

Atelier du groupe probabilités-statistique

Annette Corpart et Nelly Lassalle

Le groupe a été amené à travailler sur l'introduction des notions de probabilités, de la 3^{ème} jusqu'au BTS.

Partant d'expériences « à la main » ou de simulations avec les TICE, de nombreuses activités ont été créées ou sont en cours de réflexion.

Nous vous proposons de les découvrir.

Contenu de l'atelier

- Activités Collège et Seconde
- Activités Lycée
- Activités BTS
- Participation à la CII (Commission Inter-IREM)

Activités Collège et Seconde

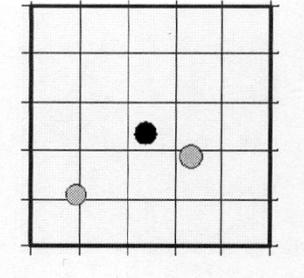
- **Probabilités :**

- Expériences à une épreuve (lancer d'un dé, d'un osselet, jeu de « Franc Carreau »)
- Expériences à deux épreuves (lancer de 2 dés, jeu des cartons, tourn'en rond)
- Comment bien choisir au hasard ?

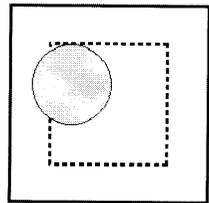
- **Statistiques :**

- lecture de diagrammes
- lecture de graphiques

Expériences à une épreuve : jeu de « Franc Carreau »



- **Principe du jeu** : on lance un palet rond sur un parquet quadrillé, on fait « Franc Carreau » si le palet s'immobilise à l'intérieur d'un carreau.
A-t-on davantage de chance de gagner que de perdre ?
- **Objectifs** : - approche d'une probabilité (sans que cette valeur soit connue au départ) à partir d'un grand nombre d'expériences.
- justification géométrique.
- **Expérimentation et démonstration** (en calculant le rapport de surfaces de 2 carrés)



Et si on ne sait pas estimer la probabilité d'un évènement ?



- **Expérimentation** : lancer d'un osselet.
- **Objectif** : approche fréquentiste de la probabilité dans un cas où il n'y a pas d'autre approche possible : lorsque rien ne permet d'estimer la probabilité d'un événement élémentaire, la stabilisation des fréquences conduit à une estimation de cette probabilité.

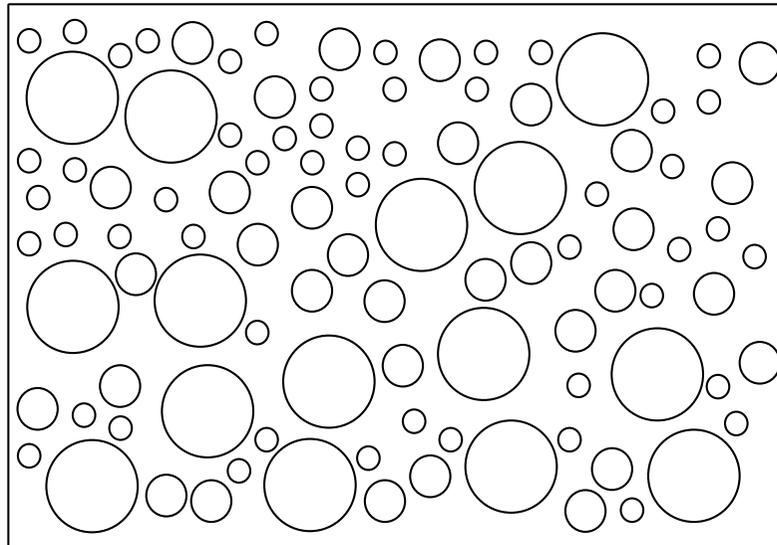
Expériences à deux épreuves : jeu des cartons

3	5	8	12	7
-1	-2	-9	-4	-6

- **Description de l'activité** : on dispose de dix cartons. Sur chacun figure un nombre. 5 de ces nombres sont positifs, les 5 autres sont négatifs. Vaut-il mieux faire le pari d'obtenir un nombre négatif en tirant un seul carton ou d'obtenir un produit négatif en tirant successivement et sans remise deux cartons ?
- **Objectifs** :
 - approche d'une probabilité à partir d'un grand nombre d'expériences.
 - calcul de cette probabilité (à partir d'un tableau à double entrée ou d'un arbre de modélisation).

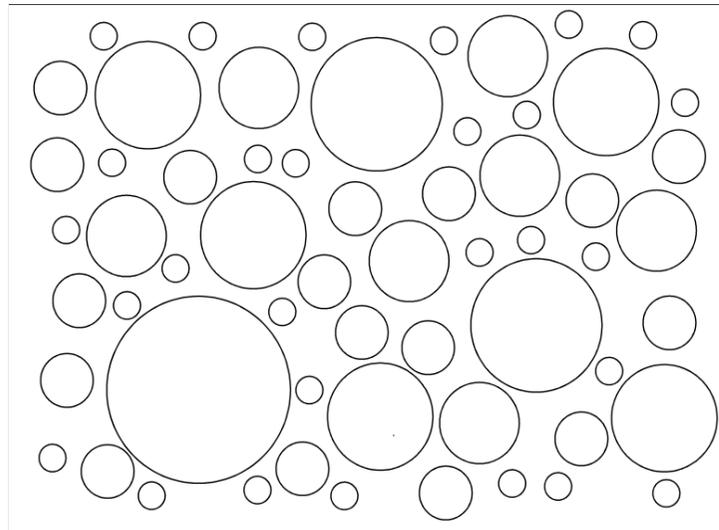
Comment bien choisir au hasard ?

