Introduction au Contrôle statistique de la qualité

Liège, février 2021

F. Farnir

Faculté de Médecine Vétérinaire

Université de Liège (Belgique)

Introduction

- Qualité ?
  - Il existe de nombreuses définitions, fonctions des exigences de ceux qui la définissent
  - Une définition possible serait: « un produit est de qualité lorsqu'il rencontre les exigences de celui qui l'achète »
- Dans les sciences du vivant ?
  - Comme dans d'autres applications (y inclus industrielles), le besoin de qualité est omniprésent dans les sciences du vivant.

Introduction

- Exemples ?
  - Agronomie
    - exemple: qualité et quantité des plants cultivés.
  - Pharmacologie et vaccinologie
    - exemple: qualité des médicaments et des vaccins
  - Contrôle de la chaine alimentaire
    - exemple: quantité de résidus (pesticides, dioxines), comptages bactériens dans la viande et le lait, …
  - Production agro-alimentaire
    - exemple: poids et aspect des portions, …
  - …

Introduction

- Normes ISO 9001 (2015)
  - Définition d'exigences concernant l'organisation d'un système de gestion de la qualité
    - Les principes actuels ont été établis en 2008, une actualisation a eu lieu en 2015
  - Les exigences couvrent 4 grands domaines:

Introduction

- Normes ISO 9001 (2015) - suite
  - Domaines
    - Responsabilité de la direction : exigences d'actes de la part de la direction en tant que premier acteur et acteur permanent de la démarche.
    - Système qualité : exigences administratives permettant la sauvegarde des acquis. Exigence de prise en compte de la notion de système.
    - Processus : exigences relatives à l'identification et à la gestion des processus contribuant à la satisfaction des parties intéressées.
    - Amélioration continue : exigences de mesure et enregistrement de la performance à tous les niveaux utiles ainsi que d'engagement d'actions de progrès efficaces.

Introduction

- Normes ISO 9001 (2015) – suite
  - Parmi les exigences de ce « Système de Gestion de la Qualité », plusieurs font appel à des techniques d'ordre statistique:
    - mise en place d'indicateurs de la qualité,
    - suivi des actions mises en place,
    - exploitation des données recueillies,
    - amélioration continue du système.
  - On parlera alors de « Contrôle statistique de la qualité (CSQ) »

Problématique statistique

- Un contrôle longitudinal de la qualité est donc nécessaire
  - traçabilité
  - maintien dans des normes acceptables des paramètres étudiés
- Sauf dans de rares cas, toute la population visée n'est pas accessible, et on devra se contenter d'un échantillon
  - Populations généralement grandes
  - Coûts échantillonnage ↔ non-conformité
  - Contrôles destructifs…

Problématique statistique (suite)

- On souhaitera donc disposer d'échantillons:
  - représentatifs…
  - de taille limitée (coûts…)
  - mais suffisante pour espérer une précision satisfaisante…

  - => rappels sur l'échantillonnage
    => rappels sur la précision

Aspects statistiques

- Dans cette partie, nous aborderons:
  - quelques rappels sur les distributions importantes pour le CSQ
    - binomiale
    - χ^2
    - normale, Student
  - quelques rappels sur l'échantillonnage
  - quelques rappels sur la précision de l'estimation

Rappels statistiques

- Rappel sur les distributions importantes pour le CSQ
  - Voir cours du début d'année...

Rappels statistiques

- Exemples d'utilisation de la binomiale
  - Si un défaut a 10% de chance de survenir sur chaque sujet testé, quelle est la probabilité d'avoir plus de 2 défauts en testant 10 sujets ?

Rappels statistiques

- Exemples d'utilisation de la binomiale
  - Si un défaut a 10% de chance de survenir sur chaque sujet testé, quelle est la probabilité d'avoir plus de 2 défauts en testant 10 sujets ?
  - Solution
    - Utiliser la loi binomiale $P(r) = C_n^r * p^r * (1-p)^{n-r}$
    - La probabilité recherchée vaut: $P = P(3) + P(4) + \cdots + P(10) = 1 - P(0) - P(1) - P(2)$
      - Calcul (excel): P = 1 – loi.binomiale(2;10; 0,1; VRAI)
      - Résultat: P = 0,0702

Rappels statistiques

- Exemples d'utilisation de la binomiale
  - Si un défaut a en principe 10% de chance de survenir sur chaque sujet testé, à partir de combien de défauts commencera-t-on à douter de ce pourcentage si on teste 10 sujets ?
  - Solution
    - On peut formuler le problème comme un test d'hypothèse (H_0: $\pi = 0.10$)
    - Si le nombre de défauts (d) est tel que:
$$P(r \geq d) < 0{,}05$$
on rejettera H_0.

Rappels statistiques

- Exemples d'utilisation de la binomiale
  - Solution
    - On peut cumuler les probabilités P(10) + P(9) + ... jusqu'à ce que la somme excède 5%
    - Avec excel, on obtient facilement:
      - $P(r \geq 3) = 0,0702$
      - $P(r \geq 4) = 0,0128$
      - On doutera donc de « l'hypothèse 10% » si $r > 3$.

Rappels statistiques

- Exemples d'utilisation de la normale
  - Si les poids des portions d'un aliment produit par une société agro-alimentaire a une distribution normale de moyenne 100 grammes et de déviation standard 5 grammes (notation: N[100;5]), combien de portions s'attend-on à voir dépasser le poids de 112 grammes dans une production journalière de 1500 portions ?

Rappels statistiques

- Exemples d'utilisation de la normale
  - Solution
    - Probabilité d'un poids de plus de 112 grammes ?
      - Avec excel: `=1-loi.normale(112;100;5;vrai)`
      - Résultat: $p = 0,008198$
    - Nombre attendu de poids de plus de 112 grammes ?
      - $x = N*p = 1500*0,008198 = 12,3$
      - En moyenne, chaque jour, on s'attend à obtenir 12.3 portions de plus de 112 grammes.

Rappels statistiques

- Exemples d'utilisation de χ^2
  - Si, dans le problème précédent, on obtient 25 portions de plus de 112 grammes sur la même journée, que peut-on en déduire ?

Rappels statistiques

- Exemples d'utilisation de χ^2
  - Solution
    - L'hypothèse nulle est que les poids des portions ont une distribution N(100;5)
    - Si cette hypothèse est vraie, on s'attend à observer 12,3 portions de plus de 112 grammes (et 1487,7 portions de moins de 112 grammes) par jour
    - En réalité, on a observé 25 portions de plus de 112 grammes (et 1475 portions de moins de 112 grammes) sur une journée

Rappels statistiques

- Exemples d'utilisation de χ^2
  - Solution (suite)
    - On peut comparer les effectifs aux attendus via un χ^2:
      - $\chi^2 = (12{,}3-25)^2/12{,}3 + (1487{,}7-1475)^2/1475 = 13{,}23$
      - Comme il y a 2 catégories, il y a (2-1) = 1 degré de liberté
      - La probabilité d'atteindre une telle valeur de chi-carré, ou de la dépasser vaut:
        - (avec excel) `=loi.khideux(13,23;1)`
        - p = 0,00028
        - Comme cette probabilité est très faible, on est conduit à rejeter l'hypothèse nulle: les poids ont augmenté !

Aspects statistiques

- Dans cette partie, nous aborderons:
  - quelques rappels sur les distributions importantes pour le CSQ
    - binomiale
    - χ^2
    - normale, Student
  - quelques rappels sur l'échantillonnage
  - quelques rappels sur la précision de l'estimation

Types de mesures

- Qualitatives
  - p.e. défectueux – non défectueux…
- Quantitatives
  - p.e. poids de la portion…

- Selon les cas, le contrôle sera abordé de manière différente:

Tableau récapitulatif

	Contrôle qualitatif	Contrôle quantitatif
Paramètre étudié dans lot de taille N	π = % défectueux	μ = moyenne pop σ = écart-type pop
Estimateur utilisé sur lot de taille N	p = % défectueux	m = moyenne lot s = écart-type lot
Lois de distr. d'échantillonnage	Binomiale (n,p)	Student (n-1) $\rightarrow \mu$ χ^2 (n-1) $\rightarrow \sigma$

Echantillonnage

- Un « bon » échantillonnage est caractérisé par certaines propriétés:
  - Représentativité de la population visée
    - Elimination des biais
    - Adaptation au contrôle visé
  - (Pseudo-)aléatoire
    - Echantillonnage simple
    - Echantillonnages stratifié, en grappe, …
  - Economiquement justifié
    - Compromis coûts - sécurité

Représentativité

- Définir la population visée
  - En routine
    - Un secteur (p.e. restauration rapide), une enseigne (p.e. Quick), un restaurant particulier
    - Produits visés
  - Lors d’une crise
    - Production d’un pays, d’une région, d’un abattoir ?
    - Mode de fabrication
    - Producteurs dans une zone délimitée
    - Produit incriminés

Représentativité

- Eliminer les biais
  - Exemples de biais
    - Effet du moment de la journée
      - p.e. changement d'équipes, nettoyage, …
    - Effet du moment dans la semaine
      - p.e. procédures de nettoyage de la chaine de production
    - Effet de la période de l'année
      - T° extérieure, transport, …
    - Effets « locaux »
      - Normes différentes de respect d'hygiène suivant l'opérateur
    - Position dans le lot
    - …

Représentativité

- Adaptation au contrôle visé
  - Contrôle à la réception d'un lot
    - Choix de n individus au hasard
  - Contrôle de fabrication
    - Choix de n individus au hasard et d'une fréquence f de prélèvement (voir « cartes de contrôle »)
      - si f et n augmentent, l'efficacité du contrôle augmente, mais le coût également…
  - Question: comment choisir n (et f) ?
    - Cfr plus loin

Exemples d'échantillonnage

- Echantillonnage simple
- Echantillonnage stratifié
- Echantillonnage en grappes
- …
  - Voir quelques exemples en annexe.

Aspects statistiques

- Dans cette partie, nous aborderons:
  - quelques rappels sur les distributions importantes pour le CSQ
    - binomiale
    - χ^2
    - normale, Student
  - quelques rappels sur l'échantillonnage
  - quelques rappels sur la précision de l'estimation

Estimation

- L'idée de l'estimation est de fournir à partir d'un échantillon un « ordre de grandeur » de la variable étudiée. Cette estimation peut prendre:
  - la forme ponctuelle d'une valeur censée représenter la valeur estimée
  - la forme d'un intervalle de confiance renfermant la valeur estimée avec une probabilité qu'on peut choisir

Estimation ponctuelle

- Contrôle par attribut (p.e. contamination)
  - On estime π, la prévalence dans la population visé, via $p = n / N$, la prévalence observée sur l'échantillon.
- Contrôle aux mesures (p.e. poids portion)
  - On estime μ, le poids moyen, et éventuellement σ, la dispersion, via m et s, estimateurs mesurés sur l'échantillon.

Estimation ponctuelle: remarques

- Les grandeurs à estimer (π, μ, σ) sont des constantes, caractéristiques de la population étudiée.
- Par contre, les estimations ponctuelles (p, m, s) sont des variables, dépendant de l'échantillon utilisé pour les calculer => l'estimation est en général imparfaite, et il existe une distribution des valeurs d'estimateurs.

Estimation ponctuelle: exemple

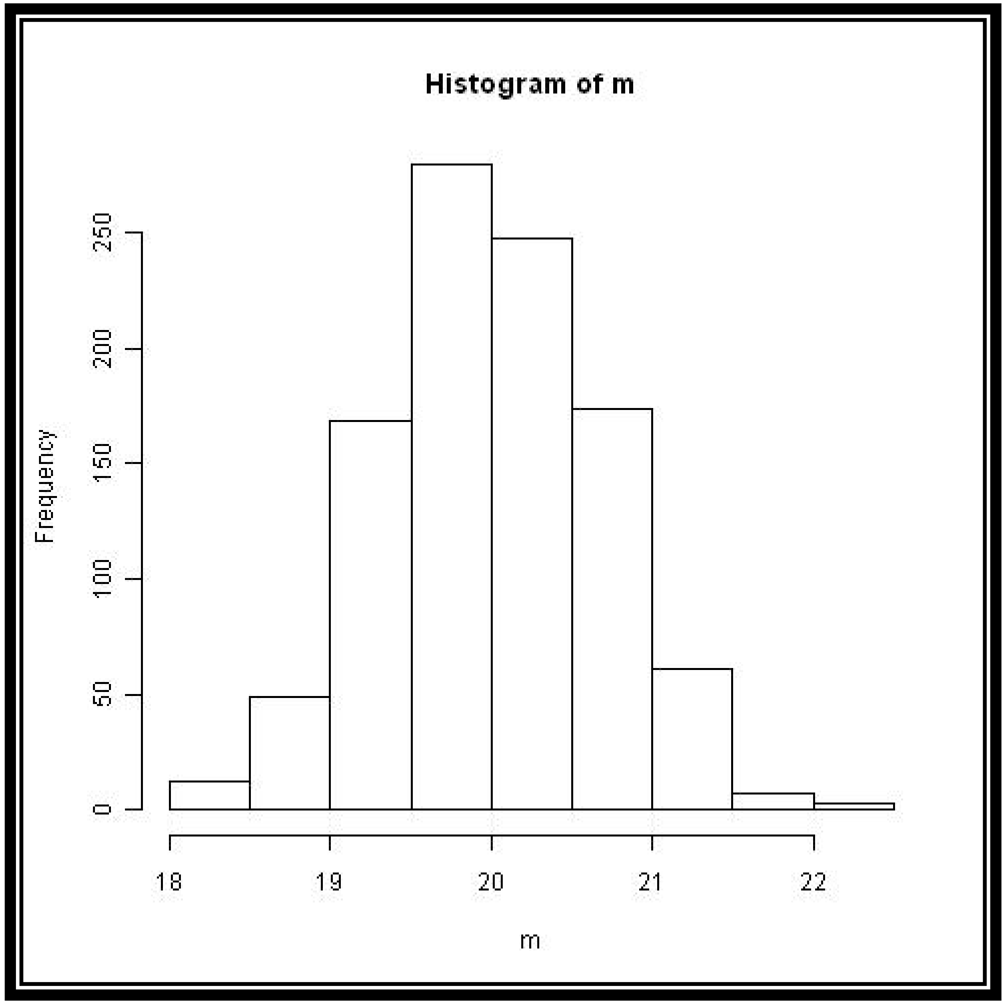
- Exemple avec R

```
> m<-rep(0,1000)
> for (i in 1:1000) {
   m[i]<-mean(rnorm(20,mean=20,sd=3))
+ }
> hist(m)
```

  - rnorm(n,mean=m,sd=s) échantillonne n valeurs au hasard dans N(m,s)
  - mean(X) calcule la moyenne de X
  - hist(m) produit un histogramme avec m

Estimation ponctuelle: exemple

- Résultat avec R



Estimation par intervalle de confiance

- Commençons par l'estimation de π
- On montre, en statistique, que:
  - quand n est grand (n*p > 5, n*(1-p) > 5)

$$z = \frac{n/N - \pi}{\sqrt{\pi*(1-\pi)/N}} \approx \frac{p-\pi}{\sqrt{p*(1-p)/N}}$$

où Z ~ N(0;1)

  - quand n est petit
    - Utilisation de la loi binomiale

Estimation par intervalle de confiance

- Supposons tout d'abord n « grand »
  - Utilisation pour le calcul d'un IC(α):
    - On choisit α: probabilité $p \in IC(\alpha) = 1 - \alpha$
      - Typiquement: $\alpha = 0.05$ (souvent aussi: 0.01, 0.001)
    - On cherche les valeurs Z_i et Z_s entre lesquelles Z a une probabilité $(1 - \alpha)$ de se situer
      - $P(-1.96 \leq Z \leq 1.96) = 0.95$
    - On remplace Z par sa valeur approchée:

$$P\left(p - 1.96 * \sqrt{p * (1 - p) / N} \leq \pi \leq p + 1.96 * \sqrt{p * (1 - p) / N}\right) = 0.95$$

Estimation par intervalle de confiance

- Supposons tout d'abord n « grand »
  - De manière plus générale:

$$P(\, p - W(\alpha) \leq p \leq p + W(\alpha)) = 1 - \alpha$$

où $W(\alpha) = t_{\alpha/2,(N-1)} * \sqrt{p*(1-p)/N}$

Dist. t de Student avec (N-1) ddl

Estimation par intervalle de confiance

- <u>Exemple</u>: si 7 lots sur 1000 examinés dépassent une norme qu'on ne doit dépasser que dans 5 cas sur 1000, doit-on conclure immédiatement à un problème ?

Estimation par intervalle de confiance

- <u>Solution</u>:

Si 7 lots sur 1000 examinés dépassent une norme qu'on ne doit dépasser que dans 5 cas sur 1000, doit-on conclure immédiatement à un problème ?

- p = 0.007, α=0.05 ⇒ z = -1.96

$$IC = 0,007 \pm 1,96 * \sqrt{0,007*0,993/_{1000}}$$

- $$= 0,007 \pm 0,00516$$

$$= [0,00183; 0,01217]$$

Comment interpréter ce résultat ?

Estimation par intervalle de confiance

- Interprétation
  - Si on part de l'hypothèse (nulle…) selon laquelle il y a une proportion $\pi = 0.005$ de valeurs au-dessus de la norme autorisée, on voit qu'une valeur de $p = 0.007$ n'est pas aberrante puisque l'intervalle de confiance contient la valeur $\pi = 0.005$
  - On dira que l'hypothèse nulle est acceptée (au seuil $\alpha = 0.05$) dans ce cas.
  - On peut donc interpréter le calcul de l'IC comme un test d'hypothèse.

