Chapitre 6

TESTS STATISTIQUES

Les tests statistiques sont des méthodes de la statistique inférentielle qui, comme l'estimation, permettent d'analyser des données obtenues par tirages au hasard. Ils consistent à généraliser les propriétés constatées sur des observations à la population d'où ces dernières sont extraites, et à répondre à des questions concernant par exemple la nature d'une loi de probabilité, la valeur d'un paramètre ou l'indépendance de deux variables aléatoires. Nous retrouverons dans certains cas la démarche utilisée en estimation.

1. GÉNÉRALITÉS SUR LES TESTS STATISTIQUES.

1.1 Notion de test statistique.

Pour introduire intuitivement la démarche générale, nous allons considérer le cas d'un tribunal examinant la culpabilité d'un prévenu : ce tribunal doit établir cette culpabilité et, si elle est avérée, infliger une peine au coupable[1].

[1] Ce parallélisme entre la décision d'un tribunal et les tests statistiques n'est pas fortuit. C'est en étudiant ce genre de problème que Condorcet, pendant la Révolution Française, a déterminé le nombre de membres d'un jury et la règle de décision (unanimité, majorité simple, majorité des deux tiers …) afin de minimiser les erreurs judiciaires. Il ne croyait vraisemblablement pas à la façon dont ses résultats étaient appliqués pendant la Terreur puisque après avoir été arrêté, il se suicida avant d'être jugé.

On sait que le prévenu bénéficie d'un a priori favorable : il est présumé innocent, et la preuve de sa culpabilité est en principe à la charge de l'accusation.

La démarche, dans un test statistique, est strictement analogue : on suppose que la variable aléatoire étudiée possède une propriété particulière, appelée hypothèse nulle (c'est la présomption d'innocence), et pour remettre en cause cette propriété, il faut apporter la preuve qu'elle est fausse (c'est la preuve de la culpabilité).

C'est l'enquête policière qui montre la culpabilité de l'accusé : par exemple, la présence d'une grosse somme d'argent inexpliquée sur un compte en banque peut être considérée comme la preuve d'une escroquerie. De même c'est l'enquête statistique qui montre que la v.a. ne possède pas la propriété supposée vraie et que l'hypothèse nulle est fausse. Cette enquête est fondée sur l'analyse des observations de la v.a. : suivant les résultats de cette analyse, l'hypothèse nulle est considérée comme vraisemblable (l'innocence dans le cas de l'enquête policière) ou presque impossible et « rejetée » : on accepte alors une autre hypothèse, appelée hypothèse alternative (la culpabilité).

Définition : on appelle test statistique une démarche de la statistique inférentielle consistant :

- à contrôler la validité d'une hypothèse considérée comme vraie a priori, appelée hypothèse nulle et notée H_0 ;
- à admettre une hypothèse différente lorsque le contrôle se révèle négatif, appelée hypothèse alternative et notée H_1.

Il existe donc deux façons de se tromper. La première consiste à accepter l'hypothèse nulle alors qu'elle est fausse : cela revient à acquitter un coupable faute de preuve. La seconde est le rejet de l'hypothèse nulle alors qu'elle est vraie : on condamne quelqu'un d'innocent.

Définition : on appelle :

- erreur de première espèce, l'erreur consistant à rejeter l'hypothèse nulle alors qu'elle est vraie ;
- erreur de seconde espèce, l'erreur consistant à accepter l'hypothèse nulle alors qu'elle est fausse.

Comment procède un tribunal ? La signification de l'expression « erreur judiciaire », dont l'inculpé est toujours la victime, montre bien qu'en pratique, on cherche à limiter le risque de condamner un innocent (du moins, on peut le souhaiter). De même, en statistique, on limite le risque de rejeter l'hypothèse nulle alors qu'elle est vraie.

Il est bien clair qu'en limitant ce risque, on augmente l'autre : plus on acquitte facilement les accusés, moins on condamne d'innocents, mais plus on acquitte de coupables. La relation n'est pas fonctionnelle ; nous y revenons en fin de chapitre.

Définition : on appelle :

- risque de première espèce la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie (erreur de première espèce). On le note α.
- risque de seconde espèce la probabilité d'accepter l'hypothèse nulle alors qu'elle est fausse (erreur de seconde espèce). On le note β.

		Hypothèse vraie	
		H_0	H_1
Hypothèse acceptée	H_0	pas d'erreur	Erreur de $2^{\text{ième}}$ espèce Risque β
	H_1	Erreur de $1^{\text{ère}}$ espèce Risque α	pas d'erreur

Tableau 1.6 : erreurs et risques suivant la décision.

1.2 Règle de décision.

Après avoir défini les hypothèses, ou simplement l'hypothèse nulle, la démarche consiste à étudier les résultats d'une enquête aléatoire : si ces résultats paraissent en contradiction avec l'hypothèse nulle, cette dernière sera considérée comme fausse. Sinon, on continuera à la supposer vraie : c'est un « raisonnement par l'absurde ».

Les résultats de l'enquête sont en général résumés par un nombre calculé sur les observations et appelé statistique.

Définition : dans un test statistique, la variable aléatoire que l'on utilise pour contrôler l’hypothèse nulle est appelée elle-même statistique.

L’hypothèse nulle sera remise en cause si cette statistique possède une propriété qu’elle ne devrait pas avoir : on retrouve l’analogie avec le solde d’un compte bancaire : s’il est anormalement élevé, c’est une forte présomption de culpabilité.

Un exemple intuitif :

Hypothèse nulle H_0 : Dominique est une femme

Hypothèse alternative H_1 : Dominique est un homme

Statistique : pointure

1) Observation : Dominique chausse du 43.

Décision : Peu de femmes chaussant du 43, on peut considérer que l’observation est en contradiction avec H_0. Donc Dominique n’est pas une femme (rejet de H_0) , c’est un homme (acceptation de H_1).

2) Observation : Dominique chausse du 40.

Décision : La pointure 40 est fréquente chez les femmes. On peut considérer que l’observation n’est pas contradictoire avec H_0. Donc Dominique peut être une femme (acceptation de H_0 et rejet de H_1).

3) Discussion : la décision statistique prise à l’issue du test est discutable. Plusieurs hypothèses ont été émises implicitement :

- *il n’y a que des hommes et des femmes passant sur la plage, pas d’enfant.*
- *il y a autant d’hommes que de femmes.*

Un raisonnement hâtif peut donc conduire à une décision erronée (cf. exercice 5 du chapitre 4). Cette discussion caractérise la problématique de la statistique bayesienne.

Définition : on appelle région critique d’un test statistique l’ensemble des valeurs observées de la statistique provoquant le rejet de l’hypothèse nulle.

Dans la pratique, c’est le risque de première espèce qui précise les bornes de la région critique.

Exemple *: supposons que Dominique chausse du 41 ou plus. La proportion de femmes chaussant du 41 ou plus étant inférieure à 5%, on rejette l’hypothèse nulle. Mais on ne peut* *pas être totalement sûr que Dominique n'est pas une femme, puisque c'est le cas de 2% environ des femmes. Dans 2% des cas, on va donc commettre l'erreur consistant à rejeter l'hypothèse nulle alors qu'elle est vraie. C'est cette faible proportion qui est le risque de première espèce, et que l'on cherche à limiter.*

Comment donc choisir ce risque α ? Le risque est en quelque sorte la faiblesse de la preuve : plus le risque est grand, plus la preuve est faible et inversement. On choisira un risque d'autant plus faible, c'est-à-dire une preuve d'autant plus forte, que l'hypothèse alternative est moins vraisemblable. La démarche scientifique générale est d'ailleurs claire : une expérience physique ou chimique est d'autant plus répétée, vérifiée et contrôlée, c'est-à-dire qu'on choisit un risque d'autant plus faible - une preuve d'autant plus forte, que cette expérience remet en cause une théorie jusque-là considérée comme satisfaisante. Le choix de ce risque est équivalent à celui du niveau de confiance dans l'estimation par intervalle de confiance (chapitre 5).

