MATHÉMATIQUES

11

Statistique inférentielle

BTS | Mathématiques | Groupements B1, B2, D1, D2

Objectifs du chapitre

- Distinguer **population** et **échantillon**, paramètres et statistiques d'échantillon.
- Comprendre et appliquer le **théorème central limite** (TCL) et l'erreur standard.
- Calculer et interpréter un **intervalle de confiance** pour une moyenne (σ connue ou inconnue).
- Construire un **IC pour une proportion** et déterminer la taille d'échantillon nécessaire.
- Choisir et lire les tables de la **loi normale standard** et de la **loi de Student**.
- Poser et conduire un **test d'hypothèse** : Z-test, T-test, test sur proportion.
- Calculer et interpréter une **p-valeur** dans un contexte industriel ou économique.
- Réaliser un **test du χ^2 d'adéquation** à une distribution théorique de référence.

Situation professionnelle — Contrôle qualité en production industrielle

Un technicien qualité dans une usine fabriquant des roulements à billes reçoit chaque matin un lot de $n = 50$ pièces prélevées sur la chaîne de production. Il mesure le diamètre intérieur de chaque roulement, dont la valeur nominale est $\mu_0 = 25{,}00$ mm avec un écart-type historique de $\sigma = 0{,}03$ mm.

Pour valider ou rejeter le lot, le technicien doit répondre à plusieurs questions :

- Le diamètre moyen du lot est-il conforme à la valeur nominale de 25 mm ?
- La proportion de pièces hors tolérance est-elle inférieure au seuil maximal de 2 % ?
- La distribution des diamètres suit-elle bien la loi normale attendue ?

Ces questions relèvent de la **statistique inférentielle** : à partir d'un échantillon partiel, on tire des conclusions sur l'ensemble de la production, avec un niveau de confiance quantifié et maîtrisé.

1. Population, échantillon et paramètres

Définition — Population et échantillon

La **population** est l'ensemble complet des individus ou objets étudiés (par exemple, toute la production annuelle de roulements). Un **échantillon** de taille n est un sous-ensemble représentatif extrait de cette population par tirage au sort.

Définition — Paramètres et statistiques

Un **paramètre** est une caractéristique de la population, généralement inconnue :

- μ : *moyenne* de la population
- σ^2 : *variance* de la population ; σ : écart-type
- p : *proportion* dans la population (individus vérifiant un critère)

Une **statistique** (estimateur) est calculée à partir de l'échantillon :

- $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$: moyenne empirique
- $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$: variance empirique corrigée
- $\hat{p} = \dfrac{\text{nombre de succès}}{n}$: fréquence empirique

Attention — Variance corrigée

La variance corrigée divise par $n-1$ (et non n) pour être un *estimateur sans biais* de σ^2. Avec le diviseur n, on sous-estime systématiquement la variance de la population. On utilise toujours s^2 pour les IC et tests, sauf si σ est connue par l'historique.

Exemple 1 — Calcul de la moyenne et de l'écart-type empiriques

On mesure le diamètre intérieur (en mm) de $n = 6$ roulements :
$x_1 = 24{,}98$; $x_2 = 25{,}01$; $x_3 = 25{,}00$; $x_4 = 24{,}99$; $x_5 = 25{,}02$; $x_6 = 25{,}00$.

Moyenne empirique :

$$\bar{x} = \frac{24{,}98 + 25{,}01 + 25{,}00 + 24{,}99 + 25{,}02 + 25{,}00}{6} = \frac{150{,}00}{6} = 25{,}000 \text{ mm}$$

Variance corrigée (écarts à la moyenne : $-0{,}02\ ;\ 0{,}01\ ;\ 0\ ;\ -0{,}01\ ;\ 0{,}02\ ;\ 0$) :

$$s^2 = \frac{(0{,}02)^2 + (0{,}01)^2 + 0^2 + (0{,}01)^2 + (0{,}02)^2 + 0^2}{5} = \frac{0{,}0004 + 0{,}0001 + 0 + 0{,}0001 + 0{,}0004 + 0}{5} = \frac{0{,}001}{5} = 2 \times 10^{-4} \text{ mm}^2$$

Écart-type empirique : $s = \sqrt{2 \times 10^{-4}} \approx 0{,}0141$ mm.