Estimation par intervalle de confiance

- Autres exemples: voir en annexe

Intervalles de confiance

- <u>Résumé</u>

Paramètre à prédire	Formule à utiliser
π (n petit)	$IC(\alpha) = [p_i; p_s]$ où pi, ps fournis par 2 lois binomiales
π (n grand)	$IC(\alpha) = [p - z*d; p + z*d]$ où $z = z(\alpha)$ et $d^2 = p*(1-p)/n$
μ (σ connu)	$IC(\alpha) = [m - z*d; m + z*d]$ où $z = z(\alpha)$ et $d^2 = \sigma^2/n$
μ (σ inconnu)	$IC(\alpha) = [m - t*d; m + t*d]$ où $t = t(\alpha,n-1)$ et $d^2 = s^2/n$

Décisions et risques

- Comme le contrôle s'apparente à un test d'hypothèse, il est exposé aux mêmes risques que ces derniers:

	H0 acceptée	H0 rejetée
H0 vraie	**OK !**	**Erreur type I** α
H0 fausse	**Erreur type II** β	**OK !**

Interprétation des risques

- Si on part de l'hypothèse (nulle) de conformité, les deux erreurs s'interprètent de la manière suivante:
  - Risque de 1° espèce (α) = probabilité de refuser à tort un lot conforme
  - Risque de 2° espèce (β) = probabilité d'accepter à tort un lot non conforme

Interprétation des risques (suite)

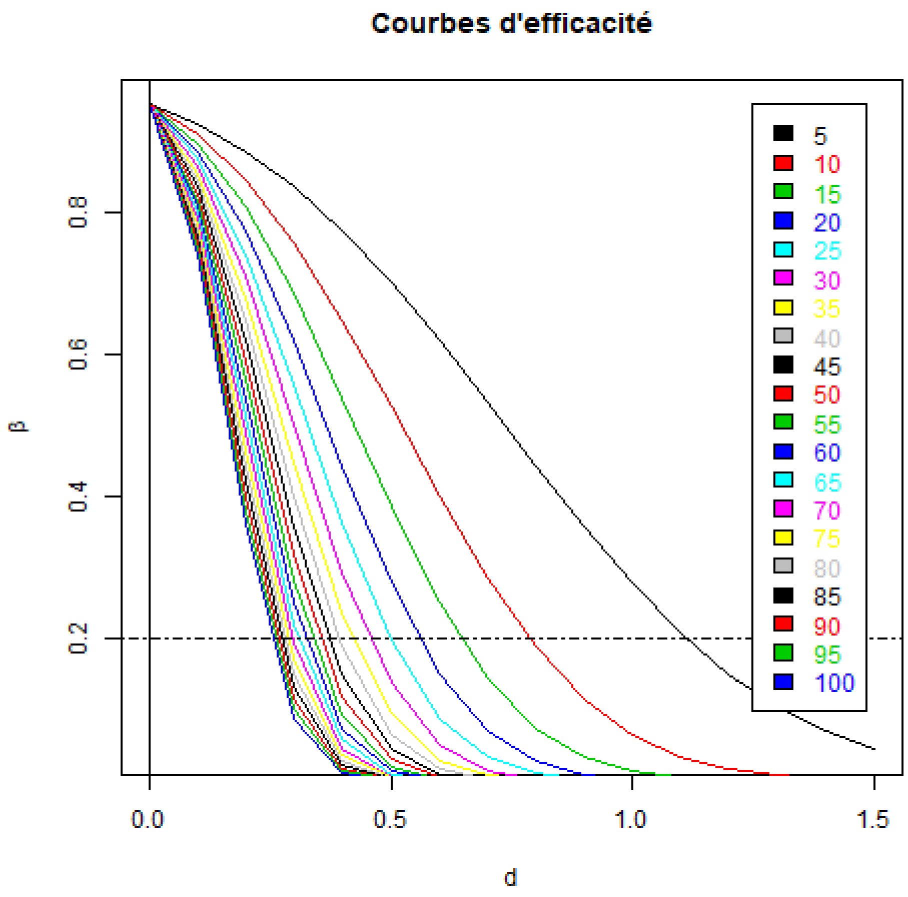
- Du point de vue de la sécurité alimentaire, le risque β est évidemment le plus dangereux. Il dépend:
  - de α
    - α = 5%, 1%, 0.1% arbitrairement (le plus souvent)
    - $\beta \uparrow$ quand $\alpha \downarrow$
  - de la taille de l'échantillon n
    - $\beta \uparrow$ quand $n \downarrow$
  - de l'écart d existant entre conformes et non-conformes
    - $\beta \uparrow$ quand $d \downarrow$

Interprétation des risques (suite)

- Illustration
  - Calculons le risque encouru dans une situation de contrôle quantitatif (gaussien)
    - $d = (\mu_R - \mu_C)/\sigma$
      - μ_R = moyenne de référence
      - μ_C = moyenne de la population contaminée
      - σ = déviation standard (supposée identique dans les 2 pop.)
    - On veut obtenir:
      - $\beta = \beta(d,n,\alpha=0.05)$ = « courbes d'efficacité » pour $\alpha = 0.05$

Interprétation des risques (suite)

- <u>Résultat</u>:



Contrôle statistique d'un processus

- SPC = *Statistical Process Control*
  - Attitude concernant le désir de tous les membres d'une organisation d'améliorer continument la qualité et la productivité.
  - La qualité d'une production est liée à la notion de variabilité (Shewhart):
    - Toute chose étant égale, la production reste stable, et par conséquent, la qualité aussi.
    - Le contrôle de la qualité est donc un exercice de contrôle de la variabilité de la production.

Contrôle statistique d'un processus

- SPC = *Statistical Process Control*
  - La statistique met des outils techniques à disposition pour atteindre les objectifs de contrôle de la qualité
    - Les « 8 magnifiques » sont les outils majeurs
      - Histogrammes
      - Feuille de vérification
      - Graphique de Pareto
      - Diagramme Cause – Effet
      - Diagramme de concentration des défauts
      - Diagramme de dispersion
      - Graphique de contrôle

Contrôle statistique d'un processus

- SPC = *Statistical Process Control*
  - Les objectifs principaux sont donc:
    - Réduire la variabilité des processus
    - Suivi et surveillance d'un processus
    - Estimation des paramètres du produit ou du processus

D'où vient la variabilité ?

- Variabilité commune (*Chance variability*)
  - Variations stochastiques, petites, inhérentes au système, laissant le processus dans un niveau de performance acceptable
    => système « sous contrôle statistique »
- Variabilité spéciale (*Assignable variability*)
  - Variation systématique, grandes, conduisant le processus dans un niveau de performance inacceptable (typiquement, ↑ du # de non-conformités)
    => système « hors de contrôle statistique »

D’où vient la variabilité spéciale?

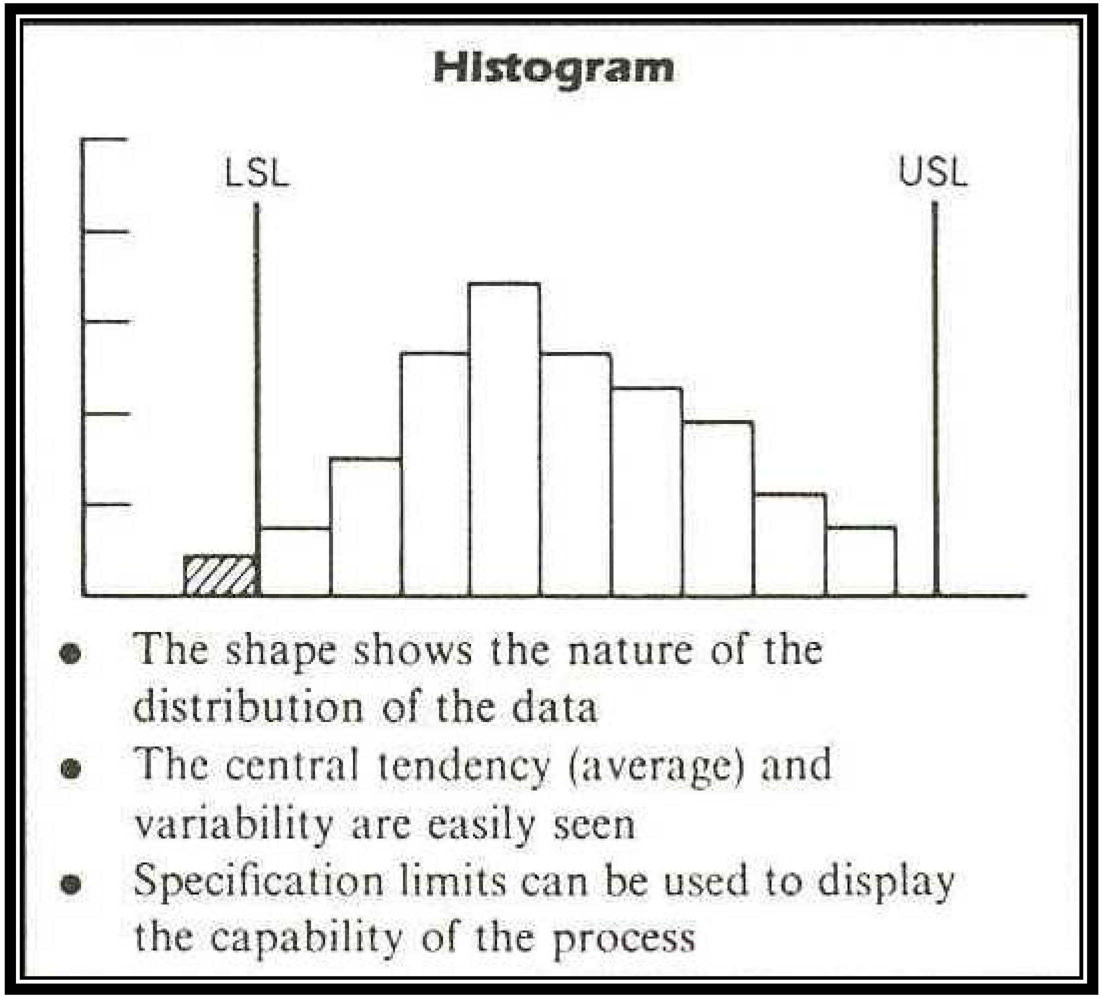
- Causes de variabilité spéciale
  - Machines mal ajustées
    - Ex: taille ou poids d’une production agro-alimentaire
    - Ex: contamination d’une machine/d’un outil dans la chaine de production
  - Erreurs d’opérateurs
    - Ex: erreur de manipulation
    - Ex: erreur de dosage
  - Défaut du matériel de base
    - Ex: présence d’antibiotiques dans la viande de porc

Outils statistiques

- 8 principaux outils statistiques sont utiles
  - l'histogramme (« *Histogram* »)
  - le tableau de contrôle (« *Check sheet* »),
  - le diagramme de Pareto (« *Pareto chart* »),
  - le diagramme des causes et des effets (« *Causes and effects diagram* »),
  - le diagramme de concentration (« *Defects concentration diagram* »),
  - le diagramme de dispersion (« *Scatter diagram* »),
  - le diagramme de contrôle (« *Control chart* »),
  - le design expérimental (« *Design of experiment* »).

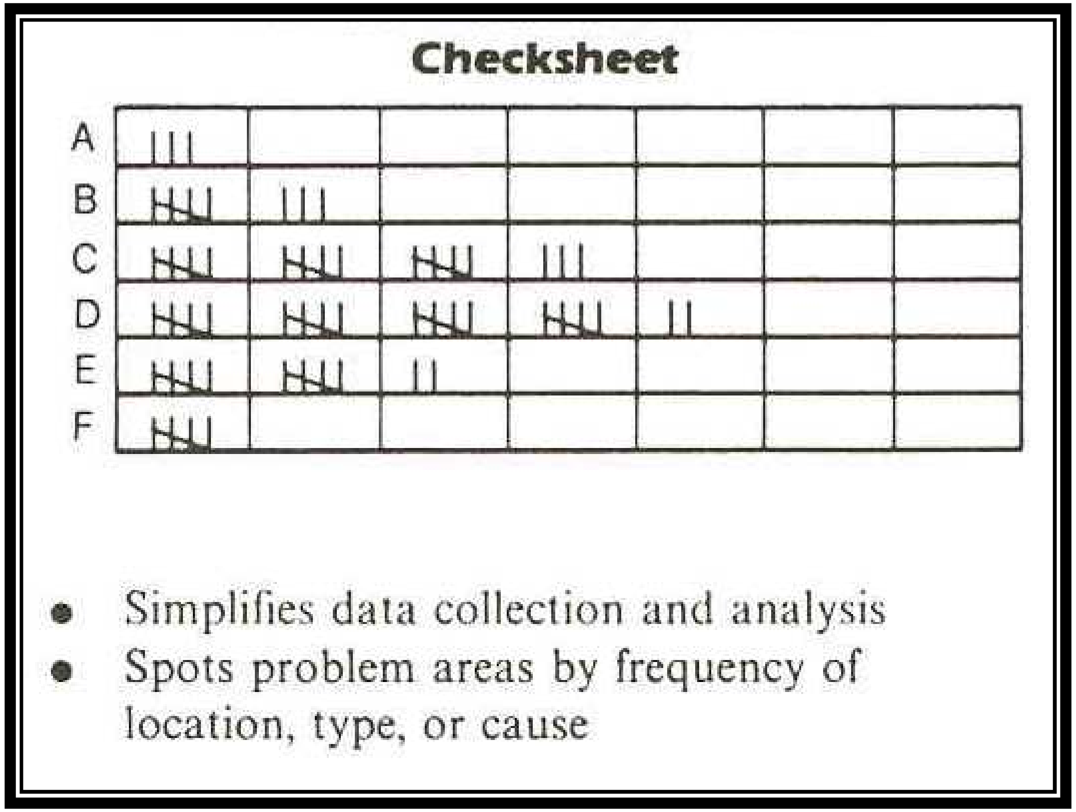
Outils statistiques

- Histogramme



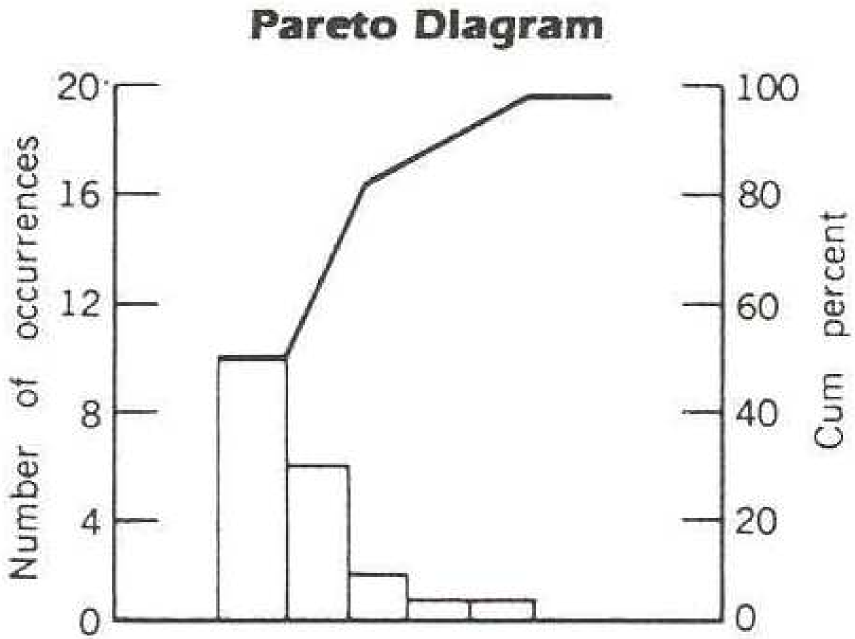
Outils statistiques

- Tableau de contrôle



Outils statistiques

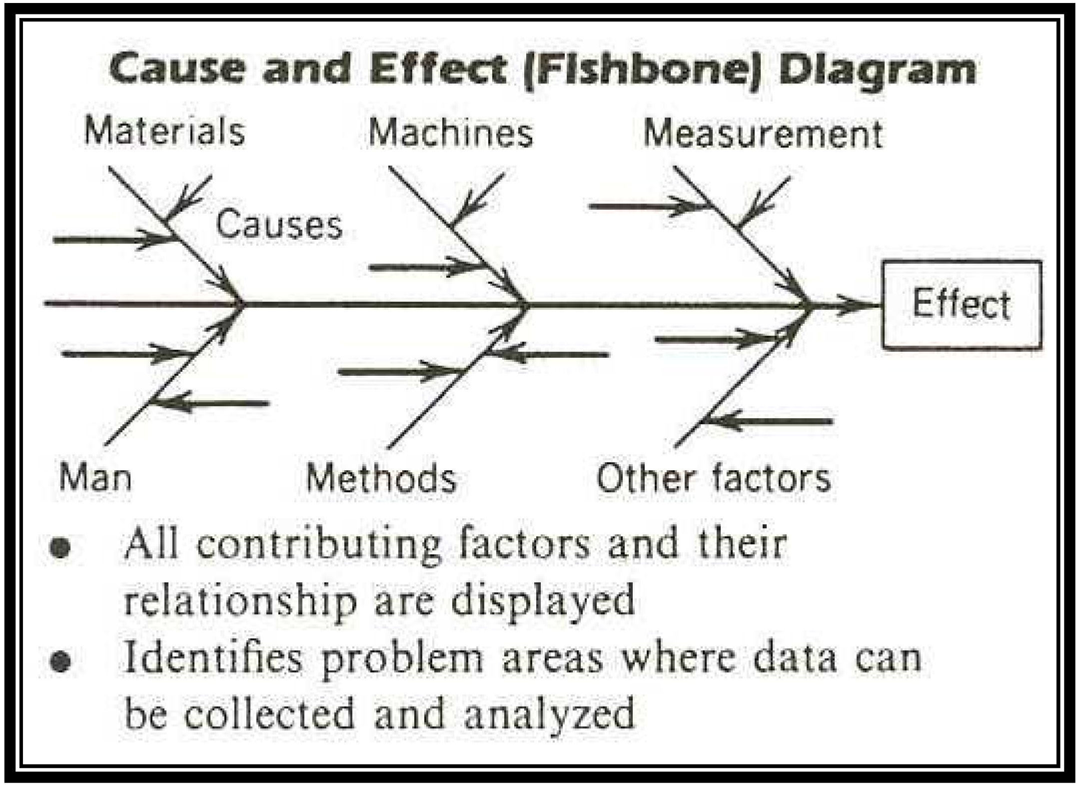
- Diagramme de Pareto



- Identifies most significant problems to be worked first
- Historically 80% of the problems are due to 20% of the factors
- Shows the vital few