Il y a des valeurs classiques :

Risque de première espèce	hypothèse alternative
0.001 (0.1%)	quasiment impossible
0.01 (1%)	très peu vraisemblable
0.05 (5%)	peu vraisemblable
0.1 (10%)	possible

Tableau 2.6 : choix du risque de première espèce.

1.3 Tests élémentaires.

Les tests sur le coefficient d'asymétrie et d'aplatissement sont très simples et fournissent de bons exemples d'application immédiate.

On définit les hypothèses suivantes :

- hypothèse nulle H_0 : la loi de la v.a. est la loi normale.
- hypothèse alternative H_1 : la loi de X n'est pas la loi normale.

Si l'hypothèse nulle est vraie, les valeurs théoriques des coefficients d'asymétrie et d'aplatissement sont $\gamma_{as} = 0$ et $\gamma_{ap} = 3$.

On étudie un échantillon (X_i) i=1, …, n, de la v.a. X. Les coefficients d'asymétrie c_{as} et d'aplatissement c_{ap} de l'échantillon observé devraient être proches de 0 et de 3 si la loi est normale : ce sont les statistiques du test.

Un coefficient d'asymétrie c_{as} très différent de 0 est donc en contradiction avec la loi normale. Pour décider s'il est très différent de 0, on choisit un risque de première espèce α, et on en déduit la région critique à l'aide d'une table statistique.

Cette table donne, pour n = 50 observations et $\alpha = 0.05$: 0.534. Cela signifie que la probabilité de l'événement $\{c_{as} < -0.534\} \cup \{c_{as} > 0.534\}$ est égale à 0.05 si la loi est normale. La région critique est :

$$\boxed{RC =]-\infty, -0.534[\cup]0.534, +\infty[}$$

De la même façon, un coefficient d'aplatissement très différent de 3 est en contradiction avec l'hypothèse de la loi normale. Ce coefficient n'étant pas réparti symétriquement, la table donne deux valeurs. Pour n = 50 et $\alpha = 0.05$, la région critique est :

$$\boxed{RC =]0, 2.15[\cup]3.99, +\infty[}$$

La décision est donc la suivante : on considère que la loi de la v.a. X n'est pas la loi normale si l'un des coefficients observés appartient à la région critique correspondante[2].

Exemple *: les coefficients d'asymétrie et d'aplatissement des variables des données Euromarket sont les suivants :*

	coefficient d'asymétrie	*coefficient d'aplatissement*
âge	*1.108*	*4.496*
revenu	*1.590*	*5.064*
achats	*1.160*	*3.859*
enfants	*-0.070*	*2.418*

Dans ces données constituées des 50 clients d'Euromarket, seule la variable nombre d'enfants peut être considérée comme normale avec un risque de première espèce de 5%. C'est une variable discrète. Cette propriété signifie ici que la v.a. X dont la densité par intervalle définie par :

$$d_i = f_i \text{ dans l'intervalle }]\, i - 0.5,\ i + 0.5\, [\qquad i \in \mathbf{N}$$

[2] Si l'on tient compte de deux coefficients à la fois, le risque de première espèce est modifié. Il faudrait en toute rigueur choisir a priori un des coefficients suivant la nature de la loi de probabilité supposée vraie sous l'hypothèse alternative (cf. exercice 2).

f_i étant la proportion observée de familles de i enfants, est approximativement une v.a. suivant la loi normale de même moyenne et de même variance que la variable nombre d'enfants.

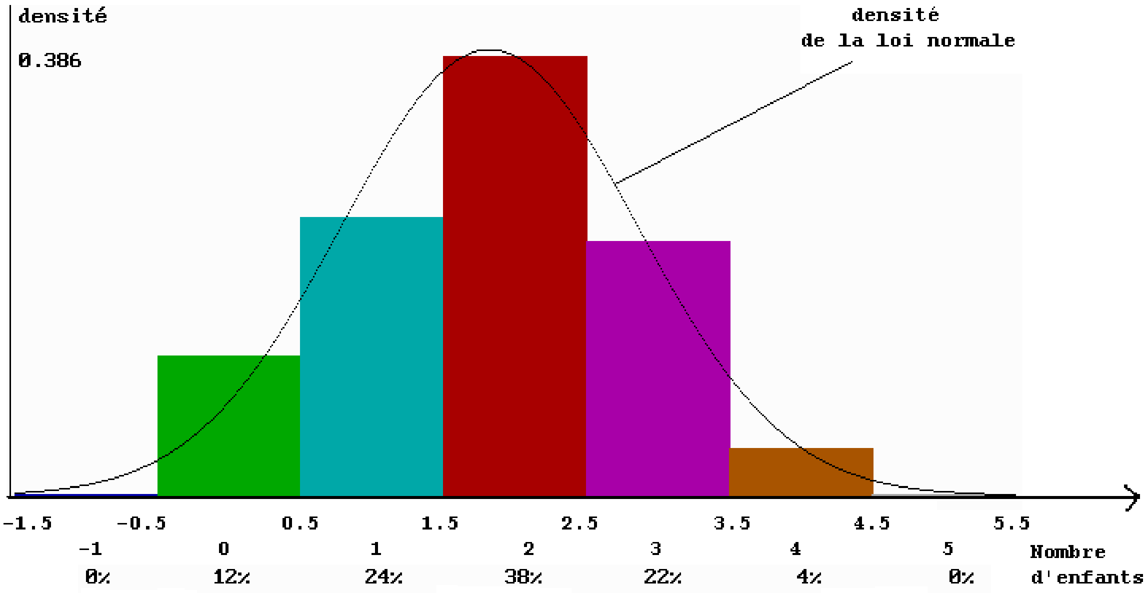

Figure 1.6 : diagramme du nombre d'enfants et densité de la loi normale

2. TEST D'AJUSTEMENT DU χ^2 DE PEARSON.

2.1 Cas d'une variable discrète.

On considère le cas d'un dé à 6 faces. Les hypothèses concernent la loi de probabilité de la face obtenue en lançant le dé.

- Hypothèse nulle H_0 : cette loi est la loi uniforme discrète sur {1, …, 6} (le dé est parfaitement équilibré).
- Hypothèse alternative : ce n'est pas la loi uniforme discrète sur {1, …, 6} (les faces n'ont pas toutes la même probabilité, et le dé est mal équilibré).

L'expérience consiste à lancer le dé n fois (100, 200 ou 1000 fois par exemple). On compare ensuite les proportions f_i (i = 1, …, 6) observées aux probabilités théoriques p_i (i = 1, …, 6) de chaque face du dé. Elle est généralisable à toutes les v.a. discrètes ou qualitatives :

- L'hypothèse nulle est définie par la loi de probabilité supposée vraie, dont la densité est définie par la suite (p_i), i = 1, …, k (chapitre IV, paragraphe 2.3).

- L'hypothèse alternative est que la loi de probabilité n'est pas égale à la précédente, sans plus de précision.

Pour contrôler l'hypothèse nulle, on compare les proportions n_i/n observées sur un échantillon de taille n aux probabilités théoriques p_i..

Exemple *: on donne le nombre de faces obtenues en lançant le dé 100 fois :*

i = 1	*i = 2*	*i = 3*	*i = 4*	*i = 5*	*i = 6*
16	*15*	*18*	*14*	*19*	*18*

Si le dé est parfaitement équilibré, on devrait obtenir des proportions $f_i = n_i/n$ de l'ordre de 1/6 = 0.1667. Il est évident que cette proportion de 1/6 ne sera jamais obtenue exactement, et que l'écart peut provenir du hasard ou d'un mauvais équilibrage du dé.

