Revue de statistique appliquée

G. Calot

Significatif ou non significatif ? Réflexions à propos de la théorie et de la pratique des tests statistiques

Revue de statistique appliquée, tome 15, nº 1 (1967), p. 5-69
<http://www.numdam.org/item?id=RSA_1967__15_1_5_0>

SIGNIFICATIF OU NON SIGNIFICATIF ?

RÉFLEXIONS A PROPOS DE LA THÉORIE ET DE LA PRATIQUE DES TESTS STATISTIQUES

G. CALOT

ancien élève de l'École Polytechnique, administrateur de l'I.N.S.E.E

La théorie des tests de NEYMAN et PEARSON présentée à la Royal Statistical Society en 1942 est très généralement admise comme fondement théorique des tests classiques.

Cette théorie envisage d'abord les tests d'une hypothèse simple contre une hypothèse simple. Ainsi dans le cas uniparamétrique, on compare, sur la base d'un échantillon d'observations indépendantes, deux valeurs particulières du paramètre : θ_1 et θ_2, ayant admis a priori que la vraie valeur θ^* du paramètre ne peut être égale qu'à l'une de ces deux valeurs. Dans ce cadre simplifié, et assez peu réaliste du point de vue pratique, la théorie de NEYMAN et PEARSON conduit à la détermination, non pas d'une règle unique optimum, mais d'une famille de règles optimum : chaque règle de la famille est caractérisée par l'un ou l'autre de trois nombres, le risque de première espèce α, le risque de seconde espèce β, le rapport de vraisemblance k sur la frontière de la région critique. La sélection d'une règle unique parmi la famille optimum est souvent effectuée en choisissant arbitrairement la valeur du risque de première espèce α, ce qui revient à faire jouer un rôle privilégié à l'hypothèse alternative.

Lorsqu'on passe ensuite à l'étude des tests d'une hypothèse simple contre une hypothèse composite, on examine le cas où il existe un test uniformément le plus puissant (uniformly most powerful ou UMP). Ce cas exceptionnel se présente lorsqu'on compare une valeur particulière θ_0 d'un paramètre contre un ensemble de valeurs Θ_1 soit uniquement supérieures, soit uniquement inférieures à θ_0 (paramètre p d'une loi binomiale, m d'une loi normale d'écart-type connu, σ d'une loi normale de moyenne connue, m d'une loi de Poisson,...). Il existe alors non pas un test UMP mais une famille de tests UMP : chaque règle de la famille des tests UMP peut être identifiée soit par le risque α de première espèce, soit par la valeur β_1 du risque de seconde espèce correspondant à une valeur particulière θ_1 de l'hypothèse alternative, soit encore par le rapport k_1 des vraisemblances correspondant à θ_1. En général, on procède à la sélection d'une règle particulière parmi la famille, en se fixant arbitrairement la valeur α du risque de première espèce.

Lorsqu'il n'existe pas de test UMP, le problème se complique et il faut faire choix d'un critère supplémentaire pour aboutir à la sélection d'une règle unique (tests de type λ ou de type γ), une fois qu'on a fixé la valeur du risque α de première espèce. C'est la situation qu'on rencontre lorsqu'on considère le cas uniparamètrique où l'hypothèse alternative comporte des valeurs du paramètre, les unes supérieures, les autres inférieures à θ_0.

Enfin, le cas des tests où les hypothèses testée et alternative sont l'une et l'autre composites présente des difficultés encore accrues. La notion de régions semblables a été introduite parce qu'on privilégie l'hypothèse testée (la fonction de risque de première espèce est alors uniformément bornée supérieurement, éventuellement constante et égale à une valeur fixée).

Toutefois, il y a lieu de se demander - et c'est là une préoccupation que doit avoir constamment le praticien - si la solution que donne le théoricien à un problème donné répond bien au problème concret qui est posé. Le praticien est en effet toujours exposé au risque d'une erreur de troisième espèce consistant à résoudre parfaitement... un problème différent de celui qu'il voudrait traiter. C'est la question fondamentale à laquelle M. CALOT essaie d'apporter une réponse à la fois critique et constructive, d'ailleurs fondée sur les résultats généraux de la théorie des tests classiques.

N. D. L. R.

SIGNIFICATIF OU NON SIGNIFICATIF ?

RÉFLEXIONS A PROPOS DE LA THÉORIE ET DE LA PRATIQUE DES TESTS STATISTIQUES

G. CALOT
ancien élève de l'École Polytechnique, administrateur de l'I.N.S.E.E

Le statisticien use abondamment du terme significatif : *telle différence est* hautement significative, *tel test est effectué avec un* niveau de signification *de 95 %, de 99 %, etc. Dans cet article, nous nous interrogeons sur le sens qu'il convient d'attacher à ces expressions. Il nous apparait en effet que la notion même de* significatif *est ambigüe et donne lieu à malentendus entre le* statisticien *chargé d'analyser des données chiffrées et d'en déduire des conclusions synthétiques et l'*homme de l'art(*)*, pris au sens très large du terme : industriel, médecin, agronome, administrateur,... qui utilise ces conclusions pour passer au plan de la décision pratique. Le premier attache à l'adjectif* significatif *un sens* probabiliste *précis qu'on peut résumer approximativement de la facon suivante : l'écart constaté entre* l'hypothèse prise pour référence *et le* reflet de la réalité *fourni par les observations est* significatif *s'il est* supérieur *à ce qu'on peut* raisonnablement *attendre du hasard, lorsqu'il y a* coïncidence parfaite *entre la réalité et l'hypothèse de référence. Au contraire, pour l'homme de l'art, l'adjectif* significatif *se rapporte à l'écart qui sépare la réalité et l'hypothèse de référence et n'a rien à voir fondamentalement avec le caractère aléatoire du flou qui environne l'information disponible : un écart significatif est un écart* substantiel, conséquent, décisif, *entre réalité et hypothèse de référence,* du point de vue du problème particulier envisagé.

Dans quelle mesure ces deux acceptions de l'adjectif significatif se rejoignent-elles ? De façon plus précise, dans quelle mesure la solution du statisticien est-elle une solution convenable au problème posé par l'homme de l'art ?

(*) Nous distinguons le statisticien de l'homme de l'art pour la commodité de l'exposé. Les malentendus évoqués ici seront d'autant mieux dissipés que statisticien et homme de l'art se confondront dans la même personne.

SOMMAIRE

Pages

I. LE PROBLÈME DU TEST DANS LE DOMAINE NON ALÉATOIRE

Avant d'envisager le problème du test statistique, c'est-à-dire le problème de la décision fondée sur des observations de nature *aléatoire*, il convient d'examiner le problème analogue, posé dans un contexte purement *déterministe*. On ne peut en effet espérer résoudre convenablement le problème aléatoire, si on ne dispose pas d'une solution satisfaisante au problème déterministe. D'une part, cette solution servira de *repère* à l'analyse du problème déterministe ; d'autre part, supposer les observations exemptes de flou aléatoire ne modifie pas *fondamentalement* l'énoncé du problème posé par l'homme de l'art.

Effectuer un test statistique, c'est réaliser une épreuve dont l'issue prendra l'une de *deux formes exclusives*, qu'on peut symboliser par OUI et par NON. Un test statistique est ainsi analogue à un réactif chimique qui virera ou qui ne virera pas, à un indicateur électrique dont la lampe s'allumera ou ne s'allumera pas.

D'une façon générale, le problème du test est celui de la décision à *deux niveaux*, c'est-à-dire, en définitive et malgré toute sa complexité, le problème de décision le plus simple qui soit.

Pour décrire une méthode de test, nous conviendrons dans la suite d'adopter le langage des instruments de mesure : une méthode de test (plan d'échantillonnage + règle de test) est un *instrument de mesure* sur lequel ne figurent que *deux graduations* : OUI et NON.

1.1 - LES PERFORMANCES DE L'INSTRUMENT

Portons notre attention sur une classe particulière d'instruments de mesure : ceux qu'on peut caractériser par un *seuil de sensibilité*. C'est par exemple tel réactif chimique qui ne vire que si l'acidité de la solution dépasse un certain seuil.

En nous plaçant dans un contexte déterministe, nous supposerons que la réponse de l'instrument est *entièrement conditionnée* par la mesure m de la grandeur faisant l'objet de l'expérience. Les performances de l'instrument sont alors entièrement *résumées* par le seuil s de sensibilité :

$$(m < s) \Longleftrightarrow \text{(l'instrument répond OUI) ;}$$
$$(m > s) \Longleftrightarrow \text{(l'instrument répond NON).}$$

Nous convenons ainsi que la réponse OUI s'applique aux valeurs relativement *faibles* de m, NON aux valeurs relativement *fortes*.

Appelons *courbe d'efficacité de l'instrument* la courbe représentative de la fonction P(m) :

P(m) = probabilité que l'instrument réponde OUI si la valeur vraie de la grandeur considérée est m.

P(m) vaut 1, lorsque m est inférieure à s ; P(m) vaut 0, lorsque m est supérieure à s (Fig. 1).

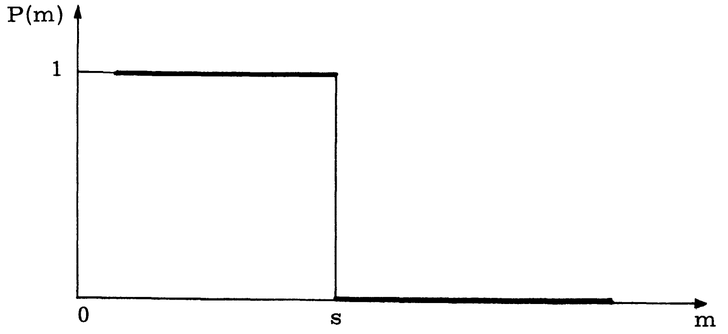

Figure 1

1.2 - LES SPECIFICATIONS DU PROBLEME POSE.

Le type particulier d'instruments qui vient d'être décrit permet de répondre à des questions du genre :

la vraie valeur m^* de la grandeur considérée est-elle *relativement faible*, ou *relativement forte ?*

Il convient évidemment de préciser ce qu'on entend par *relativement faible* et *relativement forte :* pour ce faire, *il faut définir à l'avance laquelle des deux réponses OUI et NON on souhaiterait que l'instrument fournisse, lorsque la vraie valeur m^* est égale à m.* L'utilisateur de l'instrument doit ainsi se livrer à une *analyse préalable des spécifications du problème particulier considéré.* Désignons par (O) l'ensemble des valeurs m tel que, si m^* appartient à (O), l'utilisateur souhaiterait la réponse OUI ; par (N), l'ensemble des valeurs m tel que, si m^* appartient à (N), l'utilisateur souhaiterait la réponse NON.

Si on admet, pour le type de problème envisagé, que les spécifications satisfont aux deux conditions générales :

$$[m \in (O) \quad \text{et} \quad m' < m] \Longrightarrow [m' \in (O)] \quad ;$$
$$[m \in (N) \quad \text{et} \quad m'' > m] \Longrightarrow [m'' \in (N)] \quad ;$$

les ensembles (O) et (N) sont de la forme (Fig. 2) :

$$\left.\begin{array}{l} (O) : m \leqslant m_1 \\ (N) : m \geqslant m_2 \end{array}\right\} \text{avec } m_1 < m_2 \, .$$

L'intervalle (m_1 , m_2) définit un ensemble (I) de valeurs m tel que, si la *vraie* valeur m^*, supposée *connue* de l'utilisateur, était comprise entre m_1 et m_2, celui-ci ne saurait trop quelle réponse il aimerait recevoir de l'instrument : sans être *substantiellement faible*, une telle valeur m^* n'est pas non plus *substantiellement forte*. Nous appellerons l'ensemble (I) *ensemble d'indécision* : lorsque m^* appartient à l'ensemble d'indécision, il est indifférent à l'utilisateur que la réponse soit OUI ou NON.

En résumé, les spécifications du problème envisagé peuvent s'énoncer de la façon suivante :

$m^* \leqslant m_1$: alors l'utilisateur voudrait recevoir la réponse OUI ;

$m^* \geqslant m_2$: alors l'utilisateur voudrait recevoir la réponse NON ;

ou encore, *de façon équivalente :*

$m^* \leqslant m_1$: alors l'utilisateur ne voudrait pas recevoir la réponse NON ;

$m^* \geqslant m_2$: alors l'utilisateur ne voudrait pas recevoir la réponse OUI.

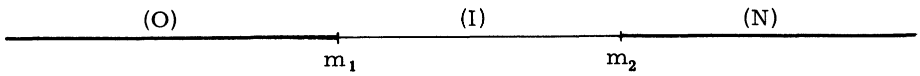

Figure 2

1. 3 - L'ADEQUATION DE L'INSTRUMENT AUX SPECIFICATIONS DU PROBLEME POSE.

Les *performances* d'un instrument appartenant à la classe envisagée sont résumées par le *seuil de sensibilité* s. Par ailleurs, les *spécifications* du problème étudié sont résumées par le couple (m_1 , m_2), limites de *l'intervalle d'indécision* (I). Dans quelle mesure un instrument donné est-il *adapté* à un problème donné ? Peut-on, sous certaines conditions, utiliser un instrument *inadapté* ?

Nous conviendrons de donner les réponses suivantes à ces deux questions fondamentales :

a. Un instrument caractérisé par le seuil de sensibilité s est *adapté* à un problème, dont les spécifications correspondent à (m_1 , m_2), si s est compris entre m_1 et m_2 :

$$m_1 \leqslant s \leqslant m_2.$$

En effet, l'instrument satisfait alors aux deux spécifications *simultanément* du problème considéré :

1/ : $[m^* \in (O)] \Longrightarrow [m^* < s] \Longleftrightarrow$ [la réponse de l'instrument est OUI] ;

2/ : $[m^* \in (N)] \Longrightarrow [m^* > s] \Longleftrightarrow$ [la réponse de l'instrument est NON];

c'est-à-dire encore, *de façon équivalente :*

1'/ : [la réponse de l'instrument est NON] $\Longleftrightarrow [m^* > s] \Longrightarrow [m^* \notin (O)$ car $s \geqslant m_1]$;

2'/ : [la réponse de l'instrument est OUI] $\Longleftrightarrow [m^* < s] \Longrightarrow [m^* \notin (N)$ car $s \leqslant m_2]$.

Un instrument *inadapté* au problème considéré correspond à :

- ou bien : $s > m_2$;
- ou bien : $s < m_1$.

Dans le premier cas, l'instrument est *insuffisamment sensible* ; dans le second cas, il est *trop sensible.*

b. Un instrument *inadapté* peut toutefois fournir des résultats intéressants. En effet, quel que soit le défaut d'adaptation, l'une des deux spécifications est satisfaite :

- *instrument insuffisamment sensible*: $s > m_2$; donc $s > m_1$;

- *instrument trop sensible* : $s < m_1$; donc $s < m_2$.

En conséquence :

$[s > m_2$ et la réponse de l'instrument est NON$] \Longleftrightarrow [m^* > s] \Longrightarrow [m^* \notin (O)$ car $s > m_1]$;

$[s < m_1$ et la réponse de l'instrument est OUI$] \Longleftrightarrow [m^* < s] \Longrightarrow [m^* \notin (N)$ car $s < m_2]$.

Ainsi la réponse NON d'un instrument insuffisamment sensible comme la réponse OUI d'un instrument trop sensible sont des réponses valables.

En revanche, la réponse OUI d'un instrument insuffisamment sensible comme la réponse NON d'un instrument trop sensible sont des réponses inutilisables.

c. La question qui se pose, lorsqu'on se trouve dans l'une des deux dernières situations, est de savoir quelle attitude adopter. Il nous semble raisonnable d'envisager l'une des trois attitudes suivantes :

- ou bien *changer d'instrument* et en retenir un qui soit adapté au problème considéré ;

- ou bien *s'abstenir de formuler une conclusion,* faute de disposer d'information valable ;

- ou bien encore, *conclure* conformément à la réponse fournie par l'instrument inadapté, en attirant l'attention de l'homme de l'art sur *l'incertitude* de la conclusion.

1.4 - CONCLUSIONS.

Cette analyse sommaire d'un cas particulier de test (que nous appellerons test de la forme "< *contre* >"), posé dans un contexte *déterministe,* met en lumière les points suivants :

● La solution d'un problème donné est conditionnée par les *spécifications propres* du problème envisagé. Tel instrument adapté à un certain problème peut être inadapté à la résolution d'un autre problème. En conséquence, il ne saurait exister d'instrument "passe-partout", adapté à toutes sortes de problèmes dont les spécifications seraient différentes.

● Faute de connaître les *performances* de l'instrument qu'on utilise, on est dans l'ignorance de ce qu'on fait. Il en est de même si on ne connaît pas les *spécifications* particulières du problème envisagé.

● Un instrument *trop* sensible est tout aussi *inadapté* qu'un instrument *insuffisamment* sensible.

● Si l'instrument dont on dispose est inadapté aux spécifications du problème posé et si on décide de maintenir les spécifications sans pouvoir recourir à un instrument adapté, il est raisonnable d'adopter une *règle à trois niveaux* de conclusion et non à deux :

- Réponse OUI ,
- Réponse NON ;
- Réponse ABSTENTION.

Il nous semble en effet préférable de *s'abstenir* au moins provisoirement, plutôt que de donner une réponse incertaine. Cette démarche relève tout simplement de l'honnêteté intellectuelle ; elle devrait figurer en bonne place dans un code de déontologie du statisticien.

● De façon analogue à la remarque précédente, si l'instrument dont on dispose est inadapté aux spécifications du problème posé tandis que le recours à un instrument adapté est impossible et si l'homme de l'art veut absolument une réponse par OUI ou NON, il convient d'attirer son attention sur les performances précises de l'instrument inadapté utilisé.

II. LE PROBLÈME DU TEST DANS LE DOMAINE ALÉATOIRE

Le problème du test posé dans un contexte *aléatoire* se distingue de celui posé dans un contexte *déterministe* à trois points de vue :

- le statisticien doit *construire* son instrument : en général, il a à sa disposition non pas *un* instrument mais *une gamme d'instruments* parmi laquelle il lui faut en sélectionner un ;

- la courbe d'efficacité de chaque instrument ne varie pas de 0 à 1 avec un seul point de discontinuité : en général même, elle est *continue* ;

- l'homme de l'art doit exprimer les spécifications de son problème en termes à la fois *déterministes* (ensemble d'indécision) et *aléatoires* (risque encouru).

D'une façon générale, le problème du test statistique comporte deux étapes :

- la première consiste à définir le *type d'instruments* le mieux adapté à un *type de problèmes* donné. Cette première étape nécessite l'adoption d'un *critère de comparaison* entre règles de test (c'est à dire entre instruments), critère qui permette d'affirmer que telle règle est *préférable* à telle autre, du moins dans certains cas, c'est à dire lorsque deux règles sont *comparables.*

- la seconde étape consiste en la *sélection*, parmi la classe des règles optimales au regard du critère de comparaison adopté, d'une *règle unique* qui satisfasse non seulement *au mieux* mais encore *de façon convenable* aux spécifications du problème envisagé.

La première étape correspond à la démarche proposée par Neyman et Pearson. Nous en étudierons de façon détaillée les modalités sur un exemple dans le paragraphe 2. 1. Il convient d'insister sur les points suivants :

- la théorie de Neyman et Pearson ne conduit pas à la détermination d'une *règle de test unique* mais au contraire à la détermination d'une *classe de règles*, chaque règle de la classe étant identifiée par un ou plusieurs *paramètres* ;

- dans le cas fréquent où la classe des règles optimales au sens de Neyman et Pearson est *uniparamétrique*, on peut identifier chaque règle de la classe par le *risque de première espèce*, désigné habituellement par α. Ceci est le cas lorsque l'hypothèse testée est *simple* ou encore lorsque, l'hypothèse testée étant *composite*, la classe des régions critiques optimales constitue une famille de régions *semblables*. Néanmoins, la sélection d'une règle unique parmi la classe des règles optimales doit être effectuée sur la base de la *courbe d'efficacité* attachée à chaque règle et non par choix plus ou moins *arbitraire* du risque α de première espèce. Neyman a fortement insisté sur ce point qui semble souvent perdu de vue par les statisticiens.