Outils statistiques

- Diagramme des causes et effets

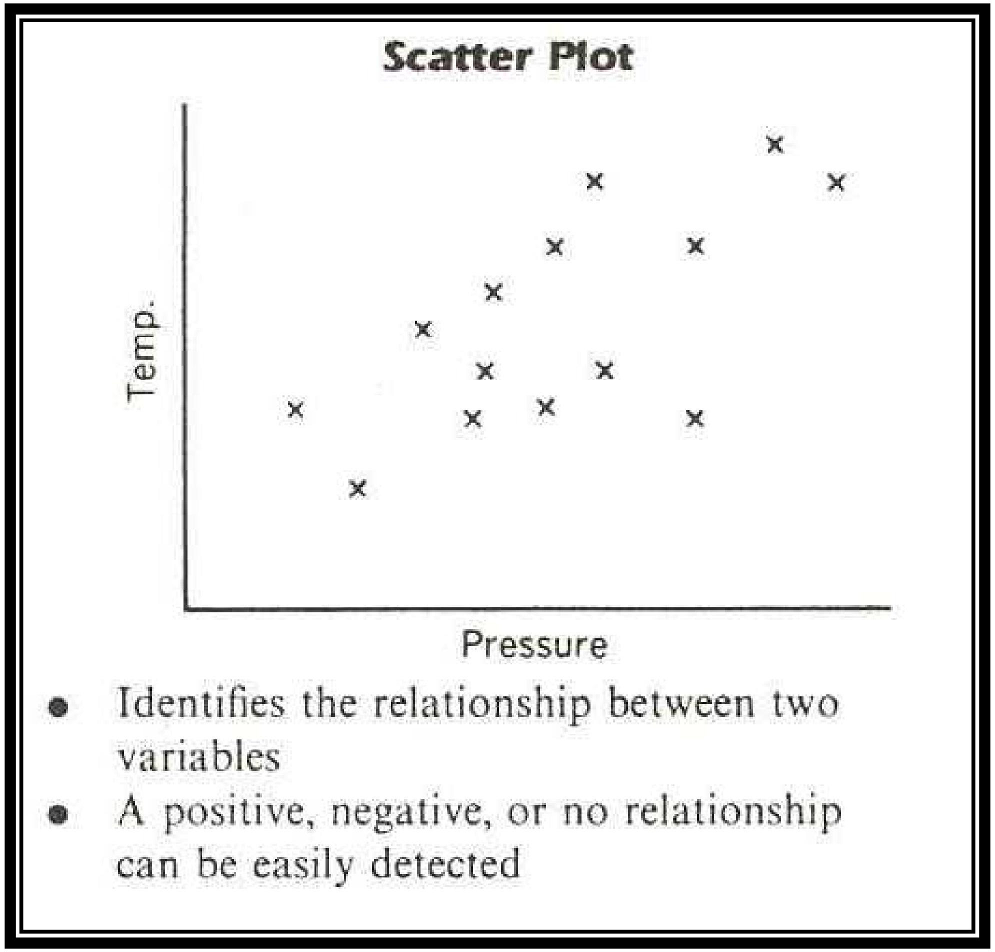


Outils statistiques

- Diagramme de concentration
  - Représentation schématique du produit et localisation des défauts sur le schéma
  - Si l'analyse porte sur un nombre suffisants d'unités, des zones sensibles émergent fréquemment, et permettent de focaliser l'attention pour le futur

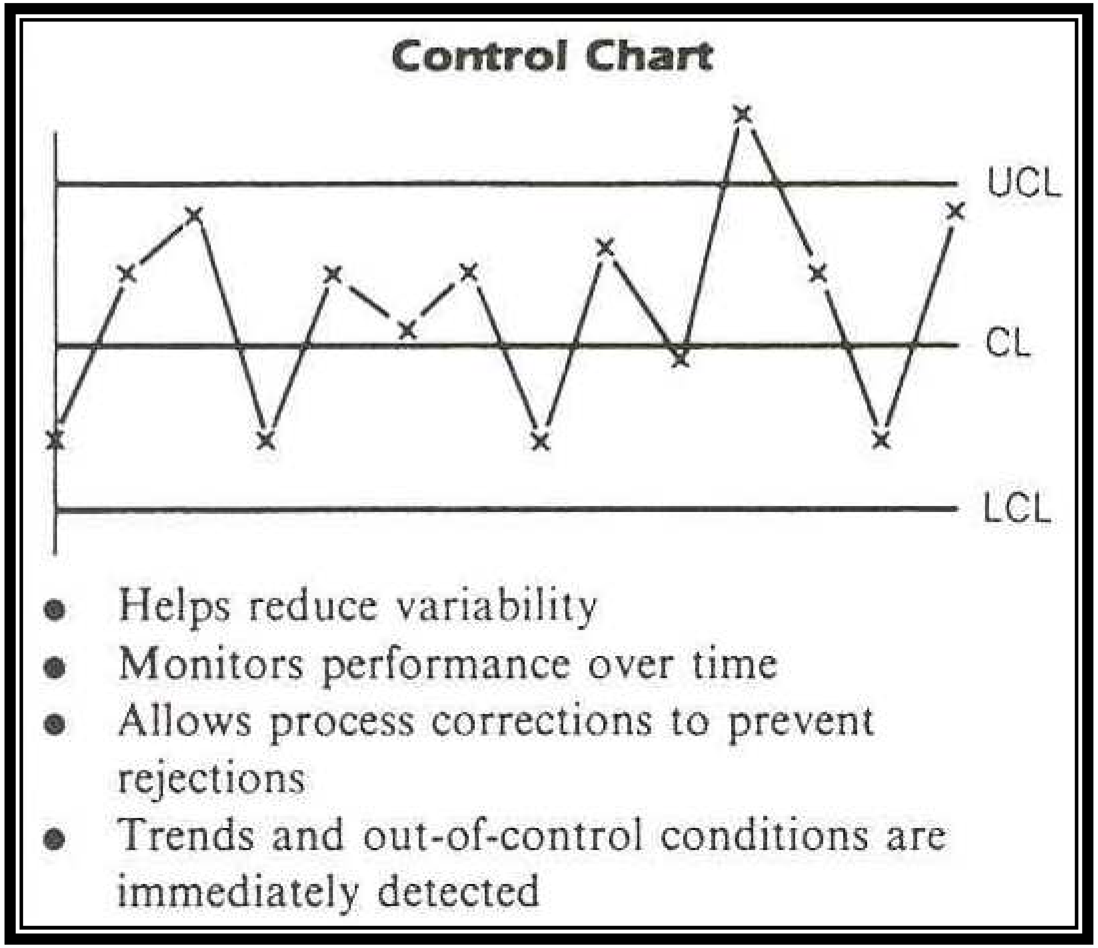
Outils statistiques

- Diagramme de dispersion



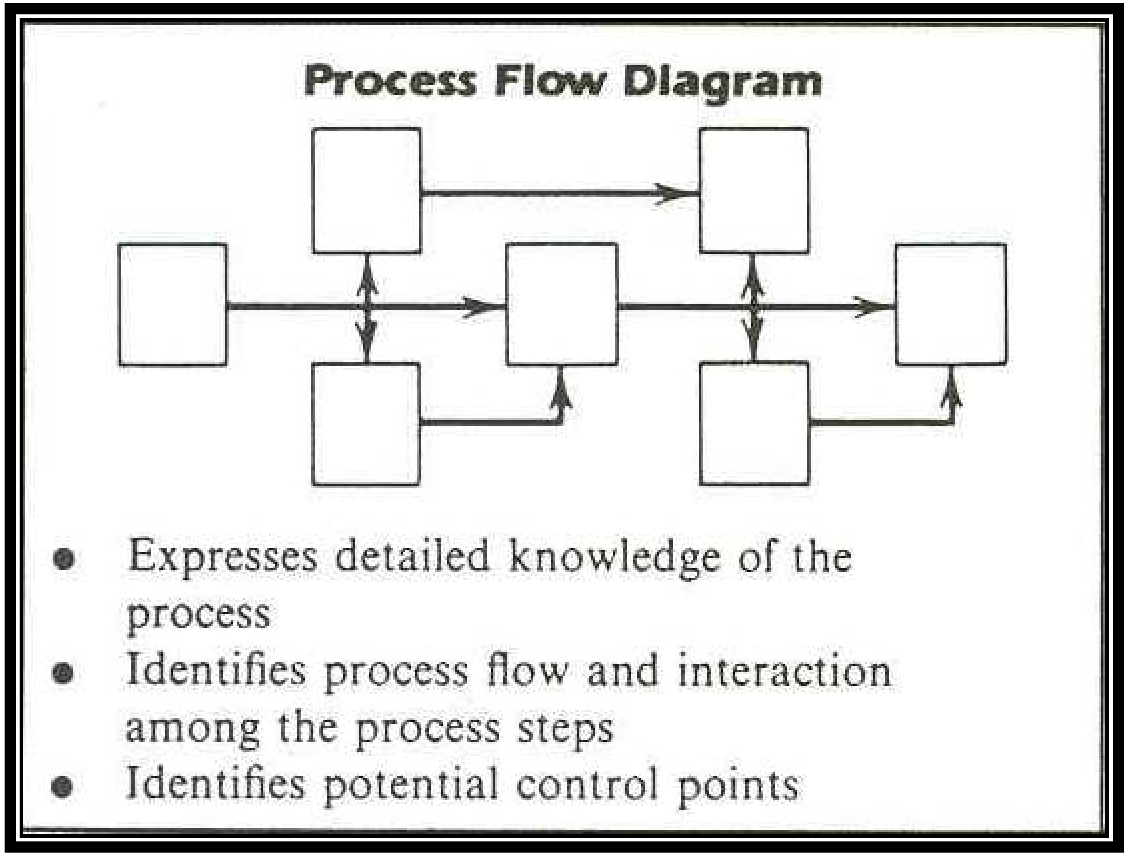
Outils statistiques

- Cartes de contrôle



Outils statistiques

- Design expérimental
  - En préalable:



Outils statistiques

- Design expérimental
  - Ensuite:

Design of Experiments (DOE)

- Useful in process development and troubleshooting
- Identifies magnitude and direction of important process variable effects
- Greatly reduces the number of runs required to perform an experiment
- Identifies interaction among process variables
- Useful in engineering design and development
- Focuses on optimizing process performance

Outils de contrôle

- Deux outils seront plus particulièrement développés ici:
  - Diagramme de Pareto: technique simple de mise en exergue des défauts les plus importants
  - Diagramme de contrôle (Control-chart): technique plus sophistiquée de suivi longitudinal d'un processus
    - aussi appelé « Cartes de contrôle »

Base statistique du graphique de Pareto

- Principe de base
  - Histogramme mettant en évidence les défauts les plus fréquents
  - La fréquence peut être pondérée par le coût lié au défaut pour faire ressortir les défauts les plus couteux (p.e. contaminations…)

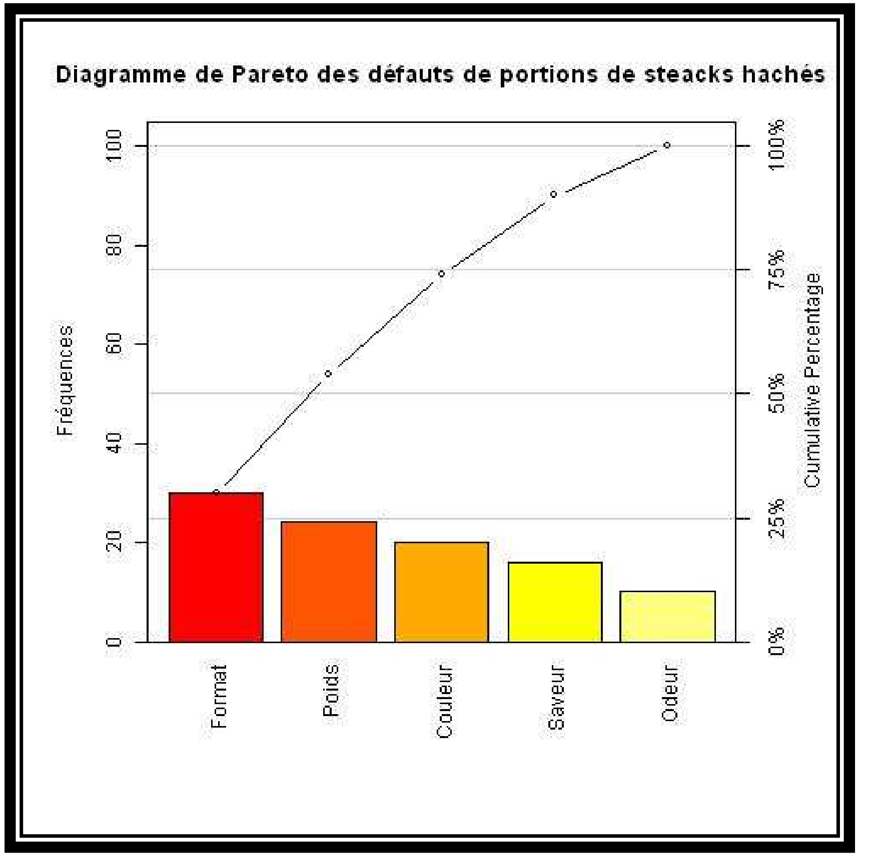
Graphique de Pareto

- Exemple
  - Production de steaks hachés

```
> library(qcc)
> defauts<-c(20,30,10,24,16)
> names(defauts)<-c(«Couleur»,«Format»,«Odeur»,
+ «Poids»,«Saveur»)
> pareto.chart(defauts, ylab="Fréquences",
+  main="Diagramme de Pareto des défauts de
+  portions de steacks hachés")
```

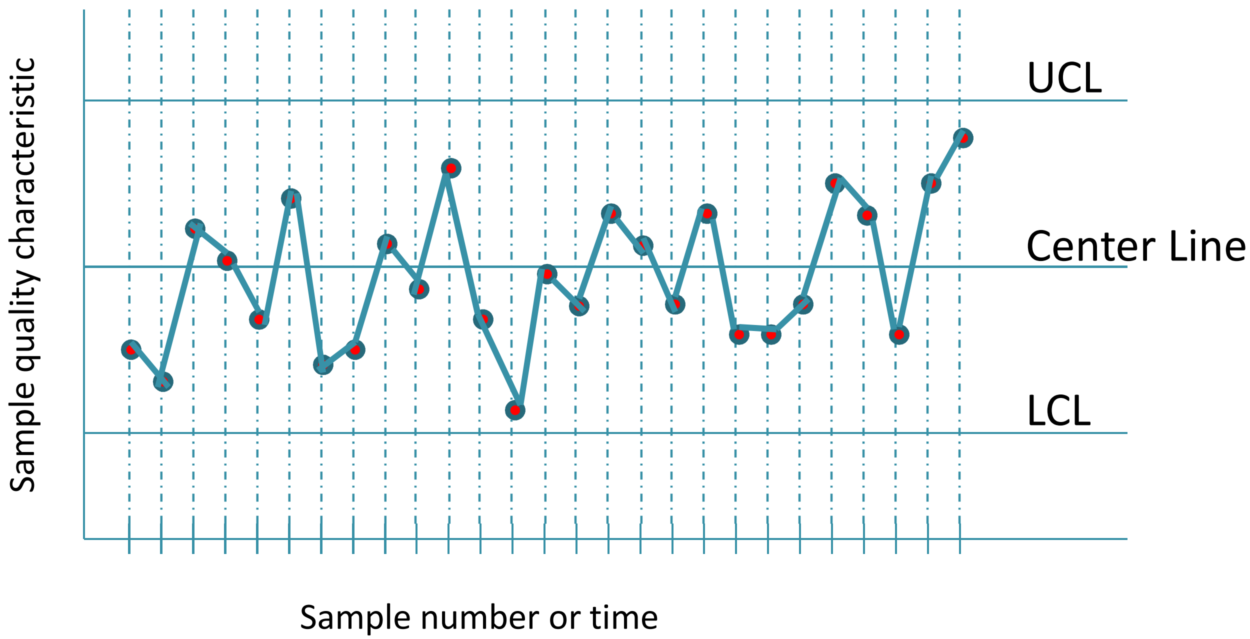
Graphique de Pareto

- Exemple
  - Production de steaks hachés
    - Résultat



Base statistique du graphique de contrôle (*control chart*)

- Principes de base



Base statistique du graphique de contrôle (*control chart*)

- Principes de base (suite)
  - CL (Center Line): représente la valeur moyenne de la caractéristique pour laquelle le processus est « sous contrôle »
  - UCL, LCL: limites (calculées) entre lesquelles la plupart des observations doivent tomber si le processus est « sous contrôle »
  - Le processus est « sous contrôle » si les points restent entre UCL et LCL et qu'il n'y a pas d'évidents écarts à un parcours aléatoire…

Base statistique du graphique de contrôle (*control chart*)

- Principes de base (suite)
  - L'exigence sur le parcours aléatoire des excursions provient de la nature supposée aléatoire des variations communes
  - Les raisons menant à un processus « hors contrôle » peuvent dès lors être détectées, expliquées, et, si possible, éliminées.

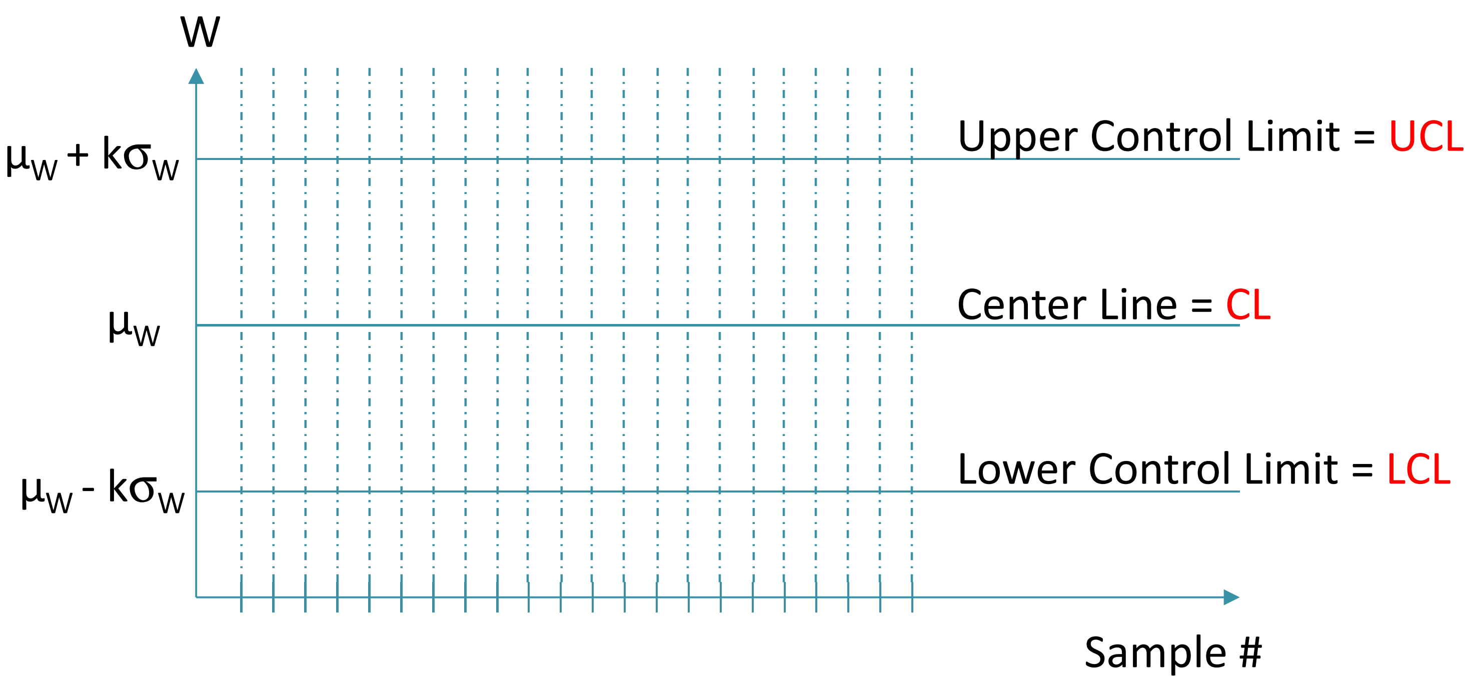
Base statistique du graphique de contrôle (*control chart*)