On doit donc mesurer l'écart entre les proportions n_i/n obtenues et les probabilités p_i, par une statistique, et donner une règle, appelée règle de décision, permettant de considérer ou non que l'écart est dû au hasard.

Définition : on appelle X^2 la statistique choisie pour comparer les proportions observées aux probabilités théoriques :

$$X^2 = n \sum_{i=1}^{k} (n_i/n - p_i)^2 / p_i = \sum_{i=1}^{k} (n_i - n\,p_i)^2 / n\,p_i$$

Reprenons l'exemple du dé : s'il est bien équilibré, les proportions n_i/n convergent vers les probabilités $p_i = 1/6$, et la valeur prise par la v.a. X^2 est faible. Inversement, si la valeur prise par X^2 est élevée, on peut penser que le dé n'est pas bien équilibré puisque les proportions sont différentes de 1/6.

Le raisonnement est exactement identique dans le cas général, et nous allons déterminer une valeur x_α^2 indiquant la limite à partir de laquelle nous considérons que la valeur de X^2 est trop élevée pour que l'hypothèse nulle soit vraie.

Pour déterminer cette valeur x_α^2, on utilise le risque de première espèce α, qui est la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie (considérer le dé pipé alors qu'il est bien équilibré).

La valeur x_α^2 est donc telle que :

$$\boxed{P(X^2>x_\alpha^2) = \alpha}$$

Cette valeur dépend bien entendu de la loi de probabilité de la v.a. X^2 : cette loi est approximativement la loi du χ^2 définie dans le chapitre 4.

Théorème : si les termes $n \times p_i$ sont tous supérieurs ou égaux à 5, la loi de probabilité de la v.a. X^2 est approximativement la loi du χ^2 de degré de liberté $\nu = k - l - 1$, où k est le nombre de valeurs possibles des observations et l le nombre de paramètres calculés à l'aide des données.

Lorsque la condition de convergence $n \times p_i > 5$ n'est pas toujours satisfaite, il faut regrouper les valeurs qui ne la vérifient pas avec d'autres, en cumulant les probabilités correspondantes.

Conséquence : la valeur x_α^2 telle que la proportion des valeurs de X^2 supérieures à x_α^2 soit égale à α est donnée dans la table statistique de la loi du χ^2.

Définition : la région critique du test d'ajustement du χ^2 est l'intervalle :

$$\boxed{RC = [x_\alpha^2, +\infty[}$$

dans lequel x_α^2 est déterminé de façon que $P(X^2>x_\alpha^2) = \alpha$.

Exemple numérique *: dans le cas du dé, le degré de liberté est égal à $\nu = 5$, et, pour $\alpha = 0.05$, la table statistique donne $x_\alpha^2 = 11.07$. Un dé bien équilibré donnera donc rarement (dans 5% des cas) une valeur x^2 de X^2 supérieure à 11.07.*

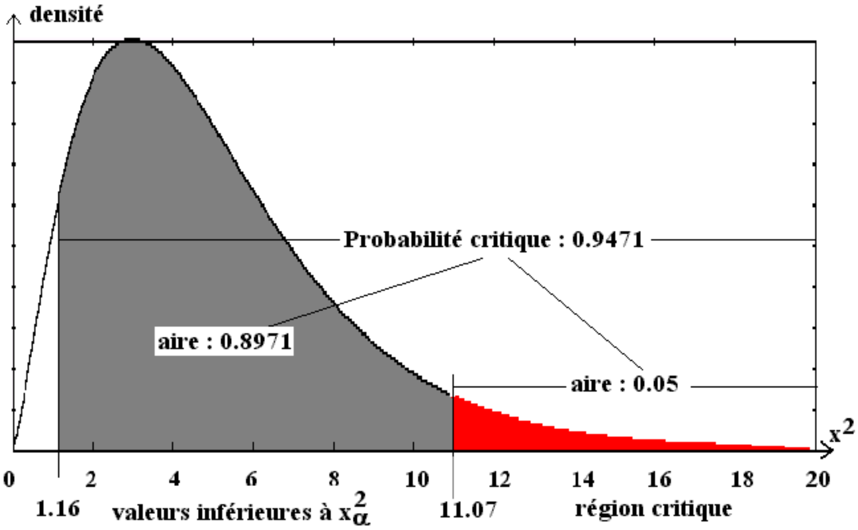

Figure 2.6 : densité de la loi du χ^2 (ν = 5) et région critique (α = 5%)

Les calculs sur le tableau précédent sont les suivants :

$$x^2 = (16 - 100/6)^2/(100/6) + (15 - 100/6)^2/(100/6) + (18 - 100/6)^2/(100/6)$$
$$+ (14 - 100/6)^2/(100/6) + (19 - 100/6)^2/(100/6) + (18 - 100/6)^2/(100/6)$$

$$\boxed{x^2 = 1.16}$$

Tous les termes de la forme n $\times$ pi sont égaux à 100 / 6 = 16.66 et donc supérieurs à 5 : la condition de convergence est satisfaite. La valeur X^2 obtenue 1.16 est inférieure à 11.07 : on peut considérer que le dé est bien équilibré. Plus exactement, les 100 lancers n'ont pas permis de montrer si le dé est mal équilibré : l'accusé n'est peut-être pas coupable.

Définition: on appelle **probabilité critique** (en anglais « p-value ») la probabilité que la valeur observée x^2 de la statistique X^2 soit dépassée.

Cette probabilité critique est l'aire de l'ensemble des points d'abscisses supérieures à la valeur observée x^2 compris entre la courbe représentant la densité, l'axe des abscisses.

Lorsqu'elle est supérieure au risque de première espèce α, cela signifie que x_α^2 est supérieure à la valeur observée x^2, qui n'appartient donc pas à la région critique (figure 2.6).

Dans le cas contraire, x^2 appartient à la région critique et on rejette l'hypothèse nulle.

Exemple *: le programme donne $P(X^2 > 1.16)= 0.94716$. La probabilité critique est donc supérieure au risque de première espèce α = 0.05. La table statistique n'est pas indis-*

pensable pour conclure puisque cela signifie que x_α^2 *est supérieur à 1.16, comme on peut le voir sur la figure 2.6 : on accepte l'hypothèse nulle.*

2.2 Cas d'une variable continue.

Considérons maintenant le cas des v.a. continues :

- L'hypothèse nulle est définie par une densité théorique de X ;
- L'hypothèse alternative est une loi non précisée de X.

La procédure est la suivante :

- On définit des intervalles I_i, i = 1, …, k, dont on calcule les probabilités théoriques $p_i = P(X \in I_i)$.
- On répartit les n observations de la v.a. X dans ces intervalles.
- On en déduit la densité observée égale à la suite des proportions n_i/n, où n_i est le nombre d'observations classées dans l'intervalle I_i.
- On applique la procédure du paragraphe 2.1 pour comparer p_i et n_i/n.

Remarques :

- Les classes seront choisies toujours a priori, avant le calcul de X^2, et de préférence de probabilité égale.[3]
- Le calcul des probabilités théoriques peut exiger préalablement l'estimation de paramètres de la densité théorique.
- Deux densités différentes peuvent donner la même densité par intervalle. L'hypothèse nulle ne les distingue pas l'une de l'autre et le test donne la même valeur de X^2 et par suite , pour un même degré de liberté, la même décision.

Exemple *: nous voulons savoir si l'âge est réparti suivant une loi normale dans la clientèle de l'hypermarché EUROMARKET. On choisit comme risque de première espèce* $\alpha = 0.05$.

La figure 3.6 permet de comparer la densité de la loi normale à l'histogramme, mais ne donne pas d'indications quantitatives et ne prend pas en compte le nombre d'observations.