- **Expérimentation** : prélever un échantillon dans une population (pour estimer des proportions).
- **Objectifs** : - illustrer l'ambiguïté de l'expression « choisir au hasard » ;
- mettre en évidence l'importance de préciser les conditions de l'expérience (comparaison des résultats pour un échantillon empirique et un échantillon aléatoire)



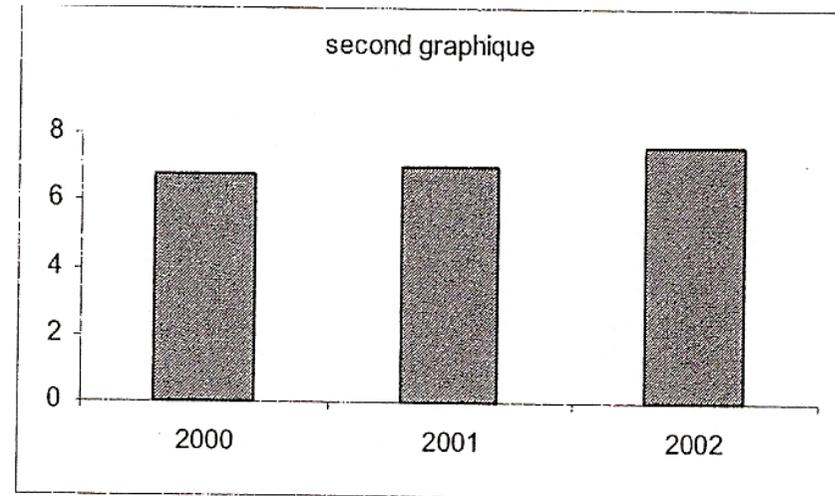
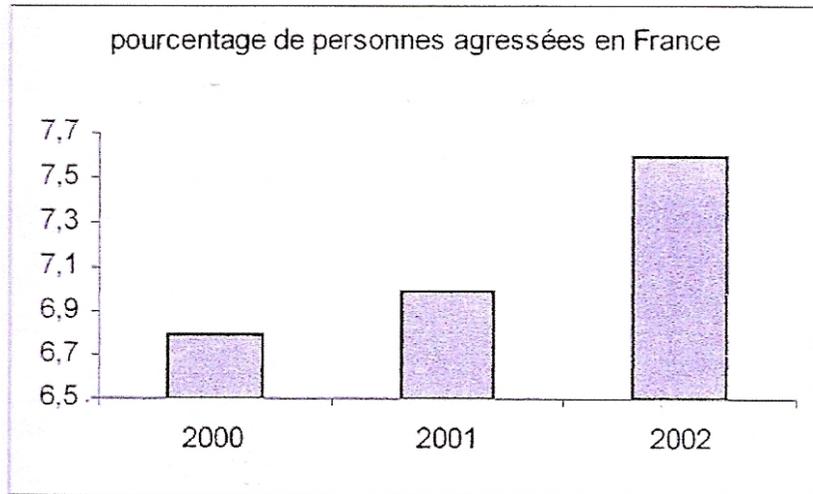
Variante (*activité plus complète pour le lycée*)

- **Expérimentation** : prélever un échantillon dans une population (pour estimer une moyenne).
- **Objectifs** : - illustrer l'ambiguïté de l'expression « choisir au hasard » ;
- mettre en évidence l'importance de préciser les conditions de l'expérience ;
- comparer différentes méthodes d'échantillonnage : échantillon empirique / échantillon aléatoire simple / échantillon aléatoire stratifié / échantillon à deux degrés.