- Du point de vue statistique:
  - Control chart $\Longleftrightarrow$ Test d'hypothèse
    H_0: processus sous contrôle
  - En particulier, le paradigme des tests d'hypothèses s'applique ici:

	Accepte H_0	Rejette H_0
H_0 OK	OK	α
H_1 OK	β	OK

Base statistique du graphique de contrôle (*control chart*)

- Graphique de contrôle de Shewhart



Base statistique du graphique de contrôle (*control chart*)

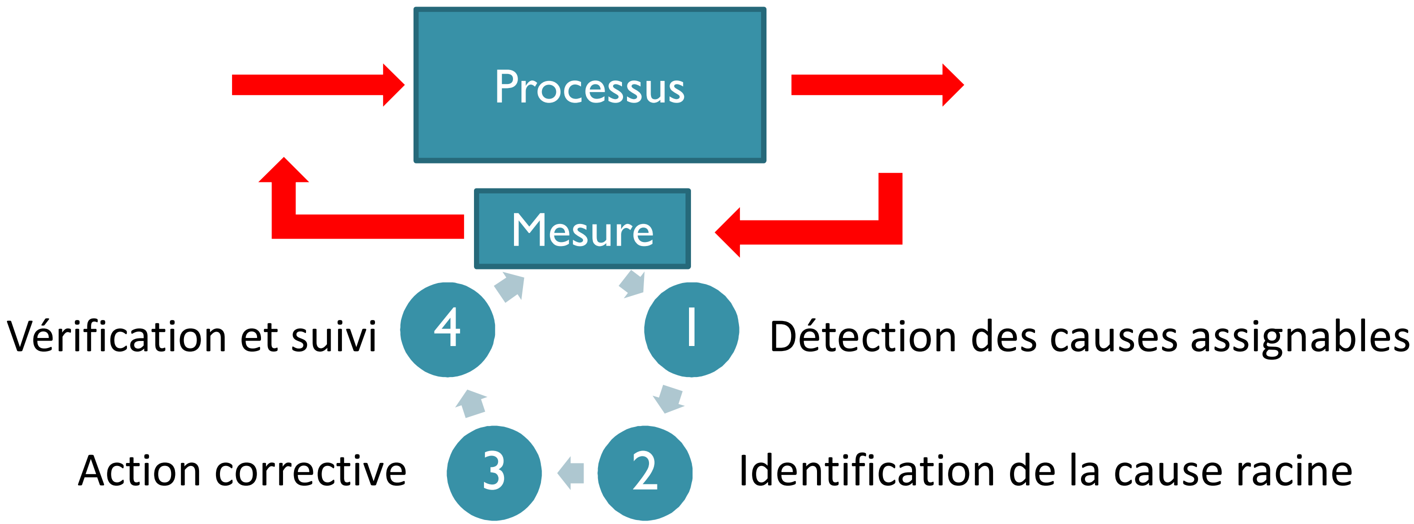
- Graphique de contrôle de Shewhart
  - Principes posés par Dr. Walter A. Shewhart (Bell labs) dans les années 1920
  - w = statistique décrivant une caractéristique de qualité d'intérêt
    - exemple 1: poids d'une portion de viande hachée
    - exemple 2: « conformité » d'une production à un standard pré-fixé.
  - μ_W = valeur attendue de la caractéristique si le processus est « sous contrôle »

Base statistique du graphique de contrôle (*control chart*)

- Graphique de contrôle de Shewhart
  - k = « distance » entre limites de contrôle et ligne du centre, en déviation standard
  - Le graphique de contrôle est un test répété de l'hypothèse nulle H_0: $E(w) = \mu_W$ à un seuil spécifié par la valeur de k.

Base statistique du graphique de contrôle (*control chart*)

- Objet du graphique de contrôle
  - Surveillance « on-line » d'un processus
    - Identification des causes assignables
      - Elimination des causes assignables (« root causes »)
      - Amélioration du processus



Base statistique du graphique de contrôle (*control chart*)

- Objet du graphique de contrôle
  - Estimation des paramètres du processus « sous contrôle »
    - exemples: moyenne, proportion, déviation standard
    - Capacités du système (« process capabilities »)
      - Construire ou acheter (« make or buy »)
      - Accords contractuels avec les clients…

Base statistique du graphique de contrôle (*control chart*)

- Types de graphiques de contrôle
  - Graphiques de contrôle sur variables
    - La caractéristique de qualité mesurée est un nombre sur une échelle continue de mesures. La mesure est alors représentée par une mesure de position et une mesure de dispersion
      - Exemple: graphique de contrôle $\overline{x}$, graphique de contrôle R
  - Graphiques de contrôle sur attributs
    - On mesure si l'unité produite est « conforme », ou éventuellement le nombre de défauts par unité
      - Exemple: graphique de contrôle p (proportion de conformes)

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  1. Choix des limites de contrôle
  2. Sélection de la taille des échantillons
  3. Fréquence d'échantillonnage
- Les choix se font sur des bases
  - statistiques ($\downarrow \beta$, $\uparrow$ P, …)
  - économiques (coût de l'échantillonnage, coût d'une perte de contrôle, …)

Base statistique du graphique de contrôle (*control chart*)

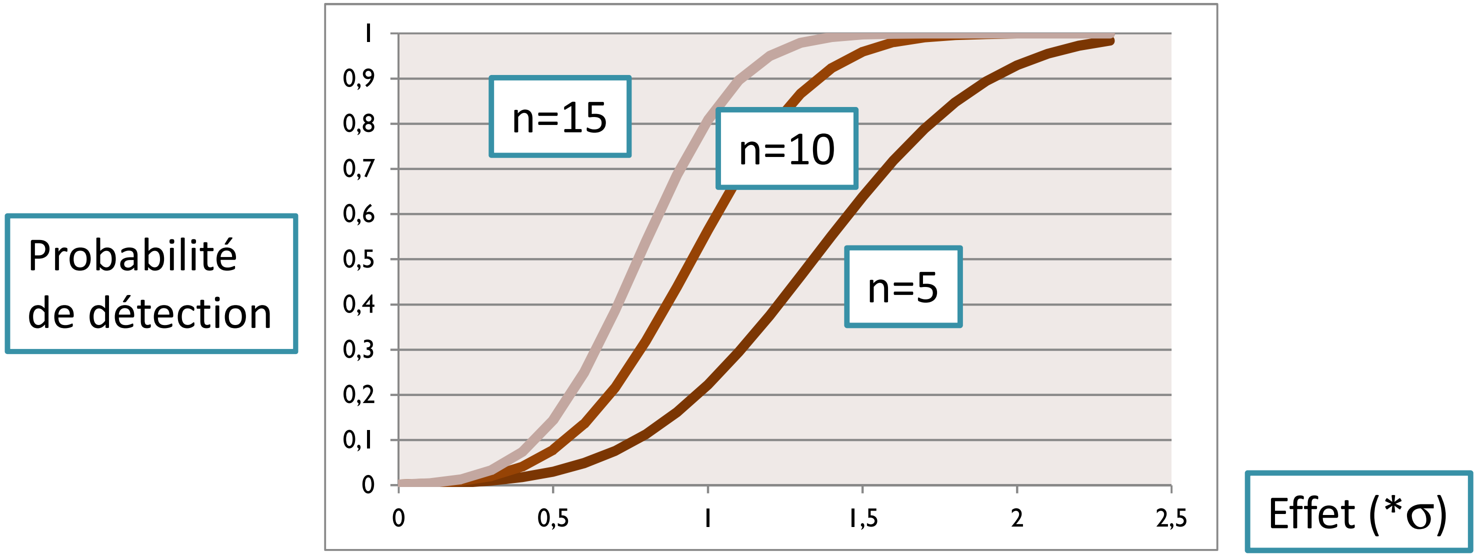
- Design de graphiques de contrôle
  - 1) Choix des limites de contrôle
    - Typiquement: $k = 3$
      - Si $W \sim N(\mu_W, \sigma_W)$, $k = 3 \Rightarrow \alpha = 0.0027$, et donc 27 fausses alarmes seront déclenchées tous les 10000 points en moyenne
    - $\uparrow(\downarrow)$ $k \Rightarrow \downarrow(\uparrow)$ α et $\uparrow(\downarrow)$ β
      - On peut également travailler directement sur une valeur donnée de α (p.e. $\alpha = 0.001 \Rightarrow k = 3.09$) plutôt que sur k
    - Si un point tombe hors des limites, on recherche la cause assignable, et on essaie de la corriger.

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Choix des limites d'avertissement
    - Typiquement: k = 2
      - Si W ~ $N(\mu_W,\sigma_W)$, k = 2 => α = 0.0455
    - Si un point tombe hors des limites d'avertissement (mais dans les limites de contrôle), on peut se poser des questions sur le bon fonctionnement du système
      - Une action pourrait être d'augmenter la fréquence d'échantillonnage

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - 2) Taille de l'échantillon
    - La probabilité de détecter une perte de contrôle augmente avec la taille de la perturbation et la taille de l'échantillon



Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Taille de l'échantillon
    - La taille de l'échantillon peut être choisie en fonction de l'amplitude potentielle de l'écart, ou en fonction de la valeur de l'écart pour laquelle une action corrective doit être impérativement menée.

Base statistique du graphique de contrôle (*control chart*)

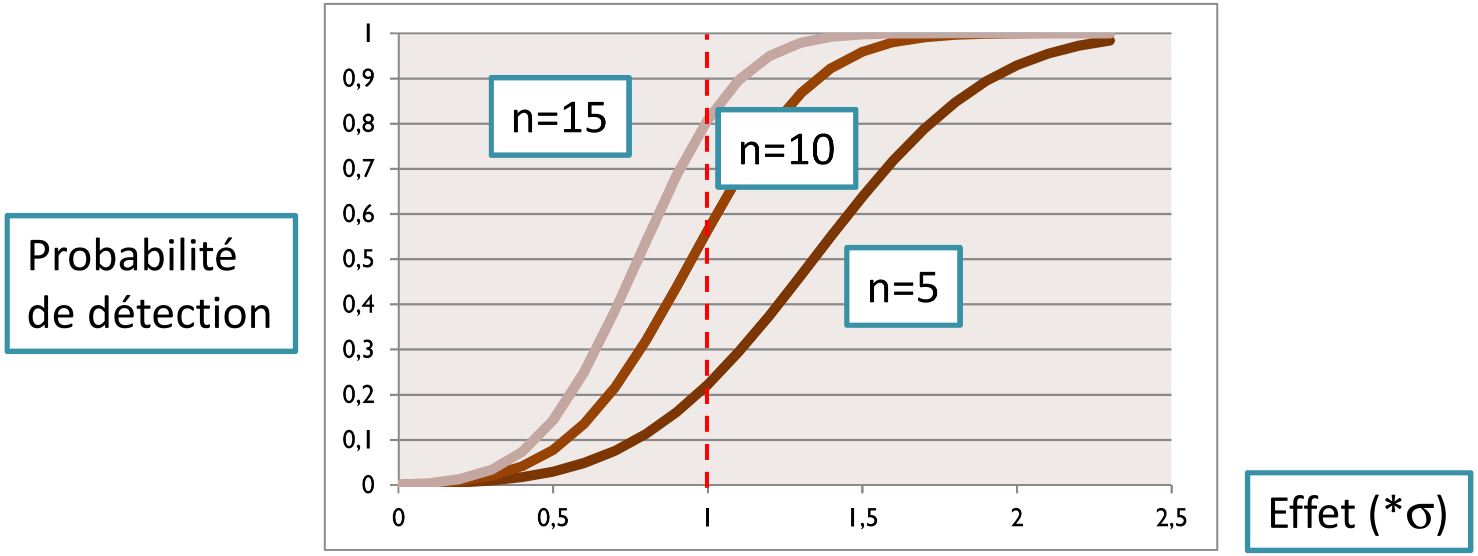
- Design de graphiques de contrôle
  - 3) Fréquence de l'échantillonnage
    - Idéalement:
      - prendre de grands échantillons fréquemment, mais cela a un coût !
    - En pratique:
      - compromis taille – fréquence
      - dépend du problème (coût de l'échantillonnage…)

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Fréquence de l'échantillonnage
    - Une approche:
      - longueur de *run* moyen (*average run length* = ARL)
        ARL est le nombre moyen de points à représenter avant qu'un point n'indique une condition de « perte de contrôle »
      - Calcul: ARL = 1/p
      - Exemple: k = 3 => p = 0.0027 => ARL = 370
      - Interprétation: en moyenne, il y aura un point indiquant une perte de contrôle tous les 370 échantillons si le processus est effectivement sous contrôle (« faux positif »)

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Fréquence de l'échantillonnage
    - Exemple: supposons un effet de $1*\sigma$



Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Fréquence de l'échantillonnage
    - <u>Exemple</u>: supposons un effet de $1^*\sigma$
      - si n = 5, p ≈ 0.2 et ARL ≈ 5
      - si n = 10, p ≈ 0.55 et ARL ≈ 2
      - si n = 15, p ≈ 0.8 et ARL ≈ 1.2
    - Donc, si les échantillons sont petits, il faut ↑ la fréquence d'échantillonnage pour ↓ le nombre d'unités produites hors contrôle
    - Mais…
      - Cfr dia suivante

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Fréquence de l'échantillonnage
    - Pour décider, il faudra tenir compte:
      - du coût d'échantillonnage
      - des pertes associées à la production d'unités dans un état du système « hors contrôle »
      - de la probabilité associées à une dérive donnée du système

Base statistique du graphique de contrôle (*control chart*)

- Design de graphiques de contrôle
  - Modalités d'échantillonnage
    - Sous-groupes « rationnels »
      - Intégration de facteurs qui maximisent la chance que les différences entre sous-groupes dues à des causes assignables soient grandes, tout en minimisant la chance que des différences intra-groupes soient importantes.
      - <u>Exemple</u>: effet du temps, de la température, de l'opérateur…
      - <u>Cfr</u>: design de l'expérience

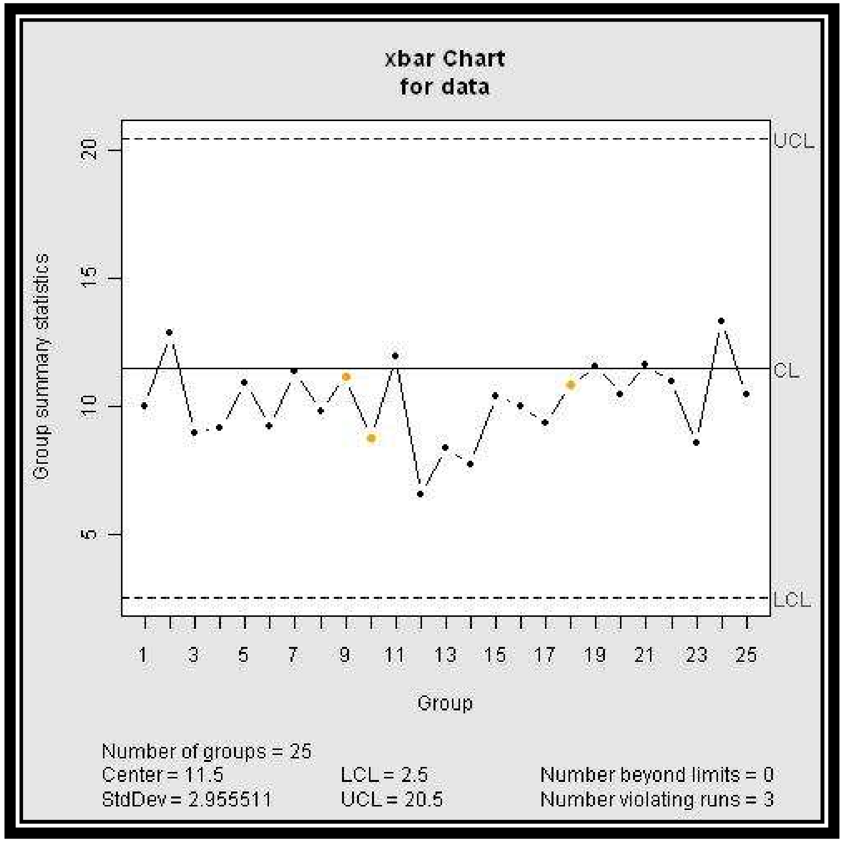
Base statistique du graphique de contrôle (*control chart*)

- L'analyse de *patterns* dans les graphiques de contrôle
  - Deux signaux de perte de contrôle
    - 1 ou >1 points hors des « limites de contrôle »
    - *pattern* non aléatoire de points du graphique de contrôle
      - présence de « runs » (séquence d'observations de même type) anormaux
      - comportement cyclique
      - …

Base statistique du graphique de contrôle (*control chart*)

- L'analyse de *patterns* dans les graphiques de contrôle
  - Exemple de « *run* »

20 points sur 25 < CL
=> $\chi^2 = 9$
=> $p[\chi^2(1) = 9] = 0.0027$

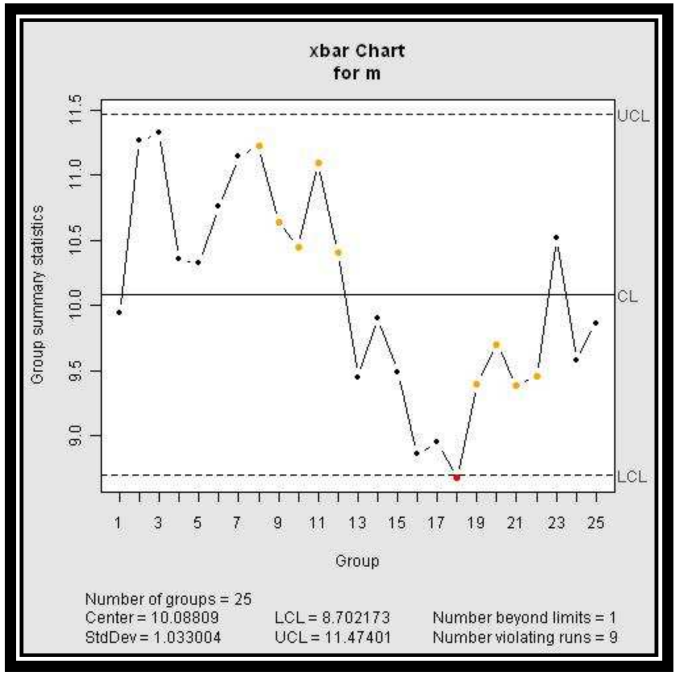


Base statistique du graphique de contrôle (*control chart*)

- L'analyse de *patterns* dans les graphiques de contrôle
  - Exemples de « *run* »
    - # anormal de points du même coté de CL (cfr dia précédente)
    - une suite de ≥ 8 points avec des valeurs croissantes (« *run up* »), décroissantes (« *run down* »), ou du même coté de CL

Base statistique du graphique de contrôle (*control chart*)

- L'analyse de *patterns* dans les graphiques de contrôle
  - Exemples de comportement cyclique



