GUY LE GARFF

1. Le statisticien, le gestionnaire et les risques. 2. Détermination du nombre de répétitions

Revue de statistique appliquée, tome 23, nº 4 (1975), p. 15-40
<http://www.numdam.org/item?id=RSA_1975__23_4_15_0>

1 – LE STATISTICIEN, LE GESTIONNAIRE ET LES RISQUES
2 – DÉTERMINATION DU NOMBRE DE RÉPÉTITIONS

Guy LE GARFF
Ingénieur agricole

Cette publication comprend 2 parties :

– La première est l'exposé de Monsieur G. Le Garff au colloque statistique des 13 et 14 février 73. Elle vise à nous faire prendre un peu de recul par rapport aux risques de première et deuxième espèce tels qu'ils sont définis par la statistique classique.

Pour cela on y examine l'adéquation du schéma classique de raisonnement au problème élémentaire de la comparaison d'un "témoin" et d'un "traité". On y envisage aussi les autres risques encourus par le statisticien.

– La deuxième partie propose une nouvelle méthode de détermination du nombre de répétitions : méthode qui intègre la notion de risque maximum et des considérations économiques.

Elle est le fruit de la première partie en ce sens que ses propositions n'auraient pu être établies sans les réflexions précédentes. La publication conjointe, permettant au lecteur de suivre le cheminement de l'auteur, facilitera sans doute la compréhension.

PLAN

1ère PARTIE

LE STATISTICIEN, LE GESTIONNAIRE ET LES RISQUES

1 – RAPPEL DU RAISONNEMENT CLASSIQUE ET DE LA DEFINITION DES RISQUES DE PREMIERE ET DEUXIEME ESPECE

Le statisticien endosse des risques quand il travaille sur échantillons et qu'il désire, à partir de ces échantillons, tirer des conclusions valables pour la population.

On peut écrire :

$$\text{échantillons} + \text{décision} \Rightarrow \text{risques}$$

Rappelons sa démarche intellectuelle :

Dans l'ignorance où il se trouve des vraies valeurs qui l'intéressent (calculées sur la population entière) il émet une hypothèse (en général l'hypothèse nulle Ho) et mesure, sur échantillons, des estimations des vraies valeurs qui l'intéressent.

S'il y a accord entre son hypothèse et ses observations, il conserve son hypothèse. S'il y a désaccord, il rejette son hypothèse.

Exemple

Est-ce que le taureau A est meilleur que le taureau B du point de vue de la valeur laitière de ses filles ? Les vraies valeurs des taureaux A et B ne pourraient être mesurées que sur les populations entières des filles qui naîtront de ces géniteurs. En toute rigueur, on peut même se demander si les vraies populations de référence ne sont pas les deux ensembles des filles qu'il serait possible de faire naître et d'élever à partir de chacun des spermatosoïdes produits par chacun des deux taureaux. Dans cette définition de la population, il est évident que le statisticien ne connaîtra jamais les vraies valeurs. Il en est souvent ainsi.

Pour répondre à sa question :

– Il émet une hypothèse : "Les deux taureaux ont même valeur laitière". (C'est l'hypothèse nulle, la différence entre les deux valeurs laitières étant nulle),

– Il mesure les valeurs laitières d'un certain nombre de fille de A et en déduit une estimation de la valeur de A. De même pour B.

– Il adopte une règle de décision :

• Si la différence entre les deux estimations est proche de zéro, il conserve son hypothèse.

• Si elle est très différente de zéro, il rejette son hypothèse initiale, ce qui revient à dire qu'un taureau est meilleur que l'autre.

La limite entre "proche de zéro" est "très différente de zéro" est le seuil de décision. Dans le raisonnement classique le problème est toujours posé en terme d'alternative : acceptation au refus de l'hypothèse. Il n'y a pas de zone d'indécision.

1.1. Risque de Première espèce

S'il rejette son hypothèse alors qu'elle est vraie, il commet une erreur. C'est l'erreur de première espèce.

Le risque de première espèce, α, est le risque qu'il a de commettre cette erreur avec la méthode de travail et la règle de décision qu'il a choisies.

On le calcule très couramment en agronomie. D'une façon très générale, on travaille au seuil de 5 %. Cela veut dire que si l'on répétait un très grand nombre de fois notre expérience, dans les mêmes conditions, et avec des échantillons issus d'une population qui vérifie notre hypothèse initiale, dans 5 % des cas on rejetterait cette hypothèse bien qu'elle soit vraie.

1.2. Risque de deuxième espèce

Quand notre statisticien conserve l'hypothèse nulle, si elle est vraie, il ne se trompe pas. Par contre, si elle est fausse, c'est alors qu'il commet l'erreur de deuxième espèce.

Le risque de deuxième espèce, β, est le risque, la probabilité, attaché à cette erreur. C'est donc le risque de conserver, c'est-à-dire de considérer comme vraie, une hypothèse qui en réalité est fausse.

On résume habituellement par le tableau 1, ci-après, ces deux définitions :

Tableau 1

		Réalité	
		H_0	H_1
Conclusion du statisticien	H_0	Pas d'erreur $(1 - \alpha)$	β_1
	$\neq H_0$	α	Pas d'erreur $(1 - \beta_1)$

On retiendra que c'est quand on réfute l'hypothèse nulle que l'on endosse α. C'est quand on la conserve qu'on endosse β.

Dans la comparaison de deux moyennes, les risques α et β sont illustrés par la figure 1 ci-après :

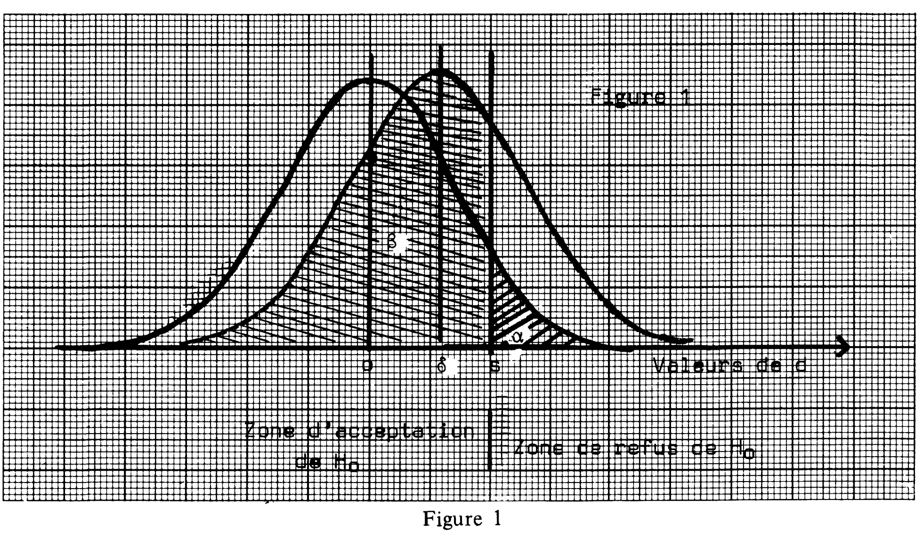

Figure 1

Soit d la différence observée entre les deux moyennes.

Soit δ la différence vraie.

Si l'hypothèse nulle Ho est vraie, d se répartir suivant la courbe noire.

S'il existe une différence vraie δ, d se répartit suivant la courbe rouge.

s est la valeur seuil de d au-delà de laquelle on réfute l'hypothèse nulle Ho.

α se mesure donc sous la roube noire (Ho vrai), au-delà du seuil, dans la zone de refus.

β se mesure donc sous la courbe rouge ($\delta = 0$), en deça du seuil, dans la zone d'acceptation.

(Remarque : Cette figure correspond à un test unilatéral, α étant d'un seul côté de la courbe noire ; elle correspond à la comparaison d'un témoin à un traité).

Sur cette figure, il est évident que pour une valeur de s (donc de α) fixée, β dépend de δ : β diminue quand δ augmente. On voit également que pour une valeur de δ fixée β augmente quand α diminue.

2 – COUT DE L'ERREUR DE PREMIERE ESPECE – ARBITRAIRE DU SEUIL DE 5 % – PROPOSITION D'UN AUTRE SEUIL

En agronomie, la technique de l'analyse de variance est très répandue et l'on teste toujours si α est supérieur ou inférieur à 5 %.

Ce calcul est tellement classique que l'on ne prend même plus la peine de formuler l'hypothèse que l'on teste. On se contente de mettre un S ou un astéristique en face des résultats significatifs (si l'on a travaillé en seuil de 1 % on met HS ou 2 astéristiques). Ainsi le lecteur non averti sait qu'on a fait un calcul statistique, mais lequel ?

Avec le développement des moyens de calcul plus rapide que celui des connaissances, on peut même se demander si dans bien des cas celui qui a fait le calcul (ou le fait faire) sait à quoi il correspond.

Bien sûr, tout le monde sait que le seuil de 5 % est arbitraire. Dans tous les stages, pour faire comprendre cet arbitraire, on dit que si la décision à prendre met en jeu des vies humaines le seuil de 5 % est intolérable ; mais, le stage passé, on continue (ou le programme continue) à calculer si la différence est significative ou non au seuil de 5 %.