● la théorie de Neyman et Pearson ne peut être appliquée que si on admet certaines hypothèses précises sur la nature analytique des lois de probabilité mises en jeu. Néanmoins, dans certains cas où de telles hypothèses ne sont pas admises, il est possible - en se référant ou non, par analogie, à la théorie de Neyman et Pearson - de définir *a priori* une classe de règles parmi lesquelles on en sélectionnera une, sur la base de la courbe d'efficacité : l'instrument qu'on construira ainsi ne présentera pas des qualités d'optimalité relativement à un critère de comparaison explicite mais il aura au moins l'avantage d'être un instrument *dont on connaît les performances*. Ce point nous paraît fondamental : *mieux vaut utiliser un instrument dont on connaît les limites, même s'il n'est pas le meilleur qui soit, plutôt que recourir à un instrument optimal dont on ignore les performances*. D'une part en effet, lorsque la quantité d'information disponible est insuffisante, le meilleur instrument est un instrument médiocre, sinon mauvais. D'autre part, il ne faut pas perdre de vue qu'un instrument trop sensible est tout aussi inadapté qu'un instrument insuffisamment sensible.

En ce qui concerne la sélection de la règle unique parmi une classe donnée de règles, nous conviendrons d'adopter la méthode suivante inspirée de la notion de *minimax* et généralement employée dans le domaine du contrôle industriel de la qualité (en particulier dans le contrôle qualitatif de réception) :

● nous *caractériserons* les performances d'une règle par *la plus grande probabilité d'erreur*, en appelant *erreur* le fait de répondre OUI alors qu'on désirerait la réponse NON ou de répondre NON alors qu'on désirerait la réponse OUI. En reprenant les notations du paragraphe 1, on aboutit à :

$$\delta^* = \max \left\{ \sup_{m \in (O)} [1-P(m)], \ \sup_{m \in (N)} P(m) \right\}$$

● nous considèrerons comme *convenable* une règle dont le risque maximum δ^* est *inférieur ou égal* à δ, seuil donné à l'avance :

$$\delta^* \leqslant \delta.$$

● nous sélectionnerons comme règle unique une règle *convenable* - s'il en existe une - : suivant le cas, ce sera ou bien *la règle convenable la plus économique* (lorsque l'avis du statisticien est demandé *avant* échantillonnage et qu'il peut recommander un plan de sondage) ou bien *la règle convenable qui possède le risque maximum le plus faible* lorsque le statisticien est consulté *après* échantillonnage).

● s'il n'existe aucune règle convenable, nous considèrerons que la *quantité d'information disponible est insuffisante* pour construire une règle à deux niveaux satisfaisant aux spécifications du problème envisagé. En reprenant les arguments présentés plus haut en 1. 3., nous conviendrons d'adopter alors l'une des trois attitudes suivantes :

- *accroître le nombre des observations* pour se ramener au cas où il existe une règle convenable ;

- *adopter une règle à trois niveaux* (OUI, NON, ABSTENTION);

- *adopter la règle à deux niveaux dont le risque maximum est le plus faible, en indiquant à l'homme de l'art la valeur (supérieure à δ) de ce risque maximum.*

Cette méthode de sélection peut être justifiée de la façon suivante : soit δ un niveau de risque qu'on accepte de courir : tout événement de probabilité inférieure à δ est *considéré comme impossible* et en conséquence *négligé*. δ est en quelque sorte *l'échelon minimum en termes de risque accepté.*

Supposons que la réponse du test est OUI. En vertu de la convention précédente et en reprenant les notations du paragraphe 1, on déduit que la vraie valeur m* est telle que la probabilité attachée à la réponse OUI dépasse δ. Dans le cas contraire en effet, l'instrument n'aurait pu fournir la réponse OUI. En conséquence, si la condition :

$$\forall m \in (N) \quad : \quad P(m) \leqslant \delta$$

est satisfaite, m* ne peut appartenir à (N) ; m* appartient donc à (O) ou à (I) et la réponse fournie par le test est *correcte*.

Inversement, si la réponse du test est NON, la vraie valeur m* ne peut appartenir à (O) si la condition :

$$\forall m \in (O) \quad : \quad 1 - P(m) \leqslant \delta$$

est satisfaite et en conséquence la réponse fournie par le test est *correcte*.

L'argument essentiel à la base de la méthode de sélection est ainsi le suivant : *la vraie valeur m* doit attacher à la réponse fournie par le test une probabilité supérieure à δ car, si ce n'était pas le cas, il faudrait admettre qu'un évènement de probabilité au plus égale à δ s'est produit, ce qui par convention est rejeté.* Cette démarche est ainsi analogue à celle qu'on emprunte, lorsqu'on construit un intervalle de confiance : une règle de test convenable - au sens où nous avons défini ce terme - est une règle qui fournit la réponse correcte avec une probabilité au moins égale à $1-\delta$, *quelle que soit la réponse correcte*. De même, une règle de construction d'un intervalle de confiance au niveau $1-\alpha$ est une règle qui fournit un intervalle recouvrant la vraie valeur du paramètre avec une probabilité égale à $1-\alpha$, *quelle que soit cette vraie valeur*.

On pourrait objecter que les erreurs de *première et de seconde espèce* étant de nature différente et par conséquent assorties de coûts différents, il serait préférable de retenir *deux seuils* δ_1 et δ_2, non nécessairement égaux, et d'assurer les deux conditions :

$$\sup_{m \in (O)} [1 - P(m)] \leqslant \delta_1 ,$$

$$\sup_{m \in (N)} P(m) \leqslant \delta_2 .$$

En réalité, tout dépend de la façon dont on a choisi les ensembles (O) et (N) pour caractériser les spécifications du problème posé. Nous admettons implicitement dans la méthode décrite ci-dessus que les ensembles (O) et (N) ont été définis de *manière équivalente*. Ainsi dans le cas du test "< contre >", m_1 représente parmi les valeurs *relativement faibles* de m une situation aussi *extrême* que m_2 parmi des valeurs *relativement fortes* de m.

Avant de juger plus en détail des avantages et inconvénients de la méthode proposée, envisageons son application pratique dans les exemples simples suivants : tests de signification, de comparaison et d'homogénéité de moyennes (dans le cas de populations normales d'écarts-types connus ou non).

2.1 - TEST DE SIGNIFICATION D'UNE MOYENNE : TEST "< CONTRE>", L'ECART-TYPE ETANT CONNU.

Considérons le problème suivant : sur la base de n observations *indépendantes* d'une loi *normale* $\mathcal{N}(m, \sigma_0)$, où σ_0 est supposé *connu*, test de l'hypothèse : *a est relativement faible* contre l'hypothèse *: m est relativement forte*.

Supposons que les spécifications du problème peuvent être résumées par le triplet (m_1, m_2, δ) :

$$\forall m \leqslant m_1 \quad : \quad 1 - P(m) \leqslant \delta$$

$$\forall m \geqslant m_2 \quad : \quad P(m) \leqslant \delta$$

Du point de vue géométrique, ce triplet de spécifications signifie que la courbe d'efficacité doit passer dans la zone hachurée de la figure 3.

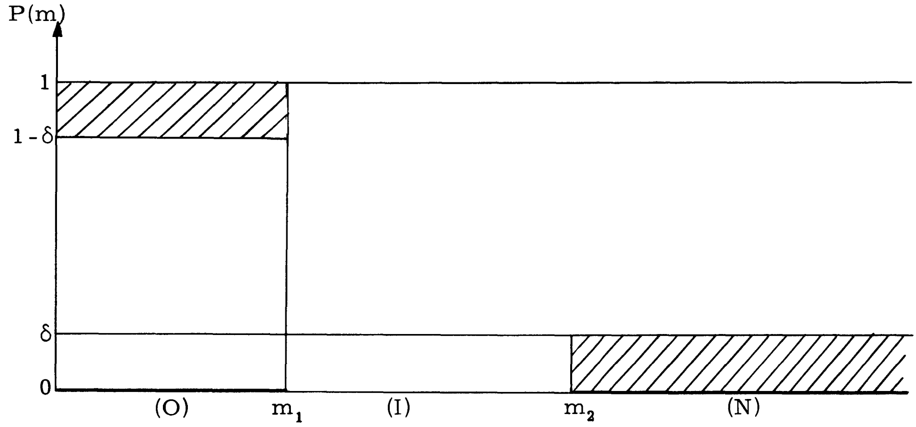

Figure 3

2.1.1 - La théorie de Neyman et Pearson.

La théorie de Neyman et Pearson repose sur le critère de comparaison suivant entre règles de test : Soit T un test correspondant à la région critique W (domaine de l'espace des observations où la réponse du test est NON). Soit $P_W(m)$ la probabilité que la réponse du test soit OUI, lorsque la vraie valeur de la moyenne est égale à m[$1-P_W(m)$ est la probabilité que la réponse du test soit NON, c'est à dire la probabilité attachée à la région critique W, lorsque la valeur vraie de la moyenne est égale à m]. Soit de même T' un test correspondant à la région critique W' et $P_{W'}(m)$, la probabilité que la réponse du test soit OUI, lorsque la vraie valeur de la moyenne est égale à m. Le test T est *préférable ou équivalent* au test T' si à la fois :

$$\forall m \in (O) \quad : \quad P_W(m) \geqslant P_{W'}(m)$$

$$\forall m \in (N) \quad : \quad P_W(m) \leqslant P_{W'}(m)$$

Le test T est *strictement préférable* au test T' (ou encore : le test T *domine* le test T') s'il existe *au moins une* valeur de m appartenant soit à (O), soit à (N) telle que l'inégalité correspondante soit stricte.

L'application de ce critère permet d'éliminer toute règle dominée par une autre et conduit à la *classe des règles optimales* au regard du critère considéré : c'est l'ensemble des règles dominées par aucune autre ; on peut dire encore qu'à toute règle n'appartenant pas à la classe, on peut associer au moins une règle de la classe qui la domine. En général, l'application de ce critère conduit à la définition d'une classe de règles optimales tellement riche que la sélection pratique d'une règle s'avère impossible.

Dans le problème particulier envisagé, on ne rencontre pas cette difficulté. En effet, la théorie de Neyman et Pearson nous apprend que si m_0, m_I, m_N sont trois valeurs de m appartenant respectivement à (O), (I) et (N), les régions critiques W satisfaisant à :

$$P_W(m_I) = \text{valeur fixée}$$

et rendant :

- problème (1) : $P_W(m_0) = \text{maximum}$

ou

- problème (2) : $P_W(m_N) = \text{minimum}$

sont de la forme :

$$W : \quad \bar{x} > c,$$

où c est une constante dépendant de la valeur fixée de $P_W(m_I)$ mais ne dépendant pas :

- problème (1) : de $m_0 \in$ (O) ;
- problème (2) : de $m_N \in$ (N) ;
- du problème (1) ou (2) envisagé.

Il en résulte que dans le problème particulier étudié, la classe des régions critiques optimales au regard du critère de comparaison considéré est de la forme :

$$W : \quad \bar{x} > c,$$

ce qui correspond à la classe des règles optimales :

- Réponse OUI si $\bar{x} \leqslant c$,
- Réponse NON si $\bar{x} > c$.

2.1.2 - Sélection d'une règle de test.

Il convient, maintenant qu'on a défini la classe des règles optimales, de sélectionner la règle *unique* qu'on appliquera, c'est à dire la valeur de c retenue.

Considérons une valeur c donnée. La probabilité $P_c(m)$ de répondre OUI, lorsque la moyenne est égale à m, est :

$$P_c(m) = \Pi\left(\frac{c-m}{\sigma_0/\sqrt{n}}\right),$$

où Π désigne la fonction cumulative de la variable normale centrée réduite :

$$\Pi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) d\xi .$$

En effet, sous les hypothèses envisagées, la loi de probabilité suivie par la moyenne $\bar{x}$ est *normale* :

$$\bar{x}/_{m} = \mathcal{N}(m , \sigma_0/\sqrt{n}) .$$

La fonction $P_c(m)$ est strictement *décroissante* par rapport à m puisque Π est une fonction strictement *croissante*. En conséquence, le triplet de spécifications (m_1, m_2, δ) sera assuré si c répond simultanément aux deux conditions :

1/ $$P_c(m_1) \geqslant 1 - \delta ,$$

2/ $$P_c(m_2) \leqslant \delta ,$$

c'est-à-dire encore, en désignant par u_δ le *quantile* d'ordre δ de la variable normale centrée réduite :

$$\Pi(u_\delta) = \delta :$$

1'/ $$c \geqslant m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} ,$$

2'/ $$c \leqslant m_2 + \frac{\sigma_0}{\sqrt{n}} u_\delta = m_2 - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} .$$

Ces deux conditions sont *compatibles* dans le seul cas où :

$$m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} \leqslant m_2 - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$$

c'est à dire encore, si n dépasse le seuil n_0 :

$$n \geqslant n_0 = \left(\frac{2 \sigma_0 u_{1-\delta}}{m_2 - m_1}\right)^2 .$$

On est amené ainsi à distinguer deux cas :

a. ou bien $n \geqslant n_0$: *la quantité d'information disponible est suffisante pour résoudre le problème posé.* En effet, le risque maximum encouru est le plus grand des deux nombres :

$$\delta_1 = \Pi\left(\frac{m_1 - c}{\sigma_0/\sqrt{n}}\right) \quad \text{et} \quad \delta_2 = \Pi\left(\frac{c - m_2}{\sigma_0/\sqrt{n}}\right) .$$

$\delta_1(c)$ est une fonction *décroissante* de c, $\delta_2(c)$ une fonction *croissante* de c. Le plus grand de ces deux nombres est minimum lorsqu'ils sont *égaux*, c'est à dire lorsque c est la moyenne arithmétique de m_1 et m_2 :

$$[\delta_1 = \delta_2] \Longleftrightarrow \left[\Pi\left(\frac{m_1 - c}{\sigma_0/\sqrt{n}}\right) = \Pi\left(\frac{c - m_2}{\sigma_0/\sqrt{n}}\right) \right] \Longleftrightarrow \left[c = \frac{m_1 + m_2}{2} \right] .$$

Le risque maximum encouru est alors :

$$\delta^* = 1 - \Pi\left(\sqrt{n}\ \frac{m_2 - m_1}{2\sigma_0}\right) \leqslant \delta .$$

D'où la régle de test à adopter lorsque n dépasse n_0 :

- Réponse OUI si $\bar{x}$ est inférieure à l'abscisse du *milieu de l'intervalle d'indécision* (I) : $\bar{x} \leqslant \frac{m_1 + m_2}{2}$

- Réponse NON dans le cas contraire.

b. ou bien $n < n_0$: *la quantité d'information disponible est insuffisante pour résoudre le problème posé, à l'aide d'une règle à deux niveaux.*

Dans ce cas, on peut adopter l'une des trois attitudes suivantes :

b_1. ou bien *accroître la taille* de l'échantillon, en prélevant $\Delta n = n_0 - n$ observations supplémentaires et se ramener au cas *a*.

b_2. ou bien *réduire la spécification du problème relative au risque* δ, en acceptant de courir un risque qui peut atteindre :

$$\delta^* = 1 - \Pi\left(\sqrt{n}\ \frac{m_2 - m_1}{2\sigma_0}\right) > \delta .$$

La règle à adopter est alors celle obtenue plus haut :

- Réponse OUI si $\bar{x} \leqslant \frac{m_1 + m_2}{2}$;

- Réponse NON dans le cas contraire.

b_3. ou bien *adopter une règle de décision à trois niveaux* : OUI, NON, ABSTENTION.

En effet, si n est inférieur à n_0, on a :

$$m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} > m_2 - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} .$$

On satisfait à la première spécification du problème :

$$\forall m \leqslant m_1 \quad : \quad P_c(m) \geqslant 1 - \delta ,$$

en convenant de ne répondre NON que dans le cas où $\bar{x}$ *dépasse* c_2, avec :

$$c_2 \geqslant m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$$

On a bien alors :

$$\forall m \leqslant m_1 \quad : \quad \Pr\left\{\text{Réponse NON}/_m\right\} - 1 - \Pi\left(\frac{c_2 - m}{\sigma_0/\sqrt{n}}\right) \leqslant 1 - \Pi(u_{1-\delta}) = \delta .$$

De la même façon, on satisfait à la seconde spécification du problème :

$$\forall m \geqslant m_2 \quad : \quad P_c(m) \leqslant \delta \ ,$$

en convenant de ne répondre OUI que dans le cas où $\overline{x}$ est *inférieur* à c_1, avec :

$$c_1 \leqslant m_2 - \frac{\sigma_o}{\sqrt{n}} u_{1-\delta} < c_2 \ .$$

On a alors :

$$\forall m \geqslant m_2 \quad : \quad Pr\left\{\text{Réponse OUI}/_m\right\} = \Pi\left(\frac{c_1 - m}{\sigma_o/\sqrt{n}}\right) \leqslant \Pi(-u_{1-\delta}) = \delta$$

En conséquence, la règle suivante à trois niveaux satisfait simultanément aux deux spécifications du problème :

- Réponse OUI si $\overline{x} \leqslant c_1$;
- Réponse NON si $\overline{x} \geqslant c_2$;
- Réponse ABSTENTION si $c_1 < \overline{x} < c_2$.

La probabilité de la réponse ABSTENTION est minimum *pour tout m* lorsque l'intervalle (c_1, c_2) est aussi court que possible, c'est à dire, compte tenu des contraintes précédentes, lorsque simultanément :

$$c_1 = m_2 - \frac{\sigma_o}{\sqrt{n}} u_{1-\delta} \ ,$$

$$c_2 = m_1 + \frac{\sigma_o}{\sqrt{n}} u_{1-\delta} \ .$$

D'où en définitive la règle à trois niveaux assortie d'un risque au plus égal à δ :

- Réponse OUI si $\overline{x} \leqslant m_2 - \frac{\sigma_o}{\sqrt{n}} u_{1-\delta}$;
- Réponse NON si $\overline{x} \geqslant m_1 + \frac{\sigma_o}{\sqrt{n}} u_{1-\delta}$;
- Réponse ABSTENTION si $m_2 - \frac{\sigma_o}{\sqrt{n}} u_{1-\delta} < \overline{x} < m_1 + \frac{\sigma_o}{\sqrt{n}} u_{1-\delta}$.

La probabilité attachée à la réponse ABSTENTION est une fonction de m dont le maximum, correspondant à $m = \frac{m_1 + m_2}{2}$, est égal à :

$$2\,\Pi\left[u_{1-\delta}\left(1 - \sqrt{\frac{n}{n_o}}\right)\right] - 1 \ .$$

2.1.3 - Conclusions.

A la suite de cette étude, il est possible de résoudre le problème posé par l'homme de l'art sous la forme (m_1, m_2, δ) :

a. Si l'avis du statisticien est demandé avant échantillonnage :

- Recommander de prélever un échantillon de n_0 observations :

$$n_0 = \left(\frac{2\,\sigma_0 u_{1-\delta}}{m_2 - m_1}\right)^2 .$$

- Recommander d'appliquer la règle à deux niveaux :

  - Réponse OUI si $\bar{x} \leqslant \frac{m_1 + m_2}{2}$;

  - Réponse NON dans le cas contraire.

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. la taille n de l'échantillon se trouve être au moins égale à n_0 :

- Recommander d'appliquer la règle à deux niveaux :

  - Réponse OUI si $\bar{x} \leqslant \frac{m_1 + m_2}{2}$;

  - Réponse NON dans le cas contraire.