2. Distribution d'échantillonnage et théorème central limite

Définition — Distribution d'échantillonnage

La **distribution d'échantillonnage** de $\bar{X}$ est la loi de probabilité suivie par la moyenne d'échantillon lorsqu'on répète l'opération de prélever un échantillon de taille n dans la population. C'est une loi différente de la loi de la population individuelle.

Théorème central limite (TCL)

Soient $X_1, X_2, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées de moyenne μ et de variance $\sigma^2 < +\infty$. Pour $n \geq 30$ (ou pour tout n si la population est normale), la variable :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

suit *approximativement* la loi normale :

$$\bar{X} \sim \mathcal{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$$

L'**erreur standard** (écart-type de $\bar{X}$) vaut :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

La variable centrée réduite associée suit :

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Distribution d'échantillonnage d'une proportion

Si X suit une loi de Bernoulli de paramètre p et si $np \geq 5$ et $n(1-p) \geq 5$, alors :

$$\hat{P} \sim \mathcal{N}\left(p,\ \frac{p(1-p)}{n}\right)$$

Exemple 2 — Calcul de l'erreur standard

La production de diamètres suit une loi normale de moyenne $\mu = 25{,}00$ mm et d'écart-type $\sigma = 0{,}03$ mm. On prélève des échantillons de différentes tailles.

Taille n	Erreur standard $\sigma/\sqrt{n}$
5	$0{,}03/\sqrt{5} \approx 0{,}0134$ mm
15	$0{,}03/\sqrt{15} \approx 0{,}00775$ mm
30	$0{,}03/\sqrt{30} \approx 0{,}00548$ mm
50	$0{,}03/\sqrt{50} \approx 0{,}00424$ **mm**

Plus n augmente, plus l'erreur standard diminue : la moyenne d'échantillon devient un estimateur de plus en plus précis de μ.

MINI-EXERCICE :

Une production a un écart-type historique $\sigma = 0{,}06$ mm. On prélève un échantillon de $n = 36$ pièces. Calculer l'erreur standard $\sigma/\sqrt{n}$. Par quel facteur faudrait-il multiplier n pour diviser cette erreur standard par 2 ?

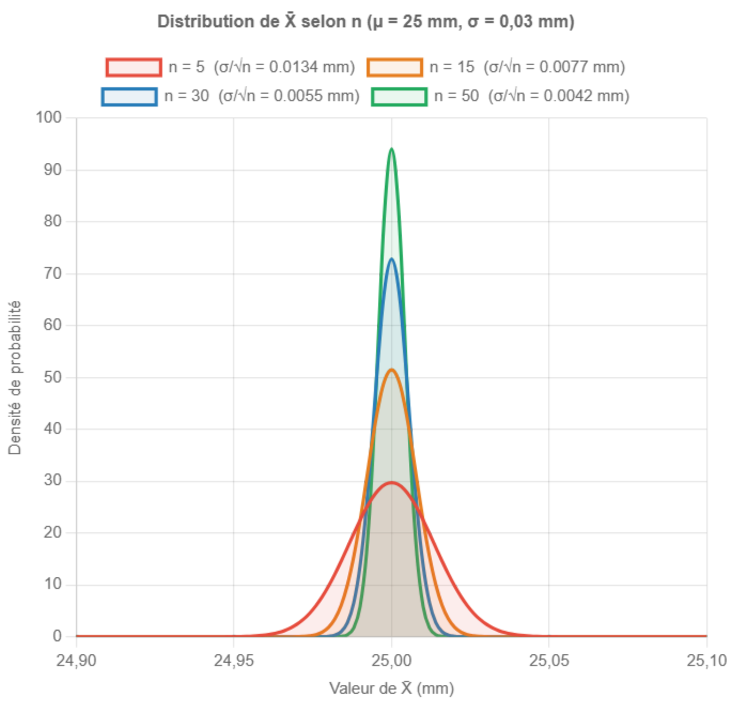

Fig. 1 — Distribution de $\bar{X}$ selon la taille de l'échantillon ($\mu = 25$, $\sigma = 0{,}03$ mm)

3. Estimation ponctuelle

Définition — Estimateur sans biais

Un estimateur $\hat{\theta}$ d'un paramètre θ est dit **sans biais** si $\mathbb{E}[\hat{\theta}] = \theta$: en moyenne sur toutes les réalisations possibles, il vise exactement la vraie valeur du paramètre.

