

Statistiques pour l'Analyste HPLC

Distribution normale • Tests d'hypothèses • Régression • Intervalles de confiance

01 Distribution

Normale & données

02 Tests

Hypothèses & t-test

03 Régression

Linéarité & résidus

04 Intervalles

Confiance & incertitude



2.1 Pourquoi l'analyste a besoin de statistique

L'œil humain est sensible aux écarts visibles — pas à leur signification réelle

« La statistique ne répond pas par intuition mais par **démonstration**. Elle examine la dispersion, la taille de l'échantillon, la cohérence interne et le niveau de confiance souhaité. »

Questions concrètes auxquelles répond la statistique :

? La dispersion entre plusieurs injections est-elle normale ou excessive ?

? La différence entre deux analystes est-elle réelle ou due au hasard ?

? Une courbe d'étalonnage linéaire est-elle vraiment exploitable sur toute la gamme ?

? Le changement après une maintenance traduit-il une dérive ou une variation naturelle ?

Protège

contre les surinterprétations

Détecte

les signaux faibles

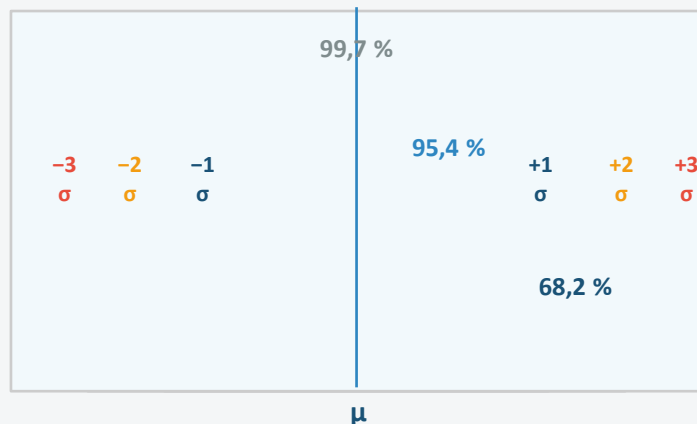
Structure

le jugement scientifique

2.2 Distribution normale — La loi de la chromatographie

En HPLC, la plupart des erreurs (pesée, injection, température) sont de nature aléatoire → Loi Normale

Distribution normale — Règle des 6 σ



Points clés pour l'analyste

Pourquoi la loi normale est partout

Les erreurs aléatoires (pesée, injection, T°) se compensent par le théorème central limite. Résultat : la somme suit une loi normale.

Signal d'alerte : non-normalité

Si vos temps de rétention ou aires ne suivent pas la normale → cause spéciale probable : fuite intermittente, instabilité lampe UV.

Règle des 3 σ en pratique

99,7 % des résultats doivent se situer dans $\pm 3\sigma$. Un résultat hors de cette zone est potentiellement un outlier à investiguer.

Exception : impuretés proches LOQ

Les données asymétriques ou proches de la limite de quantification peuvent ne pas suivre la normale → vérifier systématiquement.

2.2.1 Tester la normalité dans Minitab

Avant d'appliquer un test, vérifier la distribution — Anderson-Darling recommandé

Procédures Minitab

1

Graphique > Diagramme de probabilité > Simple

Visualisation graphique — points alignés = normale ✓

2

Stat > Outils de la qualité > Identification de loi individuelle

Teste plusieurs lois — choisit la meilleure

3

Stat > Statistiques élémentaires > Test de normalité

Anderson-Darling ou Shapiro-Wilk ($n < 50$)

Règle de décision — Valeur p

$p \leq \alpha (0,05)$

Les données ne suivent pas une loi normale → Rejeter H_0

Utiliser Box-Cox ou transformation logarithmique

$p > \alpha (0,05)$

Impossible de rejeter H_0 → Normalité non rejetée ✓

Les tests paramétriques (t-test, ANOVA) sont applicables

Checklist Analyste

- ✓ Anderson-Darling $p \geq 0,05$ avant tout test
- ✓ Visualiser histogram + probability plot
- ✓ Box-Cox si $p < 0,05$ (note dans rapport)

2.3 Lire les données avant de calculer — Le rôle des graphiques

Un bon graphique révèle en quelques secondes ce qu'un tableau de chiffres peine à faire apparaître

Histogramme

- Forme de la distribution (normale, asymétrique)
- Détection valeurs extrêmes
- Dispersion globale