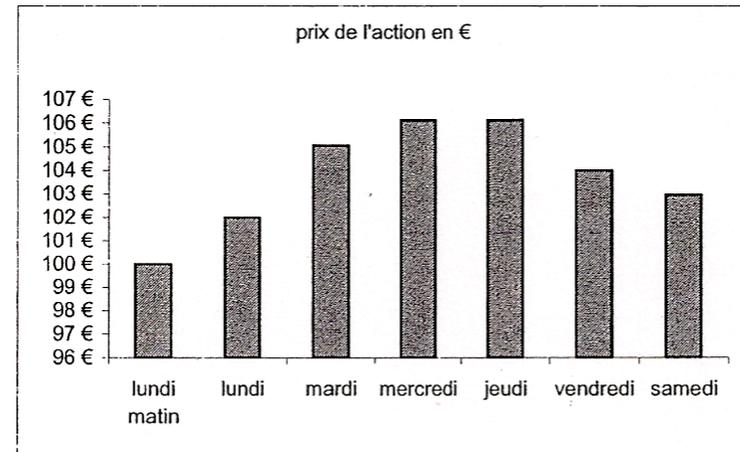
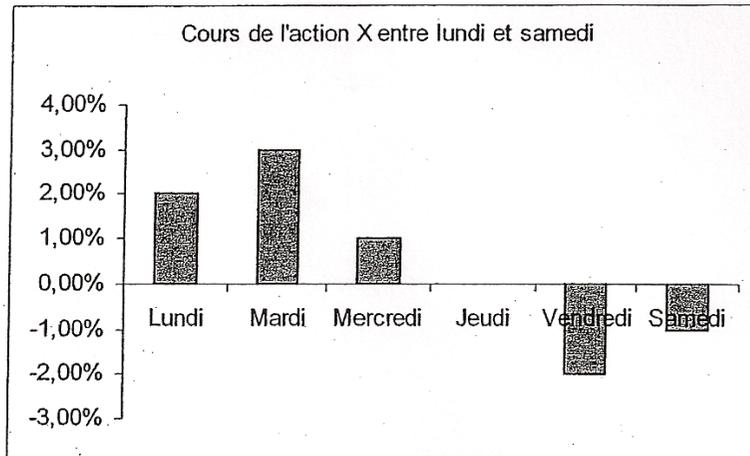
Statistiques :

Lecture de diagrammes



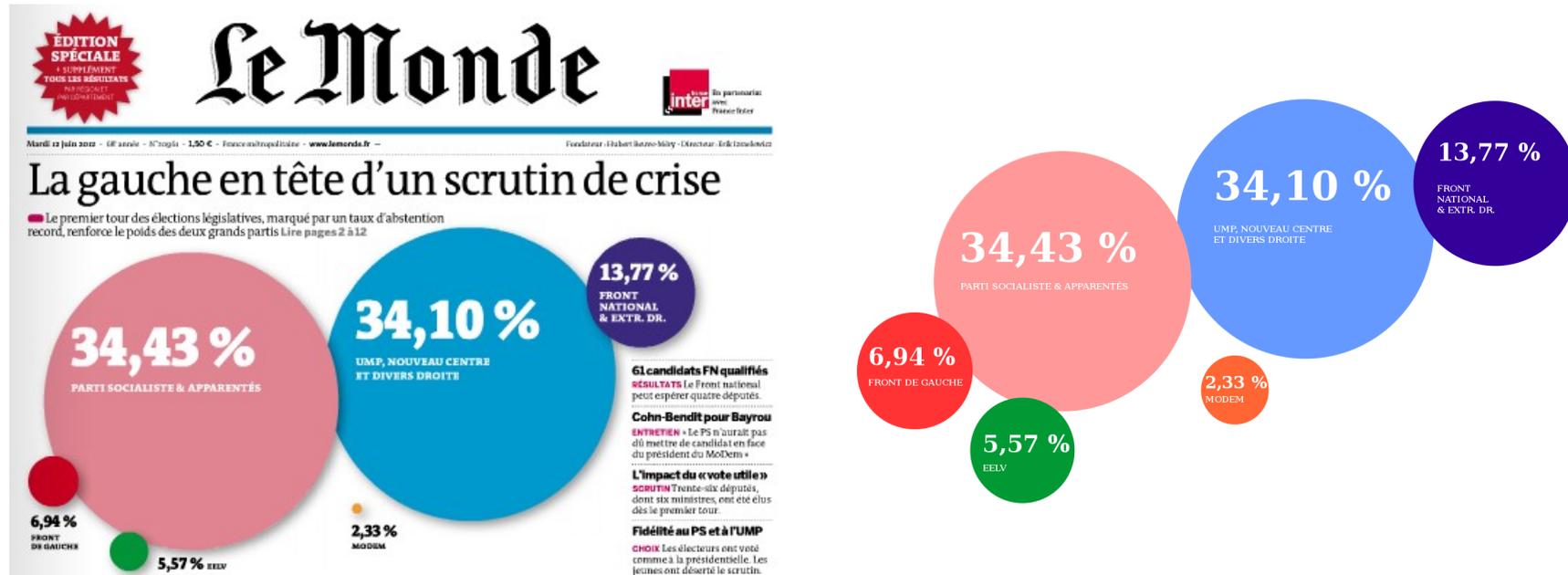
- **Objectif** : développer un regard critique
- **Enoncé** : Lu dans un magazine : « L'explosion de l'insécurité en France (...) Depuis 2000, le pourcentage de personnes agressées augmente de manière dramatique. Le graphique ci-dessus, montre bien l'explosion de violence dont nous sommes aujourd'hui victimes ».
- Si on refait un histogramme dont l'axe des ordonnées est gradué à partir de 0 l'explosion de l'insécurité est beaucoup moins nette !

Lecture de diagrammes



- **Enoncé** : Un graphique d'une revue boursière donne les variations du cours d'une action, au jour le jour depuis quelques jours. Question : quel est ce mystère d'un cours négatif, en plus en pourcentages ?
- Si on refait un histogramme donnant le prix de l'action en fonction des jours de la semaine (en supposant que l'action coûte à l'ouverture le lundi 100 €);
- et si on calcule le coût de l'action à la fermeture le samedi, il est de 103 € : l'action est donc plus chère à la fermeture le samedi que le lundi à l'ouverture...