Reprenons notre exemple des deux taureaux : j'endosse le risque de première espèce quand je réfute l'hypothèse nulle, donc quand je dis que B, par exemple, est meilleur que A. Et si par malchance la réalité était "B = A" ? Quelles conséquences pratiques cela aurait-il ? Aucune si le prix de l'insémination est le même pour A et B, ce qui est souvent le cas.

L'éleveur n'aurait pas eu le bénéfice qu'il escomptait, mais mon erreur ne lui aurait pas fait perdre d'argent.

Il en est de même quand on compare deux variétés qui ont le même prix. Dans ces cas-là, le risque de première espèce est sans conséquence pratique. Peu importe donc qu'il soit de 5 ou de 50 % !

Il semblerait donc, après un tel raisonnement, que le seul inconvénient de l'erreur de première espèce soit, dans ces cas-là, de jeter le discrédit sur la personne (physique ou morale) qui publie, et qu'un grand nombre de répétitions n'ait d'autre but que de soutenir une réputation. Il ne faut pas aller jusque-là, car il existe toujours le risque de conclure $B > A$ alors que $B < A$. C'est ce qu'on appelle parfois le risque γ. Lui n'est pas sans incidence économique ; nous en reparlerons donc dans la deuxième partie.

Dans la théorie classique on omet généralement ce risque, ou alors, pour justifier son silence, on démontre qu'il est infiniment petit (10^{-7}) pour des valeurs de α et β petites (5 %).

Le risque de première espèce ne sera cependant pas toujours sans conséquence économique : dans les exemples pris jusqu'à présent, les 2 modalités du facteur étudié induisaient chez l'utilisateur 2 dépenses, sinon égales, tout au moins de même ordre de grandeur.

Parfois on expérimente sur l'efficacité d'un traitement qui coûte très cher.

Le risque de première espèce, α, nous est-il toujours indifférent ? Quelle est la répercution économique de notre erreur ?

Le risque α est le risque de rejetter l'hypothèse nulle alors qu'elle est vraie. On l'endosse donc quand on rejette l'hypothèse nulle, c'est-à-dire quand on traite. Si on traite alors que l'hypothèse nulle est vraie, on engage une dépense, le coût du traitement, qui n'est assortie d'aucune recette supplémentaire.

Donc, le coût de l'erreur de première espèce, c'est le coût du traitement.

Notre méthode expérimentale étant telle que, si l'on applique un très grand nombre de fois sur une différence vraie nulle, dans α % des cas on traite pour rien, plutôt que de s'intéresser à α l'expérimentateur devrait s'intéresser au produit :

α x coût de l'erreur de première espèce = α x coût du traitement

Exemple :

le statisticien qui travaillerait au seuil de risque $\alpha = 20$ % sur un traitement qui coûte 40 F par ha, ne paraîtrait pas sérieux.

Pourtant, il endosserait un risque économique moindre que celui du statisticien (oh combien sérieux !) qui travaillerait au seuil de 2 % sur un traitement qui coûterait 500 F par hectare.

3 – L'HYPOTHESE NULLE ET LA REALITE – COUT DE L'ERREUR DE DEUXIEME ESPECE – INADAPTATION DU RAISONNEMENT CLASSIQUE A LA COMPARAISON D'UN "TEMOIN" ET D'UN "TRAITE"

Tous nos calculs ont été basés sur l'hypothèse nulle, mais pratiquement, en biologie, est-elle vraisemblable ?

Elle l'est probablement si l'on travaille sur oligo-éléments avec un témoin non carencé, mais si l'on travaille sur des niveaux énergétiques ou azotés, sur des dates de semis, etc. à fortiori dans un essai factoriel, elle est sûrement fausse !

Et l'on endosse le risque β à chaque fois que l'on conserve l'hypothèse nulle !

De là à dire que le risque β nous coûte toutes les expériences qui n'ont pas permis de mettre en évidence une différence significative, il n'y a qu'un pas.

Tenir un tel raisonnement, c'est admettre que toute différence entre les deux modalités du facteur que l'on étudie nous intéresse. C'est se placer en attitude de recherche. Nous étant placés en attitude pratique pour examiner le risque α, regardons également β avec les yeux du praticien.

Intéressons-nous à l'expérience très générale où l'on compare un lot traité à un lot témoin.

Supposons que nous ayons adopté la règle de décision suivante : "on traitera si d, différence observée, est supérieure à s, différence qui couvre juste tous les frais engendrés par le traitement".

• La figure 1 nous montre que si δ est inférieure à s, β est très grand (supérieur à 50 %). Notre méthode de travail et de décision semble donc très mauvaise sur la plan théorique puisque, pour toute valeur de d comprise entre 0 et s, on conserve l'hypothèse nulle ($\delta = 0$). Quand on conserve cette hypothèse nulle, la réalité est probablement différente puisque d est la meilleure estimation que l'on ait de δ.

Mais pratiquement, quand bien même aurait-on connu δ avec certitude, pour toute valeur de δ comprise entre 0 et s, on n'aurait pas effectué le traitement. En effet, si j'appelle R le résultat économique du traitement, (soit le supplément de recettes moins le supplément de dépenses), pour toute valeur de $\delta < s$, R est négatif par définition de s. Conserver Ho est donc la bonne décision sur le plan économique et la valeur prise par β nous est absolument indifférente.

• Par contre, quand $\delta > s$ et qu'on ne traite pas parce qu'on a observé $d < s$, alors β ne nous est pas indifférent car on commet l'erreur de ne pas effectuer un traitement qui serait bénéfique, puisque R est ici positif. Notre méthode est telle qu'on commet cette erreur dans β % des cas. Le coût de

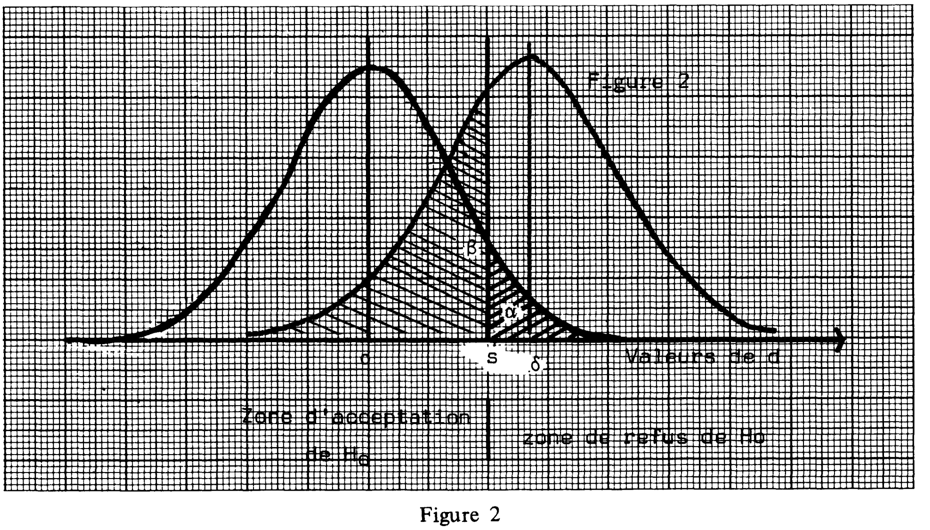

Figure 2

cette erreur est précisément égal à R. Il faut donc déterminer la taille de l'expérimentation pour que β R soit supportable.

• Enfin si on traite parce qu'on observe $d > s$ alors qu'en réalité $\delta < s$, on commet également une erreur sur le plan économique : le coût du traitement est supérieur à la recette qu'il engendre. Dans ce cas R est négatif. Cette erreur est d'autant plus grave que R, la perte engendrée par le traitement, est plus grande. Le coût de l'erreur est ici égal à la valeur absolue de R. Notre méthode est telle qu'on le commet dans $(1 - \beta)$ % des cas. Il faut donc également que $(1 - \beta)\,|R|$ soit supportable.

Par la suite, puisque nous avons appelé R le résultat économique du traitement, nous appellerons $|R|$ le coût de l'erreur, quelle qu'elle soit.

R est proportionnel à $(\delta - s)$. Ecrivons $R = k(\delta - s)$ où k apparaît comme un prix unitaire.

Insensiblement, mais inéluctablement, en nous interrogeant sur le coût de l'erreur de deuxième espèce nous avons changé de schéma de raisonnement. Peu nous importe de savoir si décision du statisticien et réalité sont égales ou différentes de l'hypothèse nulle. Ce qui compte c'est de savoir si d, différence observée sur laquelle est basée la décision du statisticien, et δ la réalité sont inférieures ou supérieures à s.

On remarque que dans ce raisonnement, beaucoup plus proche des problèmes concrets que le raisonnement du statisticien classique, α n'intervient pas. C'est seulement la valeur limite de $(1 - \beta)$ pour $\delta = 0$.