[3] Kendall et Stuart, The advanced theory of statictics, vol. 2, p. 431, **30.22** (Griffin, London, 1961).

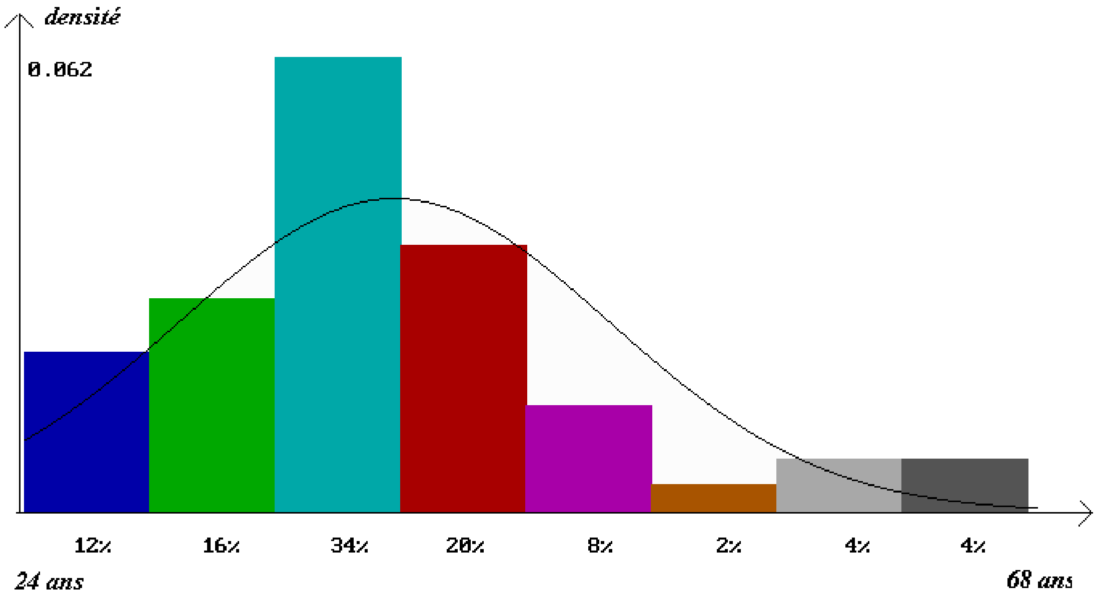

Figure 3.6 : histogramme de l'âge (8 classes)
et densité de la loi normale de mêmes paramètres

Nous effectuons ci-dessous l'ajustement en considérant la répartition de l'âge des 50 clients d'EUROMARKET en 8 classes de même longueur.

	Classe	Effectifs en %
1	[24.0, 29.5[	12
2	[29.5, 35.0[	16
3	[35.0, 40.5[	34
4	[40.5, 46.0[	20
5	[46.0, 51.5[	8
6	[51.5, 57.0[	2
7	[57.0, 62.5[	4
8	[62.5, 68.0[	4

Répartition des 50 observations dans les huit classes

Pour calculer les probabilités théoriques p_i de chaque intervalle, il faut connaître les paramètres de la densité théorique, c'est-à-dire la moyenne et l'écart type dans le cas de la loi normale. Les valeurs calculées sur les données individuelles sont m = 40.06 et s = 9.341 : ce sont ces valeurs qui seront utilisées pour calculer les probabilités théoriques.

Une loi normale peut prendre des valeurs allant de $-\infty$ à $+\infty$. Il faut donc considérer comme première classe $]-\infty, 29.5[$ et comme dernière classe $[62.5, +\infty[$. On obtient :

	Classes	probabilités	condition
1	[-∞, 29.5[	0.12914	6.46
2	[29.5, 35.0[	0.16488	8.24
3	[35.0, 40.5[	0.22477	11.24
4	[40.5, 46.0[	0.21879	10.94
5	[46.0, 51.5[	0.15208	7.60
6	[51.5, 57.0[	0.07547	3.77 *
7	[57.0, 62.5[	0.02673	1.34 *
8	[62.5, + ∞ [	0.00815	0.41 *

Classes avant regroupement

*L'étoile * indique que dans les classes 6, 7 et 8 la condition de convergence* $n \times p_i \geq 5$ *n'est pas vérifiée. On doit donc réunir ces classes de façon à vérifier cette condition :*

	Classe	probabilités	condition	proportions
1	[-∞, 29.5[	0.12914	6.46	0.12
2	[29.5, 35.0[	0.16488	8.24	0.16
3	[35.0, 40.5[	0.22477	11.24	0.34
4	[40.5, 46.0[	0.21879	10.94	0.20
5	[46.0, 51.5[	0.15208	7.60	0.08
6 = 6+7+8	[51.5, + ∞ [	0.11035	5.52	0.10

Classes après regroupement

Les probabilités des classes étant proches les unes des autres, la répartition paraît satisfaisante. Le nombre de paramètres estimés est égal à 2 (moyenne et écart-type). Le degré de liberté est donc fixé à $\nu = 6 - 2 - 1 = 3$*. On en déduit la région critique* ($\alpha = 0.05$) *:*

$$RC = [7.815, +\infty[$$

Le calcul de X^2 *donne :*

$$x^2 = 50 \times [(0.12 - 0.12914)^2 + (0.16 - 0.16488)^2 + (0.34 - 0.22477)^2 + (0.20 - 0.21879)^2 + (0.08 - 0.15208)^2 + (0.10 - 0.11035)^2]$$

Soit :

$$\boxed{x^2 = 4.8305}$$

La valeur observée x^2 *de* X^2 *n'appartient pas à la région critique RC. On accepte donc l'hypothèse que l'âge est réparti dans la clientèle totale suivant une loi normale. Cela signifie plus précisément que les observations effectuées ne remettent pas en cause cette hypothèse.*

La probabilité critique de la valeur observée est supérieure au risque α *choisi :*

$$P(X^2 > 4.8305) = 0.18289 > 0.05$$

La décision est évidemment la même.

3. TEST D'INDÉPENDANCE DU χ^2 DE PEARSON.

Le test d'indépendance du χ^2 de Pearson permet d'étudier la liaison entre deux variables qualitatives X et Y que l'on a observées sur des unités statistiques que l'on suppose tirées au hasard au sein d'une population. Ces variables aléatoires se présentent sous la forme de questions proposant k et *l* réponses possibles, une seule étant choisie. Nous avons étudié dans le chapitre 4 la loi de probabilité du couple (X, Y). Nous en reprenons les notations.

3.1 Tableau des effectifs théoriques.

La démarche est la suivante :

- L'hypothèse nulle est définie par l'indépendance des deux v.a. X et Y ;
- L'hypothèse alternative n'est pas précisée ;

L'expérience consiste à tirer au hasard un échantillon d'unités statistiques et à en comparer la répartition à la répartition « théorique » déduite de l'indépendance supposée des variables.

Définition : on appelle répartition théorique des unités statistiques d'un échantillon suivant deux critères la répartition que l'on aurait si ces deux critères étaient indépendants.

La procédure consiste à comparer ces répartitions :

- si la répartition théorique et la répartition observée sont voisines l'une de l'autre, on peut considérer que la répartition observée ne remet pas en cause l'indépendance des v.a. X et Y considérées.
- si les deux répartitions présentent de grandes différences, c'est que l'indépendance des deux v.a. X et Y étudiées est contestable.

Le calcul de la répartition théorique découle de l'hypothèse d'indépendance : les réponses données à chaque question X ou Y sont réparties théoriquement de la même façon quelle que soit la réponse donnée à la seconde. L'effectif théorique $n_{i,j}'$ est donné par la formule :

$$\boxed{n_{i,j}' = n\, p_{i.}\, p_{.j} = n_{i.}\, n_{.j} / n}$$

dans laquelle les termes $n_{i\bullet}$, $n_{\bullet j}$ sont les effectifs marginaux calculés sur le tableau de données et $p_{i\bullet}$ et $p_{\bullet j}$ les proportions marginales.