Estimateurs usuels sans biais

Moyenne μ

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Variance σ^2

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Proportion p

$$\hat{p} = \frac{\text{nombre de succès}}{n}$$

Convergence

La variance de l'estimateur est proportionnelle à $1/n$: doubler n divise l'incertitude par $\sqrt{2}$.

4. Intervalles de confiance

Définition — Intervalle de confiance

Un **intervalle de confiance (IC)** au niveau $1-\alpha$ pour un paramètre θ est un intervalle aléatoire $[A\,;\,B]$, calculé à partir de l'échantillon, tel que :

$$\mathbb{P}(A \leq \theta \leq B) = 1 - \alpha$$

Le niveau de confiance est en général 90 %, 95 % ou 99 %. Le **seuil** α est le risque d'erreur : probabilité que l'IC construit ne contienne pas θ.

Interprétation fréquentiste : si l'on construisait un grand nombre d'IC de même niveau à partir d'échantillons différents, environ $(1-\alpha) \times 100\,\%$ d'entre eux contiendraient la vraie valeur de θ.

4.1 IC pour la moyenne (σ connue)

Formule — IC pour μ avec σ connue

Lorsque σ est connue et que $n \geq 30$ (ou que la population est normale), la loi normale donne :

$$\boxed{IC_{1-\alpha} = \left[\bar{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\ ;\ \bar{x} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right]}$$

avec $z_{\alpha/2}$ le quantile de $\mathcal{N}(0,1)$ tel que $\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$.

Niveau $1-\alpha$	α	$z_{\alpha/2}$
90 %	10 %	**1,645**
95 %	5 %	**1,960**
99 %	1 %	**2,576**

Exemple 3 — IC sur le diamètre moyen (σ connue)

Un ingénieur qualité mesure $n = 50$ roulements et obtient $\bar{x} = 25{,}012$ mm. On sait historiquement que $\sigma = 0{,}03$ mm. Construire un IC à 95 % pour le diamètre moyen.

Erreur standard : $\dfrac{0{,}03}{\sqrt{50}} = \dfrac{0{,}03}{7{,}071} \approx 0{,}00424$ mm

Marge d'erreur : $1{,}960 \times 0{,}00424 \approx 0{,}00831$ mm

$$IC_{95\,\%} = [25{,}012 - 0{,}0083\ ;\ 25{,}012 + 0{,}0083] = [25{,}004\ ;\ 25{,}020]\ \text{mm}$$

Conclusion : On est confiant à 95 % que le diamètre moyen de la production se situe entre 25,004 mm et 25,020 mm. La valeur nominale 25,000 mm n'est *pas* dans l'IC — signe d'une possible dérive.

MINI-EXERCICE :

Sur $n = 64$ bobines, on mesure une longueur moyenne $\bar{x} = 120{,}5$ m. L'écart-type de production, connu, est $\sigma = 1{,}6$ m. Construire l'intervalle de confiance à 95 % de la longueur moyenne ($z_{\alpha/2} = 1{,}96$).

4.2 IC pour la moyenne (σ inconnue — loi de Student)

Formule — IC pour μ avec σ inconnue (loi de Student)

Lorsque σ est inconnue et que la population est normale (ou n pas trop petit), on remplace σ par s et la loi normale par la **loi de Student à $n-1$ degrés de liberté** :

$$\boxed{IC_{1-\alpha} = \left[\bar{x} - t_{n-1;\,\alpha/2}\,\frac{s}{\sqrt{n}}\ ;\ \bar{x} + t_{n-1;\,\alpha/2}\,\frac{s}{\sqrt{n}}\right]}$$

La loi de Student a des queues plus épaisses que la normale : l'intervalle obtenu est plus large, ce qui reflète l'incertitude supplémentaire liée à l'estimation de σ. Quand $n \to \infty$, $t_{n-1;\,\alpha/2} \to z_{\alpha/2}$.