Minitab : Graph > Histogram > With Fit

Boxplot (Boîte à moustaches)

- Comparaison de groupes (analystes, instruments, lots)
- Médiane + dispersion interquartile
- Identification points atypiques

Minitab : Graph > Boxplot > One Y, With Groups

Graphique chronologique

- Détecter les dérives progressives dans le temps
- Visualiser l'ordre de production
- Mettre en évidence tendances non visibles sur moyenne

Minitab : Graph > Time Series Plot

2.4 Tests d'hypothèses — Comparer sans se tromper

Distinguer une différence réelle d'une fluctuation aléatoire normale

Principe fondamental

H_0 (Hypothèse nulle) :

Pas de différence réelle entre les groupes
— ex: les 2 analystes donnent le même résultat

H_1 (Hypothèse alt.) :

Il y a une différence réelle — ex: les résultats des 2 analystes diffèrent significativement

La valeur p — clé d'interprétation

$p < 0,05 \rightarrow$ Différence statistiquement significative — Rejeter H_0

$p \geq 0,05 \rightarrow$ Différence non significative — Ne pas rejeter H_0

⚠ Un résultat significatif statistiquement n'est pas toujours analytiquement critique — toujours replacer dans le contexte.

Tests courants en HPLC pharmaceutique

Test t à 1 échantillon

- ▶ Comparer une moyenne à une valeur cible
- ▶ Ex : dosage mesuré vs 100 % théorique

Minitab : Stat > Basic Stat > 1-Sample t

Test t à 2 échantillons

- ▶ Comparer deux analystes, deux colonnes, deux sites
- ▶ Ex : colonne A vs colonne B (même méthode)

Minitab : Stat > Basic Stat > 2-Sample t

ANOVA à 1 facteur

- ▶ Comparer > 2 groupes simultanément
- ▶ Ex : 3 instruments HPLC — même méthode

Minitab : Stat > ANOVA > One-Way

2.4 Exemple concret — Comparer deux analystes


Analyste A : 98,9 % vs Analyste B : 99,6 % | La différence est-elle réelle ?

Scénario : Deux analystes préparent chacun 6 fois le même échantillon. A obtient 98,9 % de moyenne, B obtient 99,6 %. La différence de 0,7 % est-elle significative ?

Mesure	1	2	3	4	5	Moyenne
Analyste A	99.2	98.8	99.1	99.0	98,9 %	98,9 %
Analyste B	99.4	99.7	99.3	99.8	99,6 %	99,6 %

Démarche d'analyse statistique avec Minitab — 3 étapes



1 Vérifier les variances

Stat > Basic Statistics > 2 Variances
→ $p \geq 0,05$ → variances égales  → utiliser
'Assume equal variances'

2 Appliquer le test t

Stat > Basic Statistics > 2-Sample t
→ Samples in one column + Subscripts
(Analyste)
→ Confidence level : 95 %

3 Interpréter le résultat

$p < 0,05$ → différence significative entre
analystes 
 $p \geq 0,05$ → écart compatible avec variabilité
normale 

2.5 Régression — La loi de l'étalonnage

$Y = aX + b$ | Sensibilité, bruit de fond et qualité de l'ajustement

Équation de la courbe d'étalonnage

$$Y = a \cdot X + b$$

Y	Réponse (Aire ou hauteur du pic)
X	Concentration de l'étalon
a (pente)	Sensibilité de la méthode — sensibilité = $\Delta \text{aire} / \Delta \text{concentration}$
b (ordonnée)	Bruit de fond — idéalement ≈ 0 (tester avec test t sur b)

⚠ $R^2 > 0,999 \neq$ Garantie absolue de qualité

Vérifier AUSSI : résidus aléatoires, ordonnée à l'origine, absence de non-linéarité locale

Procédure Minitab + Analyse des résidus

Stat > Regression > Regression > Fit Regression Model

→ Response : [Aire_Pic]

→ Continuous predictors : [Concentration]

→ Graphs : ☒ Residual plots : Four in one

→ Storage : ☒ Residuals, ☒ Fits

Assistant : Menu Assistant > Analyse de régression

→ Vérifie automatiquement la linéarité

→ Identifie les points influents

Critères résidus valides :

Aléatoires ☒ | Normaux ☒ | Homoscédastiques ☒ | Pas d'autocorrélation ☒

2.6 Intervalles de confiance — Mesurer avec modestie

Toute moyenne observée sur un nombre limité de répétitions n'est qu'une estimation

Dire 99,2 % est utile — **Dire 99,2 % avec IC95% [98,8 – 99,6]** est plus honnête et plus informatif. L'IC dit non seulement où se situe la moyenne, mais avec quelle précision elle a été estimée.