- Indiquer à l'homme de l'art le risque maximum encouru δ^*, inférieur ou égal à δ .

$$\delta^* = 1 - \Pi\left(\sqrt{n}\,\frac{m_2 - m_1}{2\sigma_0}\right) \leqslant \delta .$$

b_2. la taille n de l'échantillon est inférieure à n_0. Dans ce cas, trois attitudes sont possibles :

Première attitude : Recommander de prélever $\Delta n = n_0 - n$ *observations complémentaires* et se ramener au cas *a*.

Seconde attitude : Recommander d'appliquer la règle à *trois niveaux :*

- Réponse OUI si $\bar{x} \leqslant m_2 - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$;

- Réponse NON si $\bar{x} \geqslant m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$;

- Réponse ABSTENTION si :

$$m_2 - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} < \bar{x} < m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta} .$$

Troisième attitude : Recommander d'appliquer la règle à *deux niveaux :*

- Réponse OUI si $\bar{x} \leqslant \frac{m_1 + m_2}{2}$

- Réponse NON dans le cas contraire.

en précisant à l'homme de l'art que le risque maximum encouru atteint :

$$\delta^* = 1 - \Pi\left(\sqrt{n}\ \frac{m_2 - m_1}{2\,\sigma_0}\right) > \delta\ .$$

2.1.4 - Application numérique.

Considérons le problème numérique suivant :

$$m_1 = 5,00\ ,$$
$$m_2 = 5,20\ ,$$
$$\sigma_0 = 0,50\ ,$$
$$\delta = 5\ \%\ .$$

Il vient :

$$n_0 = \left(\frac{2 \times 0,50 \times 1,645}{0,20}\right)^2 = 67,7\ .$$

D'où :

a. Si l'avis du statisticien est demandé avant échantillonnage :

● Recommander de prélever $n_0 = 68$ observations et d'appliquer la règle à deux niveaux :

- Réponse OUI si $\bar{x} \leqslant 5,10$;
- Réponse NON dans le cas contraire.

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. Si, par exemple, l'échantillon comporte 100 observations :

● Recommander d'appliquer la règle à deux niveaux précédente ;

● Indiquer à l'homme de l'art que le risque maximum encouru est égal à :

$$\delta^* = 1 - \Pi(2) = 2,3\ \%.$$

b_2. Si, par exemple, l'échantillon comporte 25 observations :

● ou bien recommander de prélever 43 observations *complémentaires* et se ramener au cas *a.* ;

● ou bien recommander d'appliquer la règle à trois niveaux :

- Réponse OUI si $\bar{x} \leqslant 5,0355$;
- Réponse NON si $\bar{x} \geqslant 5,1645$;
- Réponse ABSTENTION si $5,0355 < \bar{x} < 5,1645$.

● ou bien recommander d'appliquer la règle à deux niveaux indiquée plus haut :

- Réponse OUI si $\bar{x} \leqslant 5,10$;
- Réponse NON dans le cas contraire,

en indiquant à l'homme de l'art que le risque maximum encouru atteint :

$$\delta^* = 1 - \Pi(1) = 15,9\ \% .$$

2.1.5 - Remarques.

1. Le problème que nous venons de traiter d'après la méthode proposée plus haut est celui du test de *signification* d'une moyenne, qu'on énonce habituellement :

$$\text{Test} \quad m \leqslant m_0 \quad \text{contre} \quad m > m_0 .$$

Il convient d'observer toutefois que nous n'avons pas introduit de valeur m_0 qui séparerait l'hypothèse testée de l'hypothèse alternative. Dans la pratique, une telle valeur m_0 est généralement dépourvue de signification : il n'y a pas un point m_0 en deçà duquel se trouverait une certaine vérité et au-delà duquel se trouverait la vérité contraire. En fait, il y a toute une gamme de dégradés que nous avons résumés par le triplet (m_1, m_2, δ) : la probabilité de fournir une réponse incorrecte est au plus égale à δ, lorsque m^* est inférieure à m_1 ou supérieure à m_2.

En conséquence, le test effectué a été dénommé : test de l'hypothèse suivant laquelle m^* est *relativement faible* contre l'hypothèse suivant laquelle m^* est *relativement forte*, les termes *relativement faible* et *relativement forte* étant synthétisés par le triplet (m_1, m_2, δ).

Par ailleurs, la notion de risque de première espèce α n'a pas été utilisée explicitement : ce n'est donc pas par un choix portant *uniquement* sur α que la règle finalement adoptée a été sélectionnée mais par un choix portant sur *l'ensemble* de la courbe d'efficacité.

2. Dans les applications pratiques de la théorie des tests, l'hypothèse testée est généralement *privilégiée*, de par le choix du risque de première espèce. Il s'ensuit que les règles correspondant aux tests :

$$m \leqslant m_0 \quad \text{contre} \quad m > m_0 ;$$
$$m > m_0 \quad \text{contre} \quad m \leqslant m_0 ;$$

sont différentes, dans la mesure où on adopte le *même* risque de première espèce.

Au contraire, avec la méthode proposée, les hypothèses testée et alternative jouent un *rôle symétrique :* permuter hypothèse testée et hypothèse alternative ne modifie pas la règle sélectionnée d'après le *même* triplet (m_1, m_2, δ).

Si on veut reprendre le langage habituellement utilisé, le test que nous avons effectué devrait plutôt s'appeler :

$$\text{Test} \quad m \leqslant m_1 \quad \text{contre} \quad m \geqslant m_2 .$$

3. Lorsqu'on applique habituellement la théorie des tests en se référant à un risque de première espèce donné (en particulier la fameuse valeur $\alpha = 5\ \%$), on a l'impression que le statisticien est à même de fournir une conclusion, *quelle que soit la quantité d'information disponible* (c'est-à-dire quelle que soit la taille de l'échantillon) et *quel que soit le problème particulier considéré.* Ainsi, dans le test $m \leqslant m_1$ contre $m \geqslant m_2$, on adoptera la règle :

- Réponse OUI si $\overline{x} \leqslant m_1 + \frac{\sigma_o}{\sqrt{n}} u_{1-\alpha}$;

- Réponse NON dans le cas contraire ;

quels que soient m_2, σ_o et n. On a ainsi l'impression que la valeur fatidique de 5 % généralement retenue pour α est douée de propriétés mystérieuses qui permettent de résoudre *en toutes circonstances* le problème posé.

Ainsi dans l'exemple numérique envisagé plus haut, on retiendrait les règles suivantes :

n = 25 :	Réponse OUI	si	$\overline{x} \leqslant 5,165$
n = 68 :	Réponse OUI	si	$\overline{x} \leqslant 5,100$
n = 100 :	Réponse OUI	si	$\overline{x} \leqslant 5,082$

Les valeurs du risque de seconde espèce qui correspondent à ces règles sont respectivement, lorsque m = 5,200 :

$$\beta = 36\ \% \ ;$$
$$\beta = 5\ \% \ ;$$
$$\beta = 0,9\ \% .$$

Le fait de ne pas profiter de l'accroissement de la taille de l'échantillon pour diminuer *à la fois* les risques de première et de deuxième espèce est tout à fait injustifié. Au contraire avec la méthode proposée on aboutit respectivement à :

$$\alpha = \beta = 15,9\ \% \ ;$$
$$\alpha = \beta = 5\ \% \ ;$$
$$\alpha = \beta = 2,3\ \%.$$

4. Lorsque le statisticien est consulté *après* échantillonnage et que la taille n de l'échantillon se trouve être *inférieure* au seuil minimum n_o, nous proposons, dans le cas où on maintiendrait les spécifications (m_1, m_2, δ), le choix entre deux attitudes :

- ou bien ne pas regarder les résultats de l'échantillonnage et prélever $\Delta n = n_o - n$ observations complémentaires pour se ramener au cas où $n = n_o$;

- ou bien adopter une règle à trois niveaux : OUI, NON, ABSTENTION.

On pourrait également envisager une procédure *progressive* ou même *séquentielle* : à chaque pas, l'abstention serait *provisoire*. Dans ces conditions, l'échantillon initial de taille n serait le point de départ d'une procédure bien définie à l'avance. Du point de vue purement conceptuel, cette variante de la méthode proposée ne présente pas de difficultés particulières et elle apparaît même plus satisfaisante. En revanche, sur le plan pratique, elle se révèle d'un emploi difficile, en raison de complications d'ordre algébrique. Aussi conviendrons-nous de nous en tenir aux deux attitudes envisagées plus haut.

5. Les règles pratiques exposées en 2.1.3. peuvent encore être présentées de la façon suivante :

L'homme de l'art indique au statisticien les spécifications techniques du problème sous la forme (m_1, m_2), c'est-à-dire sous la forme de l'intervalle d'indécision (I).

a. S'il est consulté *avant échantillonnage* le statisticien demande à l'homme de l'art quel risque maximum δ il accepte de courir. Il en déduit alors :

- la taille n_0 de l'échantillon à prélever :

$$n_0 = \left(\frac{2\,\sigma_0\, u_{1-\delta}}{m_2 - m_1}\right)^2 \quad ;$$

- la règle à appliquer :

  - Réponse OUI si $\bar{x} \leqslant \dfrac{m_1 + m_2}{2}$;

  - Réponse NON dans le cas contraire.

b. S'il est consulté *après échantillonnage*, le statisticien demande à l'homme de l'art s'il accepte de courir un risque maximum δ^* égal à :

$$\delta^* = 1 - \Pi \left[\frac{\sqrt{n}}{2}\left(\frac{m_2 - m_1}{\sigma_0}\right)\right]$$

b_1. Si la réponse de l'homme de l'art est affirmative, la règle à appliquer est celle indiquée en *a.*

b_2. Si la réponse de l'homme de l'art est négative, les deux attitudes possibles sont déterminées par le risque maximum δ que l'homme de l'art veut bien accepter de courir :

Première attitude : prélever $n_0 - n$ observations *complémentaires* et se ramener au cas *a.*

Seconde attitude : adopter la *règle à trois niveaux* de conclusion.

- Réponse OUI si $\bar{x} \leqslant m_2 - \dfrac{\sigma_0}{\sqrt{n}}\, u_{1-\delta}$;

- Réponse NON si $\bar{x} \geqslant m_1 + \dfrac{\sigma_0}{\sqrt{n}}\, u_{1-\delta}$;

- Réponse ABSTENTION dans les autres cas.

6. Les autres tests présentés dans la suite de cet article sont tous effectués d'après la même démarche (2.2 à 2.4). Le lecteur pressé pourra en omettre le détail et passer directement au chapitre 3, page 65.

2.2 - TEST DE SIGNIFICATION D'UNE MOYENNE : TEST "= CONTRE ≠", L'ECART-TYPE ETANT CONNU.

Envisageons maintenant le problème suivant : sur la base de n observations *indépendantes* d'une loi *normale* $\mathcal{N}(m, \sigma_0)$, où σ_0 est supposé *connu*, test de l'hypothèse : *m est relativement voisine de* m_0 contre l'hypothèse : *m est relativement différente de* m_0.

Pour définir les spécifications du problème, il faut associer, de façon analogue au cas précédent, à un niveau δ de probabilité, deux ensembles (O) et (N), où on voudrait respectivement obtenir du test la réponse OUI et la réponse NON.

L'ensemble (O) est un ensemble de valeurs entourant m_0. L'ensemble (N) est formé de deux intervalles infinis, l'un situé à droite et l'autre à gauche de m_0. Nous désignerons par (m'_1, m''_1) les bornes de l'ensemble (O), par (m'_2, m''_2) les bornes de l'ensemble (N) (Fig. 4).

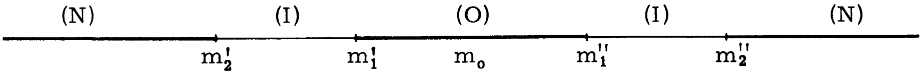

Figure 4

Le problème est donc celui de la sélection d'une courbe d'efficacité qui dans (O) et dans (N) soit comprise à l'intérieur des zones hachurées de la figure 5.

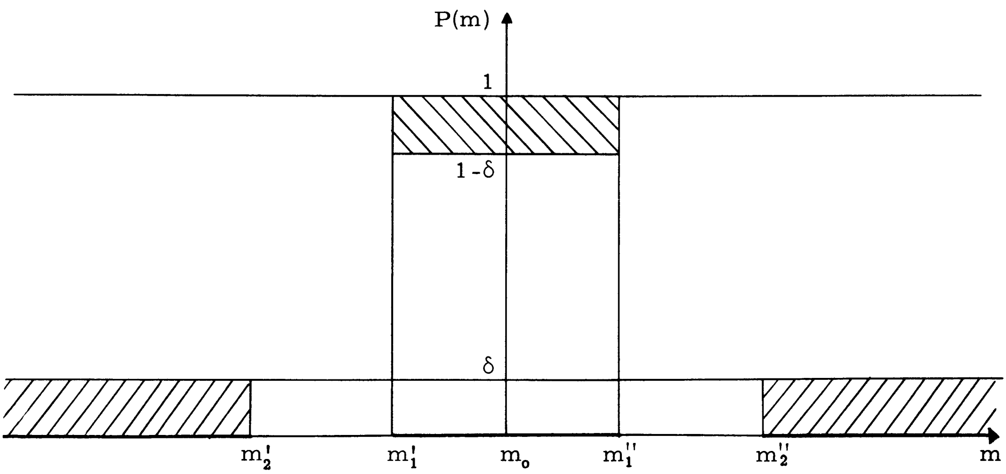

Figure 5

En général, les bornes m'_1 et m''_1 de (O) et m'_2 et m''_2 de (N) ne sont pas deux à deux *symétriques* par rapport à m_0 car un même écart *absolu* à m_0 peut avoir un sens différent - et notamment être assorti d'un coût différent - suivant qu'il correspond à un écart *positif* ou *négatif*. Toutefois, pour simplifier la solution du point de vue algébrique, nous supposerons le problème symétrique : les intervalles (m'_1, m''_1) et (m'_2, m''_2) ont même milieu que nous désignerons par m_0.

2.2.1 - La théorie de Neyman et Pearson.

Le critère de comparaison entre règles de test envisagé en 2.1.1. ne fournit pas de façon pratique la forme des règles optimales.

Comme on cherche à se prémunir contre des valeurs de m substantiellement *différentes* de m_0 *aussi bien à droite qu'à gauche*, on va retenir pour classe de régions critiques la classe obtenue par *réunion* des deux classes optimales respectivement pour les tests "< contre >" et "> contre <", c'est-à-dire la classe des régions critiques de la forme :

$$W : \quad \bar{x} > a \quad \text{ou} \quad \bar{x} < b,$$

ou encore, en changeant de notations, la classe des règles :

- Réponse OUI si $d \leqslant \bar{x} - m_0 \leqslant c$;
- Réponse NON dans le cas contraire.

Le problème que nous étudions étant *symétrique*, il convient de lui donner une solution *symétrique*, c'est-à-dire de se limiter à la classe des règles :

- Réponse OUI si $|\bar{x} - m_0| \leqslant c$;
- Réponse NON dans le cas contraire.

C'est en définitive parmi les règles de cette classe que nous rechercherons la règle finalement adoptée.

2.2.2 - Sélection d'une règle de test.

La fonction $P_c(m)$: probabilité que le test fournisse la réponse OUI, lorsque la vraie valeur m^* de la moyenne est égale à m, est, pour c donné :

$$P_c(m) = \Pi\left[\frac{m_0 + c - m}{\sigma_0/\sqrt{n}}\right] - \Pi\left[\frac{m_0 - c - m}{\sigma_0/\sqrt{n}}\right].$$

Cette fonction de m est symétrique par rapport à m_0. Elle peut en effet s'écrire :

$$P_c(m) = \Pi\left[\frac{c + (m_0 - m)}{\sigma_0/\sqrt{n}}\right] + \Pi\left[\frac{c - (m_0 - m)}{\sigma_0/\sqrt{n}}\right] - 1 .$$

Elle est *décroissante* pour $m > m_0$, *croissante* pour $m < m_0$. En conséquence, les spécifications du problème seront assurées si simultanément :

1/ $\qquad P_c(m''_1) \geqslant 1 - \delta$;

2/ $\qquad P_c(m''_2) \leqslant \delta$;

c'est-à-dire :

1/ $$\Pi\left[\frac{c + m_0 - m''_1}{\sigma_0/\sqrt{n}}\right] + \Pi\left[\frac{c - m_0 + m''_1}{\sigma_0/\sqrt{n}}\right] \geqslant 2 - \delta ;$$

2/ $$\Pi\left[\frac{c + m_0 - m''_2}{\sigma_0/\sqrt{n}}\right] + \Pi\left[\frac{c - m_0 + m''_2}{\sigma_0/\sqrt{n}}\right] \leqslant 1 + \delta .$$

De façon analogue au cas envisagé en 2.1., on peut montrer que ces inégalités ne définissent des valeurs de c (satisfaisant *simultanément* à chacune de ces deux inégalités) que si n est assez grand. Le seuil n_0 qui donne une valeur unique pour c est la racine en n du système d'équations :

$$\Pi\left[\frac{c + m_o - m''_1}{\sigma_o/\sqrt{n}}\right] + \Pi\left[\frac{c - m_o + m''_1}{\sigma_o/\sqrt{n}}\right] = 2 - \delta$$

$$\Pi\left[\frac{c + m_o - m''_2}{\sigma_o/\sqrt{n}}\right] + \Pi\left[\frac{c - m_o + m''_2}{\sigma_o/\sqrt{n}}\right] = 1 + \delta$$

Alors la solution c_o en c de ce système est la valeur correspondante unique associée à n_o.

Ce système d'équations se présente sous une forme compliquée du point de vue algébrique. Pour le résoudre aisément lorsque $\delta = 5\ \%$ et $\delta = 10\ \%$, nous avons construit une table numérique qu'on utilise de la façon suivante : en fonction de r :

$$r = \frac{m''_1 - m_o}{m''_2 - m_o} ,$$

la table 1 indique les valeurs L_2 et z solutions du système :

$$\Pi[L_2(z+r)] + \Pi[L_2(z-r)] = 2 - \delta ,$$

$$\Pi[L_2(z+1)] + \Pi[L_2(z-1)] = 1 + \delta ;$$

dont on déduit :

$$n_o = \left(\frac{\sigma_o L_2}{m''_2 - m_o}\right)^2 ,$$

$$c_o = z(m''_2 - m_o) .$$

La règle du test s'énonce alors :

- Prélever un échantillon de n_o observations ;
- Réponse OUI si $|\bar{x} - m_o| \leqslant z(m''_2 - m_o)$;

  Réponse NON dans le cas contraire.

Remarque.

On observera que, *si r est assez grand,* le système d'équations en (L_2, z) se réduit pratiquement à :

$$\Pi[L_2(z - r)] = \Pi[L_2(1 - z)] = 1 - \delta ;$$

dont les solutions sont :

$$L_2 = 2\ \frac{u_{1-\delta}}{1-r} ;$$

$$z = \frac{1+r}{2} .$$

D'où la règle du test correspondante :

- Prélever un échantillon de $n_o = \left(\frac{2\ \sigma_o\ u_{1-\delta}}{m''_2 - m''_1}\right)^2$ observations ;

Table 1

r	$\delta = 0,05$		$\delta = 0,10$	
	L_2	z	L_2	z
0,00	3,605	0,5437	2,927	0,5621
0,01	3,606	0,5438	2,928	0,5622
0,02	3,610	0,5443	2,930	0,5625
0,03	3,616	0,5451	2,933	0,5629
0,04	3,625	0,5462	2,938	0,5636
0,05	3,637	0,5477	2,945	0,5647
0,06	3,651	0,5494	2,952	0,5657
0,07	3,668	0,5515	2,962	0,5672
0,08	3,687	0,5538	2,973	0,5688
0,09	3,709	0,5565	2,985	0,5705
0,10	3,734	0,5595	2,999	0,5725
0,11	3,762	0,5628	3,015	0,5748
0,12	3,792	0,5662	3,033	0,5773
0,13	3,824	0,5698	3,052	0,5799
0,14	3.859	0,5737	3,073	0,5828
0,15	3,897	0,5779	3,096	0,5859
0,16	3,937	0,5822	3,120	0,5891
0,17	3,979	0,5866	3,147	0,5926
0,18	4,023	0,5911	3,175	0,5962
0,19	4,070	0,5958	3,206	0,6001
0,20	4,118	0,6006	3,238	0,6041
0,25	4,387	0,6250	3,428	0,6260
0,30	4,700	0,6500	3,664	0,6500
0,35	5,062	0,6750	3,944	0,6750
0,40	5,483	0,7000	4,272	0,7000
0,45	5,982	0,7250	4,660	0,7250
0,50	6,580	0,7500	5,126	0,7500
0,60	8,225	0,8000	6,408	0,8000
0,70	10,967	0,8500	8,544	0,8500
0,80	16,450	0,9000	12,816	0,9000
0,90	32,900	0,9500	25,632	0,9500
1,00	∞	1	∞	1

Formules asymptotiques (r grand) :

$$L_2 \sim 2 \frac{u_{1-\delta}}{1-r},$$

$$z \sim \frac{1+r}{2}.$$

- Réponse OUI si $\frac{m'_1 + m'_2}{2} \leqslant \bar{x} \leqslant \frac{m''_1 + m''_2}{2}$;

- Réponse NON dans le cas contraire.

