_j}\right)^2 + \frac{1}{n}\left(\sum_{j=1}^{k} t_j\sqrt{P_j}\right)^2 - \frac{1}{2n}\sum_{j=1}^{k} t_j^2 + \frac{1}{n}\varepsilon\left(\frac{1}{n}\right)$$

$$(1) = 1 - \frac{1}{2n}\left[\sum_{j=1}^{k} t_j^2 - \left(\sum_{j=1}^{k} t_j\sqrt{P_j}\right)^2\right] + \frac{1}{n}\varepsilon\left(\frac{1}{n}\right)$$

Lorsque n tend vers l'infini, la puissance n de ce produit tend vers :

$\exp\left(-\frac{1}{2}\left[\sum_{j=1}^{k} t_j^2 - \left(\sum_{j=1}^{k} t_j\sqrt{P_j}\right)^2\right]\right)$ *Qui est la fonction caractéristique d'un vecteur d'une loi normale N(0,1)*

<u>Conclusion :</u>

Lorsque n tend vers l'infini, U_j *tend vers N(0,1) .*

Et puisque $\sum_{j=1}^{k} N_j = n$, *donc* $x^2(\theta_0)$ *est la somme des carrés de variable aléatoire centrée réduites, indépendantes asymptotiquement normale liées par une relation linéaire.*

Alors la loi asymptotique de $x^2(\theta_0)$ *est* x_{k-1}^2.

Sweet

Chapitre III:

Le test du khi-deux

Test de conformité

1- Généralité :

Les tests de conformité sont destinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée ou représentatif de cette population, vis-à-vis d'un paramètre comme la moyenne, la variance ou la fréquence observée. Ceci implique que la loi théorique du paramètre est connue avec niveau de la population.

En théorie, si l'on suppose connu la valeur θ_0 d'un paramètre relatif à la population (par exemple σ^2) et $\theta^{\wedge}$ un estimateur absolument correct de ⍰ *(par exemple :S^2), obtenu à partir d'un échantillon aléatoire simple de taille n, on cherche à savoir si l'échantillon est représentatif de la population pour le paramètre considère H_0 : ϑ=θ_0*

*principe du test :

On connait une distribution observée (résultat d'une expérience)

Valeur du caractère	**C_1**	**C_2**	**…**	**Ci**	**…**	**C_n**

Effectifs observés	O_1	O_2	…	Oi	..	O_n

On veut comparer cette distribution à une loi connue (binomiale, poisson, gausse, etc…) qui donnerait les effectifs théoriques suivantes :

Valeur du caractère	C_1	C_2	**…**	C_i	**…**	C_n
Effectifs théoriques	T_1	T_2	…	T_i	…	T_n

Pour faire un test, il faut passer par les étapes suivantes :

**** Première étape :***

Soit H_0 l'hypothèse nulle « les observation suivent la loi théorique »et H_1 l'hypothèse alternative contreH_0.

****Deuxième étape :***

On doit calculer les effectifs théorique T_i par la relation suivante :T_i=$\sum_{i=1}^{n}\frac{O_i}{N}$ telle que N=le nombre totale des classes et O_i est l'effectifs observé.

****Troisième étape :***

On mesure l'écart entre la distribution théorique et la distribution observée, on calcule pour chaque classe, le nombre

$$\frac{(O_i \ -T_i)^2}{T_i}$$

Ce nombre est appelé écart quadratique relatif

On définit alors la variable aléatoire S telle que :

S=$\sum_{i=1}^{n}\frac{(O_i \ -T_i)^2}{T_i}$ cette somme sera faible si les écarts entre les valeurs théoriques et les valeurs observées sont petits, telle sera grande dans le cas contraire.

2-Position du problème:

*Exemple:

On a croisé deux races de plantes différant par deux caractère A et B .

La première génération est homogène .la seconde génération fait apparaitre 4 types de plantes; dont les phénotypes est noté : AB, Ab, aB, ab.

Si les caractères se transmettent selon les lois de « Mendel », la proposition théorique: des 4 phénotypes sont 9/16,3/16,3/16 et 1/16.

Dans une expérience un échantillon de 160 plantes a donné:

AB/100, Ab/18, aB/24, ab/18.

Cette répartition est –elle conforme aux lois de « Mendel » au seuil de signification de 5%?

**** La réponse:***

On pose l'hypothèse nulle.