Base statistique du graphique de contrôle (*control chart*)

- Résumé des règles d'analyse des CC
  - ≥ 1 point hors des limites de contrôle ($CL \pm 3*\sigma$)
  - Un *run* de ≥ 8 points
  - 2 points parmi 3 consécutifs hors des limites d'avertissement ($CL \pm 2*\sigma$)
  - 4 points parmi 5 consécutifs hors des limites ($CL \pm 1*\sigma$)
  - Pattern inhabituel ou non-aléatoire
  - …

Cartes de contrôles pour attributs

- Trois types de CC pour attributs
  - p-chart: CC pour fraction non-conforme
  - c-chart: CC pour nombre de non-conformités
  - u-chart: CC pour nombre de non-conformités par unité
- Dans ce cours, nous développerons essentiellement les p-charts

CC pour fraction non-conforme (*p-chart*)

- Utilisation de la loi binomiale
  - Dans un système sous-contrôle, la proportion non-conforme vaut p
  - Si n unités sont produites, r seront non-conformes:
    - $r \sim$ Binomiale(n,p)
      - $E(r) = n*p$ et $V(r) = n*p*(1-p)$
    - $p_{est} = r/n$
      - $E(p_{est}) = E(r)/n = p$ et $V(p_{est}) = p*(1-p)/n$

CC pour fraction non-conforme (*p-chart*)

- Design du CC – p connue
  - On utilise les limites habituelles:
    - UCL $= \mu_W + k^*\sigma_W$
    - CL $= \mu_W$
    - LCL $= \mu_W - k^*\sigma_W$
  - Dans le cas présent, en utilisant k = 3:
    - UCL $= p + 3^*[p^*(1\text{-}p)/n]^{0.5}$
    - CL $= p$
    - LCL $= p - 3^*[p^*(1\text{-}p)/n]^{0.5}$

CC pour fraction non-conforme (*p-chart*)

- Design du CC – p <u>connue</u>
  - <u>Remarques</u>:
    - La valeur de p peut être un objectif (« target value » = « valeur cible ») plutôt qu’une valeur connue.
    - Il est possible que le système soit « hors contrôle » pour la valeur cible, mais « sous contrôle » pour une autre valeur

CC pour fraction non-conforme (*p-chart*)

- Design du CC – p inconnue
  - On utilise la même approche, mais p est estimée sur m échantillons préliminaires
    - $\overline{p}$ = mean(p_i)
  - Dans le cas présent, en utilisant k = 3:
    - UCL $= \overline{p} + 3*[\overline{p}*(1-\overline{p})/n]^{0.5}$
    - CL $= \overline{p}$
    - LCL $= \overline{p} - 3*[\overline{p}*(1-\overline{p})/n]^{0.5}$
  - Ces limites sont appelées « trial CL »

CC pour fraction non-conforme (*p-chart*)

- Design du CC – p inconnue
  - <u>Remarques</u>:
    - lors de la construction du CC, il est possible que certains des échantillons préliminaires tombent en dehors de l'intervalle admissible
      - Si le nombre de ces points est faible, élimination des points « hors contrôle » puis re-calcul des « trial control limits »
      - Si le nombre de ces points est élevé, recherche d'une cause assignable, élimination, puis re-calcul des « trial control limits »

CC pour fraction non-conforme (*p-chart*)

- Design du CC
  - Taille des échantillons et fréquence d'échantillonnage
    - Guidés en général par des considérations économiques
      - Souvent, on inspecte 100% de la production sur une période donnée représentative pour servir de base au CC
      - Certaines règles ont été suggérées pour le choix de n (cfr 2 dias suivantes)

CC pour fraction non-conforme (*p-chart*)

- Design du CC
  - Règle de Duncan (1974) - Taille des échantillons
  - « ...choose n such that the probability of detecting a process shift of a specified amount is 50%... »
    - <u>Exemple</u>: shift de 0.01 à 0.05

P=0.5 => UCL = CL’
= CL + k * √ 0.01*0.99/n

=> UCL-CL = 0.04 = k* √ 0.01*0.99/n

=> n = 56 si k = 3

CC pour fraction non-conforme (*p-chart*)

- Design du CC
  - Si p est petit (p.e: $p = 0.01$)
    - n petit peut mener à des signaux intempestifs

      <u>exemple</u>: $n = 8 \Rightarrow UCL = 0.01+3*\sqrt{0.01*0.99/8}$
      $= 0.1155$
      => 1 seul non-conforme dans l'échantillon donne un signal ($p_{est} = 1/8 = 0.125 > UCL$...)

    - <u>Une solution</u>: choisir n tel que $LCL > 0$, pour être sûr d'avoir la possibilité d'investiguer des situations avec un nombre trop faible de non-conformes

CC pour fraction non-conforme (*p-chart*)

- Design du CC
  - Si p est petit (suite)
    - LCL = $p - k * \sqrt{[p*(1-p)/n]} > 0$
      => $n > k^2 * (1-p)/p$
    - <u>exemple</u>: $p = 0.05$ et $k = 3$ => $n > 171$

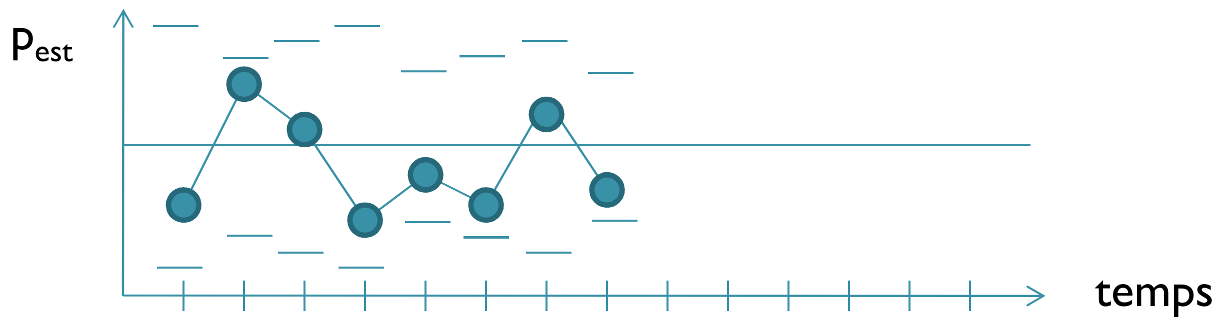
CC pour fraction non-conforme (*p-chart*)

- Remarques sur l'utilisation des *p-charts*
  - Le choix d'une binomiale implique qu'on suppose que:
    - la probabilité d'occurrence de défauts est constante
    - les unités successives de production sont indépendantes…(clustering)

CC pour fraction non-conforme (*p-chart*)

- Remarques sur les *p-charts* (suite)
  - Il arrive aussi que la taille des échantillons varie, ce qui conduit à des limites variables en fonction de chaque échantillon

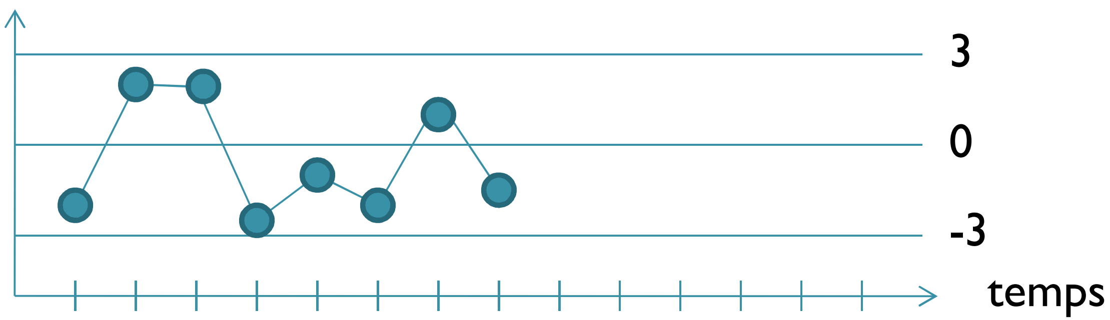
$$[LCL;UCL] = CL \pm k^*\sqrt{p^*(1-p)/n_i}$$



CC pour fraction non-conforme (*p-chart*)

- Remarques sur les *p-charts* (suite)
  - Toutefois, l'interprétation (runs, notamment) est malaisée, des écarts plus petits par rapport à CL pouvant correspondre en réalité à des excursions plus importantes
    => utilisation de p-chart standardisé
    où l'écart est exprimé en déviations standards

$$z = \frac{p_i - p}{\sqrt{\frac{p*(1-p)}{n}}}$$



CC pour fraction non-conforme (*p-chart*)

- Une alternative aux *p-charts*: *np-charts*
  - Plutôt qu'utiliser la « fraction non-conforme », il est parfois plus parlant d'utiliser le « nombre non-conformant »
  - On utiliserait alors, en prenant k = 3:
    - UCL $= n*p + 3*[n*p*(1-p)]^{0.5}$
    - CL $= n*p$
    - LCL $= n*p – 3*[n*p*(1-p)]^{0.5}$

CC pour fraction non-conforme (*p-chart*)

- Exercice:
  - Suivi des défauts dans des plants de tomates

```
> data<-read.table(file=«tomates.txt»,
+                          sep=«\t»,
+                          head=T)
> names(data)
[1] "sample"    "non.conf" "fraction"
```

  - Estimation de p sur les 30 premiers échant.

```
> attach(data)
> p<-mean(non.conf[1:30])/50
> p
[1] 0.2313333
```

CC pour fraction non-conforme (*p-chart*)

- Exercice (suite):
  - Calcul de LCL et UCL

```
> ucl<-p+3*sqrt(p*(1-p)/50)
> lcl<-p-3*sqrt(p*(1-p)/50)
> c(lcl,ucl)
[1] 0.05242755 0.41023912
```

  - Points « hors contrôle » ?

```
> trial<-non.conf[1:30]/50
> trial[trial>ucl]
[1] 0.44 0.48
> trial[trial<lcl]
numeric(0)
```

CC pour fraction non-conforme (*p-chart*)

- Exercice (suite):
  - Affichage du p-chart
    - Installation du package « qcc » (si nécessaire)

```
> utils:::menuInstallPkgs()
```

- Un premier menu apparait avec une liste de « miroirs » (sites avec un dépôt des packages)
- Après sélection du « miroir », une deuxième liste apparait avec la (longue) liste des packages disponibles. Choisir « qcc »
- Le package s'installe alors automatiquement !

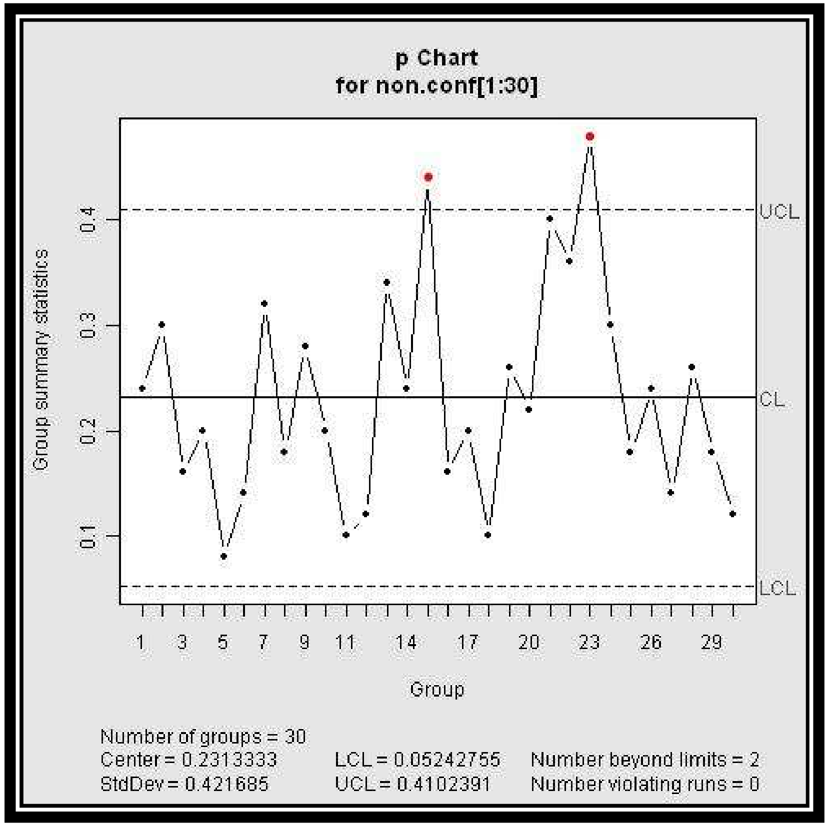
CC pour fraction non-conforme (*p-chart*)

- Exercice (suite):
  - Affichage du p-chart

```
> sizes<-rep(50,30)
> qc<-qcc(non.conf[1:30],sizes,type=«p»,plot=«true»)
```

CC pour fraction non-conforme (*p-chart*)

- Exercice (suite):
  - Résultat



CC pour fraction non-conforme (*p-chart*)

- « Operating characteristic »

$$OC \equiv \beta = f(p)$$

d'une p_chart dont les paramètres sont:

  - n = 50, LCL = 0.0303, UCL = 0.3697

CC pour fraction non-conforme (*p-chart*)

- « Operating characteristic »
  - Il s'agit d'une mesure de la sensibilité du CC. par définition:

$$OC \equiv \beta = f(p)$$

    - On utilise les distributions cumulées pour calculer

$$\beta = P(p_{est} < UCL|p) - P(p_{est} \leq LCL|p)$$

ou $$\beta = P(D < n^*UCL|p) - P(D \leq n^*LCL|p)$$

Exemple avec EXCEL

CC pour fraction non-conforme (*p-chart*)

- « Operating characteristic » - exemple
  - Reprenons les données vues plus haut

```
> Dmax<-50*ucl
> Dmin<-50*lcl
> c(Dmin,Dmax)
[1]  2.621377 20.511956
```

  - Les valeurs acceptables pour D vont donc de 3 à 20, et:

$$\beta = P(D \leq 20) - P(D \leq 2)$$

CC pour fraction non-conforme (*p-chart*)

- « Operating characteristic » - exemple
  - Calculons $\beta = f(p)$ pour p dans [0;1]

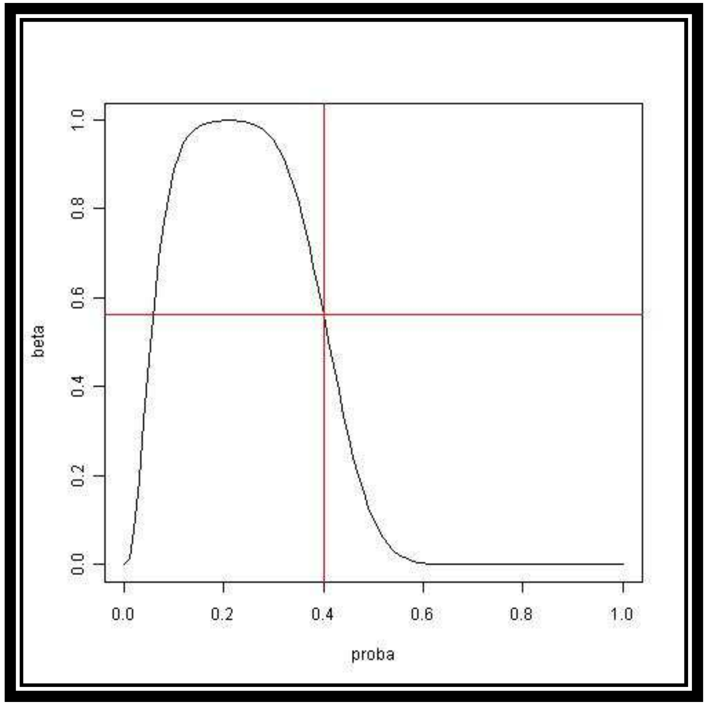
```
> proba<-seq(0,1,0.01)
> beta<-pbinom(20,50,proba)-pbinom(2,50,proba)
> plot(proba,beta,type="l")
> abline(v=0.4,col=«red»)
> abline(h=beta[41],col=«red»)
```

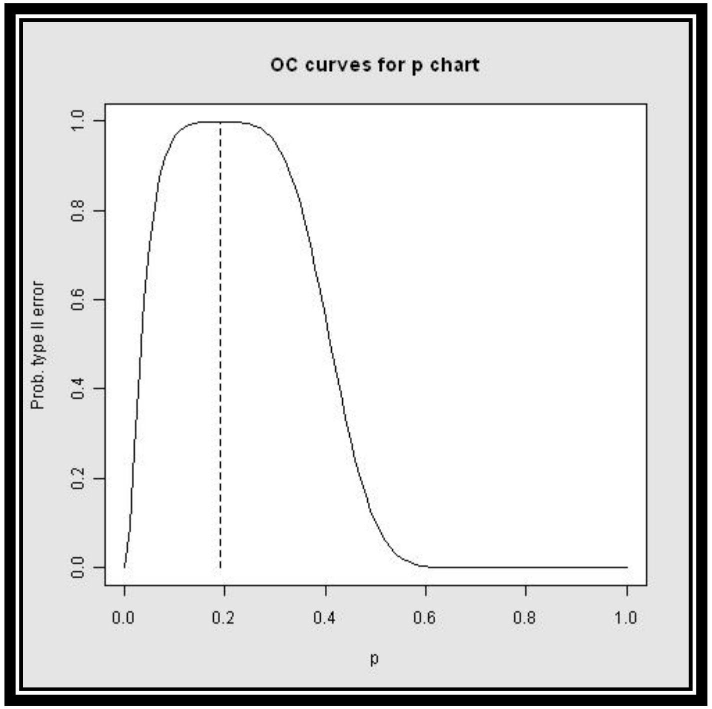
  - Plus simplement…

```
> oc.curves(qc)
```

CC pour fraction non-conforme (*p-chart*)

- « Operating characteristic » - exemple
  - Solution 1 Solution 2





CC pour fraction non-conforme (*p-chart*)

- « Average Run Length » - sous contrôle
  - Par définition:

    ARL = 1/P(signal « out-of-control ») = $1/\alpha$

  - C'est donc le nombre de points moyens avant d'obtenir un signal de perte de contrôle (faux positif)

Exemple avec EXCEL

CC pour fraction non-conforme (*p-chart*)

- « Average Run Length » - hors contrôle
  - Par définition:

$$ARL = 1/P(\text{signal « out-of-control »}) = 1/(1-\beta)$$

  - C'est donc le nombre de points moyens avant d'obtenir un signal de perte de contrôle (vrai positif)

Exemple avec EXCEL

CC pour nombre de non-conformités (*c-chart*)

- Utilisation de la loi de Poisson
  - Dans un système sous-contrôle, le nombre moyen de produits non-conformes vaut c
  - Dans une unité d'inspection (p.e. jour, lot, …), le nombre x de produits non-conformes varie de 0 à un nombre arbitraire (théoriquement ∞).
  - Si on suppose:
    - p(défaut) petite (et donc, P(x → ∞) → 0)
    - p(défaut) constante

CC pour nombre de non-conformités (*c-chart*)

- Utilisation de la loi de Poisson (suite)
  - Alors, x peut être modélisé par une loi de Poisson:

    $$P(x) = \exp(-c)*c^x/x!$$

  - Le CC, appelé c-chart, est alors défini par:
    - UCL $= c + 3*c^{0.5}$
    - CL $= c$
    - LCL $= c - 3*c^{0.5}$
    - Comme pour le p-chart, si c est estimé, on parlera alors de « trial control limits »

CC pour nombre de non-conformités (*c-chart*)

- Exemple: production de circuits imprimés
  - *c-chart*

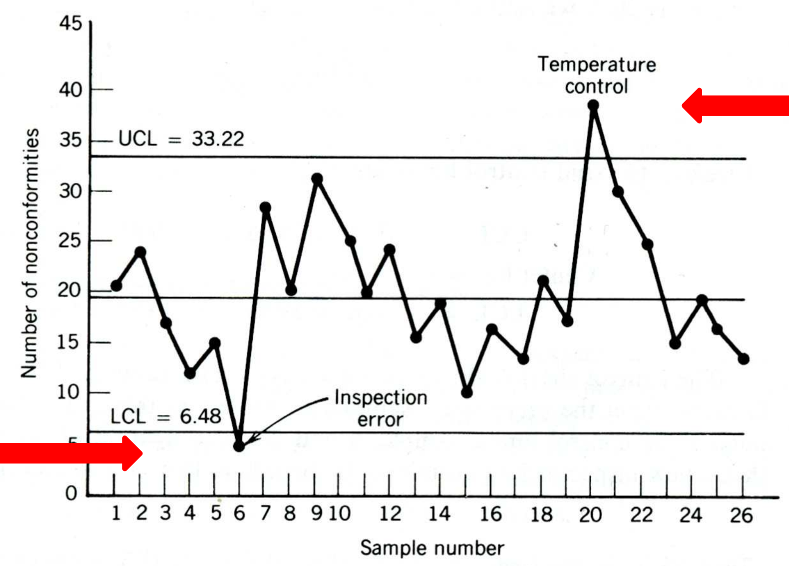


Figure 5-10
Control chart for nonconformities for Example 5-3.