A la comparaison d'une hypothèse nulle à une hypothèse alternative simple je propose donc de substituer le schéma suivant :

Tableau 2		Réalité	
		$\delta \leqslant s$	$\delta > s$
Conclusion du statisticien	$d \leqslant s$ On ne traite pas	Bonne décision	Erreur dont le coût $\lvert R \rvert$ est un manque à gagner et dont la fréquence est β[1]
	$d > s$ On traite	Erreur dont le coût $\lvert R \rvert$ correspond à l'exécution d'un traitement non rentable. Sa fréquence est $(1 - \beta)$	Bonne décision

Dans ce mode de raisonnement on ne compare plus une hypothèse simple à une hypothèse alternative simple. On compare deux familles d'hypothèses. On est donc plus capable de fixer une valeur unique pour α et β, mais il est très facile de tracer les variations, en fonction de δ, du produit du risque par son coût.

----

(1) Remarque : En réalité, puisque nous avons abandonné l'hypothèse nulle à laquelle était attaché ce risque β, il serait logique de ne plus parler de β. Nous ne ferons le changement de symbole qu'au début de la deuxième partie. Nous appellerons alors ce risque γ car il s'agit tout simplement du risque d'inverser les conclusions tel qu'il a été présenté précédemment page 20.

La figure 5 (page 29) montre 2 telles courbes.

Leur tracé est très rapide avec une table de la loi normale. On a donc très vite fait, sur un problème concret, de connaître le risque maximum encouru.

4 – LES RISQUES ET LA DECISION ECONOMIQUE :

Comme précédemment, intéressons nous à l'expérience très générale et très simple où l'on compare un lot traité à un lot témoin.

Ayant calculé si la différence était significative ou non (après une hypothèse nulle jamais formulée) combien d'expérimentateurs ne sont-ils pas tentés de conclure simplement munis de α !

L'erreur la plus grossière, heureusement très rare, consiste à prôner le traitement dès que l'expérimentation a mis en évidence une différence nettement significative, sans regarder l'importance de cette différence ni si elle couvre les frais de traitement : on peut alors être amené à conseiller, avec une bonne sécurité, l'application d'un traitement qui a fait ses preuves techniquement mais qui n'est pas rentable.

– L'autre erreur, fréquente, consiste à oublier l'existence même du risque de deuxième espèce, β, le risque de manque de gagner.

Combien de fois n'ai-je pas entendu dire : "la différence n'est pas significative, donc on n'a pas le droit de diffuser" ; ceci veut dire, implicitement qu'on n'a pas le droit de traiter.

Ceux qui tiennent de tels raisonnements semblent oublier que, aussi petite que soit δ (différence vraie entre "témoin" et "traité"), pourvu qu'elle ne soit pas strictement nulle, on peut toujours la mettre en évidence d'une façon significative en augmentant le nombre de répétitions.

La décision économique de traiter ou de ne pas traiter dépend uniquement de δ et non pas du nombre de répétitions.

Est-ce dire que les risques sont totalement étrangers à toute décision économique ?

Non. puisqu'en remplaçant δ par son estimation on risque de commettre une erreur, soit celle de 1ère espèce, soit celle de 2ème espèce, que ces erreurs ont une incidence économique donc un coût, et que les risques en mesurent les probabilités.

A ce stade de nos réflexions, tiraillés entre le risque d'engager une dépense pour rien (ou pour une recette moindre) et celui d'un manque à gagner, il vient à l'esprit de prendre comme seuil de décision celui qui rend minimum la somme des pertes possibles. Cela revient à rechercher le minimum de $(1-\beta)|R| + \beta|R|$, soit le minimum de $|R|$.

Un simple coup d'œil sur la figure 5 nous montre que ce coût s'annule pour $\delta = s$, et seulement pour $\delta = s$. C'est du reste ainsi qu'a été défini s.

s, défini comme le seuil de rentabilité, apparaît donc ici comme un seuil de décision optimum.

Nous remarquerons que c'est celui qu'aurait choisi tout homme de bon sens n'ayant jamais fait de statistique. En effet, puisque d est la meilleure estimation que l'on ait de δ, et que cette estimation est dans biais, il est logique d'admettre, si $d > s$ que $\delta > s$ et donc qu'on a intérêt à traiter si l'on veut maximiser notre espérance de gain.

Nous avons admis que R est une fonction linéaire de δ. C'est vrai à chaque fois que le traitement va provoquer une augmentation de production qui sera vendue au même prix que les unités précédemment produites. C'est le cas très général, et notamment en agriculture puisque l'augmentation de production est faible par rapport à la production du témoin et que les producteurs sont nombreux.

Il doit exister des cas où les coûts des erreurs de première et de deuxième espèce ne sont pas identiques, mais, même dans ces cas là, pour toute variation continue de R en fonction de δ, en s, point d'équilibre, $R = 0$. Les phénomènes discontinus par contre doivent être étudiés différemment.

5 – L'ERREUR DE TROISIEME ESPECE

Plusieurs auteurs (Kimball, Calot ...) définissent l'erreur de troisième espèce comme le fait de résoudre parfaitement un problème différent de celui qu'il aurait fallu résoudre.

Chacun connaît l'histoire du mathématicien à qui l'on a posé la question : "De ces deux pendules, l'une qui est arrêtée, l'autre qui retarde de une minute par jour, laquelle marque le plus souvent l'heure exacte ?".

La réponse du mathématicien fut : "c'est la pendule arrêtée qui marque le plus souvent l'heure exacte". C'était la réponse correcte au problème posé. Mais le consultant avait-il correctement formulé sa question ?

Pour étudier les circonstances qui favorisent l'apparition de cette erreur, et donc chercher à l'éviter, nous pourrions utilement nous reporter à l'article de A.N. Kimball : "ERREURS DE TROISIEME ESPECE, COOPERATION ENTRE STATISTICIENS – CONSEILS ET CHERCHEURS" : L'essentiel de son message est que les erreurs de 3ème espèce ont pour cause une liaison insuffisante, ou de mauvaise qualité, entre le statisticien-conseil et le chercheur.

Il est difficile évidemment de chiffrer le coût de cette erreur et donc de déterminer le montant des moyens à mettre en œuvre pour l'éviter. Le fait même de dire qu'il y a erreur de troisième espèce est souvent subjectif. Un même problème peut toujours être posé de 2 façons différentes par un chercheur et par un praticien. Chacun peut donc accuser l'autre d'erreur de troisième espèce.

Ce qu'il faut, à mon avis, c'est être suffisamment sensibilisé à son existence pour se poser, à chaque nouveau problème, la question de savoir si l'on ne commet pas l'erreur de troisième espèce. On doit surtout y penser quand on applique un schéma tout fait. Par exemple, en calculant son risque de première espèce, le statisticien ne commet-il pas l'erreur de troisième espèce ?

6 – L'ERREUR DE QUATRIEME ESPECE

Le statisticien qui a résolu le bon problème a-t-il fini ?

Non. Il lui faudra faire part de ses conclusions à l'utilisateur, et le risque d'extrapolation abusive, d'application à un problème apparemment voisin, mais différent, est grand.

Ce risque n'est pas du tout propre au statisticien mais, comme il manie une discipline hermétique au français moyen il y est particulièrement exposé. Si l'erreur est commise, il n'est pas juste de lui en attribuer l'entière responsabilité, mais il en a souvent une part.

On y remédie en précisant clairement le problème auquel on a apporté une réponse et en précisant les conditions d'utilisation de la réponse.

Exemple 1 :

Le rendement de telle plante augmente de tant à chaque fois qu'on apporte 1 unité d'azote. Ceci jusqu'à quelles limites ?

Les risques de mauvais usage d'une régression en dehors des limites entre lesquelles elle a été calculée sont bien connus.

Plus subtile est l'erreur que l'on commet parfois en employant une formule de régression dans un intervalle de variation plus restreint que celui pour laquelle elle a été calculée.

Exemple 2 :

Un chercheur en travaillant sur des maîs dont le stade végétatif varie de quelques jours après la floraison femelle jusqu'à la maturation, trouve entre la teneur en matière sèche de la plante entière et celle de l'épi un coefficient de corrélation de 0,990.

Isolé de ses conditions d'obtention, un tel coefficient ne peut qu'inciter le praticien à supprimer les analyses de plantes entières pour ne faire que des analyses d'épi plus simples et moins coûteuses.

Le drame est que dans les conditions pratiques on récolte les maîs à des stades végétatifs variant du stade laiteux-pâteux au stade vitreux. Dans ces conditions la liaison est beaucoup plus faible et surtout plus variable. Elle est inutilisable.

7 – CONCLUSION DE LA PREMIERE PARTIE

Le remède évident contre les erreurs de troisième et de quatrième espèce est d'améliorer les relations entre statisticiens et utilisateur. Un minimum de formation statistique chez les utilisateurs était indispensable pour que le dialogue puisse avoir lieu. Ce fut l'objet des stages effectués ; mais ces stages, où l'utilisateur était en position trop scolaire, ne suffisaient pas. C'est pourquoi ce colloque fut organisé. Puisse-t-il n'être que le début d'un fructueux dialogue[1].