Lecture de graphiques



- **Énoncé** : Le journal *Le Monde*, dans son édition du 12 juin 2012, a consacré sa Une aux résultats du premier tour des législatives à l'aide de l'infographie ci-dessus.
- Si on respecte une superficie des cercles (censés symboliser le poids respectif de chaque parti) proportionnelle aux suffrages obtenus par les différents partis politiques, on obtient un graphique différent...
- Dans *Le Monde*, ce sont les diamètres des cercles qui sont (à peu près) proportionnels aux résultats. Par cette bévue, *Le Monde* a amplifié visuellement le poids réel du bipartisme en France.

Activités Lycée

- **Statistiques** : construction d'histogrammes
- **Probabilités** :
 - les cartes de contrôle
 - les anniversaires
 - sondage difficile
- **Statistiques inférentielles** :
 - un biberon
 - prises de décision (naissances, jugement...)
 - deux types d'erreur dans une prise de décision
 - surbooking dans les transports aériens et prise de risque

Statistiques : construction de diagrammes

Série A :

Valeurs	14	25	37	50	62	78	85	103	118
Effectifs	2	7	10	17	24	15	12	9	4

Série B :

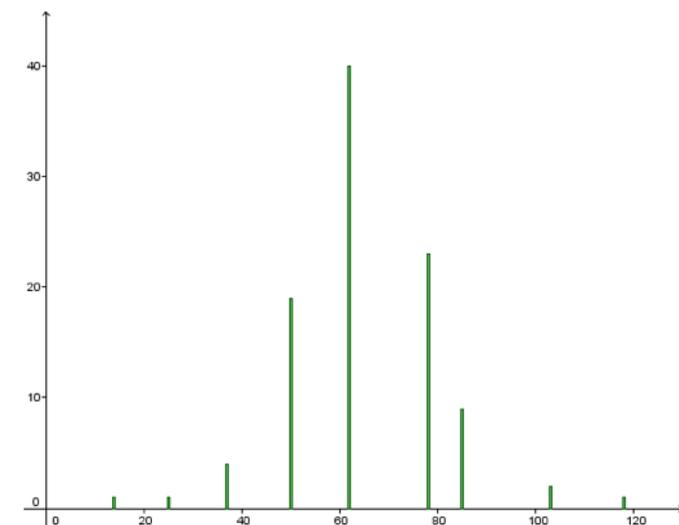
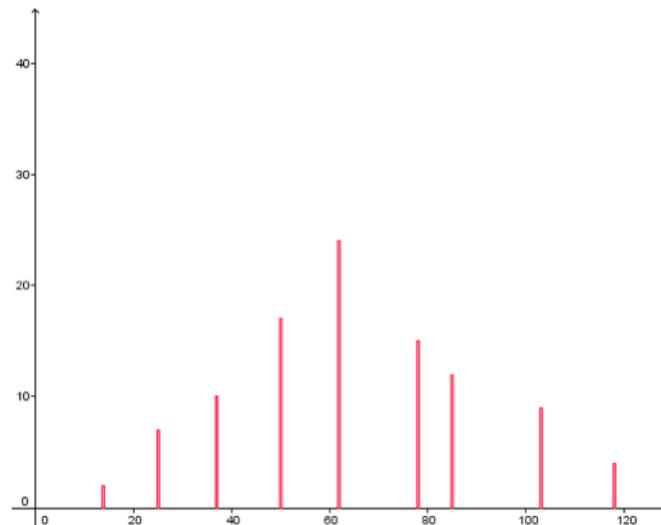
Valeurs	14	25	37	50	62	78	85	103	118
Effectifs	1	1	4	19	40	23	9	2	1

Caractéristiques

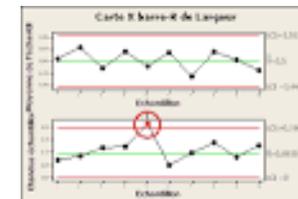
Effectif :
Etendue :
Mode :
Moyenne :
Médiane :
Quartile $Q_1 =$
Quartile $Q_3 =$

- **Objectif** : introduire l'écart type.
- **Énoncé** : Deux séries différentes sont données. Pour chacune des deux, on calcule tous les différents paramètres statistiques (sauf l'écart-type) ; puis on construit les diagrammes en bâtons correspondants.

Séries A ou B	
Effectif : 100	
Etendue : 104	
Mode : 62	
Moyenne : 65	
Médiane : 62	
Quartile $Q_1 = 50$	
Quartile $Q_3 = 78$	
Série A	
Ecart type : 24,46	
Série B	
Ecart type : 15,86	



Les cartes de contrôle



- **Énoncé** : Lors de contrôles de qualité dans l'industrie, des cartes de contrôle reposent sur la procédure suivante : la moyenne de la cote surveillée est calculée sur des échantillons aléatoires prélevés en cours de fabrication. Ces moyennes sont reportées sur une carte de contrôle. Si une série de **7** points consécutifs se trouve du même côté de la « moyenne attendue », le processus de fabrication doit être surveillé pour déceler une éventuelle « dérive ».
- *L'explication du choix du nombre 7 se trouve dans la résolution du problème de probabilités suivant : une pièce de monnaie équilibrée est lancée 7 fois, quelle est la probabilité de l'événement A : « la pièce est tombée 7 fois sur pile »*
- **Simulation** avec une pièce de monnaie et/ou avec un tableur, **estimation** de la valeur de cette probabilité.
- **Calcul** de la probabilité de l'événement A.