Exemple *: nous étudions le tableau donnant la répartition de 200 étudiants suivant le sexe et la couleur des cheveux, en supposant qu'ils ont été tirés au hasard dans l'ensemble des étudiants de l'université. Le tableau est le suivant :*

	Cheveux blonds ($j = 1$)	*Cheveux bruns ($j = 2$)*	*Autre couleur ($j = 3$)*	*Effectifs marginaux*
Masculin ($i = 1$)	*25*	*51*	*17*	$n_{1.} = 93$
Féminin ($i = 2$)	*62*	*31*	*14*	$n_{2.} = 107$
Effectifs marginaux	$n_{.1} = 87$	$n_{.2} = 82$	$n_{.3} = 31$	*200*

S'il y a indépendance entre le sexe et la couleur des cheveux, la répartition théorique des étudiants est la suivante :

	Cheveux blonds	*Cheveux bruns*	*Autre couleur*
Masculin	*40.46*	*38.13*	*14.42*
féminin	*46.54*	*43.87*	*16.58*

Par exemple, l'effectif théorique d'étudiantes au cheveux blonds est $107 \times 87 / 200 = 46.54$.

3.2 Test d'indépendance du χ^2 de Pearson.

Pour comparer les effectifs théoriques et les effectifs observés, on utilise la même statistique que dans le cas du test d'ajustement.

Définition : la statistique X^2 utilisée pour comparer les répartitions théoriques et observées est définie par :

$$X^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} (n_{i,j} - n_{i,j}')^2 / n_{i,j}'$$

L'hypothèse d'indépendance est contestable lorsque les effectifs observés $n_{i,j}$ sont très différents des effectifs théoriques $n_{i,j}'$, donc lorsque X^2 prend de grandes valeurs. Il reste à décider à partir de quelle valeur X^2 peut être considéré comme grand. Pour cela, on utilise la loi de X^2 sous l'hypothèse d'indépendance qui est la loi du χ^2 de degré de liberté ν :

$$\nu = (p - 1)\,(q - 1)$$

Ce degré de liberté est calculé comme le précédent, par la formule $\nu = k - l - 1$:

- le nombre de valeurs possibles est $k = p \times q$
- les paramètres estimés sont les lois de probabilités marginales : $p - 1$ termes pour la loi de X, $q - 1$ pour la loi de Y puisque la somme des probabilités marginales est égale à 1. On a donc $l = (p - 1) + (q - 1)$
- Le degré de liberté est égal à : $p \times q - (p - 1) - (q - 1) - 1 = (p - 1)(q - 1)$

Définition : la région critique du test d'indépendance du χ^2 est l'intervalle $[\chi_\alpha^2, +\infty[$, χ_α^2 étant le nombre auquel une proportion α de X^2 est supérieure si l'hypothèse d'indépendance est vraie.

Les observations remettent donc en cause l'hypothèse d'indépendance si X^2 prend une valeur supérieure à x_α^2 ; on rejette alors l'hypothèse d'indépendance.

Supposons maintenant que nous ayons rejeté l'hypothèse d'indépendance. Pour expliquer la liaison entre les variables, on examine l'observation x^2 de la statistique X^2, et l'on recherche, parmi les termes dont il est la somme, ceux qui sont les plus grands : les indices i et j correspondants indiquent les modalités des questions X et Y dont les effectifs théoriques et observés sont les plus différents. Ce sont ces modalités qui provoquent la liaison entre les deux riables.

Exemple : *chaque terme du tableau ci-dessous indique la valeur du terme correspondant dans la somme donnant le X^2 appelé parfois « contribution au x^2 » :*

$5.907 = (25-40.46)^2 / 40.46$	$4.344 = (51-38.13)^2 /38.13$	$0.462 = (17-14.42)^2/14.42$
$5.136 = (62-46.54)^2/46.54$	$3.776 = (31-43.87)^2/43.87$	$0.401 = (14-16.58)^2/16.58$

La valeur x^2 de X^2 est la somme des termes du tableau. On obtient :

$$\boxed{x^2 = 20.02}$$

Une liaison entre la couleur des cheveux et le sexe n'étant pas du tout invraisemblable, nous choisissons un risque raisonnable α égal à 5%.

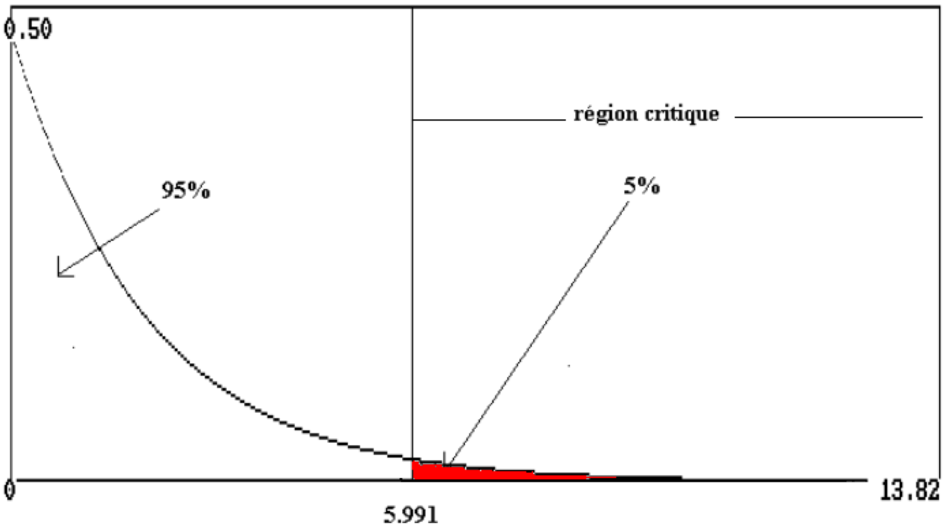

Figure 4.6 : densité de la loi du χ^2 *(ν = 2)*

Le degré de liberté est égal à (2-1) x (3-1), soit 2, et la région critique est définie par [5.991, + ∝ [(on notera la différence entre les densités du χ^2 *de degré de liberté 2 et 5, cf. figure 2.6 et 4.6). La valeur* x^2 *de X² appartient à la région critique. On rejette donc l'hypothèse d'indépendance.*

Les valeurs les plus grandes du tableau indiquent que la liaison est due à la couleur brune ou blonde et le sexe des étudiants. La modalité « autre couleur » n'intervient pas.

4. TEST SUR LE COEFFICIENT DE CORRÉLATION LINÉAIRE.

Nous avons constaté chez les 50 clients du supermarché une relation entre l'âge et le revenu. La question que l'on se pose maintenant est de savoir si cette relation existe sur l'ensemble des clients, ou si elle n'est due qu'aux tirages au hasard que l'on a effectués. Les données sont ici quantitatives, et la liaison entre les deux variables est mesurée par le coefficient de corrélation linéaire.

4.1 Hypothèses et erreurs.

La procédure consiste d'abord à vérifier que l'échantillon observé vérifie les propriétés présentées dans le chapitre 4 : théoriquement, le couple (âge, revenu) doit suivre la loi binormale pour que le test sur le coefficient de corrélation soit valide. L'équivalence entre la nullité du coefficient de corrélation et l'indépendance n'est assurée en effet que sous cette condition.

On se limite en pratique à vérifier que les répartitions de l'âge et du revenu ne sont pas trop différentes de la loi normale, par un test d'ajustement du χ^2 ou l'étude des coefficients d'asymétrie et d'aplatissement que nous avons expliquée dans le paragraphe 1.3. Dans le cas contraire, il est possible de transformer les données, en passant par exemple aux logarithmes.