---

(1) Cette première partie fut exposée au cours d'un colloque entre utilisateurs ayant suivi un cycle statistique et formateurs.

DEUXIEME PARTIE
DETERMINATION DU NOMBRE DE REPETITIONS

1 – QUE FAIT-ON ACTUELLEMENT ?

Il arrive parfois que le nombre de répétitions soit le fruit d'une coutume dont on ignore les fondements. Cette méthode "pifométrique" ne serait pas mauvaise si le "pifomètre" était étalonné correctement et régulièrement. Malheureusement il arrive qu'on ne se soucie pas de savoir quand et comment il a été étalonné.

Parfois ce nombre de répétitions est sous la dépendance exclusive d'une contrainte, par exemple le nombre d'animaux dont on dispose ou le nombre de places dans l'étable ; plus généralement la contrainte est celle d'un budget fixé à priori. Cette situation est souvent catastrophique : des frais sont engagés alors qu'un calcul sommaire aurait montré qu'ils l'étaient en pure perte, car on était à peu près sûr au départ de ne rien pouvoir conclure.

Enfin, et de plus en plus souvent compte tenu du développement considérable des connaissances statistiques depuis quelques années, on fait "le" calcul : On se fixe une différence d à mettre en évidence (ou une précision, ce qui revient au même) et des valeurs de α et β, on se renseigne sur la variance attendue, et on détermine le nombre de répétitions n en fonction de ces 4 paramètres. Parfois le résultat nous effraie ; alors on augmente un peu β et d pour trouver un compromis entre la première valeur trouvée et ce qu'on avait l'habitude de faire.

Avec cette méthode on risque fort, si l'on y prend garde, de n'introduire aucun élément économique alors que la décision à prendre est essentiellement économique. En effet une expérimentation est un investissement et le nombre de répétitions en détermine le montant.

Seul le seuil de rentabilité du traitement est parfois introduit dans le calcul par le biais de d ; mais les seuils de risques α et β sont en général arbitraires et étrangers à toute considération économique. Le résultat l' est donc également.

On répond correctement à la question "combien faut-il de répétitions pour mettre en évidence telle différence, sur telle variable, avec tels risques ?".

C'est le problème que le statisticien sait résoudre ; mais est-ce là le vrai problème ?

2 – QUEL PROBLEME INTERESSE LE GESTIONNAIRE ?

Vis à vis d'une technique nouvelle, d'un traitement nouveau (au sens le plus large du terme) son seul problème est de savoir s'il faut l'appliquer ou non.

L'expérimentation est parfois le seul moyen qu'il ait de prendre une décision rationnelle. Dans ce cas il veut savoir combien il faut investir dans l'expérimentation en question.

En effet, le coût de l'expérimentation représente alors le coût de sa décision et il est aussi mauvais :

– de payer très cher une "bonne" décision si elle est sans grande conséquence économique.

– que de trop économiser sur le coût de la décision si elle est lourde de conséquences.

Entre ces 2 extrêmes il doit exister un optimum. C'est ce que nous nous proposons de déterminer.

Remarque : Ce n'est pas toujours une seule et même personne qui finance l'expérimentation et qui applique les résultats. Par exemple, l'expérimentation peut être conduite dans le cadre de l'organisation professionnelle alors que c'est chaque entrepreneur individuellement qui applique le résultat. C'est mon cas (secteur agricole). L'optimum recherché l'est alors au niveau de la collectivité. Ce faisant je me place exclusivement en gestionnaire d'un budget collectif.

3 – COUT DE LA PRISE DE DECISION

C'est le budget de l'expérimentation. C'est une fonction croissante du nombre de répétitions n. Dans toute expérience il y a des frais fixes et des frais proportionnels au nombre de répétitions. On peut facilement et sans trop d'arbitraire les distinguer et donc exprimer ce coût par une fonction linéaire de n.

Coût de l'investissement : $I = a + b\, n$.

n = nombre de répétitions

a et b sont 2 constantes calculées à partir d'un budget prévisionnel dans lequel on a distingué des charges fixes et des charges proportionnelles à n, ou encore calculées à partir de 2 budgets établis pour 2 valeurs de n.

Le coût de notre décision est à comparer aux conséquences économiques d'une mauvaise décision, d'une erreur.

4 – CONSEQUENCES ECONOMIQUES D'UNE MAUVAISE DECISION

Comme nous l'avons vu dans la première partie, pour chiffrer les conséquences économiques d'une erreur, il faut 2 éléments :

– le coût de cette erreur si elle se produit,

– sa probabilité d'apparition.

C'est au produit de ces 2 éléments que nous nous intéressons. De la même façon un assureur calcule sa prime en multipliant le coût de l'accident assuré par sa probabilité d'apparition (et en ajoutant les frais généraux).

4.1. Coût de l'erreur si elle se produit

Soit R le résultat économique du traitement. R est égal à la différence de rendement entre le traité et le témoin, que multiplie le prix unitaire de cette différence, moins le coût du traitement. R est donc une fonction linéaire de δ, différence vraie entre le témoin et le traité.

Si l'on appelle c le coût du traitement et s son seuil de rentabilité, (valeur de δ pour laquelle la recette supplémentaire engendrée par le traitement est égale à son coût) :

R peut s'écrire : $$R = c\left(\frac{\delta}{s} - 1\right)$$

Si $\delta > s$ l'erreur consiste à ne pas traiter ; il y a un manque à gagner = R.

Si $\delta < s$ l'erreur consiste à traiter alors que R est négatif.

Dans tous les cas le coût de l'erreur, si elle se produit, est $|R|$

4.2. Probabilité de l'erreur

Si δ est supérieur à s, l'erreur consiste à ne pas traiter. Si l'on a pris s comme seuil de décision, on commet cette erreur dans γ % des cas. *(Voir tableau et remarque page 22).*

Si δ est inférieur à s, l'erreur consiste à traiter. Avec la même règle de décision, on la commet dans $(1 - \gamma)$ % des cas.

γ est fonction de s, δ, n et σ.

n est le nombre de répétitions.

σ est l'écart-type de la population.

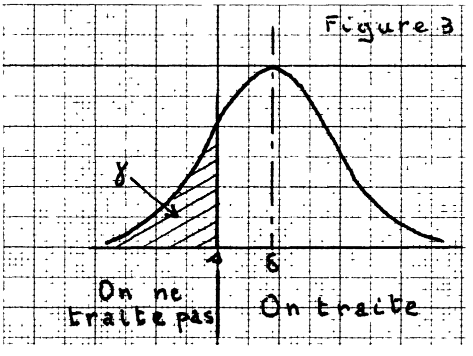

Figure 3

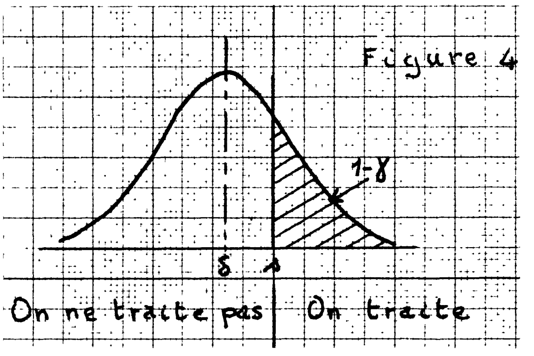

Figure 4

4.3. Coût "moyen" de l'erreur

C'est : – soit le produit $\gamma\,|R|$

– soit le produit $(1 - \gamma)\,|R|$

C'est donc une fonction de s, δ, n, σ et c

Car γ est fonction de s, δ, n et σ

et $|R|$ est fonction de δ, c et s.

Que cette complexité apparente (puisqu'il y a 5 paramètres) ne nous rebute pas, le problème se simplifie :

On peut calculer les variations de $\gamma\,|R|$ et de $(1 - \gamma)\,|R|$ en fonction de δ à condition d'exprimer δ en écart-type de la différence des moyennes du témoin et du traité. En effet, on tient ainsi compte de σ et de n. Quant à c et s, ce ne sont que des points portés par les axes $\gamma\,|R|$ et δ.