Maîtrise Statistique des Procédés



GUIDE MSP			
	<p>PROCESSUS MAITRISE</p> <p>Répartition homogène des points entre les limites de contrôle</p>		<p>ACTIONS</p> <p>ATTENDRE LE PROCHAIN PRELEVEMENT</p>
	<p>PROCESSUS ANORMAUX</p> <p>Points au-delà des limites de contrôle</p>		<p>Alerter le chef d'unité qui :</p> <ul style="list-style-type: none"> - consultera la liste des paramètres influents pour rechercher l'élément responsable des points au-delà des limites de contrôle. <p>Pendant l'ANALYSE mettre en place la procédure de verrouillage ou l'arrêt de l'installation.</p>
	<p>Changement de niveau (7 points consécutifs)</p>		<p>TENDANCES INHABITUELLES</p> <p>Alerter le chef d'unité pour faire une ANALYSE des causes de cette tendance</p>
	<p>7 points consécutifs en augmentation ou en diminution</p>		<p>VARIATIONS IMPORTANTES</p> <p>Alerter le chef d'unité pour ANALYSER les causes de cette dégradation.</p>
	<p>La variabilité des points croît</p>		<p>AMELIORATION DU PROCESSUS</p> <p>Alerter le chef d'unité pour rechercher les causes de cette amélioration.</p>
	<p>La variabilité des points décroît</p>		<p>AMELIORATION DU PROCESSUS</p> <p>Alerter le chef d'unité pour rechercher les causes de cette amélioration.</p>

POINT HORS LIMITE DE CONTROLE OU AMORCE DE DERIVE

Suivre ou verrouiller les caractéristiques processus plutôt que celles du produit est le principe de base de MSP.

L'évolution de l'un de ces facteurs explique le franchissement des limites de contrôle ou les tendances inhabituelles du processus.

Aussitôt alerté, le chef d'unité recherche l'élément responsable en consultant la liste des paramètres influents, et vérifie ensuite la présence ou non des verrous physiques ou l'évolution des cartes de contrôle des caractéristiques soupçonnées et pilote les actions correctives.

PARAMETRES INFLUENTS		
INDICATEUR	FACTEUR	VERROU
Moyenne	F1 F2	Verrou physique Carte de contrôle
Etendue	F3	Verrou physique
Les 2 à la fois	F4	Carte de contrôle

Probabilités :

Les anniversaires



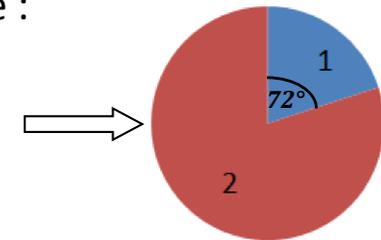
- **Énoncé** : Dans une classe, il y a 30 élèves. Marc affirme : « il y a largement plus d'une chance sur 2 que deux élèves de la classe aient le même jour anniversaire (mais pas forcément le même âge) ». Comment savoir si son affirmation est correcte ?
- **Objectifs** : - approche d'une probabilité à partir de simulations (deux **simulations** possibles, avec deux urnes ou avec la calculatrice).
 - **calcul** de cette probabilité (en calculant la probabilité de l'événement contraire).
 - travail sur les expressions « **au moins** » et « **au plus** ».



Sondage difficile

- **Objectif** : Eviter les fausses réponses grâce au hasard
- Si on pose la question « Avez-vous déjà volé dans un supermarché ? » aux clients, on peut supposer que les réponses seront biaisées !! En 1965, le statisticien américain Warner propose une méthode pour obtenir un résultat non biaisé à cette enquête :
- L'enquêteur donne à chaque client interrogé la carte suivante :

1 : *j'ai déjà volé dans un supermarché*
2 : *je n'ai encore jamais volé dans un supermarché*
- Pour répondre à cette question, il propose de réaliser **en secret** la procédure suivante :
*Faire tourner la roulette et lire la proposition indiquée en face de l'index après arrêt.
Donner la réponse « Vrai » ou « Faux » à cette proposition.*



A la fin de la journée, l'enquêteur détermine les proportions de réponses « Vrai » et « Faux » et estime alors la fréquence réelle p de vols dans les supermarchés.

- **Modélisation** du problème avec un arbre pondéré.
- **Calcul** de la fréquence p (en résolvant une équation du premier degré).