La taille n_0 est ainsi *asymptotiquement identique à celle obtenue* en 2.1.2. En outre, les bornes de l'intervalle où on conclut à la réponse OUI sont asymptotiquement, de façon analogue à ce qu'on a obtenu en 2.1.2., les *milieux des intervalles d'indécision* constituant l'ensemble (I).

Cette approximation est valable dans la mesure où r dépasse 0,20, quand δ est de l'ordre de 5 % à 10 %.

La table 1 permet ainsi de traiter le cas où le statisticien est consulté *avant échantillonnage*. Dans le cas contraire, la valeur c à adopter est la racine de l'équation :

$$\Pi\left[\frac{\sqrt{n}}{\sigma_0}(c+m''_1-m_0)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}(c+m''_2-m_0)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}(c+m_0-m''_1)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}(c+m_0-m''_2)\right] = 3$$

Lorsque n est *assez grand*, la solution de cette équation est approximativement :

$$c = \frac{m''_1 + m''_2}{2} - m_0 ,$$

qui correspond à la règle :

- Réponse OUI si $\frac{m'_1 + m'_2}{2} \leqslant \bar{x} \leqslant \frac{m''_1 + m''_2}{2}$;

- Réponse NON dans le cas contraire.

La racine exacte de l'équation peut être obtenue par tâtonnements autour de $\frac{m''_1 + m''_2}{2} - m_0$. On notera que le membre de gauche de l'équation est une fonction *croissante* de c, ce qui d'une part assure *l'unicité* de la solution et d'autre part guide la recherche de la solution.

De la racine c, on déduit la règle :

- Réponse OUI si $|\bar{x} - m_0| \leqslant c$;
- Réponse NON dans le cas contraire.

Le risque maximum encouru δ^* est alors donné par :

$$\delta^* = \Pi\left[\frac{\sqrt{n}}{\sigma_0}(c+m''_2-m_0)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}(c+m_0-m''_2)\right] - 1 .$$

Ce risque δ^* est inférieur ou égal à δ si n est supérieur ou égal au seuil n_0 déterminé ci-dessus ; il est supérieur à δ dans le cas contraire.

Lorsque n est inférieur à n_0, la règle à trois niveaux qui conduit à un risque au plus égal à δ est :

- Réponse OUI si $|\bar{x} - m_0| \leqslant c_1$;
- Réponse NON si $|\bar{x} - m_0| \geqslant c_2$;
- Réponse ABSTENTION si $c_1 < |\bar{x} - m_0| < c_2$.

Les seuils c_1 et c_2 sont les racines des équations :

$$\Pi\left[\frac{\sqrt{n}}{\sigma_o}(c_1+m_2''-m_o)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_o}(c_1+m_o-m_2'')\right] = 1+\delta ,$$

$$\Pi\left[\frac{\sqrt{n}}{\sigma_o}(c_2+m_1''-m_o)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_o}(c_2+m_o-m_1'')\right] = 2-\delta .$$

Table 2

L_2	C_1	
	$\delta = 0,05$	$\delta = 0,10$
0,0	0,0627	0,1257
0,1	0,0631	0,1263
0,2	0,0641	0,1282
0,3	0,0656	0,1314
0,4	0,0679	0,1360
0,5	0,0711	0,1424
0,6	0,0751	0,1504
0,7	0,0801	0,1605
0,8	0,0863	0,1729
0,9	0,0940	0,1881
1,0	0,1033	0,2066
1,1	0,1147	0,2291
1,2	0,1286	0,2563
1,3	0,1455	0,2900
1,4	0,1662	0,3284
1,5	0,1916	0,3755
1,6	0,2226	0,4307
1,7	0,2603	0,4949
1,8	0,3061	0,5678
1,9	0,3609	0,6487
2,0	0,4252	0,7360
2,5	0,8589	1,2185
3,0	1,3552	1,7184
3,5	1,8551	2,2184
4,0	2,3551	2,7184
L_2	$L_2 - 1,645$	$L_2 - 1,282$

L_1	C_2	
	$\delta = 0,05$	$\delta = 0,10$
0,0	1,960	1,645
0,1	1,970	1,653
0,2	1,999	1,677
0,3	2,045	1,717
0,4	2,107	1,772
0,5	2,181	1,839
0,6	2,265	1,916
0,7	2,356	2,001
0,8	2,450	2,093
0,9	2,548	2,187
1,0	2,646	2,284
1,1	2,745	2,383
1,2	2,845	2,482
1,3	2,945	2,582
1,4	3,045	2,682
1,5	3,145	2,782
1,6	3,245	2,882
1,7	3,345	2,982
1,8	3,445	3,082
1,9	3,545	3,182
2,0	3,645	3,282
2,5	4,145	3,782
3,0	4,645	4,282
3,5	5,145	4,782
4,0	5,645	5,282
L_1	$L_1 + 1,645$	$L_1 + 1,282$

On trouvera dans la table 2, pour $\delta = 5\,\%$ et $\delta = 10\,\%$, en fonction de :

$$L_2 = \sqrt{n}\,\frac{m_2'' - m_o}{\sigma_o} ;$$

et

$$L_1 = \sqrt{n}\,\frac{m_1'' - m_o}{\sigma_o} ;$$

les valeurs de :

$$C_1 = \sqrt{n}\,\frac{c_1}{\sigma_o} \ ;$$

et

$$C_2 = \sqrt{n}\,\frac{c_2}{\sigma_o} \ ;$$

dont on déduit c_1 et c_2 :

$$c_1 = \frac{\sigma_o}{\sqrt{n}}\,C_1 \ ,$$

$$c_2 = \frac{\sigma_o}{\sqrt{n}}\,C_2 \ .$$

Lorsque L_1 et L_2 sont grands, on obtient approximativement :

$$C_1 = L_2 - u_{1-\delta} \ ;$$
$$C_2 = L_1 + u_{1-\delta} \ ;$$

ce qui correspond à la règle :

- Réponse OUI si $|\bar{x}-m_o| \leqslant m_2''-m_o-u_{1-\delta}\,\frac{\sigma_o}{\sqrt{n}}$;

- Réponse NON si $|\bar{x}-m_o| \geqslant m_1''-m_o+u_{1-\delta}\,\frac{\sigma_o}{\sqrt{n}}$;

- Réponse ABSTENTION si

$$m_2'' - m_o - u_{1-\delta}\,\frac{\sigma_o}{\sqrt{n}} < |\bar{x} - m_o| < m_1'' - m_o + u_{1-\delta}\,\frac{\sigma_o}{\sqrt{n}} \ .$$

2.2.3 - Conclusions.

On peut résumer les résultats qui viennent d'être obtenus de la façon suivante, pour résoudre le problème posé par l'homme de l'art sous la forme $(m_1', m_1'', m_2', m_2'', \delta)$, avec $m_1' + m_1'' = m_2' + m_2''$:

a. Si l'avis du statisticien est demandé avant échantillonnage :

- Recommander de prélever un échantillon de n_o observations :

$$n_o = \left(\frac{2\,\sigma_o\,L_2}{m_2'' - m_2'}\right)^2 \ ,$$

où L_2 est donné par la table 1 en fonction de r :

$$r = \frac{m_1'' - m_1'}{m_2'' - m_2'} \ .$$

● Recommander d'appliquer la règle à deux niveaux :

- Réponse OUI si $\left|\bar{x} - \frac{m_1' + m_1''}{2}\right| \leqslant z\left(\frac{m_2'' - m_2'}{2}\right)$;

- Réponse NON dans le cas contraire.

où z est donné par la table 1 en fonction de r.

● Si r dépasse 0,20 et si δ est de l'ordre de 5 % à 10 %, appliquer les formules :

$$n_0 = \left(\frac{2\sigma_0 u_{1-\delta}}{m_2'' - m_1''}\right)^2 \text{ ; Réponse OUI si } \frac{m_1' + m_2'}{2} \leqslant \bar{x} \leqslant \frac{m_1'' + m_2''}{2} .$$

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. la taille n de l'échantillon se trouve être au moins égale au seuil n_2 défini en a.

● Recommander d'appliquer la règle à deux niveaux :

- Réponse OUI si $\left|\bar{x} - \frac{m_1' + m_1''}{2}\right| \leqslant c$;

- Réponse NON dans le cas contraire ;

c étant la racine de l'équation :

$$\Pi\left[\frac{\sqrt{n}}{\sigma_0}\left(c + \frac{m_1'' - m_1'}{2}\right)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}\left(c + \frac{m_2'' - m_2'}{2}\right)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}\left(c - \frac{m_1'' - m_1'}{2}\right)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}\left(c - \frac{m_2'' - m_2'}{2}\right)\right] = 3.$$

● Indiquer à l'homme de l'art le risque maximum δ^* encouru, inférieur ou égal à δ :

$$\delta^* = \Pi\left[\frac{\sqrt{n}}{\sigma_0}\left(c + \frac{m_2'' - m_2'}{2}\right)\right] + \Pi\left[\frac{\sqrt{n}}{\sigma_0}\left(c - \frac{m_2'' - m_2'}{2}\right)\right] - 1 \leqslant \delta.$$

b_2. la taille n de l'échantillon se trouve être inférieure au seuil n_0 défini en a. :

Dans ce cas, trois attitudes sont possibles :

Première attitude : Recommander de prélever $\Delta n = n_0 - n$ *observations complémentaires* et se ramener au cas *a.* .

Seconde attitude : Recommander d'appliquer la règle à *trois niveaux :*

- Réponse OUI si $\left|\bar{x} - \frac{m_1' + m_1''}{2}\right| \leqslant \frac{\sigma_0}{\sqrt{n}} C_1$;

- Réponse NON si $\left|\bar{x} - \frac{m_1' + m_1''}{2}\right| \geqslant \frac{\sigma_0}{\sqrt{n}} C_2$;

- Réponse ABSTENTION si

$$\frac{\sigma_0}{\sqrt{n}} C_1 < \left|\bar{x} - \frac{m_1' + m_1''}{2}\right| < \frac{\sigma_0}{\sqrt{n}} C_2 ;$$

C_1 et C_2 étant fournis par la table 2 en fonction de L_1 et L_2 :

$$L_1 = \sqrt{n}\ \frac{m_1'' - m_1'}{2\,\sigma_0} \quad ; \quad L_2 = \sqrt{n}\ \frac{m_2'' - m_2'}{2\sigma_0} \ .$$

Troisième attitude : Recommander d'appliquer la règle à *deux niveaux* décrite en b_1. en précisant à l'homme de l'art le risque maximum encouru δ^*, supérieur à δ.

2.2.4 - Application numérique.

Soit le problème numérique suivant :

$$m_1' = 5{,}90 \ , \quad m_1'' = 6{,}10 \ ,$$
$$m_2' = 5{,}50 \ , \quad m_2'' = 6{,}50 \ ,$$
$$\sigma_0 = 2{,}00 \ ,$$
$$\delta = 5\ \% \ .$$

Il vient :

$$r = 0{,}20 \ .$$

La table 1 indique alors :

$$L_2 = 4{,}118 \ ,$$

dont on déduit :

$$n_0 = \left(\frac{2 \times 2{,}00 \times 4{,}118}{1{,}00}\right)^2 = 272.$$

D'où :

a. Si l'avis du statisticien est demandé avant l'échantillonnage.

● Recommander de prélever 272 observations et d'appliquer la règle :

- Réponse OUI si $5{,}700 \leqslant \overline{x} \leqslant 6{,}300$;
- Réponse NON dans le cas contraire.

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. Si, par exemple, l'échantillon comporte 900 observations :

● Recommander d'appliquer la *règle à deux niveaux* précédente ;

● Indiquer à l'homme de l'art que le risque maximum encouru est seulement égal à :

$$\delta^* = \Pi\left[\frac{\sqrt{900}}{2{,}00} \times 0{,}80\right] - \Pi\left[\frac{\sqrt{900}}{2{,}00} \times 0{,}20\right] = 0{,}14\ \% \ .$$

b_2. Si, par exemple, l'échantillon comporte 100 observations. Les trois attitudes possibles sont alors :

● ou bien recommander de prélever 172 *observations complémentaires* et se ramener au cas *a*.

● ou bien adopter une règle à *trois niveaux*.

On a $L_1 = 0,5$ et $L_2 = 2,5$. D'où, en utilisant la table 2 :

$\delta = 5\ \%$: $C_1 = 0,8589$, $C_2 = 2,181$;

soit $c_1 = 0,1718$; $c_2 = 0,4362$.

$\delta = 10\ \%$: $C_1 = 1,2185$, $C_2 = 1,839$;

soit $c_1 = 0,2437$; $c_2 = 0,3678$.

Les règles à trois niveaux sont ainsi :

$\delta = 5\ \%$:

- Réponse OUI si $5,828 \leqslant \bar{x} \leqslant 6,172$;
- Réponse NON si $\bar{x} \leqslant 5,564$ ou $\bar{x} \geqslant 6,436$;
- Réponse ABSTENTION si

$$5,564 < \bar{x} < 5,828 \quad \text{ou} \quad 6,172 < \bar{x} < 6,436.$$

$\delta = 10\ \%$:

- Réponse OUI si $5,756 \leqslant \bar{x} \leqslant 6,244$;
- Réponse NON si $\bar{x} \leqslant 5,632$ ou $\bar{x} \geqslant 6,368$;
- Réponse ABSTENTION si

$$5,632 < \bar{x} < 5,756 \quad \text{ou} \quad 6,244 < \bar{x} < 6,368.$$

● ou bien adopter une règle à *deux niveaux :*

La racine de l'équation donnant c est 0,3085.

D'où la règle :

- Réponse OUI si $5,6915 \leqslant \bar{x} \leqslant 6,3085$;
- Réponse NON dans le cas contraire.

Le risque maximum δ^* encouru est alors égal à :

$$\delta^* = 10\ \% :$$

Les différents résultats obtenus lorsque n = 100 sont représentés par la figure 6.

2.3 - TESTS DE COMPARAISON ENTRE MOYENNES, LES ECARTS-TYPES ETANT CONNUS.

Appliquons la méthode exposée en 2.1 et 2.2 au problème du test de *comparaison de deux moyennes*. Soit deux échantillons *indépendants* d'observations *normales indépendantes*. Les tailles des échantillons sont désignées par n_1 et n_2, les écarts-types supposés *connus* par σ_1 et σ_2 et les moyennes *inconnues* par m_1 et m_2.

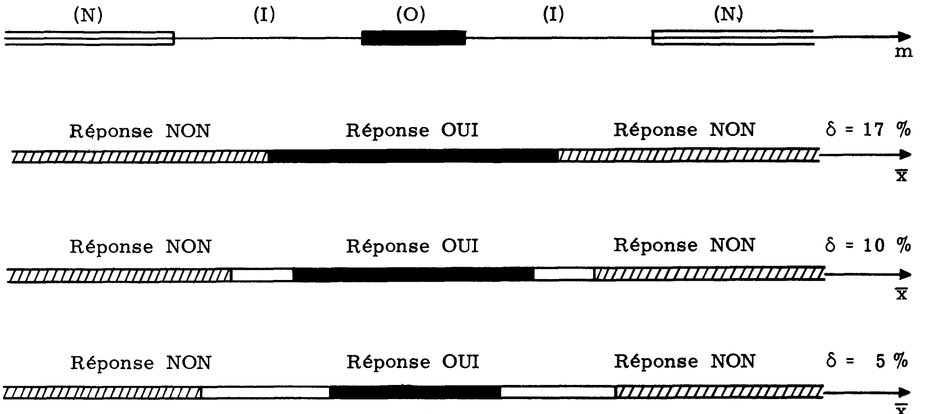

Figure 6

2.3.1 - Test de l'hypothèse : la moyenne m_1 est substantiellement plus faible que la moyenne m_2

Examinons d'abord le test *unilatéral* : la moyenne m_1 est *substantiellement plus faible* que la moyenne m_2 : l'ensemble (O) est l'ensemble des couples (m_1, m_2) tels que la différence $m_1 - m_2$ est inférieure ou égale à une quantité d_1, l'ensemble (N) est l'ensemble des couples tels que la différence $m_1 - m_2$ est supérieure ou égale à une quantité d_2 (Fig. 7).

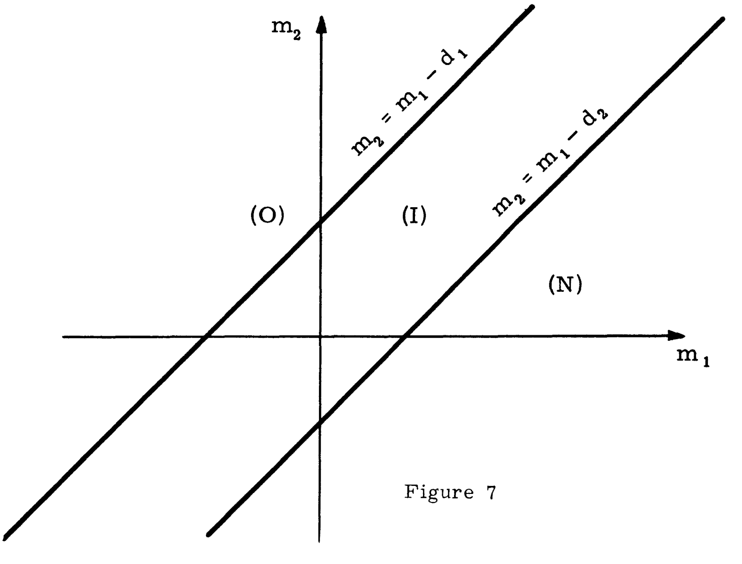

Figure 7

On peut montrer que, dans ces conditions, le critère de comparaison entre deux règles : le test T est *préférable ou équivalent* au test T' si à la fois :

$$\forall (m_1, m_2) \in (O) \quad : \quad P_W(m_1, m_2) \geqslant P_{W'}(m_1, m_2) \quad ;$$
$$\forall (m_1, m_2) \in (N) \quad : \quad P_W(m_1, m_2) \leqslant P_{W'}(m_1, m_2) \quad ;$$

conduit à la classe des régions critiques optimales :

$$W \quad : \quad \bar{x}_1 - \bar{x}_2 > c \ ,$$

c'est-à-dire à la classe des règles optimales :

- Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant c$;
- Réponse NON dans le cas contraire ;

$\bar{x}_1$ et $\bar{x}_2$ désignant les moyennes empiriques c'est-à-dire les moyennes des échantillons n° 1 et n° 2 respectivement.