H_0: »la répartition observée est conforme aux loi de Mendel »

Pour calculer le khi-deux on établit le tableau suivant:

Phénotype	**AB**	**Ab**	**aB**	**ab**	**totale**
Proposition théorique	*9/16*	*3/16*	*3/16*	*1/16*	*1*
Effectif calculé(C_i)	*90*	*30*	*30*	*10*	*160*
Effectif observé(O_i)	*100*	*18*	*24*	*18*	*160*

On obtient :

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - C_i)^2}{C_i} = \frac{10^2}{90} + \frac{(-12)^2}{30} + \frac{(-6)^2}{30} + \frac{8^2}{10}$$

=12.51

Tq:

O_i *=les effectifs observés.*

C_i *=les effectifs calculent.*

Ici le nombre de degré de liberté=n-1=4-1=3.

Sur la table de khi-deux : on lit $X_{0.05}^2$*=7.815 avec α=0.05=le risque d'erreur. D'où* $X^2 > X_\alpha^2$*, l'hypothèse de conformité* H_0 *doit être rejetée au seuil de signification de 5%.*

Test d'homogénéité:

1. Généralité:

On prélève au hasard k échantillons de taille $n_1, n_2, n_3, \ldots n_k$ *de k populations. Les résultats du caractère observé dans chaque population sont ensuite classés Selon n modalités. Dans ce cas, les* n_i *associés aux k échantillons sont fixés .il s'agit de savoir si elles ont un comportement semblable en regard du caractère étudié. On rassemble les données dans un tableau suivant:*

	Population échantillonnées

Caractère observe selon r modalités		$j=1$	$j=2$	…	$j=i$	…	$j=k$
	$i=1$	n_{11}	n_{12}		n_{1i}		n_{1k}
	$i=2$	n_{21}	n_{22}		n_{2i}		n_{2k}
	…						
	$i=j$	n_{j1}	n_{j2}		n_{ji}		n_{jk}
	…						
	$i=r$	n_{r1}	n_{r2}		n_{ri}		n_{rk}
		$n_1=\sum_{i=1}^{r}$	$n_2=\sum_{i=1}^{r}$		$n_j=\sum_{i=1}^{r}$		$n_k=\sum_{i=1}^{r} n_{ik}$

2. Définition de test d'homogénéité:

Il s'agit de comparer les effectifs observés pour chaque modalité du caractère avec les effectifs théoriques sous l'hypothèse d'un répartition équivalente entre les k populations et ceci pour chaque modalité du caractère si nous notons P_{ij} la probabilité théorique pour qu'une unité statistique choisie au hasard dans la population présente la modalité i du caractère étudié ,on peut alors préciser les hypothèses de la façon suivante:

***Première étape**:*

$H_0: P_{i1}=P_{i2}=\ldots=P_{ik}$ pour $i=1\ldots r$

Soit encore les propositions d'individus présentant chaque modalité du caractère sont les mêmes dans les k populations .

H_1:$P_{ij1} \neq P_{ij2}$ pour au moins un i parmi 1,2,…,r et pour au moins deux j_1 et j_2 différents choisis parmi 1,2,…,k.

Soit encore : les propositions d'individus présentant chaque modalité du caractère ne sont pas identiques pour les populations pour au moins une modalité du caractère .

* **Deuxième étape**:

Sous l'hypothèse d'homogénéité deux populations, on doit comparer les effectifs observés aux effectifs théoriques.

Pour calculer les effectifs théoriques, il nous faut déterminer P_i la proposition d'individus associées à la modalité i et que l'on suppose identique dans les k populations.

On obtiendra une estimation de cette proportion en utilisant l'ensemble des données collectées ;on choisit donc:

$P_i = \frac{\sum_{j=1}^{k} n_{ij}}{\sum_{j=1}^{k} n_j}$, on déduire les effectifs théoriques de chaque classe grâce à la relation $n_{t_{ij}} = P_i . n_j$. Pour comparer les écarts entre ce qu'on observe et ce qui se passe sous l'hypothèse H_0, on considère la somme des écarts réduits de chaque classe ,à savoir la quantité :

$X^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(N_{ij} - n_{t_{ij}})^2}{n_{t_{ij}}}$ Cette variable aléatoire suit une loi de khi-deux mais quel est donc nombre de degrés de liberté ?