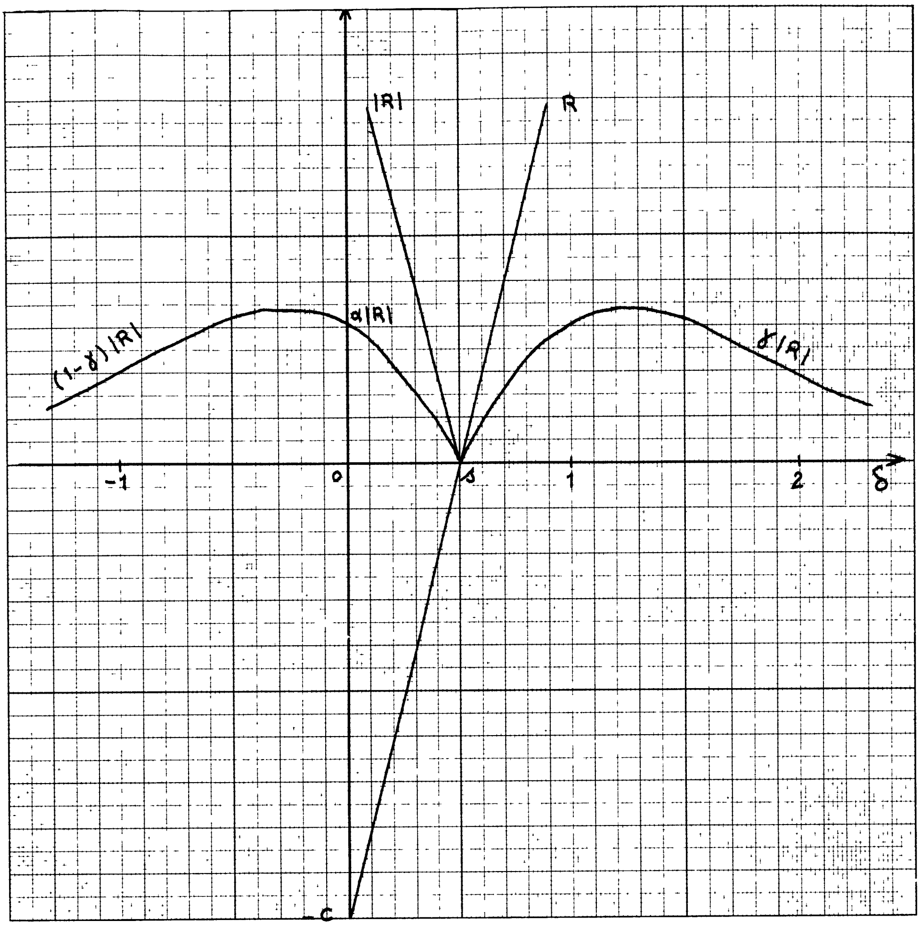

Figure 5 – Variations du coût moyen de l'erreur, $\gamma\,|R|$ ou $(1 - \gamma)\,|R|$ en fonction de δ

La figure 5 ci-dessus montre les variations de $\gamma|R|$ et de $(1 - \gamma)\,|R|$ en fonction de δ. Pour la construire nous avons supposé (comme toujours !) que d était distribué normalement.

On remarque que :

– ces 2 courbes s'annulent pour $\delta = s$ puisqu'alors $R = 0$.

– Les coûts moyens des erreurs croissent quand δ s'éloigne de s, passent par un maximum unique puis décroissent pour tendre vers zéro quand δ est très différent de s (car alors γ, ou $(1 - \gamma)$, tendent vers zéro. Ceci veut dire tout simplement que si la différence entre le témoin et le traité est très grande, on a peu de chance de se tromper).

– Les courbes $\gamma\,|R|$ et $(1 - \gamma)\,|R|$ sont symétriques par rapport à la verticale $\delta = s$: le maximum du coût moyen de l'erreur est donc unique quelque soit la nature de l'erreur que l'on commet (ne pas traiter quand il le faudrait ou traiter quand ce n'est pas rentable).

4.4. Coût maximum de l'erreur

Comme δ est inconnu il faut s'en affranchir. (Après l'expérience on en connaîtra une estimation d ; au moment de déterminer le nombre de répétition, donc avant l'expérience, on ne connaît même pas son estimation d).

Pour s'affranchir de δ je propose de considérer le maximum de $\gamma\,|R|$ par rapport à δ.

Remplacer, résumer, toutes les variations de $\gamma\,|R|$ par le maximum de la courbe semble assez logique puisque nous le faisons pour déterminer le nombre de répétitions, donc, en quelque sorte pour nous assurer contre ce risque.

Le résumé, auquel nous sommes contraints, même s'il est critiquable, me semble meilleur que celui que l'on faisait dans la théorie classique en considérant le risque de première espèce. En effet, $\alpha\,|R|$ (égale α c) n'est qu'un point particulier de cette courbe ; c'est sa valeur pour $\delta = 0$. Le maximum a lieu pour une valeur variable de δ ; $\alpha\,|R|$ constitue donc un bien mauvais résumé de cette courbe.

Par contre, le maximum correspond à une valeur fixe de $\gamma = 0{,}2239$.

5 – VARIATIONS DU COUT MAXIMUM DE L'ERREUR EN FONCTION DE n

Nous avons vu au paragraphe 43 que $\gamma\,|R|$ était fonction de 5 variables ou paramètres. Nous y avons vu également qu'on pouvait s'affranchir de l'écart type en le prenant comme unité pour δ et donc pour s. Nous venons de nous affranchir de δ inconnu en considérant le maximum de $\gamma\,|R|$ par rapport à δ. On peut également éliminer c en choisissant c comme unité de R.

Ainsi le maximum de $\gamma \frac{|R|}{c}$ ne dépend plus que de s et de n. On peut donc fixer s et tracer dans un plan les variations du maximum de $\gamma \frac{|R|}{c}$ en fonction de n. On peut le faire pour différentes valeurs de s.

Le résultat est l'abaque ci-contre (figure 6).

6 – PRINCIPE D'UTILISATION DE L'ABAQUE

Dans ce même plan (maximum de $\gamma \frac{|R|}{c}$, n) il est facile de tracer la droite de coût de l'expérimentation. Il faut évidemment prendre également σ comme unité.

Il faut aussi diviser l'investissement par le champ d'application des résultats. En effet, R sera un résultat économique par unité de production, par exemple, par hectare ou par animal etc. Si on fait une expérimentation, ce n'est pas pour appliquer le résultat à 1 ha ou 1 animal, mais pour l'appliquer à u ha ou u animaux. C'est u que j'appelle le champ d'application des résultats de l'essai.

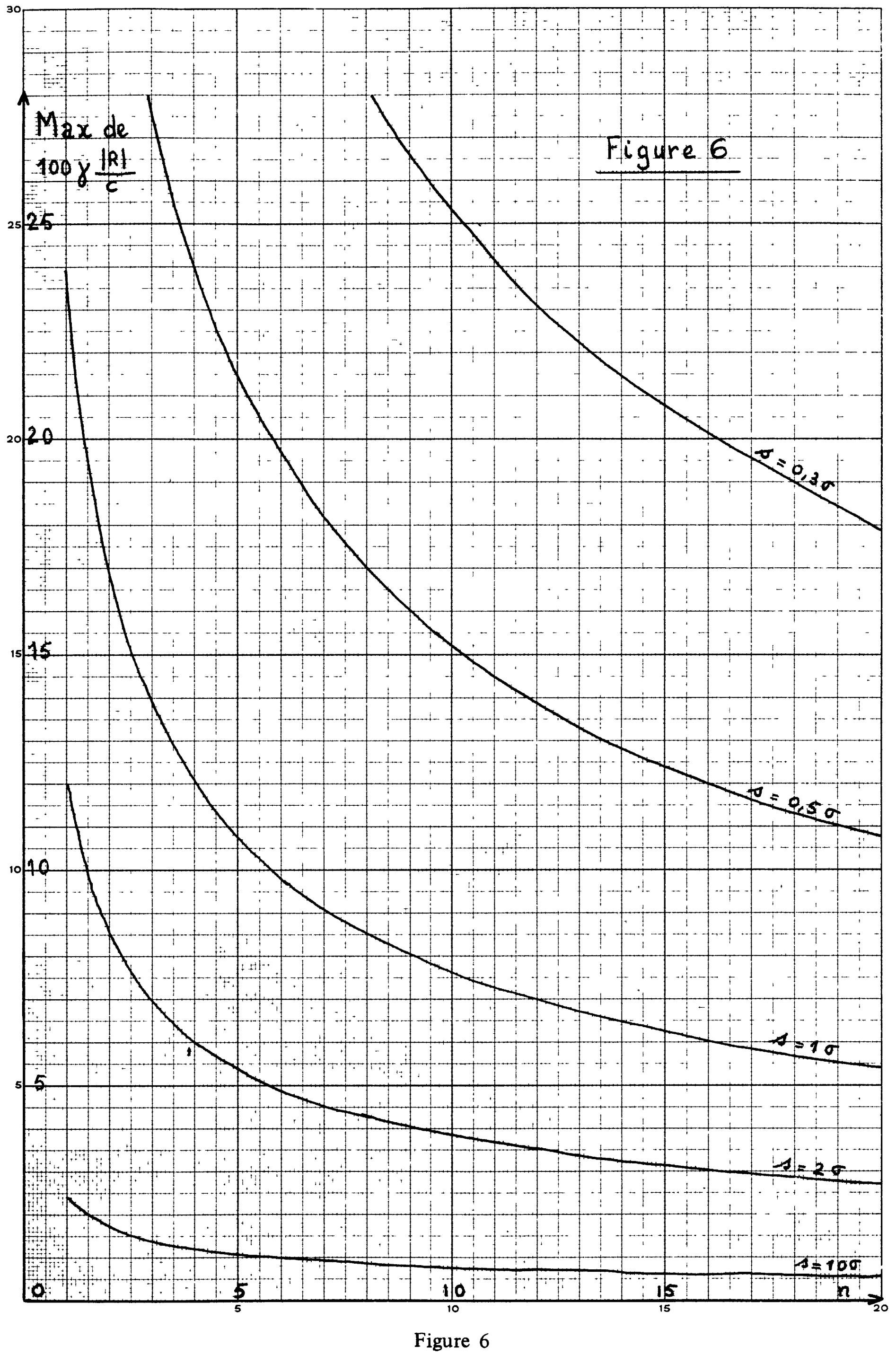

Figure 6

(Pour certains la fixation plus ou moins arbitraire du champ d'application de la technique expérimentée sera un inconvénient de la méthode.