Statistiques inférentielles

- Les statistiques inférentielles sont avant tout un mode de pensée qui permet de passer :
 - du connu* : données fournies par **des statistiques** sur un ensemble nécessairement restreint d'individus,
 - à l'inconnu* : résultats applicables à l'ensemble des individus concernés,
 - par un raisonnement logique, le raisonnement inférentiel, reposant sur des bases théoriques : le calcul des probabilités.*
- Trois thèmes sont étudiés au lycée : **fluctuation des échantillons** (programme de seconde), **prise de décision** (test d'hypothèse, programme de première) et **estimation** d'une proportion (intervalle de confiance ou « fourchette de sondage », programme de terminale)

Le biberon



- Pour permettre une bonne appréhension des notions difficiles de statistiques inférentielles, nous proposons une activité découpée en trois parties, chacune d'elles introduisant une des trois notions vues précédemment.
- Nous sommes parties d'une situation concrète rencontrée en lycée agricole : l'étude de la répartition de graines dans une semence de gazon. Pour effectuer les tirages, nous utilisons des bouteilles opaques (des « biberons ») et nous exploitons les résultats observés avec le logiciel géogébra.
- Ainsi (par exemple), à partir d'[intervalles de fluctuation](#) obtenus pour différentes valeurs d'une proportion p , nous déterminons des [intervalles de confiance](#) de p au niveau 95%.

Prise de décision

Les deux activités qui suivent sont extraites du document ressource des nouveaux programmes de lycée professionnel



Naissances à pile ou face

- **Énoncé** : Les données statistiques suivantes ont été relevées :
 - en 2000, dans le village de Xicun, en Chine, il est né 20 enfants, parmi lesquels 16 garçons,
 - dans la réserve indienne d'Aamjiwnaag, située au Canada à proximité d'industries chimiques, il est né entre 1999 et 2003, 132 enfants dont 46 garçons.

Ces observations sont-elles le fruit du hasard ?

- **Simulation** avec une pièce de monnaie
- **Calcul des intervalles de fluctuation** au seuil 0,95 d'une proportion p égale à 0,5 (pour $n = 20$ et $n = 132$)
- **Conclusion** : doit-on accepter ou rejeter l'hypothèse que la distribution des sexes des 20 enfants nés en 2000 à Xicun (ou des 132 enfants nés à Aamjiwnaag) n'est due qu'au seul hasard ?

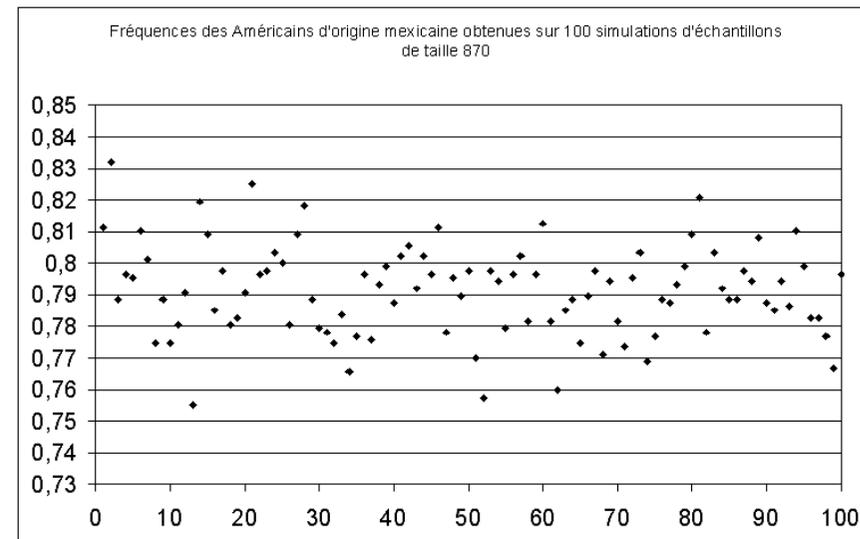
Contester un jugement



- **Enoncé** : En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population de ce comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eut que 339 personnes d'origine mexicaine.

- **Simulation** sur tableur : prélèvement d'échantillons aléatoires de taille $n = 870$ dans une population avec une proportion $p = 0,791$; représentation dans un graphique des fréquences observées.

- **Conclusion** : doit-on accepter ou rejeter l'hypothèse que la constitution des jurys dans ce comté du Texas est totalement aléatoire ?



Etude des deux types d'erreurs dans une prise de décision

- **Objectif** : étudier les risques de première espèce et de seconde espèce intervenant dans un test d'hypothèse.
- **Énoncé** : Un examinateur doit faire passer une épreuve de type QCM à des étudiants. Ce QCM est constitué de vingt questions indépendantes. Pour chaque question, il y a trois réponses possibles dont une seule correcte. L'examineur veut déterminer une valeur critique k telle que l'étudiant est reçu si le nombre de réponses correctes est supérieur ou égal à k et recalé dans le cas contraire.
- A l'issue des résultats de l'épreuve, quatre cas sont possibles :

On a donc deux risques d'erreur correspondant aux cas (3) et (4) :

- l'erreur α : on rejette l'idée que l'étudiant n'a pas travaillé alors que c'est vrai (c'est le « risque professeur »)
- l'erreur β : on pense que l'étudiant n'a pas travaillé alors que c'est faux (c'est le « risque étudiant »)

1) l'étudiant n'a pas travaillé et il est recalé



2) l'étudiant a travaillé et il est reçu



3) l'étudiant n'a pas travaillé et il est reçu



4) l'étudiant a travaillé et il est recalé



Surbooking dans les transports aériens

- Les sociétés de transport aérien considèrent que, en moyenne, seules 90% des personnes prévues au départ d'un avion de 100 places se présentent à l'embarquement. Ces sociétés prennent-elles un risque important de surbooking en vendant 106 billets pour un avion de 100 places ?
- **Objectifs** : - [simulation](#) sur tableur de 1000 vols de 106 personnes
 - calcul du nombre de vols où il y a eu surbooking
 - conclure à la question posée dans l'énoncé.