Calcule du nombre de degrés de liberté du khi-deux :

A priori, on a k r cases dans notre tableau donc (k r) degrés de liberté.

Mais il faut retirer à cette valeur ,le nombre de paramètres estimés ainsi que le nombre de relations entre les différents élément des cases.

On a estimé probabilité théorique à l'aide des valeurs du tableau (P_1, P_2,…, P_r) mais seulement (r-1) sont indépendantes puisqu'on impose la restriction:$\sum_{i=1}^{r} P_i = 1$ *.Par ces estimations, on a donc supprimé (r-1) degrés de libertés.*

Les effectifs de chaque colonne sont toujours liés par les relations: $\sum_{i=1}^{r} n_{ij} = n_j$ *et ces relations sont au nombre de k.*

Finalement, le nombre de degrés de libertés du khi-deux est:

n=kr-(r-1)=(k-1)(r-1)

****Troisième étape**:*

On impose à la zone d'acceptation de H_0 concernant valeur du khi-deux d'être un intervalle dont θ *est la borne inferieure.*

Il nous faut donc déterminer dans la table, la valeur maximale $x_{\alpha,n}^2$ *de l'écart entre les deux distributions imputable aux variations d'échantillonnage au seuil de signification* α *, c'est –à-dire vérifiant :*

$$P(x^2 > x_{\alpha,n}^2) = \alpha$$

****Quatrième étape**:*

On calcule la valeur x_0^2 *prise par* x^2 *dans l'échantillon*

-si la valeur x_0^2 *se trouve dans la zone de rejet, on dira que l'écart observé entre les k distributions est statistiquement significatif au seuil* α *.Cet écart est anormalement élèvé et ne permet pas d'accepter H_0*

-si la valeur x_0^2 *se trouve dans la zone d'acceptation, on dira que l'écart est imputable aux fluctuations d'échantillonnage.*

3. position de problème :

***Exemple**:

Un maladie est traité dans quatre hôpitaux différent .On a fait les observations suivantes:

	Cas de guérison	Cas de non guérison	Nombre total des maladies traites	Pourcentage de guérison
Hôpital 1	123	28	151	81.4
Hôpital 2	95	19	114	83.3
Hôpital 3	152	63	2151	70.6
Hôpital 4	132	53	185	71.3
Effectif total	502	163	665	75.6

Peut – on considérer que l'efficacité des 4 traitements est sensiblement la même? Autrement dit –peut-on attribuer au seuil hasard, Les divergences observées entre les pourcentages de guérison au taux de sécurité de 95%?

***La réponse**:*

On pose l'hypothèse nulle H_0 :«l'efficacité des 4 traitements est la même ».

Dans ce cas, on prend pour estimation du pourcentage théorique de guérison, le pourcentage global correspondant a l'ensemble des maladies traités:

$$P_0 = \frac{(123 + 95 + 152 + 132)}{(151 + 114 + 215 + 185)} = 0.756$$

Les effectifs théoriques des différentes classes au moyen de cette valeur P_0 dans le tableau suivant:

	Cas de guérison	**Cas de non guérison**	**Nombre total Des maladies traitées**	**Pourcentage de guérison**
Hôpital 1	*114*	*37*	*151*	*75.6*
Hôpital 2	*86*	*28*	*114*	*75.6*
Hôpital 3	*162*	*53*	*215*	*75.6*
Hôpital 4	*140*	*45*	*185*	*75.6*

Effectif totale	*502*	*163*	*665*	*75.6*

Par exemple :

L'effectif théorique des cas de guérison pour l'hôpital 1 est: 151×0.756=114.

L'effectif théorique des cas de non –guérison est:151-114=37

Nous avons :

$$X^2 = \frac{(123-114)^2}{114} + \frac{(95-86)^2}{86} + \frac{(152-162)^2}{162} + \frac{(132-140)^2}{140} + \frac{(28-37)^2}{37} + \frac{(19-28)^2}{28} + \frac{(63-53)^2}{53} + \frac{(53-45)^2}{45}$$

$$X^2 = \frac{81}{114} + \frac{81}{162} + \frac{106}{126} + \frac{64}{140} + \frac{81}{37} + \frac{81}{28} + \frac{106}{53} +$$

$X^2 = 11.11$

Le nombre de degré de liberté d.d.l= (n-1) (k-1) c'est-à-dire d.d.l= (4-1) (2-1)=3

La table de X2 indique pour d.d.l =3 et au taux de sécurité de 95%, La valeur $X^2_{0.05}$ =7.8147…

Nous avons donc $X^2 > X^2_{0.05}$.