Problème de contrôle de la T° (corrigé)

Inspecteur inexpérimenté (corrigé)

CC pour nombre de non-conformités (*c-chart*)

- Exemple: production de circuits imprimés
  - *c-chart*

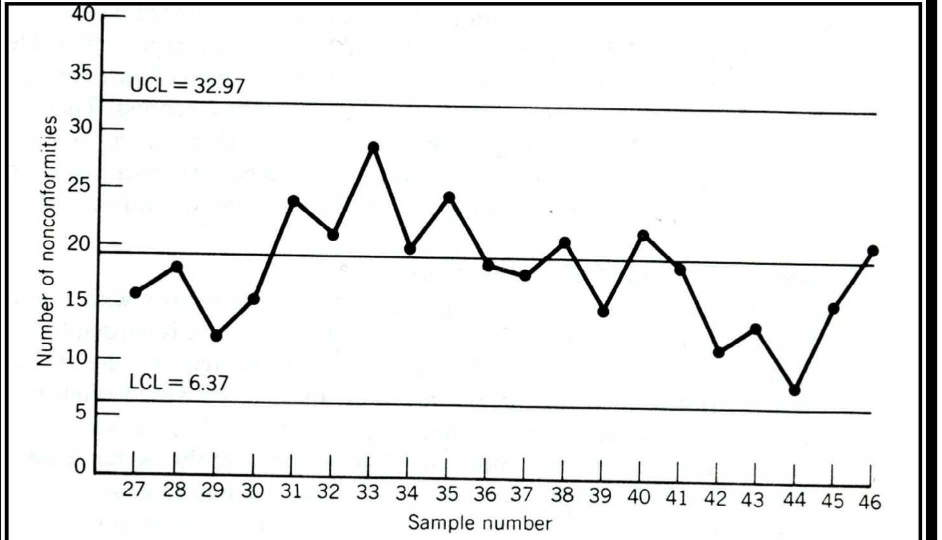
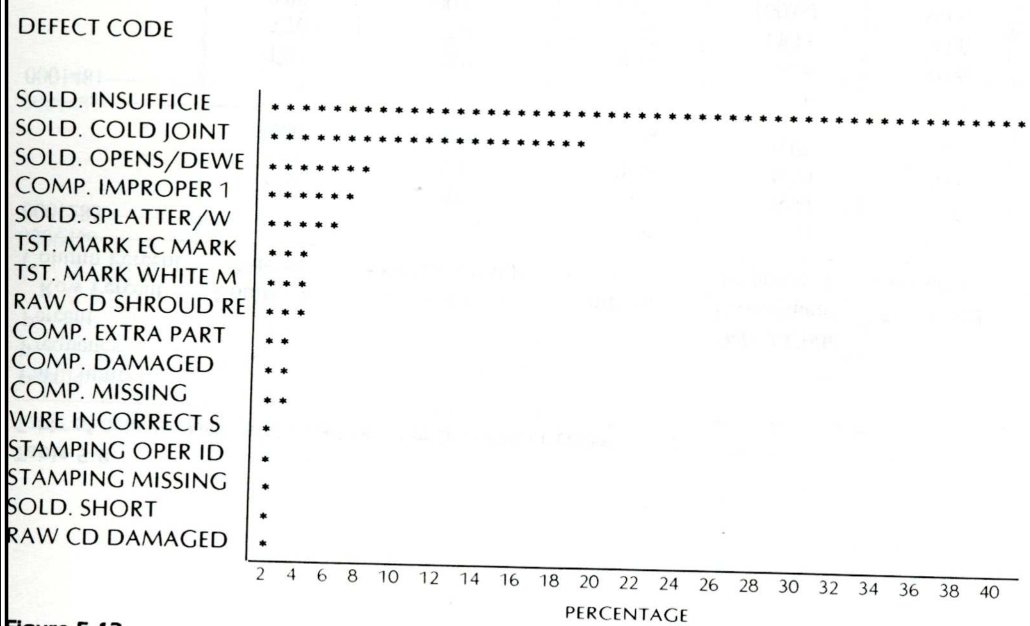


Figure 5-11

Sous-contrôle, mais c inacceptable => recherche de causes assignables

CC pour nombre de non-conformités (*c-chart*)

- Exemple: production de circuits imprimés
  - *Pareto*



DEFECT CODE	FREQ	CUM. FREQ	PERCENT	CUM. PERCENT
SOLD. INSUFFICIE	40	40	40.82	40.82
SOLD. COLD JOINT	20	60	20.41	61.23
SOLD. OPENS/DEWE	7	67	7.14	68.37
COMP. IMPROPER 1	6	73	6.12	74.49
SOLD. SPLATTER/W	5	78	5.10	79.59
TST. MARK EC MARK	3	81	3.06	82.65
TST. MARK WHITE M	3	84	3.06	85.71
RAW CD SHROUD RE	3	87	3.06	88.78
COMP. EXTRA PART	2	89	2.04	90.82
COMP. DAMAGED	2	91	2.04	92.86
COMP. MISSING	2	93	2.04	94.90
WIRE INCORRECT S	1	94	1.02	95.92
STAMPING OPER ID	1	95	1.02	96.94
STAMPING MISSING	1	96	1.02	97.96
SOLD. SHORT	1	97	1.02	98.98
RAW CD DAMAGED	1	98	1.02	100.00

Figure 5-12
Pareto analysis of nonconformities for the printed circuit board process.

Cause majeure de défauts: problèmes de soudure

CC pour nombre de non-conformités (*c-chart*)

- Exemple: production de circuits imprimés
  - *Table de défauts*

Table 5-9
Table of defects classified by part number and defect code

Part Number Frequency Percent Row Percent Column Percent	Component Missing	Component Damaged (NO)	Component Extra Part	Defect Code Component Improper I	Raw Card Shroud RE	Raw Card Damaged	Solder Short	Solder Opens/DEWE	Solder Cold Joint
0001285	1 1.02 1.41 50.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 1.02 1.41 100.00	0 0.00 0.00 0.00	5 5.10 7.04 71.43	20 20.41 28.17 100.00
0001481	1 1.02 3.70 50.00	2 2.04 7.41 100.00	2 6.12 22.22 100.00	6 3.06 11.11 100.00	3 0.00 0.00 100.00	0 0.00 0.00 0.00	1 1.02 3.70 100.00	2 2.04 7.41 28.57	0 0.00 0.00 0.00
0006429	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00
Total	2 2.04	2 2.04	2 2.04	6 6.12	3 3.06	1 1.02	1 1.02	7 7.14	20 20.41

CC pour nombre de non-conformités (*c-chart*)

- Exemple: production de circuits imprimés
  - *Table de défauts*

Table 5-9 *(continued)*

Part Number Frequency Percent Row Percent Column Percent	Solder Insufficiencies	Solder Splatter	Stamping Missing	Defect Code Stamping Operator ID	Test Mark White *M*	Test Mark EC Mark	Wire Incorrect 5	Good Unit(s)	Total
0001285	40 40.82 56.32 100.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 2.04 2.82 66.67	1 1.02 1.41 33.33	1 1.02 1.41 100.00	0 0.00 0.00 0.00	71 72.45
0001481	0 0.00 0.00 0.00	5 5.10 18.52 100.00	1 1.02 3.70 100.00	1 1.02 3.70 100.00	1 1.02 3.70 33.33	2 2.04 7.41 66.67	0 0.00 0.00 0.00	0 0.00 0.00 0.00	27 27.55
0006429	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00
Total	40 40.82	5 5.10	1 1.02	1 1.02	3 3.06	3 3.06	1 1.02	0 0.00	98 100.00

Un composant rassemble la majorité des défauts de soudure…

CC pour nombre de non-conformités (*c-chart*)

- Exemple: production de circuits imprimés
  - *Diagramme causes-effets*

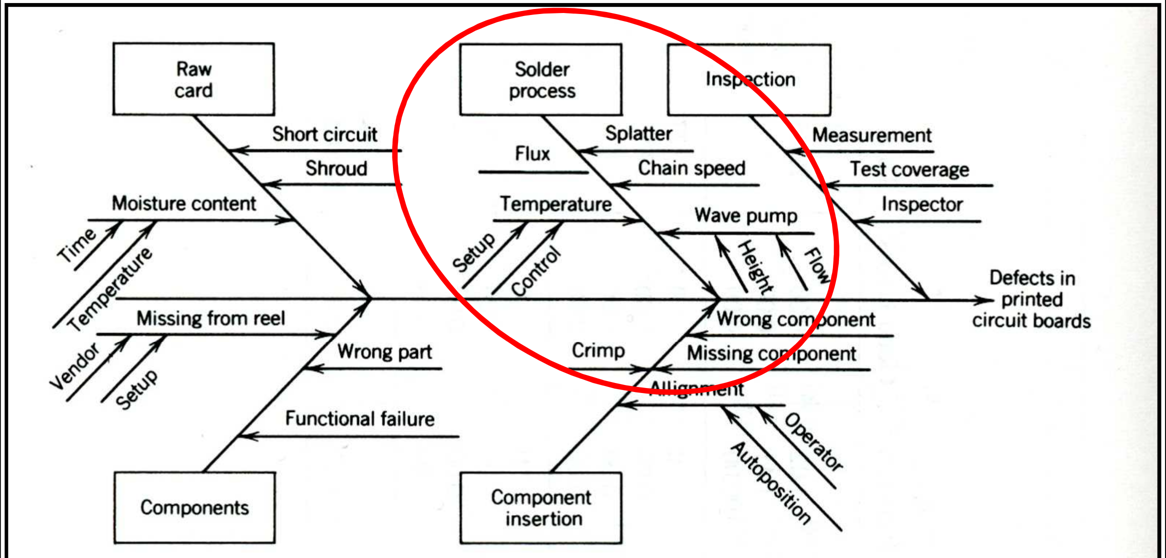


Figure 5-13
Cause and effect diagram.

Ce diagramme montre les facteurs sur lesquels concentrer la recherche

CC pour nombre de non-conformités par unité d'inspection (*u-chart*)

- Même situation, mais n unités d'inspection sont utilisées plutôt qu'1
- Utilisation du nombre moyen de non-conformités par unité d'inspection

$$u = c/n$$

  - Le CC, appelé u-chart, est cette fois défini par:
    - UCL $= u + 3*(u/n)^{0.5}$
    - CL $= u$
    - LCL $= u - 3*(u/n)^{0.5}$
    - Comme pour le p-chart, si u est estimé, on parlera alors de « trial control limits »

Cartes de contrôles pour variables

- Trois types de CC pour les variables
  - $\bar{x}$-chart: CC pour les moyennes
  - S-chart: CC pour les déviations standard
  - R-chart: CC pour l'étendue (*range*)
  - Le contrôle simultané de la position ($\bar{x}$-chart) et de la variation (S-chart, plus souvent R-chart) est essentiel

CC pour $\bar{x}$ (*$\bar{x}$-chart*)

- Utilisation de la loi normale
  - En théorie (TLC):
    - UCL = $\mu + 3*\sigma/n^{0.5}$
    - CL = μ
    - LCL = $\mu - 3*\sigma/n^{0.5}$
  - En pratique, μ et σ sont rarement connus et doivent être estimés. On utilise:
    - $\bar{\bar{x}} = (\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_m)/m$
      - $\bar{x}_i$ = moyennes d'échantillon de taille n chacun
    - $\bar{R} = (R_1 + R_2 + \cdots + R_m)/m$
      - R_i = étendue (x_{max}-x_{min}) de l'échantillon i

CC pour $\bar{x}$ (*$\bar{x}$-chart*)

- Utilisation de la loi normale (suite)
  - Il existe une relation entre σ et R:

    $$\sigma = R/d_2$$

    où d_2 = constante (tabulée) dépendant de n
  - <u>Exemple</u>: calcul de d_2 pour n = 5

```
> n<-10000
> ranges<-rep(0,n)
> for (i in 1:n) {
+  x<-rnorm(5)                          # => σ = 1
+  ranges[i]<-max(x)-min(x)
+ }
> mean(ranges)
[1] 2.322911                            # => d2 = 2.323
```

CC pour $\bar{x}$ ($\bar{x}$-*chart*)

- Utilisation de la loi normale (suite)
  - Et donc, en pratique:
    - UCL = $\overline{\overline{x}} + A_2 * \overline{R}$
    - CL = $\overline{\overline{x}}$
    - LCL = $\overline{\overline{x}} - A_2 * \overline{R}$
      - $A_2 = 3/(d_2 * n^{0.5})$ est une constante (fonction de n)

CC pour R (*R-chart*)

- On utilise encore:
  - $W = R/\sigma$
    - $E(W) = d_2(n)$ => $\sigma = E(R)/d_2$
    - $V(W) = d_3^2(n)$ => $\sigma_R = \sigma_W * \sigma = d_3 * \sigma$
  - On obtient donc:
    - UCL = $\overline{R} + 3 * d_3 * \overline{R} / d_2 = \overline{R} * (1 + 3 * d_3 / d_2) = \overline{R} * D_4$
    - CL = $\overline{R}$
    - LCL = $\overline{R} - 3 * d_3 * \overline{R} / d_2 = \overline{R} * (1 - 3 * d_3 / d_2) = \overline{R} * D_3$
  - Les valeurs de D_3 et D_4 sont constantes, pour n donné, et tabulées

CC pour R (*R-chart*)

- Exercice: calculer d_3,D_3 et D_4 pour n = 5

```
> n<-10000
> ranges<-rep(0,n)
> for (i in 1:n) {
+   x<-rnorm(5)
+   ranges[i]<-max(x)-min(x)
+ }
> d2<-mean(ranges)
> d3<-sd(ranges)
> D3<-1-3*d3/d2
> D4<-1+3*d3/d2
> c(d2,d3,D3,D4)
[1]  2.3233665  0.8612102 -0.1120203  2.1120203
```

  - $D_3 < 0$, est mis à 0 dans les calculs

CC pour $\bar{x}$ et R

- Capacité du processus (*process capability*)
  - Mesure de la performance du processus
    1. Probabilité de produits non-conformes
       - Exemple: spécifications = 74.000 ± 0.05
         => (LSL,USL) = (73.95,74.05)

         $\bar{\bar{x}}$ = 74.001

         $\hat{\sigma} = \bar{R} / d_2$ = 0.0099

```
> m<-74.001
> s<-0.0099
> 1-(pnorm(74.05,mean=m,sd=s)-pnorm(73.95,mean=m,sd=s))
[1] 5.012267e-07
```

CC pour $\bar{x}$ et R

- Capacité du processus (*process capability*)
  - Mesure de la performance du processus
    2. Taux de capacité du processus: PCR = (USL-LSL)/6σ
       - En pratique, σ est estimé par $\hat{\sigma} = \overline{R} / d_2$
       - Des valeurs > 1 indiquent que le processus utilise moins de 100% de la bande de variation tolérée, ce qui est évidemment une situation favorable.

```
> PCR<-(74.05-73.95)/(6*s)
> PCR
[1] 1.683502
```