Étude d'exactitude

9 préparations → recouvrement moyen 100,4 %

IC95% étroit → méthode juste ✓

IC95% large → précision insuffisante ⚠

Critère : IC95% \subset [98 – 102 %]

Comparaison deux colonnes

IC95% de la différence contient 0 → pas de biais significatif ✓

IC95% ne contient pas 0 → différence réelle à investiguer

Critère : IC95% de Δ contient 0

Validation — Répétabilité

Répétabilité : IC95% du %RSD

Si borne supérieure < 2 % → conforme ICH Q2



Utiliser STDEV et SE Mean dans Minitab

Critère : IC95% %RSD < 2 %

Minitab : Stat > Basic Statistics > Display Descriptive Statistics → ☒ SE Mean ☒ CI of mean | Options : Confidence level 95 %

L'IC est calculé automatiquement dans tout test t — vérifier systématiquement que l'IC est cohérent avec les spécifications.

2.7 Exemples appliqués au laboratoire HPLC

4 situations réelles — du problème à la décision statistique



Répétabilité — 6 injections standard

Problème : Une aire semble légèrement plus faible. L'analyste s'inquiète.

Analyse : Histogramme + boxplot → aucun outlier apparent. Statistiques descriptives : %RSD < 1 %. Variation compatible avec le comportement normal du système.

✓ Aucune investigation requise — documenter le %RSD



Courbe d'étalonnage — $R^2 > 0,999$

Problème : R^2 excellent mais résidus non aléatoires → structure détectée.

Analyse : Examen graphique des résidus (Residuals vs Fits) → structure en U visible. Non-linéarité locale sur les extrémités de la gamme.

⚠ Restreindre la gamme ou utiliser un modèle quadratique



Précision intermédiaire — 2 analystes

Problème : Moyennes légèrement différentes. Risque d'investigation inutile.

Analyse : Test t à 2 échantillons → $p \geq 0,05$. Les dispersions se recouvrent fortement. Pas de différence significative entre analystes.

✓ Pas d'effet analyste — documenter le résultat dans le rapport



Suivi temps de rétention — 4 semaines

Problème : Moyenne globale acceptable mais quelque chose change.

Analyse : Représentation chronologique → tendance à l'augmentation détectée. Régression temporelle : pente positive significative ($p < 0,05$).

⚠ Signal précoce — action préventive avant l'échec SST

2.9 Ce que l'analyste doit retenir

La statistique n'est pas opposée à l'expertise — elle en est le prolongement naturel

Checklist praticien — À appliquer systématiquement

✓ Je calcule RSD + IC 95 % dans TOUS mes rapports — une moyenne seule ne suffit jamais

✓ Je vérifie la normalité (Anderson-Darling) avant d'appliquer tout test paramétrique

✓ Si $p < 0,05 \rightarrow$ je mentionne et contextualise dans la conclusion (significance \neq criticité)

✓ Si $Cpk < 1,33 \rightarrow$ alerte QA immédiate + investigation cause racine

✓ Si Anderson $p < 0,05 \rightarrow$ Box-Cox ou transformation log avant de continuer

✓ Je réalise un Gage R&R annuel (objectif : $\%GRR < 15 \%$)

✓ Si $\%RSD > 1,5 \%$ \rightarrow analyse de régression temporelle pour détecter une dérive

En résumé — Chapitre 2

Distribution

La loi normale est le point de départ — vérifiez-la toujours avec Anderson-Darling avant tout calcul.

Tests

$p < 0,05$ = différence significative. Mais significatif \neq critique. Remplacez toujours dans le contexte analytique.

Régression

R^2 seul ne suffit pas. Analysez les résidus. Testez l'ordonnée à l'origine. Vérifiez la linéarité locale.

Intervalles

Présentez toujours un IC95% avec votre moyenne. C'est la façon honnête de communiquer une incertitude.

« La statistique transforme les données en compréhension, et la compréhension en décision. » — HPLC sous Contrôle, Rachid BERKANI