Attention — Conditions d'application

- La loi de Student est valable si la population suit *approximativement* une loi normale.
- Pour $n \geq 30$, la différence entre t et z devient négligeable (moins de 5 %).
- Toujours utiliser la variance corrigée s^2 (diviseur $n-1$).

Exemple 4 — IC sur la résistance de soudures (petit échantillon)

Un contrôleur qualité mesure la résistance à la rupture (en MPa) de $n = 12$ assemblages soudés :

```
420   415   430   418   425   412   422   428   416   424   419   427
```

Calculs : $\bar{x} = 421{,}33$ MPa, $s = 5{,}61$ MPa. Construire un IC à 95 %.

$\nu = n - 1 = 11$; table de Student : $t_{11;\,0{,}025} = 2{,}201$.

Marge : $2{,}201 \times \dfrac{5{,}61}{\sqrt{12}} = 2{,}201 \times 1{,}620 \approx 3{,}57$ MPa

$$IC_{95\,\%} = [421{,}33 - 3{,}57\ ;\ 421{,}33 + 3{,}57] = [417{,}8\ ;\ 424{,}9]\ \text{MPa}$$

4.3 IC pour une proportion

Formule — IC pour p

Si $n\hat{p} \geq 5$ et $n(1-\hat{p}) \geq 5$ (conditions de validité de l'approximation normale) :

$$\boxed{IC_{1-\alpha} = \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ;\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]}$$

La **marge d'erreur maximale** (atteinte pour $\hat{p} = 0{,}5$) :

$$e_{\max} = \frac{z_{\alpha/2}}{2\sqrt{n}} \qquad \text{ex. pour IC 95 \textbackslash\% : } e_{\max} \approx \frac{1{,}96}{2\sqrt{n}}$$

Exemple 5 — Proportion de pièces non conformes

Sur $n = 200$ roulements inspectés, 8 sont déclarés non conformes. $\hat{p} = 8/200 = 0{,}040$. Construire un IC à 95 %.

Vérification : $200 \times 0{,}04 = 8 \geq 5$ ✓ $\quad 200 \times 0{,}96 = 192 \geq 5$ ✓

$$\sqrt{\frac{0{,}04 \times 0{,}96}{200}} = \sqrt{0{,}000192} \approx 0{,}01386$$

$$IC_{95\,\%} = [0{,}04 - 1{,}96 \times 0{,}01386\ ;\ 0{,}04 + 1{,}96 \times 0{,}01386] \approx [0{,}013\ ;\ 0{,}067]$$

Conclusion : Le taux de non-conformité est estimé entre 1,3 % et 6,7 % avec une confiance de 95 %. Comme l'IC contient des valeurs supérieures à 2 %, il faut augmenter le contrôle.

MINI-EXERCICE :

Sur un échantillon de $n = 500$ clients, 300 se déclarent satisfaits. Donner $\hat{p}$, vérifier les conditions de validité, puis construire l'intervalle de confiance à 95 % du taux de satisfaction ($z_{\alpha/2} = 1{,}96$).

4.4 Taille d'échantillon nécessaire

Taille minimale pour une précision donnée

Pour une marge d'erreur $\leq e$ sur une proportion (niveau $1 - \alpha$) :

$$n \geq \frac{z_{\alpha/2}^2\,\hat{p}(1-\hat{p})}{e^2} \qquad \text{(sans info sur } p : n \geq \frac{z_{\alpha/2}^2}{4e^2}\text{)}$$

Pour une marge $\leq e$ sur une moyenne (σ connue) :

$$n \geq \left(\frac{z_{\alpha/2}\,\sigma}{e}\right)^2$$

Exemple 6 — Taille d'échantillon pour un sondage

Un auditeur veut estimer le taux de satisfaction des clients d'une entreprise avec une marge d'erreur de $\pm 3\,\%$ au niveau 95 %. Aucune info préalable sur p.

$$n \geq \frac{(1{,}96)^2}{4 \times (0{,}03)^2} = \frac{3{,}8416}{0{,}0036} \approx 1068$$

Il faut interroger au moins **1068 clients**.