Dans ces conditions, la fonction $P_c(m_1, m_2)$ ne dépend que de la différence $d = m_1 - m_2$:

$$P_c(m_1, m_2) = \Pr\left\{\text{Répondre OUI}\Big/ m_1, m_2\right\} = \Pi\left[\frac{c-(m_1-m_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right] ,$$

puisque, sous les hypothèses envisagées, la loi de la différence $\bar{x}_1 - \bar{x}_2$ est normale :

$$\bar{x}_1 - \bar{x}_2 = \mathcal{N}\left(m_1 - m_2 \ , \ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

Le problème à résoudre est tout à fait analogue à celui traité en 2.1, moyennant le remplacement de :

$$\bar{x} \quad \text{par} \quad \bar{x}_1 - \bar{x}_2 \ ;$$

$$\frac{\sigma_0}{\sqrt{n}} \quad \text{par} \quad \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \ .$$

La quantité d'information disponible est suffisante pour répondre au problème défini par les spécifications (d_1, d_2, δ) si n_1 et n_2 satisfont à :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \leqslant \left(\frac{d_2-d_1}{2\,u_{1-\delta}}\right)^2 .$$

Cette condition ne suffit pas à déterminer n_1 et n_2 lorsqu'on cherche le plan de sondage le moins coûteux (n_1 et n_2 minimum revient à saturer la contrainte mais ne détermine pas n_1 et n_2). On peut alors chercher à optimiser le coût total d'échantillonnage. Si p_1 et p_2 sont proportionnels aux coûts unitaires d'échantillonnage supposés *constants* pour l'échantillon n° 1 et l'échantillon n° 2 respectivement, on est conduit à rendre minimum :

$$n_1 p_1 + n_2 p_2 \ ,$$

sous la contrainte (qu'il faut alors saturer) :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \leqslant \left(\frac{d_2 - d_1}{2\,u_{1-\delta}}\right)^2 \quad .$$

On obtient alors (méthode de Lagrange par exemple) :

$$n_1 = \frac{\sigma_1}{\sqrt{p_1}} \; \frac{\sqrt{p_1}\,\sigma_1 + \sqrt{p_2}\,\sigma_2}{(d_2 - d_1)^2} \; (2\,u_{1-\delta})^2 \; ;$$

$$n_2 = \frac{\sigma_2}{\sqrt{p_2}} \; \frac{\sqrt{p_1}\,\sigma_1 + \sqrt{p_2}\,\sigma_2}{(d_2 - d_1)^2} \; (2\,u_{1-\delta})^2 \; ;$$

dont on déduit :

$$n_1 p_1 + n_2 p_2 = \left(\frac{2\,u_{1-\delta}(\sqrt{p_1}\,\sigma_1 + \sqrt{p_2}\,\sigma_2)}{d_2 - d_1}\right)^2 .$$

En particulier, si on cherche à rendre minimum la taille totale $n_1 + n_2$, on aboutit (en faisant $p_1 = p_2$) à :

$$n_1 = \sigma_1(\sigma_1 + \sigma_2)\left(\frac{2\,u_{1-\delta}}{d_2 - d_1}\right)^2 \; ;$$

$$n_2 = \sigma_2(\sigma_1 + \sigma_2)\left(\frac{2\,u_{1-\delta}}{d_2 - d_1}\right)^2 \; ;$$

$$n_1 + n_2 = \left(\frac{2\,u_{1-\delta}(\sigma_1 + \sigma_2)}{d_2 - d_1}\right)^2 .$$

D'où la conduite à tenir pour résoudre le problème envisagé :

a. Si l'avis du statisticien est demandé avant échantillonnage :

- Recommander de prélever des échantillons de tailles n_1 et n_2 données par les formules précédentes.
- Recommander d'appliquer la règle :
  - Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant \frac{d_1 + d_2}{2}$;
  - Réponse NON dans le cas contraire.

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. les tailles n_1 et n_2 satisfont à :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \leqslant \left(\frac{d_2 - d_1}{2\,u_{1-\delta}}\right)^2$$

- Recommander d'appliquer la même règle qu'en *a* ;
- Indiquer à l'homme de l'art le risque maximum δ^* encouru, au plus égal à δ :

$$\delta^* = 1 - \Pi\left[\frac{d_2 - d_1}{2\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}\right] \leqslant \delta .$$

b_2. les tailles n_1 et n_2 satisfont à :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} > \left(\frac{d_2 - d_1}{2\,u_{1-\delta}}\right)^2 .$$

Trois attitudes sont possibles :

● Recommander de prélever Δn_1 et Δn_2 *observations complémentaires* de façon à se ramener au cas *a* :

$$\frac{\sigma_1^2}{n_1 + \Delta n_1} + \frac{\sigma_2^2}{n_2 + \Delta n_2} = \left(\frac{d_2 - d_1}{2\, u_{1-\delta}}\right)^2$$

On pourra faire en sorte que le coût additionnel d'échantillonnage:

$$p_1\, \Delta n_1 + p_2\, \Delta n_2$$

soit rendu minimum.

● Appliquer la règle à *trois niveaux :*

- Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant d_2 - u_{1-\delta}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$;

- Réponse NON si $\bar{x}_1 - \bar{x}_2 \geqslant d_1 + u_{1-\delta}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$;

- Réponse ABSTENTION dans le cas contraire.

● Appliquer la règle à *deux niveaux* définie en *a.* et indiquer à l'homme de l'art le risque maximum encouru δ^*, supérieur à δ :

$$\delta^* = 1 - \Pi\left[\frac{d_2 - d_1}{2\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}\right] > \delta\ .$$

Application numérique.

Soit les données numériques suivantes :

$$d_1 = 0{,}20\ ;\quad d_2 = 0{,}50\ ;$$
$$\sigma_1 = 0{,}80\ ;\quad \sigma_2 = 0{,}40\ ;$$
$$\frac{p_1}{p_2} = 0{,}5\ ;\quad \delta = 5\ \%\ .$$

a. Si l'avis du statisticien est demandé avant échantillonnage :

● Recommander de prélever :

n_1 = 131 observations ;

n_2 = 46 observations.

● Recommander d'appliquer la règle :

- Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant 0{,}35$;
- Réponse NON dans le cas contraire.

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. Si, par exemple, $n_1 = 100$, $n_2 = 200$.

On a alors :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = 0,0072 < \left(\frac{d_2 - d_1}{2\, u_{0,95}}\right)^2 = 0,0083 .$$

● Recommander d'appliquer la règle précédente.

● Indiquer à l'homme de l'art que le risque maximum δ^* encouru est égal à :

$$\delta^* = 3,9\ \% .$$

*b*₂. Si, par exemple, $n_1 = 200$, $n_2 = 10$.

On a alors :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = 0,0192 > 0,0083.$$

Première attitude. Recommander de prélever Δn_2 observations complémentaires ($\Delta n_1 = 0$ car $n_1 > 131$) dans la population n° 2 :

$$\Delta n_2 = 21\ ;$$

et appliquer la règle définie en *a*.

Seconde attitude. Adopter la règle à *trois niveaux* :

- Réponse OUI si $\overline{x}_1 - \overline{x}_2 \leqslant 0,3355$;
- Réponse NON si $\overline{x}_1 - \overline{x}_2 \geqslant 0,3645$;
- Réponse ABSTENTION si $0,3355 < \overline{x}_1 - \overline{x}_2 < 0,3645$.

Troisième attitude. Adopter la règle à *deux niveaux* définie en *a*. et indiquer à l'homme de l'art que le risque maximum δ^* encouru est égal à :

$$\delta^* = 14\ \% .$$

2.3.2 - Test de l'hypothèse : la moyenne m_1 ne diffère pas substantiellement de la moyenne m_2 .

Ce test *bilatéral* est analogue à celui examiné en 2.2. Admettons les spécifications suivantes (Figure 8) :

- Répondre OUI avec une probabilité au plus égale à δ, lorsque $|m_1 - m_2| \geqslant d_2$;

- Répondre NON avec une probabilité au plus égale à δ, lorsque $|m_1 - m_2| \leqslant d_1$.

Nous n'en indiquerons que le résultat :

a. Si le statisticien est consulté avant échantillonnage :

● Recommander de prélever n_1 et n_2 observations :

$$n_1 = \frac{\sigma_1}{\sqrt{p_1}} \frac{\sqrt{p_1}\, \sigma_1 + \sqrt{p_2}\, \sigma_2}{d_2^2} L_2^2$$

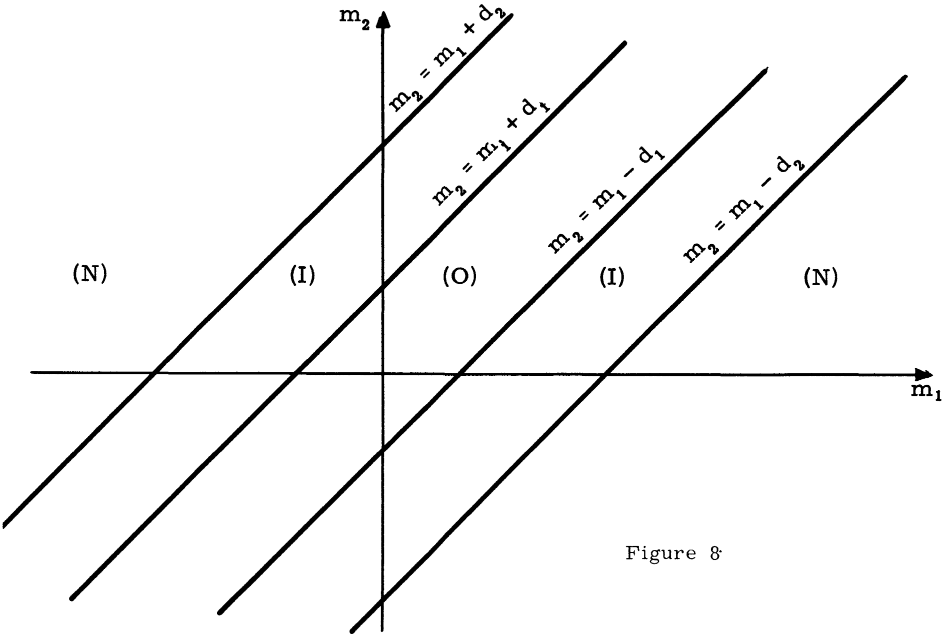

Figure 8

$$n_2 = \frac{\sigma_2}{\sqrt{p_2}} \; \frac{\sqrt{p_1}\,\sigma_1 + \sqrt{p_2}\,\sigma_2}{d_2^2} \, L_2^2$$

où L_2 est fourni par la table 1 en fonction de r :

$$r = \frac{d_2}{d_1} .$$

- Recommander d'appliquer la règle :
  - Réponse OUI si $|\overline{x}_1 - \overline{x}_2| \leqslant z\, d_2$;
  - Réponse NON dans le cas contraire ;

z étant donné par la table 1 en fonction de $r = \dfrac{d_2}{d_1}$.

b. Si le statisticien est consulté après échantillonnage :

b$_1$. Si les valeurs de n_1 et n_2 satisfont à :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \leqslant \left(\frac{d_2}{L_2}\right)^2 \quad :$$

- Appliquer la règle :
  - Réponse OUI si $|\overline{x}_1 - \overline{x}_2| \leqslant c$;
  - Réponse NON dans le cas contraire ;

c étant la racine de l'équation :

$$\Pi\left[\frac{c+d_1}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right]+\Pi\left[\frac{c+d_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right]+\Pi\left[\frac{c-d_1}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right]+\Pi\left[\frac{c-d_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right]=3.$$

Lorsque n_1 et n_2 sont grands, cette racine est voisine de

$$c = \frac{d_1 + d_2}{2} .$$

La solution précise peut être obtenue autour de cette valeur (cf. ce qui a été indiqué plus haut page 32.

● Indiquer à l'homme de l'art le risque maximum δ^* encouru, égal à :

$$\delta^* = \Pi\left[\frac{c+d_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right]+\Pi\left[\frac{c-d_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\right]-1\leqslant\delta .$$

b$_2$. Si les valeurs de n_1 et n_2 satisfont à :

$$\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}>\left(\frac{d_2}{L_2}\right)^2 :$$

Trois attitudes sont possibles :

Première attitude : Prélever des observations *complémentaires* de telle sorte que :

$$\frac{\sigma_1^2}{n_1+\Delta n_1}+\frac{\sigma_2^2}{n_2+\Delta n_2}=\left(\frac{d_2}{L_2}\right)^2 ;$$

et appliquer la règle définie en *a.* .

Seconde attitude : Adopter la règle à *trois niveaux* :

- Réponse OUI si $|\bar{x}_1 - \bar{x}_2| \leqslant C_1\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$;

- Réponse NON si $|\bar{x}_1 - \bar{x}_2| \geqslant C_2\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$;

- Réponse ABSTENTION si $C_1 < \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < C_2$;

C_1 et C_2 étant donnés par la table 2 en fonction de :

$$L_1 = \frac{d_1}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} ;\quad L_2 = \frac{d_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} .$$

Troisième attitude : Adopter la règle à *deux niveaux* définie en *b_1.* en indiquant à l'homme de l'art le risque maximum encouru δ^*, supérieur à δ.

Application numérique.

$$d_1 = 0,10 \;;\quad d_2 = 0,50 \;;$$
$$\sigma_1 = 1,00 \;;\quad \sigma_2 = 2,00 \;;$$
$$\frac{p_1}{p_2} = 2 \;;\quad \delta = 5\,\% .$$

La table 1 fournit en fonction de $r = 0,2$:

$$L_2 = 4,118 \;,\quad z = 0,6006.$$

a. Si l'avis du statisticien est demandé avant échantillonnage :

- Recommander de prélever :

$$n_1 = 164 \quad \text{et} \quad n_2 = 463 \text{ observations.}$$

- Recommander d'appliquer la règle :
  - Réponse OUI si $|\bar{x}_1 - \bar{x}_2| \leqslant 0,3003$;
  - Réponse NON dans le cas contraire.

b. Si l'avis du statisticien est demandé après échantillonnage :

b_1. Si, par exemple, $n_1 = 150$ et $n_2 = 1000$:

On a alors :

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = 0,0107 < \left(\frac{d_2}{L_2}\right)^2 = 0,0148$$

L'équation qui détermine c a pour racine :

$$c = 0,300.$$

D'où la règle à recommander :

- Réponse OUI si $|\bar{x}_1 - \bar{x}_2| \leqslant 0,300$;
- Réponse NON dans le cas contraire.

Le risque maximum δ^* encouru est égal à :

$$\delta^* = 2,5\,\%.$$

b_2. Si, par exemple, $n_1 = 100$, $n_2 = 200$:

Les trois attitudes possibles sont alors :

Première attitude : Prélever $\Delta n_1 = 64$ et $\Delta n_2 = 263$ *observations complémentaires* et se ramener au cas *a.*

Seconde attitude : Adopter la règle à *trois niveaux :*

- Réponse OUI si $|\bar{x}_1 - \bar{x}_2| \leqslant 0,215$;

- Réponse NON si $|\bar{x}_1 - \bar{x}_2| \geqslant 0,389$;

- Réponse ABSTENTION si $0,215 < |\bar{x}_1 - \bar{x}_2| < 0,389$.

Troisième attitude : Adopter la règle à *deux niveaux :*

- Réponse OUI si $|\bar{x}_1 - \bar{x}_2| \leqslant 0,305$;

- Réponse NON dans le cas contraire ;

Indiquer à l'homme de l'art que le risque maximum δ^* encouru est égal à :

$$\delta^* = 13 \%.$$

Remarque.

Les résultats obtenus en 2.1., 2.2., et 2.3. qui supposent à la fois la *normalité* des populations échantillonnées et la *connaissance* des écarts-types valent encore *de façon approchée* lorsque ces hypothèses sont abandonnées, du moins si les tailles d'échantillon sont *assez grandes.* C'est ainsi que si on abandonne l'hypothèse de normalité en maintenant celle suivant laquelle on connaît les écarts-types, les résultats valent encore en vertu du *théorème central limite* (pourvu, en toute rigueur que les écarts-types soient finis). On pourra les appliquer pratiquement dès que n dépasse la dizaine pour les tests de signification (distributions supposées sensiblement symétriques), dès que n_1 et n_2 dépassent 30 à 50 dans le cas des tests de comparaison. Lorsque l'hypothèse suivant laquelle on connait les écarts-types est à son tour abandonnée, on pourra remplacer les écarts-types inconnus par leurs estimations déduites de l'échantillon, si n dépasse la cinquantaine dans le cas des tests de signification, si n_1 et n_2 dépassent la centaine dans le cas des tests de comparaison. Si l'hypothèse de normalité est maintenue mais si les écarts-types sont inconnus, on appliquera les formules présentées ci-dessous en 2.4, lorsque les tailles d'échantillon sont faibles.

2.4 - TESTS RELATIFS A DES MOYENNES DE VARIABLES NORMALES LORSQUE LES ECARTS-TYPES SONT INCONNUS.

Lorsque les écarts-types sont inconnus, les lois de probabilité mises en jeu ne sont plus des variables normales (centrées ou non) mais des variables de Student (centrées ou non). Dans le cas des tests *d'homogénéité,* on a affaire à des variables du χ^2 (centrées ou non) quand les écarts-types sont connus, à des variables de Snedecor (centrées ou non) quand les écarts-types sont inconnus. Avant d'appliquer la méthode décrite plus haut au cas de ces tests, il a semblé utile de rappeler quelques résultats concernant les variables aléatoires non centrées.

2.4.1 - Les lois du χ^2, de Student et de Snedecor non centrées.

Loi du χ^2 non centrée.

Soit ν variables *normales centrées réduites indépendantes* $U_1, \ldots, U_i, \ldots, U_\nu$ et ν *quantités certaines* $v_1, \ldots, v_i, \ldots, v_\nu$. On sait que la loi de la quantité :

$$\sum_{i=1}^{\nu} U_i^2 ,$$

est la *loi du* χ^2 *centrée* à ν degrés de liberté :

$$\sum_{i=1}^{\nu} U_i^2 = \chi^2(\nu).$$

Ses premiers moments sont :

$$E[\chi^2(\nu)] = \nu \qquad V[\chi^2(\nu)] = 2\nu.$$

La loi de la variable :

$$\sum_{i=1}^{\nu} (U_i + v_i)^2$$

est appelée *loi du* χ^2 *non centrée*. Elle ne dépend que de ν et de la somme λ :

$$\lambda = \sum_{i=1}^{\nu} v_i^2.$$

ν s'appelle le *nombre de degrés de liberté* et λ le *paramètre d'excentricité* ou encore *de non-centralité* :

$$\sum_{i=1}^{\nu} (U_i + v_i)^2 = \chi^2\left(\nu , \sum_{i=1}^{\nu} v_i^2 \right).$$

La variable χ^2 centrée correspond ainsi à $\lambda = 0$:

$$\chi^2(\nu , 0) \equiv \chi^2(\nu) .$$

Les premiers moments de la variable χ^2 non centrée sont respectivement :

$$E[\chi^2(\nu , \lambda)] = \nu + \lambda \qquad V[\chi^2(\nu , \lambda)] = 2(\nu + 2\lambda)$$

On montre que la fonction cumulative de la variable χ^2 non centrée peut s'exprimer en fonction des fonctions cumulatives de la variable χ^2 centrée :

$$\Pr\left\{\chi^2(\nu, \lambda) < x\right\} = \sum_{z=0}^{\infty} \frac{e^{-\frac{\lambda}{2}}(\lambda/2)^z}{z!} \Pr\left\{\chi^2(\nu + 2z) < x\right\}.$$

Par ailleurs, lorsque le nombre de degrés de liberté est *pair*, on a quels que soient x et λ :

$$\Pr\left\{\chi^2(\nu, \lambda) < x\right\} = \Pr \left\{\mathcal{P}\left(\frac{x}{2}\right) - \mathcal{P}\left(\frac{\lambda}{2}\right) \geqslant \frac{\nu}{2}\right\} ,$$

$\mathcal{P}\left(\frac{x}{2}\right)$ et $\mathcal{P}\left(\frac{\lambda}{2}\right)$ désignant deux variables de POISSON indépendantes de paramètres respectivement $\frac{x}{2}$ et $\frac{\lambda}{2}$.