On suppose a priori qu'il n'existe pas de liaison entre les séries : les variables sont supposées « indépendantes ». Le coefficient de corrélation exact et inconnu, que l'on appelle coefficient de corrélation théorique (noté ρ), est alors nul. Nous choisissons comme hypothèse alternative que le coefficient de corrélation théorique est différent de 0 (on pourrait choisir strictement positif, ou strictement négatif).

- Hypothèse nulle H_0 : $\rho = 0$.
- Hypothèse alternative H_1 : $\rho \neq 0$.

Une valeur approchée du coefficient de corrélation théorique est donnée par la valeur observée r de son estimateur empirique. La démarche consistant à définir cet estimateur est strictement la même que celle qui a abouti à la définition des estimateurs empiriques de la moyenne et de la variance.

Définition : l'estimateur empirique du coefficient de corrélation théorique ρ est la v.a. notée R dont la valeur observée sur un échantillon de couples est le coefficient de corrélation observé r.

Le coefficient de corrélation observé n'a évidemment n'a aucune raison d'être exactement égal à 0 même si l'indépendance des v.a. est vraie. Deux cas peuvent se produire :

- le coefficient de corrélation r est proche de 0. Les données ne contredisent pas l'hypothèse d'indépendance : on accepte l'hypothèse d'indépendance.
- le coefficient de corrélation r est très différent de 0. Il est alors peu vraisemblable que la valeur théorique ρ soit nulle : on rejette l'hypothèse d'indépendance.

4.2 Région critique.

Les statisticiens utilisent pour des raisons mathématiques et historiques une v.a. notée F déduite de l'estimateur empirique R du coefficient de corrélation ρ par la formule suivante :

$$\boxed{F = (n-2)\,\frac{R^2}{(1 - R^2)}}$$

La loi théorique de la v.a. F est la loi de Fisher Snedecor que nous avons définie dans le chapitre 4, et qui dépend de deux degrés de liberté, ici $\nu_1 = 1$, $\nu_2 = n - 2$. Les valeurs utiles sont données dans la table statistique appelée table du F.

Pour n = 50, on trouve $\nu_2 = 48$ et on trouve dans la table $f_\alpha = 4.04$. Un calcul simple donne la valeur du coefficient de corrélation ρ_α correspondant :

$$\rho_\alpha^2 = \frac{f_\alpha}{(n-2) + f_\alpha}$$

On trouve :

$$\rho_\alpha^2 = 0.078 \qquad\qquad \rho_\alpha = \pm\, 0.28.$$

La décision peut être aussi prise en fonction de la probabilité critique donnée fréquemment par les programmes.

<u>Exemple</u> *: on sait que la liaison entre l'âge et le revenu ne vérifie pas les propriétés nécessaires pour que l'on puisse effectuer un test de Fisher sur le coefficient de corrélation : la liaison n'est pas linéaire et la répartition du revenu ne ressemble pas à la loi normale (cf. figure 2 du chapitre 3). Nous allons limiter notre étude aux clients en activité, et éliminer des données les clients retraités 25, 31 et 43 : il reste 47 unités statistiques.*

En ce qui concerne les revenus, nous en considérons ici les logarithmes de façon à obtenir une distribution un peu plus symétrique. Nous admettrons la normalité des lois de probabilité après l'élimination de ces trois u.s..

L'hypothèse nulle considérée est la nullité du coefficient de corrélation théorique ρ entre l'âge et le revenu. On choisit donc un risque de première espèce α égal à 0.05. On en déduit la région critique sur la statistique F :

$$RC = [4.05, +\infty[.$$

Le coefficient de corrélation entre l'âge et le logarithme des revenus calculé sur les 47 observations est égal à $r = 0.6846$, ce qui donne $r^2 = 0.469$ et $f = 39.746$.

On constate évidemment que f appartient à cette région critique. La liaison entre l'âge et le revenu constatée sur l'échantillon observé ne peut donc pas être due au hasard : elle *existe très vraisemblablement dans l'ensemble des clients de l'hypermarché et on peut considérer que ρ est différent de 0.*

5. TESTS SUR LA MOYENNE ET LA VARIANCE.

Nous supposerons comme dans le chapitre précédent que les observations sont réparties régulièrement et symétriquement par rapport à leur moyenne, plus précisément qu'elles obéissent à la loi normale.

5.1 Tests sur la moyenne.

Le premier test que nous étudions consiste à décider entre les hypothèses suivantes :

$$H_0 : \mu = \mu_0 \qquad\qquad H_1 : \mu \neq \mu_0$$

La question posée est la suivante : les observations remettent-elles en cause l'égalité de la moyenne théorique μ à la valeur spécifiée μ_0 ?

Nous avons déjà répondu à cette question dans le chapitre précédent : l'estimation par intervalle de confiance donne l'ensemble des valeurs possibles de la moyenne théorique μ compte tenu des observations effectuées et du niveau de confiance choisi.

Règle de décision : pour un risque de première espèce α,

- On accepte l'hypothèse nulle si la valeur μ_0 appartient à l'intervalle de confiance de niveau de confiance $1 - \alpha$;
- On rejette l'hypothèse nulle sinon.

Exemple *: l'objectif fixé par les responsables nationaux de l'enseigne Euromarket est un montant moyen des achats égal à 420F. Le directeur commercial s'inquiète du montant moyen observé (316.95F) dans son hypermarché et veut donc vérifier si cette valeur montre effectivement une différence. Pour un niveau de confiance de 95%, l'intervalle de confiance est le suivant :*

[257.173, 376.717]

On rejette donc l'hypothèse nulle avec un risque de première espèce de 5% et le montant moyen des achats des clients d'EUROMARKET peut être considéré comme nettement inférieur à la valeur fixée à 420F. L'objectif n'est pas atteint.

Une autre façon (équivalente) d'effectuer le test est de déterminer la région critique. On sait que, si l'hypothèse nulle est vraie, la v.a. T ci-dessous suit la loi de Student de degré de liberté n-1 (cf. chapitre 5):

$$T = \frac{M - \mu_0}{S/\sqrt{(n-1)}}$$

Lorsque la v.a. T prend une très grande valeur, la moyenne μ est vraisemblablement différente de μ_0 : on rejette l'hypothèse d'indépendance. La région critique est donc de la forme :

$$\boxed{]-\infty, \mu_0 - t_\alpha\, s/(n - 1)^{1/2}[\cup]\, \mu_0 + t_\alpha\, s/(n - 1)^{1/2}, +\infty\, [}$$

les bornes de la région critique t_α étant définies par la relation :

$$P(|T| > t_\alpha] = \alpha$$

Le calcul précédent demande que l'on connaisse la valeur vraie de la moyenne, ce qui n'est pas toujours le cas. Par exemple, la moyenne des achats d'un autre hypermarché (410F) peut être connue elle aussi par sondage auprès d'une partie de la clientèle : le test précédent n'est pas applicable pour comparer ces deux moyennes.

On considère maintenant deux v.a. X_1 et X_2, observées sur deux échantillons de taille n_1 et n_2. Les hypothèses sont alors les suivantes, en notant μ_1 et μ_2 les moyennes théoriques des achats :

$$H_0 : \mu_1 = \mu_2 \qquad\qquad H_1 : \mu_1 \neq \mu_2$$

Le test consiste à déterminer si les observations remettent en cause l'égalité des moyennes μ_1 et μ_2 .

Le calcul est plus compliqué. On note M_1 et S_1^2 les estimateurs empiriques de la moyenne et la variance dans le premier échantillon de taille n_1, M_2 et S_2^2 dans le second échantillon de taille n_2. Le problème est de comparer M_1 et M_2. On calcule pour cela la valeur u de la statistique U :

$$U = \frac{M_1 - M_2}{[S_1^2/(n_1-1) + S_2^2/(n_2-1)]^{1/2}}$$

Une grande valeur absolue de la v.a. U signifie évidemment que M_1 et M_2 ont pris des valeurs très différentes : on rejettera l'hypothèse H_0 d'égalité si la valeur absolue $|u|$ est trop grande pour que l'égalité des moyennes soit vraisemblable.