CC pour $\bar{x}$ et R

- Eléments du design
  - Operating characteristic pour la $\bar{x}$-chart
    - On peut calculer assez facilement cette caractéristique pour un $\bar{x}$-chart
    - On supposera σ connu dans cet exemple
    - Si on suppose un déplacement (*shift*) de $\mu_0 \rightarrow \mu_1 = \mu_0 + t^*\sigma$, on peut calculer $\beta = \beta(t|n)$
    - Dans notre CC:

$$\beta = P\left[LCL \leq \bar{x} \leq UCL \middle| \mu = \mu_1\right]$$

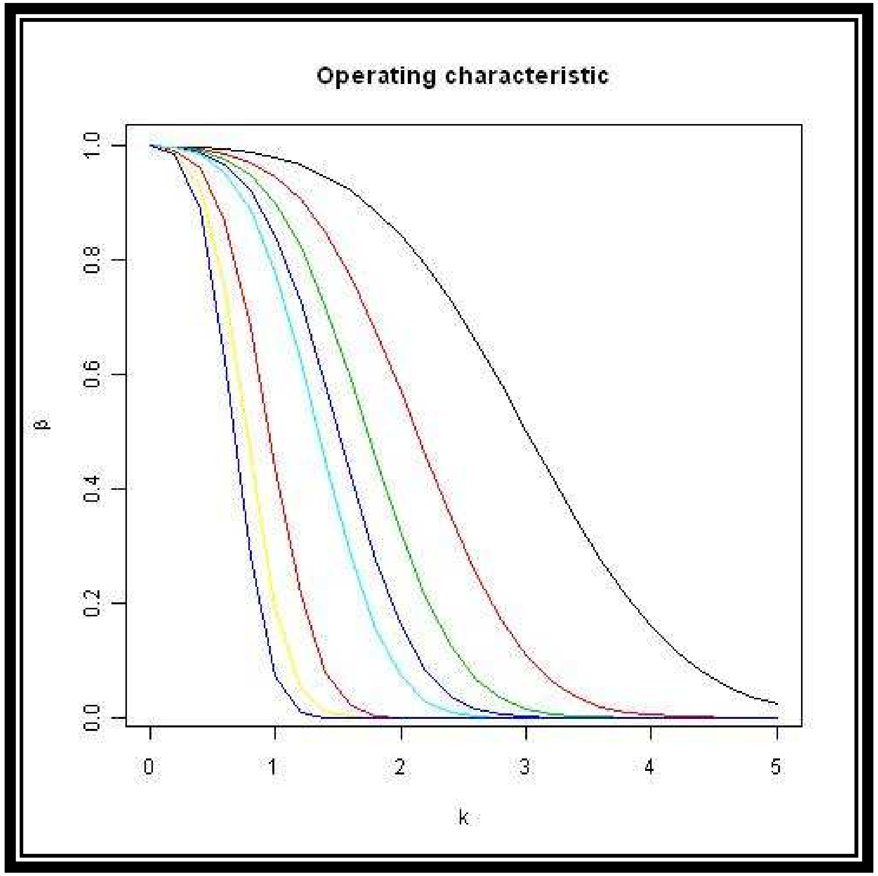
CC pour $\bar{x}$ et R

- Eléments du design
  - Operating characteristic pour la $\bar{x}$-chart

```
> k<-seq(0,5,0.2)
> oc<-pnorm(3-k*sqrt(1))-pnorm(-3-k*sqrt(1))
> plot(k,oc,main="Operating characteristic",
+      xlab="k",ylab=expression(beta),type="l")
> for (n in c(2,3,4,5,10,15,20)) {
+  oc<-pnorm(3-k*sqrt(n))-pnorm(-3-k*sqrt(n))
+  lines(k,oc,col=n)
+ }
```

CC pour $\bar{x}$ et R

- Eléments du design
  - Operating characteristic pour la $\bar{x}$-chart

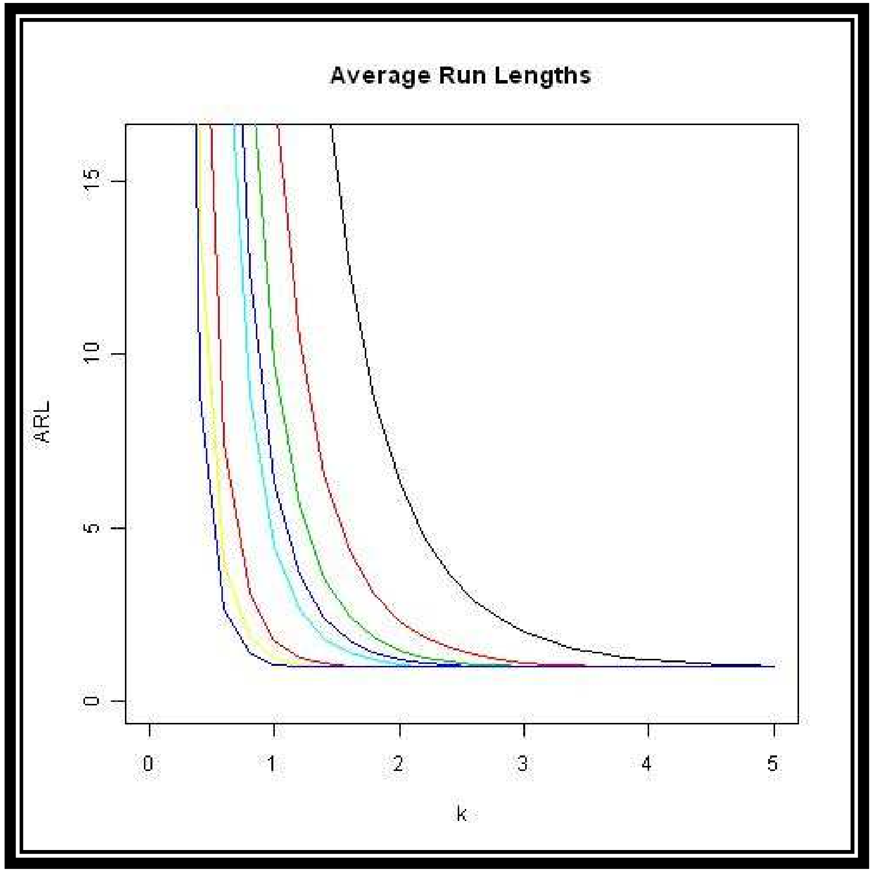


CC pour $\bar{x}$ et R

- Eléments du design
  - Operating characteristic pour la $\bar{x}$-chart
    - On constate que pour $n < 10$, $\beta \uparrow$ quand $k > 1.0$
      - exemple: $n = 5$, $k = 1.0 \Rightarrow \beta \approx 0.78$
    - Proba de détection après p échantillons:
      - $\beta_p = \beta^{p-1} * (1-\beta)$
      - exemple: $p = 3 \Rightarrow \beta = 0.78^2 * 0.22 = 0.13$
    - Nombre moyen d'échantillons pour détecter
      - $ARL = \sum_{k=1}^{\infty} k * \beta^{k-1} * (1-\beta) = \frac{1}{1-\beta}$
      - Dans notre exemple, ARL = 4
        - => acceptable (?)
        - => utilisation de n $\downarrow$ avec une fréquence $\uparrow$

CC pour x̄ et R

- Eléments du design
  - Average Run Lengths pour la x̄-chart



CC pour $\bar{x}$ et R

- Eléments du design
  - Operating characteristic pour la R-chart
    - A faire comme exercice:

      $$\beta = \beta(\lambda \mid n) \quad \text{où} \quad \lambda = \sigma_1/\sigma_0$$

    - Que constatez-vous pour des valeurs de n faibles (p.e. $n < 10$) ?

Autres *control-charts*

- De nombreuses autres possibilités existent
  - Parmi elles:
    - Cumulative-sum CC (CUSUM)
    - Exponentially Weighted Moving Average CC (EWMA)
      - Le but est d'utiliser l'information de plusieurs points simultanément dans le graphique plutôt que le dernier, comme dans les autres cartes de Shewhart
      - Il est ainsi possible de récupérer une partie de l'information que ne capture pas les autres CC (comme celle incluse dans la notion de « runs », par exemple)

Amélioration des processus

- Dans toute « production », animale ou pas, un objectif est d'avoir une quantité et une qualité du produit la meilleure possible
- Cet objectif est poursuivi:
  - au moment du *design* de nouveaux produits
  - dans l'amélioration de produits existants
  - dans l'amélioration de la chaine de production

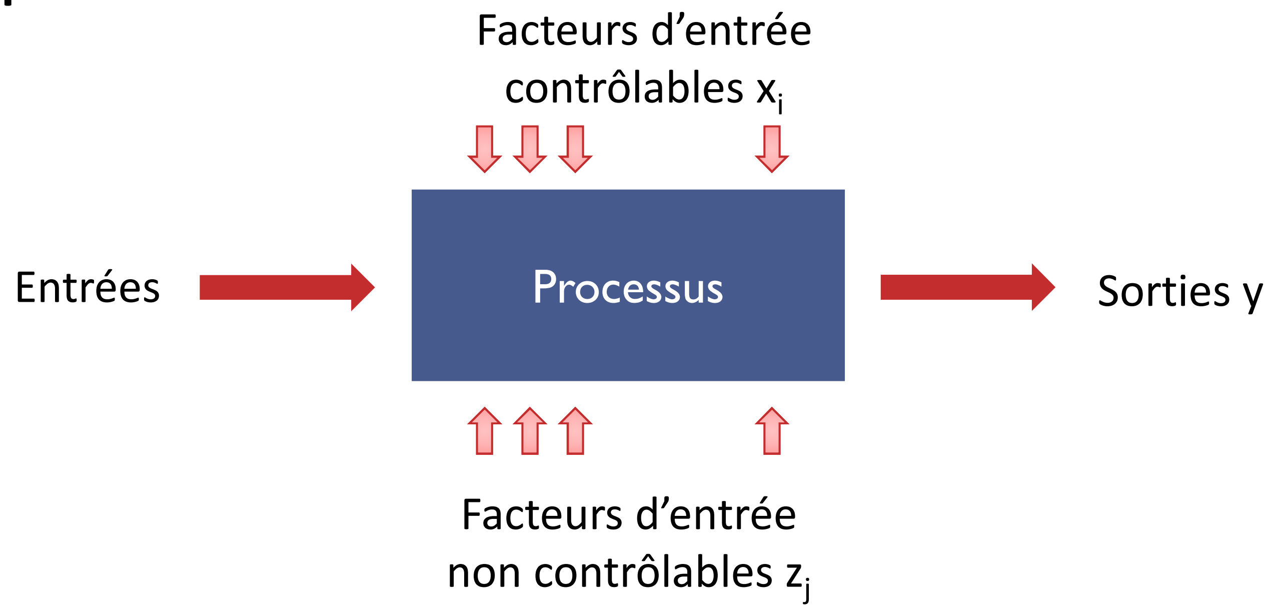
  par l'introduction de méthodes de design expérimental

Amélioration des processus

- L'utilisation effective du design expérimental peut conduire à:
  - des produits plus faciles à manufacturer
  - des produits plus fiables
  - de meilleures performances du produit
  - un développement du processus de production
  - une amélioration dans les activités de recherche d'élimination de problèmes (recherche des causes assignables…)

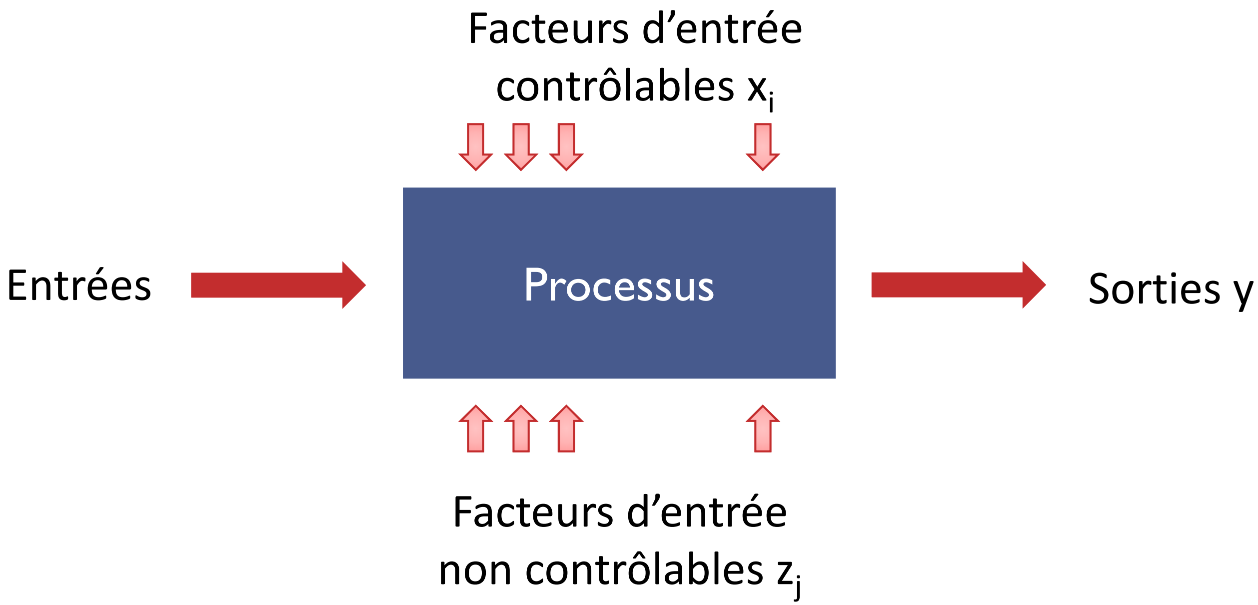
Design expérimental

- Les objectifs d'une expérience provenant d'un design expérimental (*designed experiment*) sont d'identifier les relations entre les variables qui affectent un processus



Design expérimental

- Objectifs:
  - déterminer les variables influençant le plus y
  - configurer x_i pour que y soit proche de l'optimum
  - configurer x_i pour que la variation de y soit ↓
  - configurer x_i pour minimiser l'effet de z_j



Design expérimental & SPC

- Les deux méthodologies sont imbriquées:
  - Design expérimental d'un processus
    - Designed experiments
  - Observation (passive) du processus (p.e. CC)
    - Sous-contrôle:
      - Satisfaisant: OK
      - Insatisfaisant: ajustement de variables influentes
        - Designed experiment
    - Hors-contrôle:
      - Recherche (active) de causes assignables
        - Designed experiments

Expériences unifactorielles

- Type le plus simple de *designed experiment*
  - Un seul facteur inclus dans l'expérience
  - ECA = expérience complètement aléatoire
    - Le but de cette randomisation est de supprimer les effets de nuisances qui pourraient affecter les mesures, et donc invalider les résultats de l'expérience

Expériences unifactorielles

- Exemple: amélioration de la résistance de sacs de transport d'aliments
  - Processus sous contrôle ($\bar{x}$ & R *charts* ok)
  - Résistance à la traction nominale T = 15 psi
  - Proportion de bois dans le papier P = 10%
  - Le chargé de fabrication suspecte que T ↑ si P ↑ mais les contraintes économique font que 5% ≤ P ≤ 20%
- Expérience:
  - Essais sur 4 concentrations (5%, 10%, 15%, 20%)
  - 6 répliques par concentration

Expériences unifactorielles

- Expérience:
  - Essais sur 4 concentrations (5%, 10%, 15%, 20%)
  - 6 répliques par concentration
  - Passage aléatoire sur la machine de mesure de tension

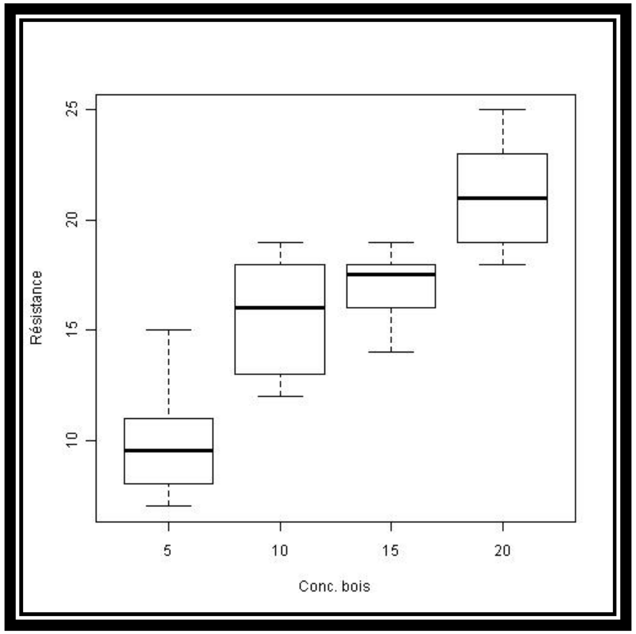
  - Résultats

[ ]	Observations						Moy
	1	2	3	4	5	6	
05 %	7	8	15	11	9	10	10.00
10 %	12	17	13	18	19	15	15.67
15 %	14	18	19	17	16	18	17.00
20 %	19	25	22	23	18	20	21.17
							15.96

Expériences unifactorielles

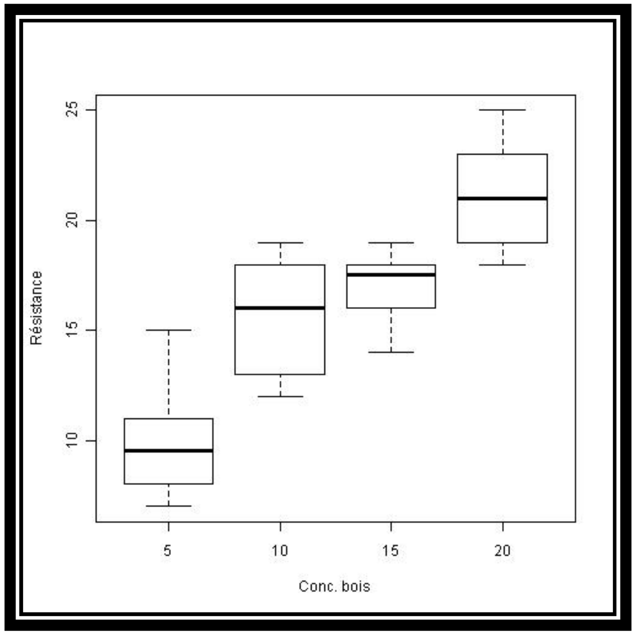
- <u>Analyse 1</u>:

```
> P<-c(rep(5,6),rep(10,6),rep(15,6),rep(20,6))
> T<-c(7,8,15,11,9,10,
+ 12,17,13,18,19,15,
+ 14,18,19,17,16,18,
+ 19,25,22,23,18,20)
> boxplot(T~P,xlab=«Conc. bois»,ylab=«Résistance»)
```



Expériences unifactorielles

- Conclusions 1:
  - Effet visible de la concentration (T ↑ quand P ↑)
  - Données ~ symétriques autour de leur moyenne
  - Variabilités similaires dans les groupes



Annexes

- Echantillonnage
- Intervalles de confiance

(Pseudo)-aléatoire

- Echantillonnage simple
  - On prélève n individus parmi les N constituant la population visée
    - Choisis de manière aléatoire
    - Choisis de manière indépendante
    - Exemple

      =ALEA.ENTRE.BORNES(1;N)

      où N désigne la cellule contenant la taille de la population (on suppose les individus numérotés de 1 à N)

Echantillons simples

(Pseudo)-aléatoire

- Echantillonnage simple
  - Version R

```
> sort(sample(1:1000,25,replace=FALSE))
```

- 1:1000 spécifie la liste des nombres de 1 à 1000
- 25 est le nombre d'échantillons à prélever
- REPLACE = FALSE indique qu'il n'y a pas de remise (c'est l'option par défaut)
- sort(X) trie la liste X (en ordre croissant)

(Pseudo)-aléatoire

- Echantillonnage stratifié
  - On prélève n_i individus parmi les N_i constituant la sous-population visée
    - Choisis de manière aléatoire
    - Choisis de manière indépendante
    - $\Sigma\, n_i = n$ et $\Sigma\, N_i = N$
    - Les sous-populations sont choisies en fonction d'une source assignable d'hétérogénéité
      - p.e. lots différents, usine différente, …

Echantillons stratifiés

(Pseudo)-aléatoire

- Echantillonnage stratifié
  - Version R

```
> echant<-rep(0,25)
> for (i in 0:24) {
   echant[i]<-sample((40*i+1):((i+1)*40),1)
+ }
> echant
[1]   26  58  97 158 197 201 268 296 360 369 411
[12] 451 511 552 592 623 643 704 754 798 840 863
[23] 884 953 962
```

(Pseudo)-aléatoire

- Echantillonnage « en grappe »
  - Exemple: palettes de conserves
    10 palettes de 4 x 5 x 3 caisses contenant chacune 8 conserves (=> 4800 en tout)
    - On tire (p.e.) 2 palettes au hasard
    - On tire ensuite 2 caisses / palette
    - Puis 4 boites / caisse

Retour

Une autre vue des tests d'hypothèses

- L'approche traditionnelle:

H_0, α $\rightarrow$ données ($\bar{X}$,p,n…) => S $\rightarrow$ distrib (z,t,χ^2…) $\rightarrow$ $P(data|H_0)$ $\rightarrow$ $P<\alpha$: rejet H_0 / $P>\alpha$: accepte H_0

- Une autre approche (équivalente):

H_0, α $\rightarrow$ données ($\bar{X}$,p,n…) => S $\rightarrow$ distrib (z,t,χ^2…) $\rightarrow$ $IC_\alpha(S)$ $\rightarrow$ $S \notin IC$: rejet H_0 / $S \in IC$: accepte H_0

Estimation par intervalle de confiance

- Supposons maintenant n « petit »
  - Utilisation pour le calcul d'un IC(α):
    - On recherche deux distributions binomiales de même n que l'échantillon, mais de paramètre p_i et p_s tels que:

$$\boxed{\sum_{i=k}^{\infty} B(i|n, p_i) = \tfrac{\alpha}{2}} \quad \text{et} \quad \boxed{\sum_{i=0}^{k} B(i|n, p_s) = \tfrac{\alpha}{2}}$$

    - où B(i | n,p) désigne la probabilité d'obtenir i fois l'événement dont la probabilité est p lors de n tirages.