5. Tests d'hypothèses

Définition — Hypothèses H0 et H1

Un **test d'hypothèse** est une procédure permettant de décider, à partir d'un échantillon, si les données sont compatibles avec une affirmation sur un paramètre de la population.

- H0 **Hypothèse nulle :** hypothèse de référence (conformité, statut quo). Ex. : $\mu = \mu_0$. On la rejette seulement si les données apportent une preuve suffisamment forte contre elle.
- H1 **Hypothèse alternative :** ce qu'on cherche à démontrer (dérive, amélioration, non-conformité). Ex. : $\mu \neq \mu_0$.

Définition — Erreurs de décision

	H_0 vraie	H_0 fausse
On rejette H_0	**Erreur de 1ère espèce** Probabilité : α (seuil)	Décision correcte (puissance du test : $1-\beta$)
On ne rejette pas H_0	Décision correcte	**Erreur de 2ème espèce** Probabilité : β

On fixe α à l'avance (souvent 5 %). Réduire α augmente β : on ne peut pas annuler simultanément les deux types d'erreur.

Méthode — Les 7 étapes d'un test d'hypothèse

1. Énoncer les hypothèses H_0 et H_1 précisément, en lien avec le contexte.
2. Fixer le seuil α (risque de 1ère espèce). Valeurs usuelles : 1 %, 5 %, 10 %.
3. Choisir la **statistique de test** adaptée à la situation (Z, T, χ^2...) et identifier sa loi sous H_0.
4. Calculer la **valeur observée** de la statistique de test sur l'échantillon.
5. Déterminer la **région critique** (valeur seuil selon le type de test : bilatéral, unilatéral) *ou* calculer la **p-valeur**.
6. Comparer : si $|T_{obs}|$ est dans la zone de rejet (ou p-valeur $< \alpha$), on **rejette H_0** ; sinon, on **ne rejette pas H_0**.
7. Formuler la **conclusion dans le contexte professionnel** de façon claire et précise.

5.1 Test sur la moyenne — Z-test (σ connue)

Statistique de test Z

Sous H_0 : $\mu = \mu_0$ (avec σ connue et $n \geq 30$) :

$$Z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \text{ sous } \mathrm{H}_0$$

Type de test	H_1	Zone de rejet (seuil α)
Bilatéral	$\mu \neq \mu_0$	$\lvert Z_{obs} \rvert > z_{\alpha/2}$
Unilatéral droit	$\mu > \mu_0$	$Z_{obs} > z_\alpha$
Unilatéral gauche	$\mu < \mu_0$	$Z_{obs} < -z_\alpha$

Exemple 7 — Dérive de production (Z-test bilatéral)

Diamètre nominal $\mu_0 = 25{,}000$ mm, $\sigma = 0{,}03$ mm. Sur $n = 50$ pièces, le technicien mesure $\bar{x} = 25{,}012$ mm. Y a-t-il une dérive ? (α = 5 %)

Ho : $\mu = 25{,}000$ | H₁ : $\mu \neq 25{,}000$ (bilatéral)

$$Z_{obs} = \frac{25{,}012 - 25{,}000}{0{,}03/\sqrt{50}} = \frac{0{,}012}{0{,}00424} \approx 2{,}83$$

$|Z_{obs}| = 2{,}83 > z_{0{,}025} = 1{,}96 \rightarrow$ zone de rejet.

On rejette Ho au seuil 5 %.

Conclusion : La dérive du diamètre moyen est statistiquement significative. Le technicien doit alerter la production pour un recalibrage immédiat de la machine.

MINI-EXERCICE :

Le poids nominal d'un sac de plâtre est $\mu_0 = 25$ kg avec $\sigma = 0{,}5$ kg (connu). Le service achats soupçonne un sous-remplissage. Sur $n = 25$ sacs, on mesure $\bar{x} = 24{,}78$ kg. Effectuer un Z-test unilatéral gauche au seuil $\alpha = 5\,\%$ ($z_{0{,}05} = 1{,}645$).