C'est une difficulté mais non pas un inconvénient. Quel est, en effet, le moindre mal ?

– le fixer d'une façon souvent contestable car cela constitue un pari sur l'avenir.

– ou bien déterminer la taille de l'expérimentation par la méthode classique avec lequelle on risque fort de ne même pas se douter qu'il faudrait prendre en considération ce facteur ?).

Ceci fait pour se ramener aux unités de l'abaque, son principe d'utilisation peut se comprendre graphiquement ou par l'algèbre :

Graphiquement, constatons que :

Toutes les courbes de l'abaque 6 sont à concavité vers le haut, et, quand n croît, la pente de chaque courbe diminue en valeur absolue.

Pour une courbe donnée (c'est-à-dire une valeur de s donnée) il existe un point pour lequel la pente est égale en valeur absolue à celle de la droite d'investissement mais de signe opposé. L'abscisse de ce point est le nombre de répétitions optimum.

En effet, en deçà de ce nombre, quand n augmente, le risque diminue plus vite que l'investissement croît : dans cette zone on a donc intérêt à augmenter n, le nombre de répétitions.

Au-delà, le risque diminue moins vite que l'investissement augmente : on n'a donc plus intérêt à augmenter n.

Par l'algèbre : Considérons un coût total (Ct) qui serait la somme d'un coût d'investissement (Ci) et d'un coût maximum d'erreur (Ce). Ce coût total sera minimum quand sa dérivée s'annulera (le raisonnement graphique nous a montré qu'il s'agit d'un minimum et non d'un maximum).

Soit pour $C't = C'i + C'e = 0$

En toute rigueur nous sommes contraints d'opérer graphiquement car, dans l'expression du coût de l'erreur, figure δ inconnu. Rappelons que nous n'avons pu nous en affranchir qu'en considérant le maximum de $\gamma\ |R|$ en fonction de δ.

En réalité, j'ai mis au point ultérieurement une formule algébrique empirique en ajustant les variations du Max de $\gamma \frac{|R|}{c}$ à une famille d'hyperboles.

7 – UTILISATION PRATIQUE

Il n'est pas utile de calculer les frais fixes de l'expérimentation qui n'entrent pas en ligne de compte dans la dérivée par rapport à n.

On pourra donc procéder de la façon suivante :

1) Déterminer sur quelle courbe de l'abaque il faudra travailler. Elle dépend de la valeur de s.

2) Calculer les frais proportionnels au nombre de répétitions (en % de c et divisés par u).

3) Tracer la droite de pente opposée à cette valeur des frais proportionnels.

4) La translater jusqu'à ce qu'elle tangente la courbe adéquate.

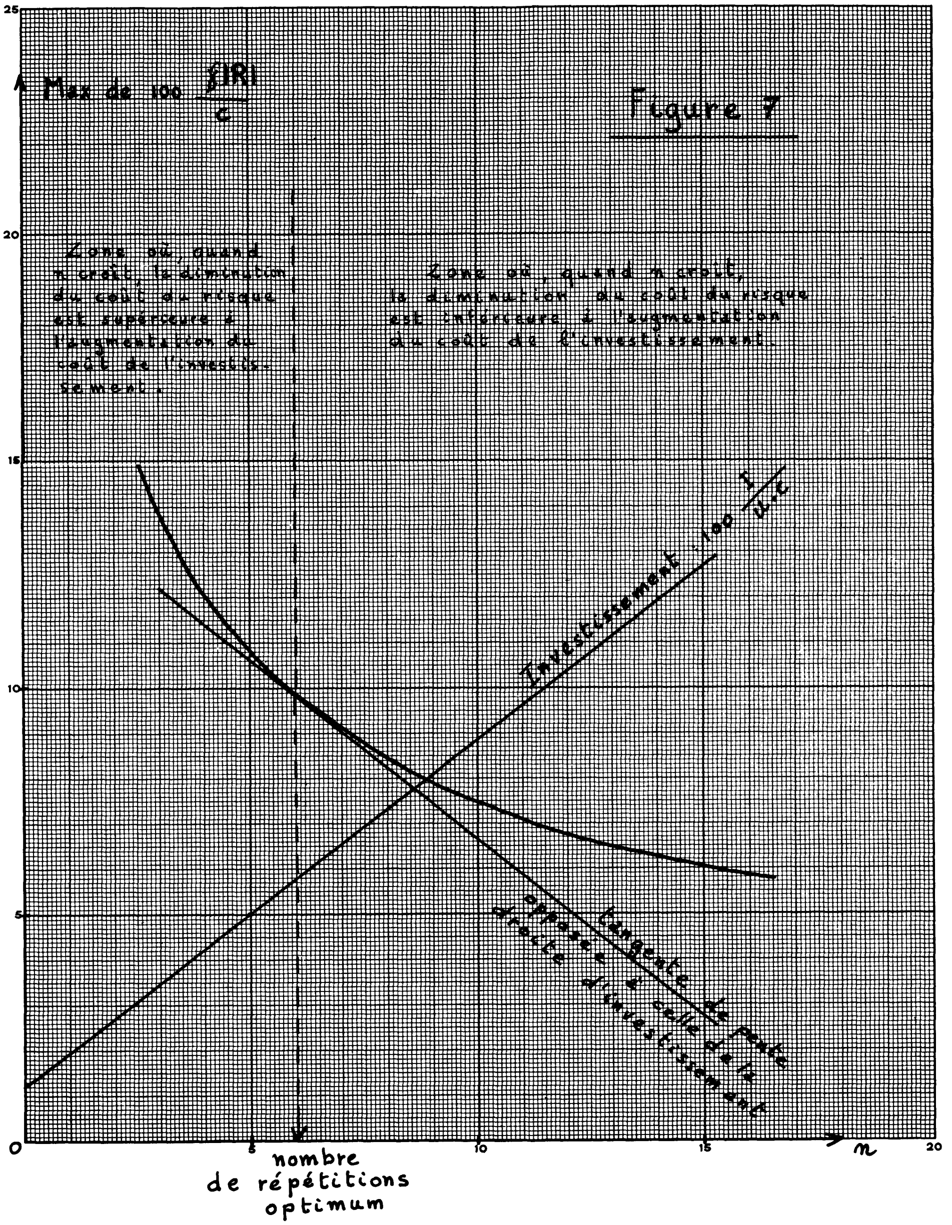
Max de 100 Σ|R|/c
Figure 7
Zone où, quand n croît, la diminution du coût du risque est supérieure à l'augmentation du coût de l'investissement.
Zone où, quand n croît, la diminution du coût du risque est inférieure à l'augmentation du coût de l'investissement.
Investissement . 100 I/n.c
tangente de pente opposée à celle de la droite d'investissement
nombre de répétitions optimum
n
0
5
10
15
20
25

Figure 7

Remarques :

1) Ce calcul prend en compte le coût du risque maximum ; on aura donc intérêt à faire un arrondi par défaut.

2) Dans les exemples ci-après nous avons cependant tracé la droite d'investissement, car c'est très rapide et ça peut faciliter la compréhension.

8 – CRITIQUE

Cette utilisation par tangente est peu précise, mais on a pas besoin d'une grande précision dans ce type de problème compte tenu du fait que les valeurs données à certains paramètres (u notamment, parfois σ) constituent un pari sur l'avenir.

L'incertitude sur u est la plus grave. Celle sur σ ne l'est que dans la mesure où l'on a jamais expérimenté sur ce type de matériel ou dans ces conditions. Les exemples ci-après donnent une idée de l'influence d'une erreur d'estimation sur ces paramètres.

On peut aussi critiquer le fait que s soit considéré comme une variable certaine. Cette critique me semble mineure car seul le rapport $\frac{R}{c}$, donc le rapport de prix entre unités supplémentaires produites et coût du traitement, entre en ligne de compte. La méthode n'est donc pas sensible à une inflation homogène. Elle ne l'est qu'à une évolution des rapports de prix.