Enfin lorsque $\nu \longrightarrow \infty$ de telle sorte que $\frac{\lambda}{\nu} \longrightarrow a$, on a, en vertu du théorème central limite :

$$\frac{\chi^2(\nu, \lambda) - \nu(1 + a)}{\sqrt{2\nu(1 + 2a)}} \sim \mathcal{N}(0, 1) .$$

Lois de Student et de Snedecor non centrées.

On appelle lois de Student et de Snedecor non centrées les lois des variables :

$$T(\nu, |\lambda|) = \frac{U + \lambda}{\sqrt{\frac{\chi^2(\nu)}{\nu}}} \quad , \quad F(\nu_1, \nu_2, \lambda) = \frac{\chi^2(\nu_1, \lambda)/\nu_1}{\chi^2(\nu_2)/\nu_2} \quad ,$$

les numérateurs et dénominateurs de ces quotients étant supposés *indépendants*. On a évidemment :

$$F(1, \nu, \lambda^2) = [T(\nu, |\lambda|)]^2 \ .$$

On trouvera dans Scheffé : *The analysis of variance* (pages 41 et 42) une bibliographie des tables et abaques existants de ces lois non centrées. Les abaques de E.S. Pearson et H.O. Hartley publiés en 1951 (*Biometrika*, volume 38, pages 115-122) ainsi que ceux de M. Fox publiés en 1956 (*Annals of Mathematical Statistics*, volume 27, pages 485-494) sont reproduits en annexe de l'ouvrage de Scheffé.

2.4.2 - Test de signification d'une moyenne (test "< contre >").

Envisageons le problème analogue à celui traité en 2.1. : sur la base de n observations *indépendantes* d'une *loi normale* $\mathcal{N}(m, \sigma)$, où m et σ sont *inconnus*, test de l'hypothèse : *la moyenne m est relativement faible* contre l'hypothèse : *m est relativement forte.*

On a vu en 2.1. que lorsque σ était connu, la règle qui donnait au risque maximum encouru δ^* la valeur δ était :

- si $n \geqslant \left(\frac{2\,\sigma_0\, u_{1-\delta}}{m_2 - m_1}\right)^2$:
  - Réponse OUI si $\bar{x} \leqslant \frac{m_1 + m_2}{2}$;
  - Réponse NON dans le cas contraire.
- si $n < \left(\frac{2\,\sigma_0\, u_{1-\delta}}{m_2 - m_1}\right)^2$:
  - Réponse OUI si $\bar{x} \leqslant m_2 - \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$;
  - Réponse NON si $\bar{x} \geqslant m_1 + \frac{\sigma_0}{\sqrt{n}} u_{1-\delta}$
  - Réponse ABSTENTION dans les autres cas.

Lorsque σ est inconnu, il n'y a pas lieu de revenir sur la règle à deux niveaux $\left(\text{OUI si } \bar{x} \leqslant \frac{m_1 + m_2}{2}\right)$, du moins si la quantité d'information disponible est suffisante. Si la taille de l'échantillon est insuffisante, on est conduit par analogie avec le cas où l'écart-type est connu, à la règle à trois niveaux (en désignant par s l'écart-type estimé et par $t_{1-\delta}(n-1)$ le quantile d'ordre $1-\delta$ de la variable de Student centrée à $n-1$ degrés de liberté) :

- Réponse OUI si $\overline{x} \leqslant m_2 - \frac{s}{\sqrt{n}} t_{1-\delta}(n-1)$;

- Réponse NON si $\overline{x} \geqslant m_1 + \frac{s}{\sqrt{n}} t_{1-\delta}(n-1)$;

- Réponse ABSTENTION dans les autres cas ;

s étant égal à :

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} .$$

Cette règle, en effet, conduit à un risque d'erreur au plus égal à δ, *quelle que soit la valeur inconnue de σ et quelle que soit la vraie valeur m^* de m située dans (O) ou dans (N) :*

• $\forall m \in (N)$, c'est-à-dire $\forall m \geqslant m_2$:

$$\begin{aligned} \Pr\left\{ \text{Répondre OUI}\big/_{m,\sigma} \right\} &= \Pr\left\{ \overline{x} \leqslant m_2 - \frac{s}{\sqrt{n}} t_{1-\delta}(n-1) \right\} \\ &= \Pr\left\{ \frac{\overline{x} - m}{s/\sqrt{n}} \leqslant - \frac{m - m_2}{s/\sqrt{n}} - t_{1-\delta}(n-1) \right\} \\ &\leqslant \Pr\left\{ \frac{\overline{x} - m}{s/\sqrt{n}} \leqslant - t_{1-\delta}(n-1) \right\} \\ &= \delta \end{aligned}$$

• $\forall m \in (O)$, c'est-à-dire $\forall m \leqslant m_1$:

$$\begin{aligned} \Pr\left\{ \text{Répondre NON}\big/_{m,\sigma} \right\} &= \Pr\left\{ \overline{x} \geqslant m_1 + \frac{s}{\sqrt{n}} t_{1-\delta}(n-1) \right\} \\ &= \Pr\left\{ \frac{\overline{x} - m}{s/\sqrt{n}} \geqslant \frac{m_1 - m}{s/\sqrt{n}} + t_{1-\delta}(n-1) \right\} \\ &\leqslant \Pr\left\{ \frac{\overline{x} - m}{s/\sqrt{n}} \geqslant t_{1-\delta}(n-1) \right\} \\ &= \delta \end{aligned}$$

Il en résulte que le critère permettant de s'assurer que la taille n de l'échantillon est suffisante est :

$$m_2 - \frac{s}{\sqrt{n}} t_{1-\delta}(n-1) \geqslant m_1 + \frac{s}{\sqrt{n}} t_{1-\delta}(n-1) ,$$

c'est-à-dire :

$$n \geqslant \left(\frac{2 s\ t_{1-\delta}(n-1)}{m_2 - m_1} \right)^2 .$$

Si cette inégalité est satisfaite, le risque maximum δ^* encouru est :

$$\delta^* = \Pr\left\{T(n-1) > \sqrt{n}\left(\frac{m_2 - m_1}{2\,s}\right)\right\} \quad .$$

Dans ces conditions, la conduite à tenir nous paraît devoir être la suivante, pour tester l'hypothèse suivant laquelle la moyenne m est relativement faible dans le cas où le problème correspond aux spécifications (m_1, m_2, δ) :

a. Si le statisticien est consulté avant échantillonnage :

a_1. Si on connaît une estimation préalable digne de foi de σ, désignée par σ_0 :

● Recommander de prélever un échantillon de taille au moins égale à :

$$n = \left(\frac{2\,\sigma_0\, u_{1-\delta}}{m_2 - m_1}\right)^2 .$$

Il pourra être prudent de *dépasser substantiellement* ce seuil et d'atteindre une valeur de l'ordre de la racine de l'équation :

$$n = \left(\frac{2\, u_{1-\delta}}{m_2 - m_1}\right)^2 \frac{\chi^2_{1-\alpha}(n-1)}{n-1}\,\sigma_0^2$$

avec $\alpha = \frac{1}{4}$ (si σ_0 est une estimation digne de foi et si le prélèvement d'observations complémentaires n'est pas très coûteux) ou $\alpha = \frac{1}{10}$ voire $\frac{1}{20}$.

● Procéder ensuite à l'analyse décrite ci-dessous en *b*.

a_2. Si on ignore tout de σ :

● Prélever un échantillon préalable (de 20 à 30 observations par exemple) de façon à estimer σ et revenir au cas a_1 ci-dessus pour déterminer la taille de l'échantillon à retenir finalement.

b. Si le statisticien est consulté après échantillonnage :

Comparer la taille n de l'échantillon à :

$$\left(\frac{2\,s\; t_{1-\delta}(n-1)}{m_2 - m_1}\right)^2 = n_0 .$$

b_1. Si n est *supérieur ou égal* à n_0 :

● Recommander d'appliquer la règle à deux niveaux :

- Réponse OUI si $\overline{x} \leqslant \frac{m_1 + m_2}{2}$;

- Réponse NON dans le cas contraire.

● Indiquer à l'homme de l'art le risque maximum δ^* encouru :

$$\delta^* = \Pr\left\{T(n-1) > \sqrt{n}\left(\frac{m_2 - m_1}{2\,s}\right)\right\} \leqslant \delta \; .$$

b_2. Si n est *inférieur* à n_0, trois attitudes sont possibles :

Première attitude : Recommander de prélever un échantillon complémentaire dont la taille est déterminée à partir de l'estimation s de l'écart-type σ inconnu (comme il est indiqué ci-dessus en *a.*). Revenir ensuite en *b.*

Seconde attitude : Adopter la règle à trois niveaux :

- Réponse OUI si $\bar{x} \leqslant m_2 - \frac{s}{\sqrt{n}} t_{1-\delta}(n-1)$;

- Réponse NON si $\bar{x} \geqslant m_1 + \frac{s}{\sqrt{n}} t_{1-\delta}(n-1)$;

- Réponse ABSTENTION dans les autres cas.

Troisième attitude : Appliquer la règle définie en b_1. ci-dessus et indiquer à l'homme de l'art le risque maximum δ^* encouru, supérieur à δ :

$$\delta^* = \Pr \left\{ T(n-1) > \sqrt{n} \left(\frac{m_2 - m_1}{2\,s}\right) \right\} > \delta .$$

2.4.3 - Test de comparaison entre moyennes : m_1 est substantiellement plus faible que m_2, les écarts-types σ_1 et σ_2 inconnus étant supposés égaux.

Le problème du test de comparaison entre moyennes : m_1 est substantiellement plus faible que m_2, lorsque les écarts-types σ_1 et σ_2 sont *inconnus* mais supposés *égaux* est analogue à celui qui vient d'être étudié. Indiquons-en rapidement la règle.

a. Si l'avis du statisticien est demandé avant échantillonnage :

a_1. Si on *connait une estimation* de l'écart-type commun σ, disons σ_0 :

● Conseiller de prélever des échantillons dont les tailles satisfont à :

$$\sigma_0^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \leqslant \left(\frac{d_2 - d_1}{2\,u_{1-\delta}}\right)^2$$

n_1 et n_2 étant dans le rapport suivant en fonction des coûts unitaires d'échantillonnage supposés constants :

$$\frac{n_1}{n_2} = \sqrt{\frac{p_2}{p_1}}$$

En particulier, lorsque $p_1 = p_2$:

$$n_1 = n_2 \geqslant 8 \left(\frac{\sigma_0\, u_{1-\delta}}{d_2 - d_1}\right)^2$$

Ici encore, il est prudent de dépasser assez nettement ces seuils pour éviter de se trouver dans le cas b_2. ci-dessous.

● Examiner le résultat de l'échantillonnage comme il est indiqué en *b*.

a_2. Si *on ne dispose pas d'information* sur la valeur commune des écarts-types, prélever 20 ou 30 observations préalables pour estimer l'écart-type commun et poursuivre l'échantillonnage comme indiqué en a_1.

b. Si l'avis du statisticien est demandé après échantillonnage :

Comparer :

$$\frac{1}{n_1} + \frac{1}{n_2} \quad \text{à} \quad \left(\frac{d_2 - d_1}{2\, s\, t_{1-\delta}(n_1+n_2-2)}\right)^2 ,$$

où

$$s = \sqrt{\frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{n_1 + n_2 - 2}} .$$

b_1. Si $\frac{1}{n_1} + \frac{1}{n_2} \leqslant \left(\frac{d_2 - d_1}{2\, s\, t_{1-\delta}(n_1+n_2-2)}\right)^2$:

● Appliquer la règle à *deux niveaux* :

- Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant \frac{d_1 + d_2}{2}$;

- Réponse NON dans le cas contraire.

● Indiquer à l'homme de l'art que le risque maximum encouru est :

$$\delta^* = \Pr \left\{ T(n_1 + n_2 - 2) > \frac{d_2 - d_1}{2\, s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right\} \leqslant \delta .$$

b_2. Si $\frac{1}{n_1} + \frac{1}{n_2} > \left(\frac{d_2 - d_1}{2\, s\, t_{1-\delta}(n_1+n_2-2)}\right)^2$:

Les tailles d'échantillon sont alors insuffisantes pour qu'on puisse satisfaire aux spécifications du problème en adoptant une règle à deux niveaux. Ici encore, trois attitudes sont possibles :

Première attitude : Recommander des *tirages complémentaires* et se ramener en a_1., en utilisant s comme estimation σ_0 de σ.

Seconde attitude : Adopter la règle à *trois niveaux :*

- Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant d_2 - s\, t_{1-\delta}(n_1+n_2-2) \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$;

- Réponse NON si $\bar{x}_1 - \bar{x}_2 \geqslant d_1 + s\, t_{1-\delta}(n_1+n_2-2) \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$;

- Réponse ABSTENTION dans les autres cas.

Troisième attitude : Adopter la règle à *deux niveaux :*

- Réponse OUI si $\bar{x}_1 - \bar{x}_2 \leqslant \frac{d_1 + d_2}{2}$;

- Réponse NON dans le cas contraire ;

en indiquant à l'homme de l'art que le risque maximum encouru est :

$$\delta^* = \Pr \left\{ T(n_1 + n_2 - 2) > \frac{d_2 - d_1}{2\, s \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \right\} > \delta .$$

2.4.4 - Test de signification d'une moyenne : test " contre ≠ ".

Envisageons le test *bilatéral* de signification d'une moyenne, analogue à celui étudié en 2.2., mais dans le cas où l'écart-type σ est *inconnu*. Admettons, comme en 2.2., que le problème est symétrique par rapport à une valeur désignée par m_0.

Lorsque la taille de l'échantillon est insuffisante, convenons d'adopter, par analogie avec le test étudié en 2.2., une règle de la forme :

- Réponse OUI si $\left| \dfrac{\bar{x}-m_0}{s/\sqrt{n}} \right| \leqslant D_1$;
- Réponse NON si $\left| \dfrac{\bar{x}-m_0}{s/\sqrt{n}} \right| \geqslant D_2$;
- Réponse ABSTENTION dans les autres cas.

Les probabilités de répondre OUI et NON sont données par les lois de Student ou de Snedecor *non centrées*.

$$\Pr \left\{ \text{Répondre OUI} \Big/_{m,\sigma} \right\} = \Pr \left\{ n \frac{(\bar{x}-m_0)^2}{s^2} < D_1^2 \right\} .$$

$$= \Pr \left\{ F\left(1,\ n-1,\ n\left(\frac{m-m_0}{\sigma}\right)^2\right) < D_1^2 \right\} .$$

$$\Pr \left\{ \text{Répondre NON} \Big/_{m,\sigma} \right\} = \Pr \left\{ n \frac{(\bar{x}-m_0)^2}{s^2} > D_2^2 \right\}$$

$$= \Pr \left\{ F\left(1,\ n-1,\ n\left(\frac{m-m_0}{\sigma}\right)^2\right) > D_2^2 \right\} .$$

Ces fonctions ne s'expriment pas en termes de $|m-m_0|$ comme dans le test étudié en 2.2., mais en termes de $\left|\dfrac{m-m_0}{\sigma}\right|$, σ étant l'écart-type inconnu. En conséquence, les spécifications du problème doivent être exprimées *dans le même langage* : ceci revient à dire que si les spécifications sont exprimées comme en 2.2. en termes de $|m-m_0|$:

(O) : $|m-m_0| \leqslant m_1''-m_0$;

(I) : $m_1''-m_0 < |m-m_0| < m_2''-m_0$;

(N) : $|m-m_0| \geqslant m_2''-m_0$;

le problème n'a de sens que *si on connait, de façon approchée au moins, la valeur de l'écart-type* σ. Admettons que les spécifications du problème peuvent être mises sous la forme :

(O) : $\left|\dfrac{m-m_0}{\sigma}\right| \leqslant \lambda_1$;

$$(I) \quad : \quad \lambda_1 < \left|\frac{m-m_o}{\sigma}\right| < \lambda_2 \ ;$$

$$(N) \quad : \quad \left|\frac{m-m_o}{\sigma}\right| \geqslant \lambda_2 \ .$$

On aboutit alors à :

$$D_1^2 = F_\delta(1,\ n-1,\ n\ \lambda_2^2)\ ;$$
$$D_2^2 = F_{1-\delta}(1,\ n-1,\ n\ \lambda_1^2)\ .$$

La condition à laquelle doit obéir n pour qu'une règle à deux niveaux existe, qui satisfasse aux spécifications (λ_1, λ_2, δ) du problème, est :

$$F_\delta(1,\ n-1,\ n\ \lambda_2^2) \leqslant F_{1-\delta}(1,\ n-1,\ n\ \lambda_1^2)\ .$$

Lorsque n est assez grand, cette condition est équivalente à :

$$n \geqslant \left(\frac{L_2}{\lambda_2}\right)^2 ,$$

où L_2 est la valeur fournie par la table 1 en fonction de $r = \lambda_1/\lambda_2$.

Le risque maximum encouru δ^* est alors la solution de :

$$F_{\delta^*}(1,\ n-1,\ n\ \lambda_2^2) = F_{1-\delta^*}(1,\ n-1,\ n\lambda_1^2)\ .$$

et la règle à deux niveaux correspond à :

- Réponse OUI si $\left|\dfrac{\overline{x}-m_o}{s/\sqrt{n}}\right| \leqslant \sqrt{F_{\delta^*}(1,\ n-1,\ n\lambda_2^2)}$
- Réponse NON dans le cas contraire.

D'où la conduite à tenir pour résoudre le problème envisagé :

a. Si le statisticien est consulté avant échantillonnage :

● Recommander de prélever un nombre n_o d'observations tel que :

$$F_\delta(1,\ n_o-1,\ n_o\lambda_2^2) = F_{1-\delta}(1,\ n_o-1,\ n_o\lambda_1^2)\ .$$

● Recommander d'appliquer la règle à deux niveaux :

- Réponse OUI si $\left|\dfrac{\overline{x}-m_o}{s/\sqrt{n_o}}\right| \leqslant \sqrt{F_\delta(1,\ n_o-1,\ n_o\lambda_2^2)}$;
- Réponse NON dans le cas contraire.

b. Si le statisticien est consulté après échantillonnage :

b_1. Si la taille de l'échantillon est *au moins égale* au seuil n_o défini en *a.* :

● Recommander d'appliquer la règle à deux niveaux :

- Réponse OUI si $\left|\dfrac{\bar{x}-m_o}{s/\sqrt{n}}\right| \leqslant D$;

- Réponse NON dans le cas contraire.

● Indiquer à l'homme de l'art le risque maximum δ^* encouru : D et δ^* sont donnés par :

$$D^2 = F_{\delta^*}(1,\ n-1,\ n\,\lambda_2^2) = F_{1-\delta^*}(1,\ n-1,\ n\,\lambda_1^2)$$

b₂. Si la taille de l'échantillon est *inférieure* au seuil défini en *a.*, trois attitudes sont possibles :

Première attitude : Prélever n_o-n observations *complémentaires*, de façon à se ramener au cas *a*.

Seconde attitude : Adopter la règle à *trois niveaux :*

- Réponse OUI si $\left|\dfrac{\bar{x}-m_o}{s/\sqrt{n}}\right| \leqslant \sqrt{F_\delta(1,\ n-1,\ n\,\lambda_2^2)}$;

- Réponse NON si $\left|\dfrac{\bar{x}-m_o}{s/\sqrt{n}}\right| \geqslant \sqrt{F_{1-\delta}(1,\ n-1,\ n\,\lambda_1^2)}$

- Réponse ABSTENTION dans les autres cas.

Troisième attitude : Adopter la règle définie en b_1. en précisant à l'homme de l'art la valeur du risque maximum δ^* encouru, supérieure à δ.