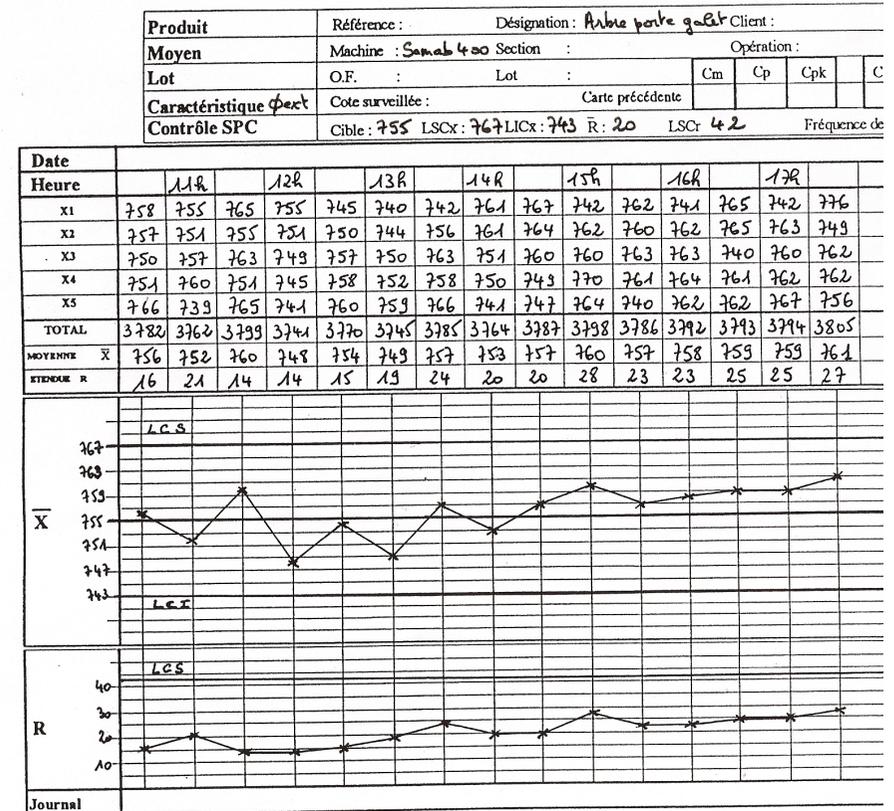
Activités BTS

Statistiques dans l'industrie

- Les cartes de contrôle
- La fiabilité : loi exponentielle, loi de Weibull
- Les plans d'expérience

Cartes de contrôle

- Une **carte de contrôle** est un outil utilisé dans le domaine du contrôle de qualité afin de maîtriser un processus.
- Les types de cartes de contrôle les plus utilisées dans l'industrie sont les cartes de contrôle de la moyenne et de l'étendue. Dans cette méthode, deux graphiques sont tracés et interprétés simultanément, ils permettent le suivi et la maîtrise des dispersions.
- A partir d'échantillons de pièces à contrôler, on calcule la moyenne des moyennes et la moyenne des étendues.
- Pour la carte des moyennes et la carte des étendues, on trace :
 - la ligne centrale (correspondant à la moyenne)
 - les limites de contrôle inférieure (LCI) et supérieure (LCS)



Fiabilité

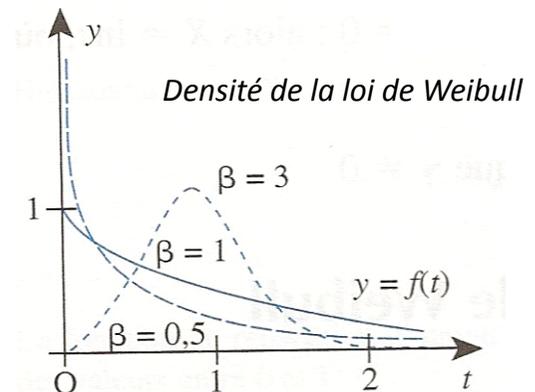
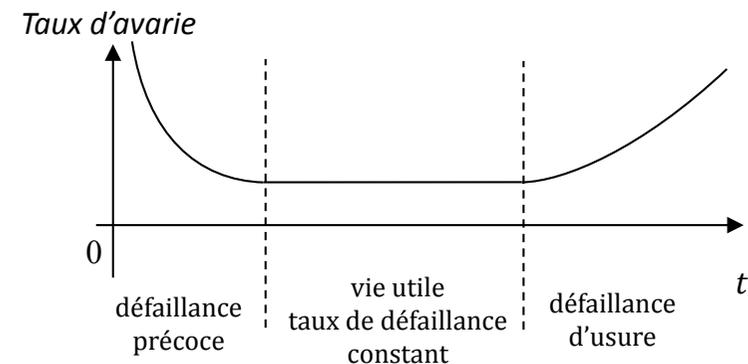
- **Définition AFNOR** : « la fiabilité est la caractéristique d'un dispositif qui s'exprime par la probabilité que ce dispositif accomplisse une fonction requise dans des conditions d'utilisation données et pour une période donnée. »
- **Objectif** : déterminer la probabilité qu'un équipement fonctionne sans défaillance pendant une période de temps déterminée et dans des conditions opérationnelles spécifiées.
- Premières notions de fiabilité :
 - **TBF** (Temps de bon fonctionnement)
 - **Taux d'avarie** $\lambda(t)$
 - Fonction de défaillance et **fonction de fiabilité**