Lorsque les moyennes théoriques sont égales et que la taille des échantillons est suffisante, on peut considérer que U suit approximativement la loi normale centrée réduite. On en déduit la valeur u_α en fonction du risque de première espèce choisi , de façon que :

$$P(|U| > u_\alpha) = \alpha$$

Définition : la région critique de la statistique U du test de comparaison de moyennes est de la forme :

$$RC =]-\infty, -u_\alpha] \cup [u_\alpha, +\infty[$$

u_α étant calculé de façon que :

$$P(|U| > u_\alpha) = \alpha$$

la v.a. U suivant la loi normale centrée réduite.

Exemple *: la moyenne des achats de l'autre hypermarché (410F) a été calculée sur 100 clients. La variance des achats calculée sur ces 100 clients est égale à 35401.01.*

On en déduit :

$$T = \frac{316.95 - 410}{[\,42902.47 / 49 + 35401.01 / 99\,]^{1/2}} = -2.65$$

Pour un risque de première espèce $\alpha = 0.05$, on a $u_\alpha = 1.96$. La valeur observée t appartient à la région critique ($|t| > 1.96$) et on rejette donc l'hypothèse nulle : la différence entre les deux moyennes n'est vraisemblablement pas due uniquement au hasard.

Répétons que l'utilisation de ce test est justifiée lorsque les variables suivent la loi normale et que leurs variances théoriques peuvent être considérées comme égales (cf. ci-dessous). Cette dernière condition doit être vérifiée surtout lorsque les échantillons sont de faibles effectifs.

5.2 Tests sur la variance.

Le test d'égalité de la variance d'une population à une valeur spécifiée est lui aussi équivalent à l'estimation par intervalle de confiance.

Il consiste à décider entre les hypothèses suivantes :

$$H_0 : \sigma^2 = \sigma_0^2 \qquad\qquad H_1 : \sigma^2 \neq \sigma_0^2$$

La question posée est la suivante : les observations remettent-elles en cause l'égalité de la variance théorique σ^2 à la valeur spécifiée σ_0^2 ?

<u>Règle de décision :</u> pour un risque de première espèce α,

- On accepte l'hypothèse nulle si la valeur σ_0^2 appartient à l'intervalle de confiance de niveau de confiance $1 - \alpha$;
- On rejette l'hypothèse nulle sinon.

Établissons la région critique du test. On rejette l'hypothèse nulle lorsque la valeur observée de la variance empirique S^2 est très différente de la valeur σ_0^2, donc lorsque la v.a. $X^2 = n\, S^2/\sigma_0^2$ prend une valeur anormalement petite (inférieure à χ_α^2) ou anormalement grande (supérieure à $\chi_{1-\alpha}^2$). Les bornes de la région critique χ_α^2 et $\chi_{1-\alpha}^2$ sont définies par la relation :

$$P(X^2 < \chi_\alpha^2) = \alpha/2 \qquad P(X^2 > \chi_{1-\alpha}^2) = \alpha/2$$

la statistique X^2 suivant la loi du χ^2 de degré de liberté n-1.

La région critique est donc de la forme :

$$\boxed{RC = [0, \chi_\alpha^2] \cup [\chi_{1-\alpha}^2, +\infty[}$$

<u>Exemple</u> *: nous supposons que la loi de probabilité de la v.a. âge est la loi normale (en éliminant les trois clients retraités) et testons l'hypothèse $H_0 : \sigma^2 = \sigma_0^2 = 50$.*

La valeur observée sur les 47 clients est $s^2 = 47.86$. On en déduit :

$$X^2 = 47 \times 47.86 / 50 = 44.99$$

Nous choisissons comme risque de première espèce $\alpha = 0.05$. La table donne directement pour le degré de liberté $\nu = 46$:

$$\chi_\alpha^2 = 29.160 \quad \text{tel que } P(X^2<\chi_\alpha^2) = 0.025$$

$$\chi_{1-\alpha}^2 = 66.617 \quad \text{tel que } P(X^2>\chi_{1-\alpha}^2) = 0.025$$

La région critique est : RC = [0, 29.160] ∪ [66.617, + ∞ [. La valeur observée n'appartient pas à la région critique et on accepte l'hypothèse nulle.

En déterminant l'ensemble des valeurs σ^2 *telles que l'on accepte l'hypothèse nulle, on retrouvera l'intervalle de confiance déterminé dans le chapitre 5. On pourra examiner aussi la figure 11 du chapitre 5.*

De la même façon que nous avons comparé deux moyennes entre elles, nous allons comparer deux variances.

On considère maintenant deux v.a. X_1 et X_2, observées sur deux échantillons de taille n_1 et n_2. On suppose que ces deux v.a. suivent la loi normale et sont indépendantes. Il s'agit de comparer leurs variances σ_1^2 et σ_2^2.

S_1^2 et S_2^2 étant les estimateurs empiriques des variances σ_1^2 et σ_2^2, on sait que les v.a. $n_1 S_1^2 / \sigma_1^2$ et $n_2 S_2^2/\sigma_2^2$ suivent une loi du χ^2 de degré de liberté n_1-1 et n_2-1. Les mathématiques nous donnent la loi du rapport :

Théorème : la loi de la v.a. F ci-dessous est la loi de Fisher de degrés de liberté n_1-1 et n_2-1.

$$F = \frac{n_1 S_1^2 / \sigma_1^2}{n_2 S_2^2 / \sigma_2^2} \times \frac{n_2 - 1}{n_1 - 1}$$

Considérons maintenant les hypothèses suivantes :

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2 \qquad\qquad H_1 : \sigma_1^2 \neq \sigma_2^2$$

Si l'hypothèse nulle est vraie, on a :

$$F = \frac{n_1 S_1^2}{n_2 S_2^2} \times \frac{n_2 - 1}{n_1 - 1} = \frac{n_1 S_1^2}{n_1 - 1} \Big/ \frac{n_2 S_2^2}{n_2 - 1}$$

et F devrait être proche de 1 puisque le numérateur et le dénominateur du rapport ci-dessus sont des estimateurs sans biais de la même variance σ^2. Si la v. a. F prend une très grande valeur ou est très proche de 0, on rejette l'hypothèse nulle. La région critique est de la forme :

$$\boxed{RC =]0, f_\alpha[\cup] f_{1-\alpha}, + \infty [}$$

les bornes f_α et $f_{1-\alpha}$ étant choisies dans la table de Ficher Snedecor de façon que :

$$P(F< f_\alpha) = \alpha/2 \qquad P(F> f_{1-\alpha}) = \alpha/2$$

Exemple *: Pour contrôler l'égalité des moyennes des achats des deux hypermarchés, nous avons supposé que les variances théoriques étaient égales. Nous le vérifions ci-dessous, en supposant que les lois sont normales . Les variances observées des achats sont égale à :*

$$s_1^2 = 42902.47\ (n_1 = 50) \qquad s_2^2 = 35401.01\ (n_2 = 100)$$

On en déduit :

$$f = \frac{50 \times 42902.47 / 49}{100 \times 35401.01 / 99} = 1.2243$$

Nous choisissons comme risque de première espèce $\alpha = 0.02$. Les degrés de liberté sont $\nu_1 = 49$ et $\nu_2 = 99$.La table donne directement $f_{1-\alpha} = 1.73$.