9 – METHODE ALGEBRIQUE

Une très bonne approximation de cette famille de courbes est obtenue par la formule empirique :

$$\text{Max de } \gamma \frac{|R|}{c} = \frac{\sigma}{s} \cdot \frac{54}{(n+2)} + 3 \,.\,[1]$$

La pente de cette famille de courbes est : $\dfrac{-54\,\sigma}{s(n+2)^2}$

On peut donc résoudre le problème algébriquement en égalant l'opposé de cette pente à celle de la droite d'investissement. Seule la racine positive a évidemment un sens, puisque c'est la partie positive de l'hyperbole qui nous a servi de modèle d'ajustement.

$$n = -2 + \sqrt{\frac{0.54\,c}{\frac{s}{\sigma}\;\frac{I'}{u}}}$$

----

(1) Cette approximation est excellente pour n compris entre 3 et 20, donc dans la zone couramment utilisée. L'écart entre les valeurs réelles et les valeurs ajustées est toujours inférieur à 2 % et la précision est donc meilleure que celle que l'on obtient en déterminant un point de tangence. Au-delà de 20 ou 25, la formule provoque une surestimation du nombre de répétitions d'autant plus accentuée que n croit. Il en est de même en-dessous de 3, ce qui est sans conséquence pratique.

Sous le radical nous trouvons 5 paramètres ; 3 d'entre eux sont toujours positifs ; ce sont : I' l'investissement supplémentaire par répétition (dérivée de I par rapport à n), u le champ d'application des résultats, et σ l'écart type. c, le supplément de coût du ' traité" par rapport au "témoin" et s, seuil de rentabilité, sont toujours du même signe. On pose généralement le problème de façon à ce que c soit positif. A la limite c peut être nul ; dans ce cas s l'est aussi et il y a indétermination.

10 – EXEMPLE 1

Dans un essai on compare différents peuplements sur betteraves fourragères. Il n'y a pas à proprement parler un témoin et un traité. Cependant on peut considérer que, par rapport au peuplement habituellement utilisé, on veuille tester un peuplement plus important.

Le coût du traitement, c, correspond à la dépense supplémentaire en semences. On l'a estimé dans ce cas à 90 F/ha.

Le seuil de rentabilité, s, est dans ce cas de 1280 kg de betteraves. (1280 kg x 0,07 F/kg = 90 F).

L'écart-type, σ, habituellement rencontré dans les essais sur betteraves dans ces conditions est de 1150 kg.

Donc s = 1,1 σ. Par la suite pour tomber sur une courbe connue de l'abaque, nous considérons s = σ.

(Ceci revient à estimer les betteraves un peu plus chères (7,8 c/kg) ou a considéré une valeur de l'écart-type attendu un peu plus pessimiste que celle enregistrée les années précédentes).

La fonction de coût d'un tel essai, a été estimée à 1600 F. + 1000 F. par répétition. Nous ne détaillerons pas l'analyse du budget de l'essai qui a abouti à cette formule, car elle ne présente ni difficulté ni intérêt.

Avec ces renseignements, et une (ou plusieurs) hypothèse sur ce champ d'application on peut facilement par 2 points, c'est-à-dire 2 valeurs de n, tracer la droite d'investissement. Nous l'avons fait pour 2 hypothèses sur la valeur de u : 500 et 1000 ha.

n	I = Investissement total		$\frac{100\,I}{c \times u}$ = Investissement par unité de production	
	en francs	% de c. (*) = $\frac{100\,I}{c}$	u = 500 ha	u = 1000 ha
1	2 600	2 890	5,8	2,9
5	6 600	7 330	14,7	7,3

(*) Pour respecter l'unité de l'abaque de la figure 6.

Après avoir tracé ces droites, on trace très facilement à l'aide d'un rapporteur une droite de pente opposée. La translation jusqu'à ce qu'elle tangente la courbe se fait très facilement avec une simple règle.

On peut aussi calculer directement la pente de la tangente.

en francs par répétition $= - 1000$

en % de c $= - \dfrac{1000 \times 100}{90} = - 1111$

en % de c par unité de production :

pour u $= 500$ ha $= - \dfrac{1111}{500} = - 2{,}22$

pour u $= 1000$ ha $= - \dfrac{1111}{1000} = - 1{,}11$

Le graphique 8 ci-contre, où les droites correspondant à cet exemple sont tracées, nous indique que pour u = 1000 ha le nombre de répétition optimum est de 5 ; pour u = 500 ha il n'est que de 3. Ces résultats montrent toute l'importance de la dimension du champ d'application, facteur habituellement négligé.

Par l'algèbre pour u = 500 ha, $n = -2 + \sqrt{\dfrac{0{,}54 \times 90}{1 \times \dfrac{1000}{500}}} \# 3$. On aurait pu ne pas faire l'approximation $\dfrac{s}{\sigma} = \dfrac{1280}{1150} \# 1$.

11 – EXEMPLE 2

Dans un autre essai on compare la croissance d'élèves bovins de 18-24 mois passant, en Normandie, l'hiver en plein air intégral (lot témoin) à celle d'élèves disposant dans le pré d'un abri sommaire ; on s'interroge donc sur la rentabilité d'un tel abri.

Le coût du traitement, c, c'est-à-dire l'amortissement de l'abri et les frais financiers, est estimé à 30 F. par animal et par hiver. (4000 F pour 18 animaux environ amorti sur 10 ans (prix 1969).

Le kg de croît vif étant estimé à 5 F. environ, le seuil de rentabilité, s, est de 6 kg de croît par animal et par hiver.

L'écart-type arrondi, σ, est de 30 kg (mesuré sur le gain total d'hiver et de printemps pour tenir compte de l'effet de la croissance compensatrice).

Donc $s = 0{,}2\,\sigma$

La fonction de coût d'un tel essai est estimée à 2 200 F + 950 F par animal, soit 2 200 F + 1 900 F par répétition. (ont été comptés dans les frais fixes les frais de surveillance des animaux : salaires et déplacements. Ont été comptés dans les frais proportionnels : les frais d'alimentation, de transport des animaux, d'assurance, vétérinaire, financiers et de pesée. A été déduite des frais proportionnels la différence "prix de vente – prix d'achat").

La dimension du champ d'application fut estimée de la façon suivante ;

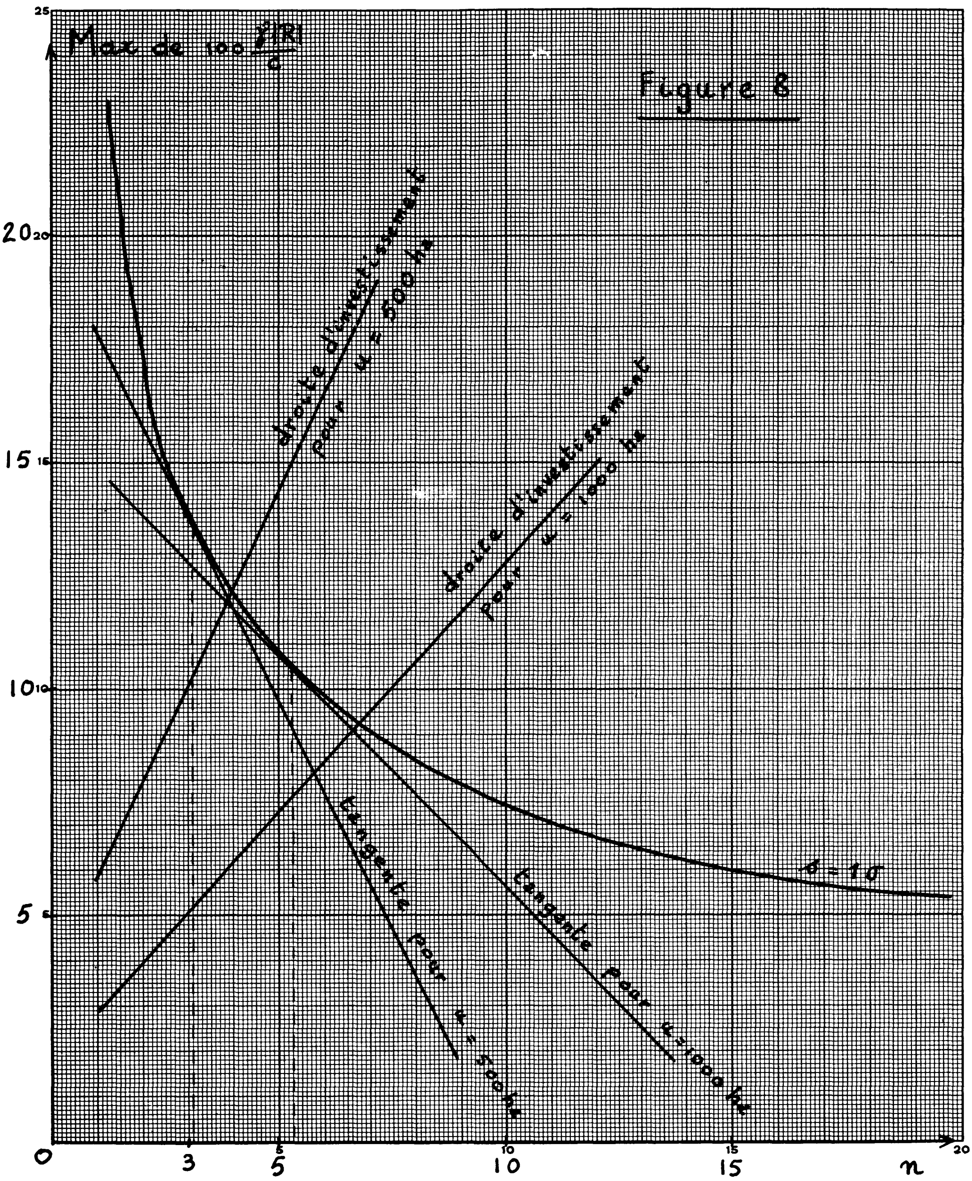
Max de 100
Figure 8
droite d'investissement
pour u = 500 ha
droite d'investissement
pour u = 1000 ha
tangente pour u = 500 ha
tangente pour u = 1000 ha
0
3
5
10
15
20
25
n

Figure 8

"L'E.D.E. qui a réalisé l'expérimentation opère dans un département (l'Eure) où il y a environ 100000 élèves de cette génération là. La moitié sont actuellement sans abri, soit environ 50000. 10 à 20 % des éleveurs sont touchés par les organismes de développement, soit très approximativement les propriétaires de 6 000 élèves sans abri.