$$R(t) = e^{-\int_0^t \lambda(u) du}$$

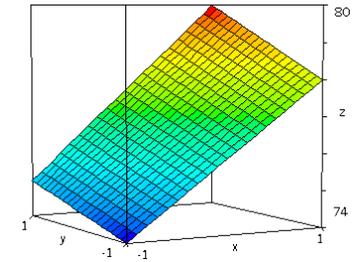
- **MTBF** (moyenne des temps de bon fonctionnement)

Ces notions sont vues d'abord sous l'aspect statistique, puis sous l'aspect probabiliste.

- Modèle à taux d'avarie constant : **loi exponentielle**, le taux d'avarie est constant, $R(t) = e^{-\lambda t}$
- Modèle à taux d'avarie variable : **loi de Weibull**, le taux d'avarie est une fonction puissance, $R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta}$

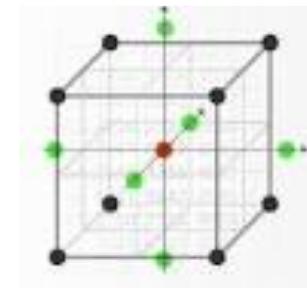


Les plans d'expérience



- Depuis que l'homme fabrique des objets, la recherche de la qualité est un objectif important. L'industriel s'est d'abord intéressé à la qualité du produit fini. Maintenant il s'intéresse à la qualité au stade de la conception ou de la production. Il est alors amené à faire des essais ou des expériences.
- Il importe de ne pas faire ces essais n'importe comment (par exemple par tâtonnements successifs) mais il s'agit au contraire de réaliser une série d'essais rigoureusement choisis, ce qu'on appelle un **plan d'expérience**.
- Un plan d'expérience permet donc de réduire le nombre d'essais à ce qui est strictement nécessaire pour prendre une décision. Il permet également une interprétation rapide et sans équivoque des résultats.
- Un plan d'expérience s'inscrit dans la démarche qualité au même titre que d'autres outils statistiques.

$$(a_0 \ a_1 \ a_2 \ a_{12}) = \frac{1}{4} \times (y_1 \ y_2 \ y_3 \ y_4) \times \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$



Ce que nous travaillons cette année :

- Les chaînes de Markov
- Puissance d'un test d'hypothèse
- L'introduction de la loi de Poisson

Participation à la CII (Commission Inter-IREM)

- La commission travaille en phase avec les évolutions de l'enseignement de la statistique et des probabilités tant du point de vue approche que du point de vue programmes de collège et de lycées.
- Elle vise un travail de veille scientifique pour les publications concernant l'enseignement de la statistique et des probabilités.

La commission se réunit trois fois par an : en novembre et mars à Paris et en mai-juin dans un des IREM représenté à la commission.

Calendrier de l'année 2014-2015 :

- 15 novembre 2014 à l'IREM de Paris 7
- 28 mars 2015 à l'IREM de Paris 7
- 12 et 13 juin 2015 à Saint-Malo

Les membres de la commission :

15 membres représentant le Comité Scientifique des IREM et 9 IREM :

- **AIX-MARSEILLE**
- **BESANCON**
- **CLERMONT-FERRAND**
- **LYON**
- **MONTPELLIER**
- **PARIS 7**
- **POITIERS**
- **RENNES**
- **TOULOUSE**

Professeurs de lycées généraux, technologiques ou agricoles, enseignants-chercheurs, inspecteurs généraux, en activité ou retraités.

Des livres de synthèse des travaux de la commission :

- *Enseigner les probabilités au lycée*, octobre 1997, 464 p.
- *Autour de la modélisation en probabilités*, avril 2001, 260 p.
- *Probabilités au lycée*, 2003, brochure n° 143, 184 p.
- *Statistique au lycée*, volume 1 : Les outils de la statistique, 2005, brochure APMEP n° 156, 314 p.
- *Statistique au lycée*. Volume 2 : Activités statistiques pour la classe, 2007, brochure APMEP n° 167, 330 p.
- *Probabilités au collège : ne pas laisser l'enseignement des probabilités au hasard...*, 2012, brochure APMEP n° 198, 120 p. (en collaboration avec la commission Collège)