Estimation par intervalle de confiance

- <u>Exemple</u>: si un lot sur 15 examinés dépasse une norme qu'on ne doit dépasser que dans 3 cas sur 100, peut-on conclure à l'adéquation du lot à la norme ?

Estimation par intervalle de confiance

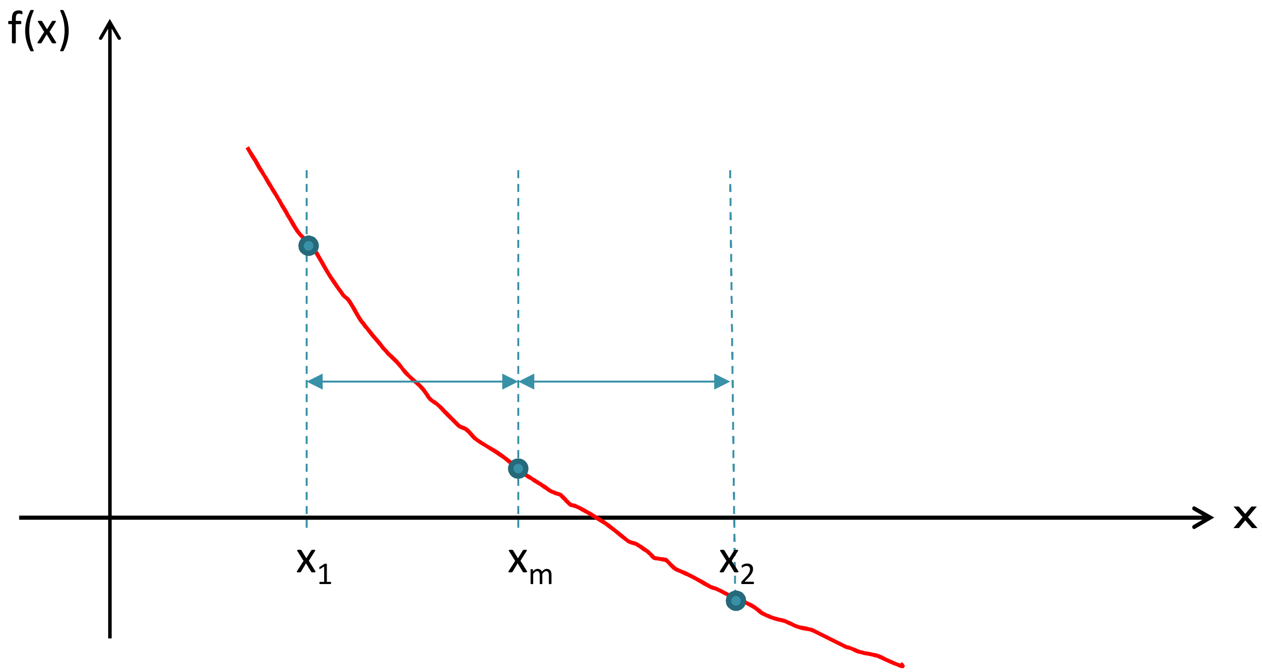
- Solution:

Si un lot sur 15 examinés dépasse une norme qu'on ne doit dépasser que dans 3 cas sur 100, peut-on conclure à l'adéquation du lot à la norme ?

- Calculons p_i:
  $B(1) + B(2) + \ldots = 0.025 \quad \Rightarrow B(0) = 0.975$
  $\Rightarrow (1 - p_i)^{15} = 0.975 \quad \Rightarrow p_i = 0.0017$
- Calculons p_s:
  $B(0) + B(1) = 0.025$
  => Ordinateur $\quad \Rightarrow p_s = 0.3195$

Estimation par intervalle de confiance

- Solution R: approche dichotomique



Estimation par intervalle de confiance

- Solution R: approche dichotomique

```
> dichot<-function(f,x1,x2) {
+  if (abs(x1-x2)<0.0001) {
+   return(x1)
+  }
+  else {
+   f1<-f(x1)
+   f2<-f(x2)
+   xm<-(x1+x2)/2
+   fm<-f(xm)
+   if (f1*fm<0) { return(dichot(f,x1,xm)) }
+          else { return(dichot(f,xm,x2)) }
+  }
+ }
```

Estimation par intervalle de confiance

- Solution R: approche dichotomique (suite)

```
> fun<-function(x) {
+  return((1-x)**15+15*x*(1-x)**14-0.025)
+ }
> dichot(fun,1/15,1.00)
[1] 0.3194824
```

Estimation par intervalle de confiance

- Interprétation
  - Si on part de l'hypothèse (nulle…) selon laquelle il y a une proportion $\pi = 0.03$ de valeurs au-dessus de la norme autorisée, on voit qu'une valeur de $p = 1/15 \approx 0.07$ n'est pas aberrante puisque l'intervalle de confiance au seuil $a = 0.05$, soit $[0.0017; 0.3195]$, contient la valeur $\pi = 0.03$
  - L'hypothèse nulle est acceptée dans ce cas.

Estimation par intervalle de confiance

- Interprétation (suite)
  - Mais l'intervalle est très large…
    IC(0.05) = [0.0017; 0.3195]
  - Ceci est du à la très petite taille de l'échantillon: par exemple, si on avait obtenu 10/150 (même proportion, mais effectif 10 x plus grand):
    IC(0.05) = [0.0324; 0.1192]
    => H0 rejetée…
  - Dilemme statistique… (cfr puissance)

Estimation par intervalle de confiance

- <u>Question</u>
  - A partir de combien de positifs (sur 15) conclura-t-on à une contamination ?

$$\sum_{i=0}^{<x} B\left(i \middle| p_i = 0.03, n = 15\right) \geq 0.975$$

$$x = 2$$

  - A partir de combien de positifs (sur 150) conclura-t-on à une contamination ?

$$\sum_{i=0}^{<x} B\left(i \middle| p_i = 0.03, n = 150\right) \geq 0.975$$

$$x = 9$$

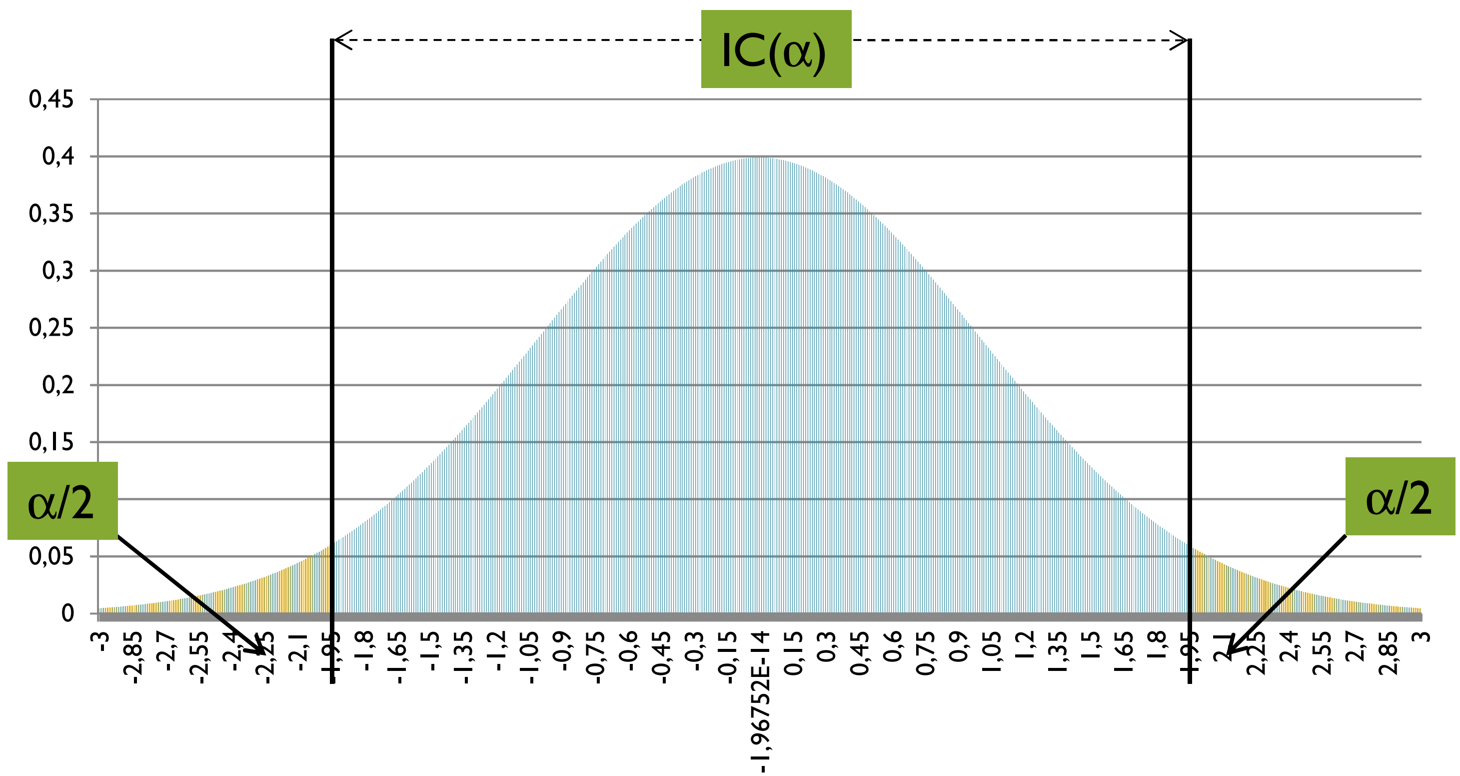
Intervalles de confiance

- <u>Passons maintenant à l'estimation de μ:</u>

Un échantillon – σ connu – H_0: μ(éch) = μ

- On sait qu'alors: m ~ $N(\mu,\sigma^2/n)$
- Donc, si l'échantillon provient bien de cette population, on peut calculer un intervalle dans lequel m a une probabilité (1-α) de figurer (voir dia suivante)

Intervalles de confiance



Intervalles de confiance

- Calculons l'intervalle:
  - On recherche $z_{\alpha/2}$ tel que $p(z_{th}>z_{\alpha/2}) = \alpha/2$
    - Exemple: $\alpha = 0.05 \Rightarrow z_{\alpha/2} = 1.96$
  - Alors, si H_0 est vraie:
    - $P(m > \mu+z_{\alpha/2}*\sigma/\sqrt{n}) = P(m < \mu-z_{\alpha/2}*\sigma/\sqrt{n}) = \alpha/2$

      $\Rightarrow P(\mu-z_{\alpha/2}*\sigma/\sqrt{n} < m < \mu+z_{\alpha/2}*\sigma/\sqrt{n}) = (1-\alpha)$
  - Cet intervalle est appelé intervalle de confiance de la moyenne (de l'échantillon).

Intervalles de confiance

- <u>Utilisations de l'intervalle de confiance (1):</u>
  - Test d'hypothèse
    - <u>Exemple</u>: supposons que $\mu = 100$, $\sigma = 10$ et qu'on a obtenu un moyenne m = 94 sur un échantillon de 25 individus. Si on veut tester H_0: $\mu(\text{éch}) = \mu$, il suffit de calculer les limites de l'IC de m au seuil a = 5% (par exemple):

      $z_{0.025} = 1.96 \Rightarrow IC = [100-1.96*10/5;100+1.96*10/5] = [96.08; 103.92]$

      => m = 94 n'est pas dans cet intervalle => rejet de H_0

Intervalles de confiance

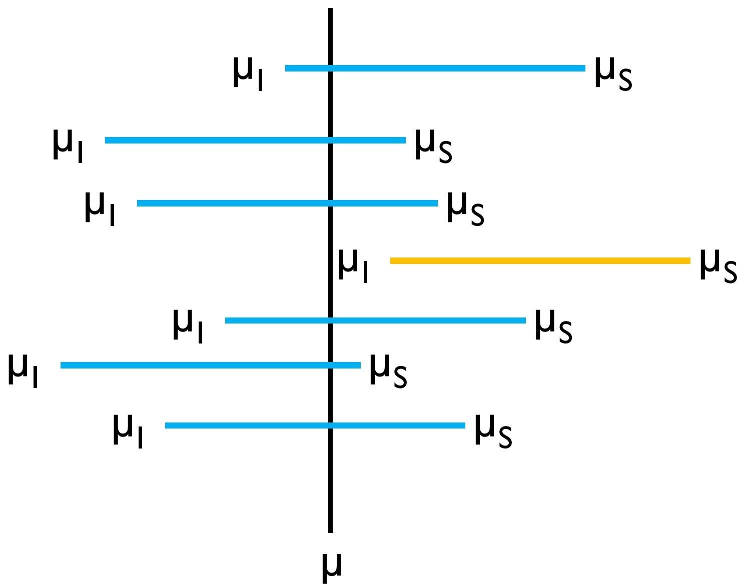
- <u>Utilisations de l'intervalle de confiance (2):</u>
  - Prédiction de la « vraie » moyenne μ
    - On pourrait employer la formule de l'IC « dans l'autre sens »:

$$=> P(m - z_{\alpha/2}*\sigma/\sqrt{n} < \mu < m + z_{\alpha/2}*\sigma/\sqrt{n}) = (1-\alpha)$$

    - Dans ce cas, on obtient les limites de « l'intervalle de confiance de la moyenne μ d'une distribution normale pour laquelle σ est connue ».

Intervalles de confiance

- <u>Utilisations de l'intervalle de confiance (2):</u>
  - Schématiquement:



Intervalles de confiance

- <u>Utilisations de l'intervalle de confiance (2):</u>
  - <u>Exemple</u>
    - On pourrait calculer les limites de l'IC de la moyenne de la pop. dont les paramètres donnés plus haut proviennent:

      => $P(m - z_{\alpha/2}*\sigma/\sqrt{n} < \mu < m + z_{\alpha/2}*\sigma/\sqrt{n}) = (1-\alpha)$
    - $m = 94$, $\sigma = 10$, $n = 25$ et $\alpha = 0.05$ => $z_{\alpha/2} = 1.96$

      => $P(90.08 < \mu < 97.92) = 0.95$

Intervalles de confiance

- Que fait-on si σ est inconnue ?
  - Réponse:
    - On remplace σ par son estimateur s, ce qui conduit à employer la distribution de t plutôt que celle de z:

$$\Rightarrow P(m - t_{\alpha/2} * s/\sqrt{n} < \mu < m + t_{\alpha/2} * s/\sqrt{n}) = (1-\alpha)$$

où $t_{\alpha/2}$ est recherché dans la distribution t de Student avec (n-1) ddl

Intervalles de confiance

- <u>Un exemple similaire: le cas des comptages</u>
  - Un cas de BSE a été reporté le mois passé. Sachant cela, et en supposant l'épidémie stabilisée, à combien de cas par mois dois-je m'attendre en moyenne ?
  - <u>Réponse</u>:
    - En moyenne, sur base de cet échantillon, on attend évidemment 1 cas par mois…
    - Il serait plus informatif de fournir une fourchette dans laquelle on a par exemple 95% de chance de trouver le vrai nombre de cas moyen !
      - Cherchons: $IC_{0.05}(\mu) = [\mu_I; \mu_S]$

Intervalles de confiance

- <u>Un exemple similaire:</u>
  - <u>Solution</u>:
    - Il s'agit d'un exemple d'utilisation de la « loi de Poisson »
    - On cherche μ_l tel que:

$$P(1|\mu_l) + P(2|\mu_l) + \ldots = \alpha/2$$

$$\Rightarrow P(0|\mu_l) = 1 - \alpha/2 = \exp(-\mu_l)$$

$$\Rightarrow \mu_l = -\ln(1 - \alpha/2) = 0.02532$$

Intervalles de confiance

- <u>Un exemple similaire:</u>
  - <u>Solution (suite)</u>:
    - On cherche ensuite μ_S tel que:
      $P(0|\mu_S) + P(1|\mu_S) = \alpha/2$
      $\Rightarrow \exp(-\mu_S)\,(1+\mu_S) = \alpha/2$

```
> fun<-function(x) { return(exp(-x)*(1+x)-0.025)}
> dichot(fun,1,100)
[1] 5.571609
```

    - On obtient ainsi IC = [0.002532;5.571609], l'intervalle de confiance pour α = 0.05 de la moyenne d'une distribution de Poisson.

Intervalles de confiance

- La prédiction d'un prochain tirage:
  - Problème:
    - Ayant échantillonné n individus dans une population (normale), est-il possible de préciser un intervalle dans lequel un (n+1)ème individu aurait une probabilité (1-α) de figurer ?
    - En d'autres termes, on recherche l'intervalle de confiance d'un prochain tirage X

Intervalles de confiance

- <u>La prédiction d'un prochain tirage:</u>
  - <u>Solution</u>:
    - On peut voir le problème comme un exemple de situations où 2 échantillons sont extraits d'une même population, un échantillon de taille n et un de taille 1.
    - Evidemment, ici, $\mu_1 = \mu_2 = \mu$!
    - On sait que:

Retour

$$t = [(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)]/\sqrt{s_C^2(1/n_1 + 1/n_2)}$$

avec $(n_1 - 1) + (n_2 - 1)$ degrés de liberté