Pour différentes raisons, tous les éleveurs touchés par la vulgarisation ne peuvent pas appliquer rapidement les conclusions de l'essai. Ultérieurement les conditions d'exploitation et la conjoncture économique peuvent changer. Pour les 2/3 des éleveurs touchés qui appliqueraient les conclusions cela ferait 4000 élèves. Calculons, par 2 points la droite d'investissement correspondant à cette hypothèse :

n	Investissement total		Investissement par unité de production pour u = 4000 élèves
	en francs	en % de c $= \frac{100\,I}{c}$	$= \frac{100\,I}{c \times u}$
5	11700	39000	9,75
30	59200	197300	49,3

Reportons la sur l'abaque pour s = 0,2 σ et traçons la tangente de pente opposée (*En réalité la courbe pour s = 0,2 σ est hors de l'abaque représentée dans la figure 6. Cet exemple pris dans le domaine des productions animales procure donc un complément de l'abaque 6).*

La figure 9 ci-contre, montre que le nombre de répétitions optimum est de 12. Ce chiffre serait habituellement considéré comme plutôt faible par une méthode plus au moins empirique mais qui de toutes façons ne tiendrait compte ni du coût du traitement c, ni du coût de l'expérimentation I (sur lequel on se contente de se lamenter) ni du champ d'application des résultats (sur lequel on ne risque pas la moindre hypothèse).

Bien entendu cet optimum dépend de u. Pour u = 6000 élèves, il passe entre 14 et 15 soit 14.

Une bonne connaissance des animaux et du problème permettrait d'effectuer une mise en couples judicieuse et diminuerait d'autant l'écart-type résiduel. On opérerait ainsi sur une courbe plus "basse" de l'abaque. Par exemple, si dans ce cas on avait pu descendre jusqu'à σ = 20 kg (hypothèse optimiste), on aurait lu le nombre de répétitions optimum sur la courbe s = 0,3 σ (car s = 6 kg) ce qui aurait donné 8 répétitions, comme le montre la figure 9 ci-contre.

Par l'algèbre, dans cette dernière hypothèse, on trouve :

$$n = -2 + \sqrt{\frac{0{,}54 \times 30}{\frac{6}{20} \times \frac{1900}{4000}}} = 8{,}7$$

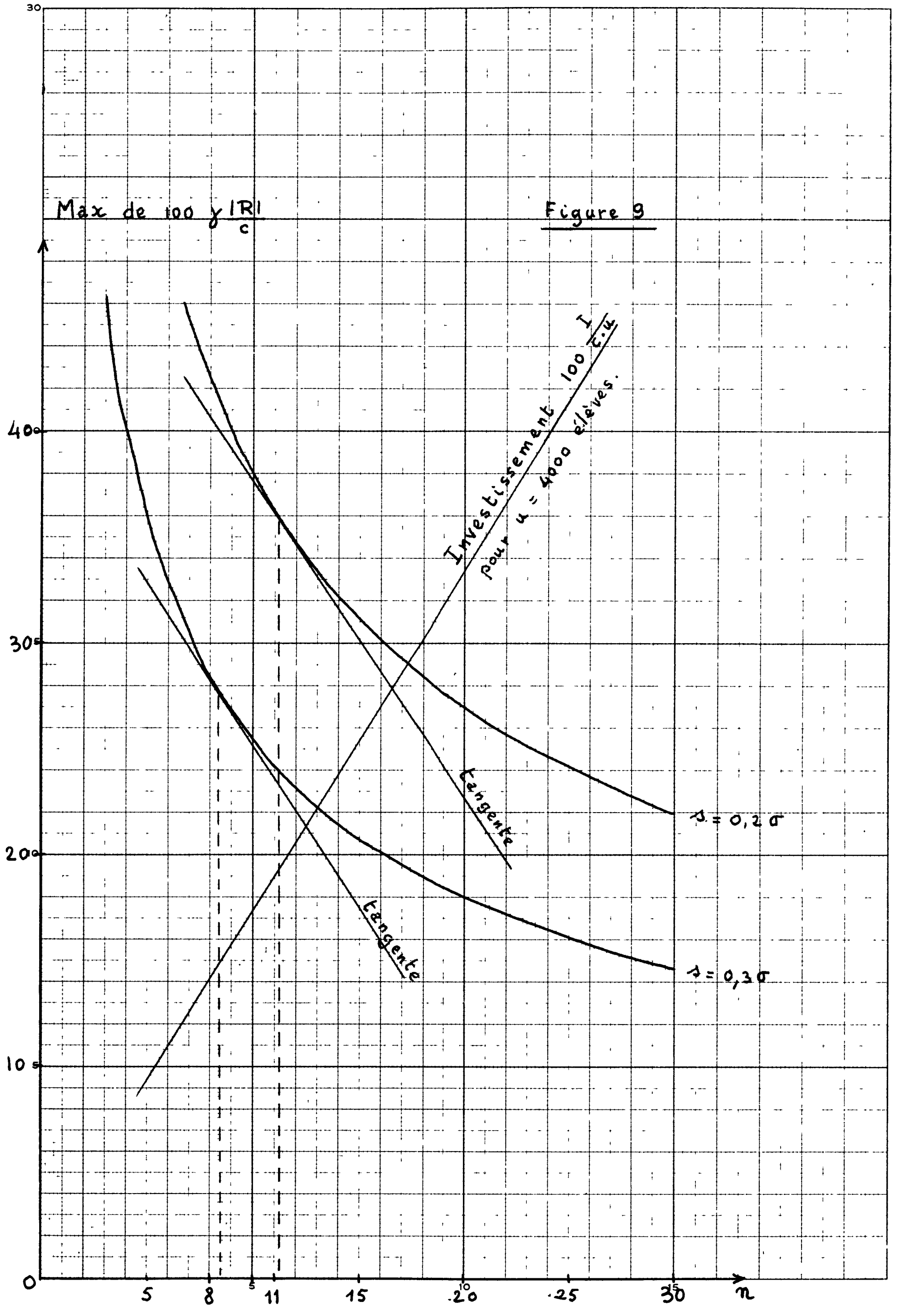
Max de 100 γ |R|/c
Figure 9
Investissement 100 I/c.u
pour u = 4000 élèves.
tangente
tangente
s = 0,2 σ
s = 0,3 σ
400
300
200
10
0
5
8
11
15
20
25
30
n

Figure 9

CONCLUSION GENERALE

Quelque soit le conférencier, je n'ai jamais vu les visages des utilisateurs s'éclairer en découvrant l'hypothèse nulle et le risque associé.

J'ai par contre relevé de nombreuses interprétations erronnées de ce fameux α.

Cela m'a amené à me demander si en calculant α, le statisticien répondait vraiment au souci de l'utilisateur qui, au moins ceux que je fréquente, posent essentiellement leurs problèmes en attitude pratique.

En centrant toute ma réflexion sur le seul problème pratique de la comparaison d'un témoin et d'un traité (problème élémentaire mais combien général) je me suis rendu compte qu'il était habituellement mal résolu, car on lui appliquait sans réfléchir un schéma de raisonnement prévu pour ce problème posé en attitude de recherche, c'est-à-dire finalement prévu pour un tout autre problème(*).

J'ai donc fait des propositions que je pense bonnes, mais dont je ne sais pas du tout si elles sont nouvelles.

Si quelque lecteur y décelait quelques erreurs, je lui serais fort reconnaissant de m'en faire part.

Si elles n'ont rien d'original, c'est preuve qu'il y a un sérieux effort de vulgarisation à faire dans mon milieu de travail.

Si elles sont bonnes et nouvelles, c'est preuve qu'un peu de réflexion sur le problème à résoudre apporte plus que l'application "bête et méchante" d'un calcul standart.

(*) Voir à ce sujet l'article de D. Schwartz et J. Lellouch "Attitude de recherche ou attitude pragmatique dans la formulation d'un problème. . ." Revue de statistique appliquée 1965 Vol XIII N° 4.