Remarque.

Si les moyens de calcul qui nous ont été promis sont effectivement mis à notre disposition, nous comptons mettre en tables numériques ou en abaques la résolution des équations précédentes, qui font intervenir la variable de Snedecor non centrée. Ces résultats feront l'objet d'une publication ultérieure dans la *Revue de Statistique Appliquée*.

2.4.5 - Test d'homogénéité entre moyennes.

Le problème du test d'*homogénéité* des moyennes de k populations *normales* de *même* écart-type *connu ou non*, sur la base d'échantillons *indépendants* dont les tailles sont désignées par $n_1, \dots, n_k$ généralise celui du test de *comparaison* - lequel correspond à k = 2. Si l'écart-type commun σ des populations est connu, les variables mises en jeu suivent la loi du χ^2 non centrée ; dans le cas contraire, la loi de Snedecor non centrée. Nous étudions ce dernier cas et nous indiquerons les modifications à apporter aux résultats pour résoudre le premier.

Désignons par $\bar{x}_h$, s_h^2 les moyennes et variances estimées de chaque échantillon :

$$\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{ih} \quad ; \quad s_h^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(x_{ih}-\bar{x}_h)^2$$

et par $\bar{\bar{x}}$, s^2 les quantités :

$$\bar{\bar{x}} = \frac{1}{n}\sum_{h=1}^{k} n_h\,\bar{x}_h = \frac{1}{n}\sum_{h=1}^{k}\sum_{i=1}^{n_h} x_{ih} \ ;$$

$$s^2 = \frac{1}{n-k}\sum_{h=1}^{k}(n_h-1)\,s = \frac{1}{n-k}\sum_{h=1}^{k}\sum_{i=1}^{n_h}(x_{ih}-\bar{x}_h)^2.$$

La loi de la quantité :

$$\frac{1}{k-1}\sum_{h=1}^{k} n_h\left(\frac{\bar{x}_h-\bar{\bar{x}}}{s}\right)^2$$

est la *loi de Snedecor non centrée :*

$$F\left[k-1,\ n-k,\ \sum_{h=1}^{k} n_h\left(\frac{m_h-\bar{m}}{\sigma}\right)^2\right],$$

où $\bar{m}$ désigne la moyenne *pondérée* des moyennes m_h :

$$\bar{m} = \frac{1}{n}\sum_{h=1}^{k} n_h\, m_h$$

et où n désigne la somme des tailles des échantillons :

$$n = \sum_{h=1}^{k} n_h.$$

En notant $\bar{\sigma}$ l'écart-type *pondéré* des moyennes vraies m_h :

$$(\bar{\sigma})^2 = \frac{1}{n}\sum_{h=1}^{k} n_h(m_h-\bar{m})^2 \quad ,$$

le résultat précédent peut s'écrire :

$$\frac{1}{k-1}\sum_{h=1}^{k} n_h\left(\frac{\bar{x}_h-\bar{\bar{x}}}{s}\right)^2 = F\left[k-1,\ n-k,\ n\left(\frac{\bar{\sigma}}{\sigma}\right)^2\right].$$

Il s'ensuit que si les spécifications du problème sont définies en termes du rapport de la variance *non pondérée* des moyennes m_h à la variance commune σ^2 des populations échantillonnéès ou en termes de la variance des moyennes m_h, lorsqu'on a une idée au moins approximative de la valeur de la variance commune σ^2 :

$$(O) \quad : \quad \frac{\frac{1}{k}\sum_{h=1}^{k}(m_h-\bar{\bar{m}})^2}{\sigma^2} \leqslant \lambda_1^2 \ ;$$

$$(N) \quad : \quad \frac{\frac{1}{k}\sum_{h=1}^{k}(m_h-\bar{\bar{m}})^2}{\sigma^2} \geqslant \lambda_2^2 \ ;$$

où $\bar{\bar{m}}$ désigne la moyenne *non pondérée* des moyennes m_h :

$$\bar{\bar{m}} = \frac{1}{k}\sum_{h=1}^{k} m_h,$$

on satisfait aux exigences $(\lambda_1^2, \lambda_2^2, \delta)$ du problème en choisissant des tailles d'échantillon n_h *égales* - et par conséquent égales à $\nu = \frac{n}{k}$ - et en retenant pour ν la valeur telle que :

$$F_\delta(k-1,\ \nu k-k,\ \nu k\lambda_2^2) = F_{1-\delta}(k-1,\ \nu k-k,\ \nu k\lambda_1^2)\ .$$

D'où la conduite à tenir pour résoudre le problème défini par les spécifications $(\lambda_1^2, \lambda_2^2, \delta)$:

a. Si le statisticien est consulté avant échantillonnage :

● Recommander de prélever des échantillons de taille *égale* ν, ν satisfaisant à la relation précédente.

● Recommander d'appliquer la règle :

- Réponse OUI si $\dfrac{k\nu}{k-1}\ \dfrac{\sum_{h=1}^{k}(\bar{x}_h-\bar{\bar{x}})^2}{\sum_{h=1}^{k} s_h^2} \leqslant F_\delta(k-1,\ \nu k-k,\ \nu k \lambda_2^2)$;

- Réponse NON dans le cas contraire.

b. Si le statisticien est consulté après échantillonnage :

b_1. Si la taille totale n satisfait à :

$$F_\delta(k-1,\ n-k,\ n\lambda_2^2) \leqslant F_{1-\delta}(k-1,\ n-k,\ n\lambda_1^2)$$

et si les spécifications du problème peuvent être mises sous la forme :

$$(O)\ :\ \frac{1}{n\sigma^2}\sum_{h=1}^{k} n_h(m_h-\bar{\bar{m}})^2 \leqslant \lambda_1^2\ ;$$

$$(N)\ :\ \frac{1}{n\sigma^2}\sum_{h=1}^{k} n_h(m_h-\bar{\bar{m}})^2 \geqslant \lambda_2^2\ ;$$

c'est-à-dire si les spécifications peuvent être exprimées en termes du rapport de la variance *pondérée* des moyennes m_h (les poids correspondant aux tailles n_h retenues) à la variance commune σ^2 des populations échantillonnées, on peut appliquer la règle :

- Réponse OUI si $\dfrac{1}{k-1}\sum_{h=1}^{k} n_h\left(\dfrac{\bar{x}_h-\bar{\bar{x}}}{s}\right)^2 \leqslant F_{\delta^*}(k-1, n-k, \lambda_2^2)$;

- Réponse NON dans le cas contraire ;

en indiquant à l'homme de l'art la valeur δ^* du risque maximum encouru : δ^* est la racine de l'équation :

$$F_{\delta^*}(k-1,\ n-k,\ n\lambda_2^2) = F_{1-\delta^*}(k-1,\ n-k,\ n\lambda_1^2)\ .$$

b_2. Si la taille totale n satisfait à :

$$F_\delta(k-1,\ n-k,\ n\lambda_2^2) > F_{1-\delta}(k-1,\ n-k,\ n\lambda_1^2)\ ,$$

la quantité d'information disponible est insuffisante pour répondre aux spécifications indiquées en *b_1.*, trois attitudes sont possibles :

Première attitude : Prélever des informations *complémentaires* pour se ramener au cas *a.* ou au cas *b_1.* suivant la forme des spécifications.

Seconde attitude : Adopter la règle à *trois niveaux :*

- Réponse OUI si $\frac{1}{k-1} \sum_{h=1}^{k} n_h\left(\frac{\overline{x}_h - \overline{\overline{x}}}{s}\right)^2 \leqslant F_\delta(k-1, n-k, n\lambda_2^2)$;

- Réponse NON si $\frac{1}{k-1} \sum_{h=1}^{k} n_h\left(\frac{\overline{x}_h - \overline{\overline{x}}}{s}\right)^2 \geqslant F_{1-\delta}(k-1, n-k, n\lambda_1^2)$;

- Réponse ABSTENTION dans les autres cas.

Ceci ne vaut que si les spécifications peuvent être énoncées sous la forme indiquée en b_1.

Troisième attitude : Adopter la règle à *deux niveaux* définie en b_1. en précisant à l'homme de l'art la valeur δ^* du risque maximum encouru, supérieure à δ.

Remarques.

1/ Si on cherche à déceler l'hétérogénéité d'un ensemble de k populations (supposées normales de même variance) et si on ne veut *privilégier* aucune des populations par rapport aux autres, c'est en termes de la variance *non pondérée* des moyennes qu'on exprimera le degré d'hétérogénéité des populations. En conséquence, c'est sur la base de k échantillons de tailles à peu près *égales* qu'il conviendra d'effectuer le test.

Lorsqu'on mélange k populations de moyennes $m_1, \ldots, m_k$ et de variances *égales*, la variance du *mélange* est égale à la variance commune σ^2 augmentée de la variance *pondérée* des moyennes m_h, les coefficients de pondération étant les proportions α_h qui définissent la structure du mélange :

$$\text{Variance du mélange} = \sigma^2 + \sum_{h=1}^{k} \alpha_h (m_h - \overline{m})^2 ,$$

où $\overline{m}$ est la moyenne pondérée des moyennes, égale à la moyenne du mélange :

$$\overline{m} = \sum_{h=1}^{k} \alpha_h m_h .$$

L'accroissement de la variance du mélange par rapport à la variance commune, accroissement dû à l'hétérogénéité des moyennes, est ainsi en valeur relative :

$$\lambda^2 = \sum_{h=1}^{k} \alpha_h \left(\frac{m_h - \overline{m}}{\sigma}\right)^2 .$$

En conséquence, si l'indicateur d'hétérogénéité des moyennes est cet accroissement relatif de variance, on retiendra des tailles d'échantillon n_h *inégales* et proportionnelles aux coefficients définissant la structure du mélange :

$$n_h = n\,\alpha_h ,$$

et on déterminera n de telle sorte que :

$$F_\delta(k-1,\ n-k,\ n\lambda_2^2) = F_{1-\delta}(k-1,\ n-k,\ n\lambda_1^2)$$

2/ Si l'écart-type commun σ de chacune des populations échantillonnées est *connu*, il faut remplacer dans ce qui précède :

$$\frac{1}{k-1}\sum_{h=1}^{k} n_h\left(\frac{\overline{x}_h-\overline{\overline{x}}}{s}\right)^2 \qquad \text{par} \qquad \frac{1}{k-1}\sum_{h=1}^{k} n_h\left(\frac{\overline{x}_h-\overline{\overline{x}}}{\sigma}\right)^2 ;$$

$$F(k-1,\ n-k,\ n\lambda^2) \qquad \text{par} \qquad \frac{1}{k-1}\ \chi^2(k-1,\ n\lambda^2)\ .$$

Alors les spécifications du problème s'expriment en termes de la variance - éventuellement pondérée - des moyennes m_h.

3/ Dans le cas d'un test de *comparaison* ($k = 2$), les formules se simplifient légèrement en ce sens que :

$$\sum_{h=1}^{k} n_h(\overline{x}_h-\overline{\overline{x}})^2 \quad \text{s'écrit} \quad \left(\frac{\overline{x}_1-\overline{x}_2}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)^2$$

Si les spécifications du problème sont :

$$(O) \quad : \quad \left|\frac{m_1-m_2}{\sigma}\right| \leqslant d_1 \ ;$$

$$(N) \quad : \quad \left|\frac{m_1-m_2}{\sigma}\right| \geqslant d_2 \ ;$$

il convient de prélever deux échantillons de tailles n_1 et n_2 telles que :

$$F_\delta\left(1,\ n_1+n_2-2,\ \frac{n_1 n_2}{n_1+n_2}\ d_2^2\right) \leqslant F_{1-\delta}\left(1,\ n_1+n_2-2,\ \frac{n_1 n_2}{n_1+n_2}\ d_1^2\right)\ .$$

Si on cherche à minimiser une fonction de coût linéaire :

$$p_1 n_1 + p_2 n_2\ ,$$

on choisira des tailles n_1 et n_2 proportionnelles à $\frac{1}{\sqrt{p_1}}$ et $\frac{1}{\sqrt{p_2}}$:

$$n_1 = n\,\frac{\sqrt{p_2}}{\sqrt{p_1}+\sqrt{p_2}} \quad , \quad n_2 = n\,\frac{\sqrt{p_1}}{\sqrt{p_1}+\sqrt{p_2}}\ .$$

La taille n *la plus économique* est celle qui satisfait à :

$$F_\delta\left(1,\ n-2,\ n\,\frac{\sqrt{p_1 p_2}}{(\sqrt{p_1}+\sqrt{p_2})^2}\ d_2^2\right) = F_{1-\delta}\left(1,\ n-2,\ n\,\frac{\sqrt{p_1 p_2}}{(\sqrt{p_1}+\sqrt{p_2})^2}\ d_1^2\right)\ .$$

La règle du test est alors :

- Réponse OUI si $\frac{n_1 n_2}{n_1+n_2}\ (\overline{x}_1-\overline{x}_2)^2 \leqslant F_\delta\left(1,\ n_1+n_2-2,\ \frac{n_1 n_2}{n_1+n_2}\ d_2^2\right)$;
- Réponse NON dans le cas contraire.

2.4.6 - Plan d'expérience factoriel avec égales répétitions.

Considérons un plan d'expérience à deux facteurs avec *égales répétitions* : l'égalité des répétitions est en effet nécessaire si on veut assurer des spécifications portant sur le rapport des variances *non pondérées*.

Supposons que par exemple le rendement d'une parcelle k ensemencée en variété i (i = 1, 2, ..., p) et fertilisée en engrais j(j = 1, 2, ..., q) soit de la forme :

$$x_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

où les ε_{ijk} sont des variables *normales centrées indépendantes* et *homoscédastiques* (variance *commune* supposée inconnue et désignée par σ^2). Soit r le nombre de répétitions (k = 1, 2, ..., r) de la combinaison (i, j). Le nombre total d'observations est pqr. Les moyennes partielles sont désignées par $\overline{x}_{ij.}$, $\overline{x}_{i..}$, $\overline{x}_{.j.}$, $\overline{x}_{...}$ suivant des notations évidentes ; de même pour $\overline{\mu}_{i.}$, $\overline{\mu}_{.j}$ et $\overline{\mu}_{..}$.

Envisageons les hypothèses suivantes :

1/ L'action simultanée de la variété et de l'engrais est *sensiblement additive* (pas *d'intéraction différentielle substantielle) :*

$$\mu_{ij} \approx \mu_o + \alpha_i + \beta_j \ .$$

α_i est l'effet différentiel *intrinsèque* de la variété i sur le rendement ; de même β_j est l'effet différentiel *intrinsèque* de l'engrais j sur le rendement :

2/ Le facteur variété n'a pas d'influence différentielle *sensible* sur le rendement :

$$\mu_{ij} \approx \mu_o + \beta_j \ .$$

3/ Le facteur engrais n'a pas d'influence différentielle *sensible* sur le rendement :

$$\mu_{ij} \approx \mu_o + \alpha_i \ .$$

4/ Aucun des deux facteurs variété et engrais n'a d'influence différentielle *sensible* sur le rendement :

$$\mu_{ij} \approx \mu_o \ .$$

Chacune des hypothèses 2/ à 4/ peut être testée dans le cadre de l'hypothèse générale (G) ou dans le cadre de l'hypothèse 1'/ :

1'/ L'action simultanée des deux facteurs est *rigoureusement* additive.

De même, l'hypothèse 4/ peut être testée dans le cadre de l'hypothèse générale (G), de l'hypothèse 1'/ ou des hypothèses 2'/ et 3'/ :

2'/ L'action du facteur variété est *rigoureusement* nulle ;

3'/ L'action du facteur engrais est *rigoureusement* nulle.

Etudions à titre d'exemple le test de l'hypothèse 2/ dans le cadre de l'hypothèse 1'/ : *le facteur variété n'a pas d'influence sensible sur le rendement, sachant que, s'il a une influence, son action se combine additivement avec celle du facteur engrais* (absence d'interaction). Désignons par R_1^2 et R_2^2 la somme des carrés des résidus après estimation des paramètres dans le cadre des modèles 1'/ et 2'/ :

$$R_1^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} [(x_{ijk} - \bar{x}_{...}) - (\bar{x}_{i..} - \bar{x}_{...}) - (\bar{x}_{.j.} - \bar{x}_{...})]^2 ,$$

$$R_2^2 = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} [(x_{ijk} - \bar{x}_{...}) - (\bar{x}_{.j.} - \bar{x}_{...})]^2 ,$$

$$R_2^2 - R_1^2 = r\,q \sum_{i=1}^{p} (\bar{x}_{i..} - \bar{x}_{...})^2 .$$

Le rapport :

$$\frac{(R_2^2 - R_1^2)/p-1}{R_1^2/pqr - p-q+1}$$

suit la loi de Snedecor non centrée :

$$F\left[p-1,\ pqr-p-q+1,\ \frac{pqr}{\sigma^2}\, Q^2(\alpha)\right] ,$$

où $Q(\alpha)$ est l'effet quadratique moyen du facteur variété :

$$Q^2(\alpha) = \frac{1}{p} \sum_{i=1}^{p} (\bar{\mu}_{i.} - \bar{\mu}_{..})^2 .$$

En conséquence, si les spécifications du problème sont définies en termes du rapport $\frac{Q^2(\alpha)}{\sigma^2}$:

$$(O) \ :\ \frac{Q^2(\alpha)}{\sigma^2} \leqslant \lambda_1^2 \ ;$$

$$(N) \ :\ \frac{Q^2(\alpha)}{\sigma^2} \geqslant \lambda_2^2 \ ;$$

on satisfait aux exigences $(\lambda_1^2, \lambda_2^2, \delta)$ du problème en choisissant un nombre de répétitions r tel que :

$$F_\delta[p-1,\ pqr-p-q+1,\ pqr\,\lambda_2^2] \leqslant F_{1-\delta}[p-1,\ pqr-p-q+1,\ pqr\,\lambda_1^2] .$$

D'où la conduite à tenir, pour tester l'hypothèse envisagée :

a. Si le statisticien est consulté avant échantillonnage :

- Recommander d'effectuer un nombre r_0 de répétitions par combinaison (variété × engrais) tel que :

$$F_\delta(p-1,\ pqr_0-p-q+1,\ pqr_0\,\lambda_2^2) = F_{1-\delta}(p-1,\ pqr_0-p-q+1,\ pqr_0\,\lambda_1^2)$$

- Recommander d'appliquer la règle :

- Réponse OUI si

$$\frac{(R_2^2-R_1^2)/p-1}{R_1^2/pqr_o-p-q+1} \leqslant F_\delta(p-1,\ pqr_o-p-q+1,\ pqr_o\lambda_2^2)\ ;$$

- Réponse NON dans le cas contraire.

b. Si le statisticien est consulté après échantillonnage :

b_1. Si le nombre r de répétitions satisfait à :

$$F_\delta(p-1,\ pqr-p-q+1,\ pqr\,\lambda_2^2) \leqslant F_{1-\delta}(p-1,\ pqr-p-q+1,\ pqr\,\lambda_1^2)\ :$$

Appliquer la règle :

- Réponse OUI si

$$\frac{(R_2^2-R_1^2)/p-1}{R_1^2/pqr-p-q+1} \leqslant F_{\delta^*}(p-1,\ pqr-p-q+1,\ pqr\,\lambda_2^2)\ ;$$

- Réponse NON dans le cas contraire ;

et indiquer à l'homme de l'art la valeur δ^* du risque maximum encouru, inférieure à δ : δ^* satisfait à :

$$F_{\delta^*}(p-1,\ pqr-p-q+1,\ pqr\,\lambda_2^2) = F_{1-\delta^*}(p-1,\ pqr-p-q+1,\ pqr\,\lambda_1^2).$$

b_2. Si le nombre r de répétitions satisfait à :

$$F_\delta(p-1,\ pqr-p-q+1,\ pqr\,\lambda_2^2) > F_{1-\delta}(p-1,\ pqr-p-q+1,\ pqr\,\lambda_1^2)\ :$$

Les trois attitudes possibles sont :

Première attitude : Effectuer un nombre *complémentaire* $\Delta r = r_o - r$ de répétitions par combinaison (variété × engrais) de façon à se ramener au cas *a*.