Pour calculer f_a, il faut considérer la v.a. 1/F , qui suit la loi de Fisher de degrés de liberté $\nu_1 = 99$ et $\nu_2 = 49$. On a :

$$P(1/F>1/f_a) = 0.01 \Leftrightarrow 1/f_\alpha = 1.82 \Leftrightarrow f_\alpha = 0.549$$

La région critique est donc : RC =]0, 0.549 [ ] 1.73, + ∞ [.

La valeur observée n'appartient pas à cette région critique et on accepte l'hypothèse d'égalité des variances.

5.3 Introduction à la fonction puissance.

Définition : on appelle puissance d'un test la probabilité π de rejeter l'hypothèse nulle quand elle est fausse.

La puissance est liée au risque de seconde espèce de façon évidente :

$$\pi = 1 - \beta$$

puisque le risque de seconde espèce est la probabilité d'accepter l'hypothèse H_0 quand elle est fausse.

Dans les tests présentés précédemment, l'hypothèse alternative n'est pas précisée, et, par suite, on ne peut déterminer analytiquement la loi de probabilité de la v.a. considérée en supposant l'hypothèse alternative vraie. On ne peut donc pas évaluer analytiquement la puissance. Le test sur la variance donne une occasion de présenter simplement la notion de fonction puissance.

Pour définir la puissance du test, nous allons modifier les hypothèses et fixer une valeur σ_1^2 à la variance dans le cas de l'hypothèse alternative. Les hypothèses sont alors les suivantes :

$$H_0 : \sigma^2 = \sigma_0^2 \qquad\qquad H_1 : \sigma^2 = \sigma_1^2$$

Nous allons pouvoir maintenant calculer la puissance du test, ou, ce qui est équivalent, le risque de seconde espèce.

On accepte l'hypothèse nulle lorsque la v.a. n S^2/σ_0^2 n'appartient pas à la région critique. Le risque β est donc la probabilité de l'événement ci-dessous :

$$\beta = P\left(\chi_\alpha^2 < \frac{n\,S^2}{\sigma_0^2} < \chi_{1-\alpha}^2\right)$$

lorsque la variance de la v.a. X est égale à σ_1^2 .

Lorsque l'hypothèse alternative est vraie, on ne connaît pas la loi de la v.a. n S^2 / σ_0^2 : c'est la v.a. n S^2 / σ_1^2 qui suit la loi du χ^2 de degré de liberté égal à n –1. Nous avons :

$$\frac{n\,S^2}{\sigma_0^2} = \frac{n\,S^2}{\sigma_1^2} \times \frac{\sigma_1^2}{\sigma_0^2}$$

On en déduit le risque de seconde espèce :

$$\beta = P\left(\chi_\alpha^2 < \frac{n\,S^2}{\sigma_0^2} < \chi_{1-\alpha}^2\right)$$

$$= P\left(\chi_\alpha^2 < \frac{n\,S^2\,\sigma_1^2}{\sigma_0^2\,\sigma_1^2} < \chi_{1-\alpha}^2\right)$$

$$= P\left(\frac{\chi_\alpha^2\,\sigma_0^2}{\sigma_1^2} < \frac{n\,S^2}{\sigma_1^2} < \frac{\chi_{1-\alpha}^2\,\sigma_0^2}{\sigma_1^2}\right)$$

On peut calculer cette probabilité lorsque l'hypothèse alternative est vraie puisque la loi de la v.a. n S^2 / σ_1^2 est connue : c'est la loi du χ^2 de degré de liberté n – 1. On en déduit évidemment la puissance $\pi = 1 - \beta$.

La puissance dépend de deux paramètres : la variance σ_1^2 choisie pour caractériser l'hypothèse alternative, et le risque de première espèce qui intervient dans la région critique. On suppose en général le second fixé, et on définit la fonction puissance comme la fonction qui associe à σ_1^2 la puissance du test.

Il y a un point particulier : pour $\sigma_1^2 = \sigma_0^2$, la fonction puissance est la probabilité de rejeter l'égalité $\sigma^2 = \sigma_0^2$ alors qu'elle est vraie puisque $\sigma_1^2 = \sigma_0^2$. On retrouve donc le risque de première espèce α.

Exemple *: nous donnons ci-dessous la fonction puissance du test sur la variance de l'âge. La valeur testée est fixée à $\sigma_0^2 = 50$, le risque de première espèce à 0.05, et le nombre d'observations est égal à 47 (cf. exemple précédent). Les valeurs sont données dans le tableau ci-dessous.*

La lecture de ce tableau donne le renseignement suivant : la probabilité de rejeter l'hypothèse $\sigma_0^2 = 50$ lorsque la vraie valeur est 33.333 est égale à 0.432 pour un risque de première espèce égal à 0.05.

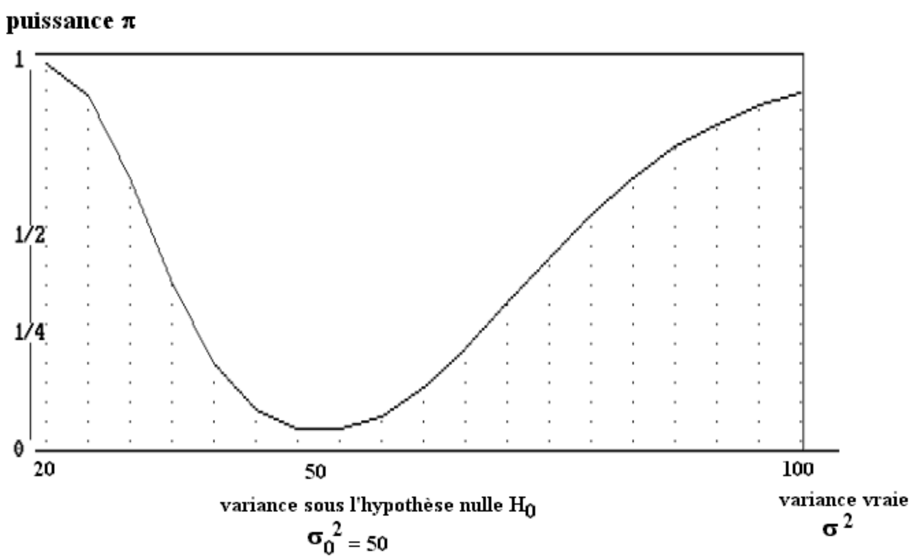

Figure 5.6 : Fonction puissance ($\sigma_0^2 = 50$, $\nu = 46$, $\alpha = 0.05$)

Rang	*variance vraie*	*puissance*	*Rang*	*variance vraie*	*puissance*
1	*20.0000*	*0.993*	*11*	*64.4444*	*0.263*
2	*24.4444*	*0.915*	*12*	*68.8889*	*0.379*
3	*28.8889*	*0.699*	*13*	*73.3333*	*0.497*

4	*33.3333*	*0.432*	*14*	*77.7778*	*0.606*
5	*37.7778*	*0.227*	*15*	*82.2222*	*0.701*
6	*42.2222*	*0.109*	*16*	*86.6667*	*0.778*
7	*46.6667*	*0.057*	*17*	*91.1111*	*0.839*
8	*51.1111*	*0.053*	*18*	*95.5556*	*0.885*
9	*55.5556*	*0.090*	*19*	*100.0000*	*0.919*
10	*60.0000*	*0.162*			

CONCLUSION

Les tests statistiques que nous avons présentés aboutissent à une règle de décision qui mérite toujours d’être examinée de façon critique. Nous n’avons présenté que les méthodes les plus simples, en tentant de montrer les difficultés de raisonnement dues à une absence de réflexion préalable sur la question posée. Une approche beaucoup plus générale des tests a été proposée par Neyman et Pearson. Fondée sur la notion de vraisemblance et sur les propriétés de la fonction puissance : elle sort du cadre de cet ouvrage, comme la statistique bayesienne à laquelle nous avons fait allusion.

TABLE DES MATIÈRES