Au taux de sécurité de 95%, nous ne pouvons donc p accepter l'hypothèse H_0 que les 4 traitements possèdent la même efficacité. Les différences observées entre les pourcentages de guérison sont significatives c'est -à-dire les 4 traitements sont indépendants. Il semble *bien que les traitements appliqués dans les hôpitaux 1 et 2 soient plus efficaces qui ceux qui sont appliqués dans les hôpitaux 3 et 4.*

Conclusion

En fin nous pouvons dire que le test de khi-deux nous permet d'obtenir des renseignements et des informations concernant les paramètres d'une population inconnue dite population mère sur la base d'un ensemble d'observation statistiques provenant de cette population

Ce test est très utilisable dans la biologie (génétique) et dans l'économie….et aussi important par rapport à les différentes type de test d'hypothèse.

Annexe:

Loi multinomiale:

*1*Paramètre:*

n>0:nombre d'épreuves (entier)

$p_1 \ldots . p_n$: probabilité des évènements ($\sum p_i = 1$)

$N_i \in \{1 \ldots m\}$

*2*Support:*

$$\sum_{i=1}^{m} N_i = n$$

*3*Densité:*

$$P(\, N_1 = n_1, \ldots , N_m = n_m \,) = n! \prod_{j=1}^{m} \frac{P_j^{\,n_j}}{n_j!}$$

*4*Espérance:*

Esp=npi

*5*Variance:*

Var=n $p_i(1-p_i)$

*6*Fonction caractéristique:*

$$()\sum_{j=1}^{n} p_j\, e^{it_j})^n$$

Sommaire:

- *La table du khi -deux:*
- *Référence:*

REMERCIEMENT

En premier lieu je tennée à remercier ALLAH mon créateur pour m'avoir donnée la force pour terminer ce petit travail.

Je tennée à exprimer nos vifs remerciement à touts les professeurs qui nous ont aidés tous au long de mon cursus universitaires en particulier mon promoteur M^{ed} Ben yelles pour ses conseils et l'aide qu'il m'a portée.

Je remercier très vivement Mr MOURID le responsable de l'option probabilité-statistique

Je tiens à remercier le chef département : Mr MABROUT

En fin, j'adresse mes remerciement les plus s'incères à tous des gents qui m'ont aidé de prés ou de loin.

Derkaoui Nadjia

REMERCIEMENT

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes que je voudrais témoigner toute ma reconnaissance.

Je voudrais tout d'abord adresser tout ma gratitude à l'encadreur de ce projet Med Ben yelles pour sa patience, sa disponibilité et surtout ses judicieux conseils.

Je désire aussi remercier les professeurs de département s de mathématiques qui m'ont fournit les outils nécessaires à la réussite de mes études universitaires.

Boumediene Ibtissam

Dédicace

Je dédis ce travail ; le fruit des années d'études à :

- *Ceux qui ont consacré toute leur vie pour la réussite de leurs fils et leurs filles, et qui ont les bougies allumant mon chemin vers la réussite, pour leur bonté, pour leur générosité et encouragement. Mes chers parents*
- *Ma chère grand-mère*
- *Mes chers frères : Ahmed, Zinelabidine, Abdeljalil, Abdelkader, wail.*
- *Ma chère sœur : Naziha*
- *Mes oncles et mes tantes*
- *Mes cousins et mes cousines*
- *Toutes Mes amies sans exception et spécialement : zahra*
- *Mon promotion 3eme année probabilité et statistique 2010*
- *Tous ceux qui m'ont aidé de prés ou de loin à réaliser ce travail.*

Derkaoui Nadjia

Dédicace:

A nos parents

Nos familles

Et nos amis…

Boumediene Ibtissame

Référence:

1. Statistique (cours & exercices)
   ADMANE O., HOANG-KY, OUAKLI N.
2. Probabilités et statistique
   ALAN RUGG
3. http:
   mathsv-univ-lyon .1.fr/cours/pdf/stat/chapitre 7.pdf
4. http:

   jourdan.ens.fr/Michel/courses/metri2_07/cours_5.pdf.
5. http:
   cours_n07_le_test_du_khi_deux.pdf.
6. http:
   Chapitre8.pdf