Seconde attitude : Adopter la règle à *trois niveaux :*

- Réponse OUI si

$$\frac{(R_2^2-R_1^2)/p-1}{R_1^2/pqr-p-q+1} \leqslant F_\delta(p-1,\ pqr-p-q+1,\ pqr\,\lambda_2^2)\ ;$$

- Réponse NON si

$$\frac{(R_2^2-R_1^2)/p-1}{R_1^2/pqr-p-q+1} \geqslant F_{1-\delta}(p-1,\ pqr-p-q+1,\ pqr\,\lambda_1^2)\ ;$$

- Réponse ABSTENTION dans les autres cas.

Troisième attitude : Adopter la règle à *deux niveaux* définie en b_1. en précisant à l'homme de l'art la valeur δ^* du risque maximum encouru, supérieure à δ.

Les autres tests envisagés plus haut sont tout à fait analogues. Dans le tableau ci-après, nous indiquons, suivant l'hypothèse testée et l'hypothèse servant de cadre, ce par quoi il faut remplacer, dans les expressions précédentes :

R_1^2 ;

$R_2^2 - R_1^2$;

p-1, désigné par ν_1 ;

p q r-p-q+1, désigné par ν_2 ;

$Q^2(\alpha)$, désigné par Q^2.

Hypothèse testée	Hypothèse de référence	$R_2^2 - R_1^2$	R_1^2	ν_1	ν_2	Q^2
1/	(G)	$R_1^2 - R_G^2$	R_G^2	p+q-1	pq(r-1)	$Q^2(\gamma)$
2/	(G)	$R_2^2 - R_G^2$	R_G^2	q(p-1)	pq(r-1)	$Q^2(\alpha)+Q^2(\gamma)$
2/	1'/	$R_2^2 - R_1^2$	R_1^2	p-1	p q r-p-q+1	$Q^2(\alpha)$
3/	(G)	$R_3^2 - R_G^2$	R_G^2	p(q-1)	pq(r-1)	$Q^2(\beta)+Q^2(\gamma)$
3/	1'/	$R_3^2 - R_1^2$	R_1^2	q-1	p q r-p-q+1	$Q^2(\beta)$
4/	(G)	$R_4^2 - R_G^2$	R_G^2	pq-1	pq(r-1)	$Q^2(\alpha)+Q^2(\beta)+Q^2(\gamma)$
4/	1'/	$R_4^2 - R_1^2$	R_1^2	p+q-2	p q r-p-q+1	$Q^2(\alpha)+Q^2(\beta)$
4/	2'/	$R_4^2 - R_2^2$	R_2^2	q-1	q(pr-1)	$Q^2(\beta)$
4/	3'/	$R_4^2 - R_3^2$	R_3^2	p-1	p(qr-1)	$Q^2(\alpha)$

Si on pose :

$$A = \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} (x_{ijk} - \overline{x}_{ij.})^2 ,$$

$$B = r \sum_{i=1}^{p} \sum_{j=1}^{q} (\overline{x}_{ij.} - \overline{x}_{i..})^2 ,$$

$$C = r \sum_{i=1}^{p} \sum_{j=1}^{q} (\overline{x}_{ij.} - \overline{x}_{.j.})^2 ,$$

$$D = qr \sum_{i=1}^{p} (\overline{x}_{i..} - \overline{x}_{...})^2 ,$$

$$E = pr \sum_{j=1}^{q} (\overline{x}_{.j.} - \overline{x}_{...})^2 ,$$

il vient :

$$R_G^2 = A ,$$
$$R_1^2 = A + B - E = A + C - D ,$$
$$R_2^2 = A + C ,$$
$$R_3^2 = A + B ,$$
$$R_4^2 = A + B + D = A + C + E .$$

Par ailleurs, les moyennes quadratiques $Q(\alpha)$, $Q(\beta)$ et $Q(\gamma)$ sont égales à :

$$Q^2(\alpha) = \frac{1}{p} \sum_{i=1}^{p} (\overline{\mu}_{i.} - \overline{\mu}_{..})^2 ,$$

$$Q^2(\beta) = \frac{1}{q} \sum_{j=1}^{q} (\overline{\mu}_{.j} - \overline{\mu}_{..})^2 \quad ,$$

$$Q^2(\gamma) = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} [(\mu_{ij} - \overline{\mu}_{..}) - (\overline{\mu}_{i.} - \overline{\mu}_{..}) - (\overline{\mu}_{.j} - \overline{\mu}_{..})]^2$$

III. CONCLUSIONS

Au cours du chapitre 2, nous avons appliqué au problème de la sélection d'une règle de test une méthode fondée à la fois sur la notion *d'ensemble d'indécision* et sur la notion de *risque maximum encouru*. Cette méthode revient dans les cas simples à se fixer a priori *deux points* de la courbe d'efficacité, correspondant aux frontières de l'ensemble d'indécision. Son application oblige à une *analyse préalable des spécifications particulières du problème étudié*. Par ailleurs, elle conduit à se poser la question de la *quantité d'information nécessaire à la solution du problème*, au moyen d'une règle à deux niveaux de conclusion. Ces différents points nous apparaissent fondamentaux sur le plan pratique. S'ils ne sont pas perdus de vue, en général, par les spécialistes du contrôle industriel de la qualité, ils sont souvent méconnus d'autres praticiens de la statistique.

3.1 - LES SPECIFICATIONS PARTICULIERES DU PROBLEME ETUDIE.

On reproche bien souvent au langage littéraire son caractère approximatif ; en revanche, on accorde au langage mathématique toutes les vertus de précision et de rigueur. Par suite, on tire l'impression qu'un problème est bien posé dès lors qu'il a été traduit dans une formulation mathématique. Cette impression peut être trompeuse si la traduction comporte une part trop grande de trahison : *Traduttore, traditore !* Lorsqu'on dit par exemple qu'un objet pèse 500 grammes, exprime-t-on *vraiment* que le poids de l'objet peut-être représenté géométriquement par le point d'abscisse 500 sur l'axe réel ? Veux-t-on dire par là qu'un objet de 501 grammes aurait un poids très différent ? Certainement pas : suivant le cas, cela voudra dire que le poids réel - abstraction mathématique du caractère pesant de l'objet considéré - est compris entre 450 et 550 grammes, ou entre 490 grammes et 510 grammes ou etc... Encore convient-il d'accorder à ce genre de précision le sens d'un *ordre de grandeur* : on n'est pas sûr qu'un objet de 500 grammes à 10 grammes près diffère de plus d'un gramme d'un autre objet qui pèserait 488 grammes à 1 gramme près. Tout au plus *est-on à peu près certain que leurs poids diffèrent*, compte tenu du grand nombre de valeurs possibles du poids d'un objet pesant environ 500 grammes : pour s'en assurer, il suffira d'effectuer une mesure assez précise de la différence des poids !

Suivant le contexte envisagé, c'est-à-dire, en définitive, le problème étudié, lorsqu'on se pose la question de savoir si un objet pèse *sensiblement* 500 grammes, ce qu'on cherche finalement à déceler c'est s'il existe ou non un écart *substantiel* entre le poids de l'objet et la valeur de référence 500 grammes. Si cet écart n'est pas substantiel, on se comportera comme si le poids réel était exactement de 500 grammes : tous les objets, dont le poids réel ne diffère pas substantiellement de 500 grammes, seront considérés comme *équivalents*, du point de vue du poids. On ne peut en conséquence se poser la question de savoir si un objet pèse ou non 500 grammes, si on n'a pas défini au préalable la *zone d'équivalence* entourant la valeur 500.

Mais la difficulté est de *quantifier* cette zone d'équivalence. En effet, en lui assignant des frontières précises, on caricature la réalité : il n'y a pas une limite en deçà de laquelle se trouverait une certaine vérité et au delà de laquelle se trouverait la vérité contraire. Il faut prendre en compte une sorte de *dégradé continu*. Ainsi dans l'exemple précédent, lorsqu'on s'éloigne de la valeur de référence 500, l'intensité de l'équivalence diminue pour finalement s'annuler. On retrouve donc aux frontières de la zone d'équivalence la même difficulté que celle rencontrée au voisinage de la valeur 500 et qui nous a conduit à la notion même de zone d'équivalence ! Cette difficulté, nous la retrouvons à nouveau si nous voulons quantifier la zone de non-équivalence. C'est pourquoi nous proposons d'introduire la notion *d'ensemble d'indécision*, sorte de *zone franche*, intermédiaire entre l'équivalence et la non-équivalence. Lorsqu'on se pose la question de savoir si le poids réel est sensiblement égal à 500 grammes, on voudrait recevoir la réponse OUI quand le poids réel appartient à la zone d'équivalence, la réponse NON quand le poids réel appartient à la zone de non-équivalence, *indifféremment* la réponse OUI ou NON quand le poids réel appartient à l'ensemble d'indécision. La discontinuité introduite par les bornes de l'ensemble d'indécision est seulement apparente dans le cas du test statistique. En effet, on se trouve alors dans le domaine aléatoire, où la probabilité de recevoir une réponse donnée (OUI ou NON) varie continûment. C'est pourquoi le choix des limites de l'ensemble d'indécision n'introduit pas une véritable discontinuité.

Nous définissons ensuite la notion de test *convenable* : une règle de test est convenable si la probabilité qu'il fournisse la réponse incorrecte est au plus égale à un seuil δ donné à l'avance, lorsque la réalité est *quelconque à l'extérieur de l'ensemble d'indécision*. Cette définition s'apparente à celle de l'intervalle de confiance : une règle de construction d'un intervalle de confiance est convenable si la probabilité qu'elle fournisse un intervalle ne recouvrant pas la vraie valeur est au plus égale à un seuil α donné à l'avance, *quelle que soit* la vraie valeur.

Bien que nous introduisions successivement les notions d'ensemble d'indécision et de risque maximum encouru, en réalité ces deux notions sont *intimement liées* : leur choix pour la résolution d'un problème pratique forme un tout. Cela signifie que les bornes de l'ensemble d'indécision constituent des situations également *extrêmes* : la frontière entre l'ensemble (O), zone d'équivalence, et l'ensemble (I), ensemble d'indécision, représente des situations analogues aussi extrêmes, en ce qui concerne la réponse OUI, que les situations, analogues entre elles, correspondant à la frontière entre l'ensemble (I) et l'ensemble (N) de non-équivalence, en ce qui concerne la réponse NON.

Lorsque le statisticien a le choix du plan de sondage, la méthode proposée consiste à retenir les tailles d'échantillon les plus économiques et une règle de test convenable à deux niveaux de réponse OUI ou NON.

Lorsque le statisticien n'a pas le choix du plan de sondage, la méthode proposée consiste :

- soit à retenir la règle à *deux niveaux de réponse* dont le risque maximum δ^* est le plus faible, à la condition que ce risque maximum soit jugé *acceptable*.

- soit à retenir la règle à *trois niveaux de réponse* : OUI, NON, ABSTENTION dont le risque maximum est égal à δ.

Il nous apparaît en effet plus sage de refuser de se prononcer lorsque la quantité d'information disponible est insuffisante ($\delta^* > \delta$ et réponse ABSTENTION) plutôt que de donner une réponse trop incertaine.

3.2 - COMPARAISON AVEC LA METHODE CONSISTANT A SE REFERER AU SEUL RISQUE DE PREMIERE ESPECE.

La méthode fréquemment utilisée par les statisticiens consiste à sélectionner, parmi un ensemble de règles jugées optimales en termes de courbe d'efficacité, celle qui donne au risque de première espèce la valeur maximum α, seuil choisi à l'avance comme le seuil δ.

Cette méthode nous semble mauvaise à plusieurs points de vue :

1/ Elle *privilégie* l'hypothèse testée alors qu'entre les réponses OUI et NON aucune considération véritablement fondée ne permet d'introduire a priori une dissymétrie entre l'hypothèse testée et l'hypothèse alternative. Si on teste l'hypothèse H_1 contre l'hypothèse H_0 avec le même risque de première espèce α, on aboutit à une autre règle que si on teste l'hypothèse H_0 contre l'hypothèse H_1. Ceci parait injustifié et contraire au bon sens.

2/ Elle ne prend pas en considération la *nature particulière du problème traité* mais tient lieu de solution "passe partout", valable pour toutes sortes de problèmes analogues dont les spécifications seraient différentes. Le recours à cette méthode représente une sorte de *démission du statisticien* devant la nature particulière du problème traité. Il ne faut évidemment pas se dissimuler la difficulté qu'il y a à définir l'ensemble d'indécision (I) et le risque δ nécessaires à l'application de la méthode que nous proposons. Il faut en effet faire un effort *d'appréciation* - inévitablement *subjectif* - qui réclame une connaissance intime des *dimensions* du problème étudié et une grande sûreté de *jugement* (notamment un bon sens des ordres de grandeur). Mais ce n'est pas parce que ce choix du doublet $[(I), \delta]$ est délicat qu'il faut l'éluder en s'en remettant à une méthode jugée "neutre" ou même "objective" en raison de son caractère "passe-partout" et en s'alignant sur tout un chacun par un choix arbitraire du risque α de première espèce : 5 %, 1 %, ... car le risque de première espèce est un mauvais indicateur des *performances* de l'instrument et plus encore un mauvais indicateur de *l'adéquation* de l'instrument au problème traité.

Il convient d'observer que, dans son principe même, une règle à deux (ou trois) niveaux de réponse est extrêmement *brutale et fruste* : ce n'est pas son allure *mathématique* qui lui confère noblesse et rigueur. Il faut voir dans un test statistique *un instrument grossier mais pratique* de décision. En outre, si la sélection d'une méthode de test repose sur des considérations *imprécises* (parce que *subjectives*) que le calcul *mathématique* oblige à formuler *avec précision*, il faut voir dans la solution obtenue *un ordre de grandeur* : la taille des échantillons nécessaire à la résolution du problème, qui résulte du calcul, doit être considérée comme une sorte de *repère indicatif* et non comme une *nécessité inéluctable*. Cette remarque atténue un peu le caractère *arbitraire* du choix portant sur le doublet $[(I), \delta]$.

3/ En caricaturant un peu, la méthode qui consiste à se référer au risque de première espèce α possède la curieuse propriété suivante : on obtient la réponse OUI lorsque la quantité d'information disponible est *faible*, la réponse NON lorsque la quantité d'information disponible est *importante*. En effet, considérons, à titre d'exemple, le test bilatéral symétrique de signification d'une moyenne, quand l'écart-type σ est connu. La méthode fondée sur le risque de première espèce α conduit à la règle :

- Réponse OUI si $\left|\dfrac{\bar{x}-m_0}{\sigma/\sqrt{n}}\right| \leqslant u_{1-\frac{\alpha}{2}}$;

- Réponse NON dans le cas contraire.

soit un écart $d = |m - m_0|$. La probabilité de répondre OUI, lorsque l'écart réel est d, est égale à :

$$\Pr\left\{\text{Réponse OUI}\Big/_{d}\right\} = \Pi\left(u_{1-\frac{\alpha}{2}} + \frac{\sqrt{n}\,d}{\sigma}\right) + \Pi\left(u_{1-\frac{\alpha}{2}} - \frac{\sqrt{n}\,d}{\sigma}\right) - 1 \,.$$

Si d est faible, lorsque n est petit, la probabilité de répondre OUI est ainsi à peine inférieure à $1-\alpha$; au contraire, *quel que soit* d, lorsque n est grand, la probabilité de répondre OUI est petite : *elle tend vers zéro lorsque n tend vers l'infini, aussi petit que soit d non nul !* En revanche avec la méthode proposée, la probabilité d'une réponse incorrecte est *toujours (c'est-à-dire quel que soit n) au plus égale à δ et elle tend vers zéro lorsque n tend vers l'infini :*

- si n est petit, la réponse *la plus probable* sera ABSTENTION ;

- si n est grand, la probabilité que la réponse soit *correcte* sera voisine de 1, lorsque la valeur correspondante de m appartient à (O) ou à (N).

4/ La question qui se pose dans ces conditions est de savoir quel genre d'information fournit la méthode fondée sur la prise en compte du risque de première espèce et à quelle sorte de problèmes elle est adaptée. En réalité, cette méthode répond à la question : *peut-on admettre, au vu des observations, si l'hypothèse testée est rigoureusement vérifiée ou non, étant entendu que si on peut déceler avec assez de certitude un écart, aussi petit soit-il, l'hypothèse est fausse et doit être refusée.* En conséquence, cette méthode ne permet que de tester des hypothèses *théoriques* : par exemple, la validité de la loi de Monsieur X, si Monsieur X avait émis une hypothèse théorique précise sur la nature des choses : tout écart à la loi, aussi faible soit-il, suffirait à prouver que la loi est fausse et à renvoyer Monsieur X à ses chères études...

Or dans le domaine concret, on ne cherche pas à savoir si telle hypothèse est *rigoureusement* vérifiée : on veut seulement savoir si entre l'hypothèse testée et la réalité la *différence n'est pas trop grande, du point de vue du problème particulier envisagé.*

Bien plus, sans craindre d'atteindre au paradoxe, on peut affirmer qu'étant donné la richesse de l'ensemble défini par l'hypothèse alternative, dans l'immense majorité des exemples de tests, *aucune hypothèse théorique ne peut être rigoureusement vérifiée* : il est *inutile* de chercher à savoir si la moyenne de telle population est *rigoureusement* égale à la valeur mathématique 3, parce qu'on est bien sûr à l'avance que cette hypothèse est *fausse*. En revanche, il est utile dans tel ou tel problème de chercher à savoir si elle ne diffère pas *substantiellement* de 3 ou, comme on dit, si elle ne diffère pas *significativement* de 3, après avoir toutefois pris la précaution de définir le terme *significativement*. Pour ce faire, nous proposons le doublet [(I), δ] .

5/ La remarque précédente a des implications extrêmement importantes en ce qui concerne les tests *non paramétriques*, tels que par exemple le test d'adéquation du χ^2. Tester l'hypothèse suivant laquelle une distribution est normale au moyen du test d'adéquation du χ^2 est un exercice de calcul numérique purement *gratuit* qui n'apporte aucune information mais qui au contraire risque d'être *nuisible* : en reprenant la remarque 3 précédente, on peut affirmer que la réponse du test sera le plus souvent OUI si l'échantillon est de petite taille, à coup sûr NON si l'échantillon est de très grande taille. Comme la notion de modèle normal n'a qu'une signification *théorique*, on peut même dire que la distribution réelle n'est *certainement pas* normale. Mais cela ne veut pas dire que le modèle normal n'est pas un excellent outil qui permettra de synthétiser les observations au moyen de deux paramètres (par exemple en vue de comparaisons avec d'autres distributions) et de tirer des estimations valables de certaines probabilités (interpolations, extrapolations, utilisation pour la construction d'un modèle plus compliqué). Admettre qu'une distribution est normale est un *postulat*, une sorte d'acte de foi fondé sur des considérations d'ordre théorique (plus ou moins simplificatrices de la réalité) qui ne saurait être ni infirmé ni confirmé par un test d'adéquation. Mieux vaut, dans ces conditions, ne pas effectuer de test d'adéquation !

Cette remarque vaut aussi bien pour le test d'adéquation du χ^2 que pour le test des signes, le test d'indépendance fondé sur un tableau de contingence, les tests de Wilcoxon, de Durbin et Watson, de Kolmogoroff-Smirnoff etc... Dans la mesure où la puissance de ces tests ne peut pratiquement pas être appréhendée, nous considérons qu'ils sont des instruments *académiques*. Leur construction est un passionnant exercice de calcul des probabilités. En revanche, leur intérêt pratique est, à notre avis, à peu près nul pour ne pas dire négatif.