5.2 Test sur la moyenne — T-test (σ inconnue)

Statistique de test T

Sous Ho : $\mu = \mu_0$ avec σ inconnue :

$$T_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim \mathcal{T}(n-1) \text{ sous } \mathrm{H}_0$$

Zone de rejet bilatérale : $|T_{obs}| > t_{n-1;\,\alpha/2}$ (lu dans la table de Student)

Exemple 8 — Vérification de résistance de soudures (T-test)

Un soudeur garantit que ses soudures résistent en moyenne à $\mu_0 = 420$ MPa. Un contrôleur mesure $n = 15$ soudures : $\bar{x} = 414{,}2$ MPa, $s = 8{,}6$ MPa. Tester au seuil α = 5 %.

Ho : $\mu = 420$ | H₁ : $\mu \neq 420$ (bilatéral)

$$T_{obs} = \frac{414{,}2 - 420}{8{,}6/\sqrt{15}} = \frac{-5{,}8}{2{,}220} \approx -2{,}61$$

Degrés de liberté : $\nu = 14$; $t_{14;\,0{,}025} = 2{,}145$ (table de Student)

$|T_{obs}| = 2{,}61 > 2{,}145 \rightarrow$ zone de rejet.

On rejette Ho au seuil 5 %.

Conclusion : La résistance moyenne observée est significativement inférieure à 420 MPa. Le procédé de soudage doit être revu.

5.3 Test sur une proportion

Statistique de test — Proportion

Sous Ho : $p = p_0$, si $np_0 \geq 5$ et $n(1-p_0) \geq 5$:

$$Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim \mathcal{N}(0,1) \text{ sous } \mathrm{H}_0$$

Exemple 9 — Test sur le taux de rebut

Le cahier des charges exige $p_0 \leq 2\,\%$ de pièces défectueuses. Sur $n = 400$ pièces contrôlées, 12 sont défectueuses. Tester au seuil α = 5 %.

Ho : $p = 0{,}02$ | H1 : $p > 0{,}02$ (unilatéral droit, seuil de non-conformité)

$\hat{p} = 12/400 = 0{,}030$

$$Z_{obs} = \frac{0{,}030 - 0{,}020}{\sqrt{0{,}020 \times 0{,}980/400}} = \frac{0{,}010}{0{,}00700} \approx 1{,}43$$

$Z_{obs} = 1{,}43 < z_{0{,}05} = 1{,}645 \rightarrow$ hors zone de rejet.

On ne rejette pas Ho au seuil 5 %.

Conclusion : On n'a pas de preuve statistique que le taux de rebut dépasse 2 %. Le lot peut être accepté, mais une surveillance continue est recommandée.

5.4 La p-valeur

Définition — p-valeur

La **p-valeur** (ou niveau de signification observé) est la probabilité, *sous Ho*, d'obtenir une valeur de la statistique de test au moins aussi extrême que celle observée :

$$\text{p-valeur} = \mathbb{P}_{H_0}(|Z| \geq |Z_{obs}|) = 2\,(1 - \Phi(|Z_{obs}|)) \quad \text{(test bilatéral)}$$

Règle : si p-valeur $< \alpha$, on rejette Ho. Plus la p-valeur est petite, plus la preuve contre Ho est forte.

Exemple 10 — Calcul de p-valeur

Dans l'exemple 7 ($Z_{obs} = 2{,}83$, test bilatéral) :

$$\text{p-valeur} = 2 \times \mathbb{P}(Z \geq 2{,}83) = 2 \times (1 - \Phi(2{,}83)) \approx 2 \times (1 - 0{,}9977) = 2 \times 0{,}0023 = 0{,}0046$$

La p-valeur est 0,46 %, **très inférieure** à α = 5 % : la preuve contre H0 est extrêmement forte. La dérive de production est hautement significative.

Interprétation pratique de la p-valeur :

- p-valeur > 0,10 → pas de preuve contre H0
- 0,05 < p-valeur ≤ 0,10 → preuve faible
- 0,01 < p-valeur ≤ 0,05 → preuve modérée (significatif)
- p-valeur ≤ 0,01 → preuve forte (très significatif)
- p-valeur ≤ 0,001 → preuve très forte (hautement significatif)

6. Test du χ^2 d'adéquation

Définition

Le **test du χ^2 d'adéquation** permet de vérifier si une distribution observée dans un échantillon est statistiquement compatible avec une distribution théorique de référence (uniforme, normale, de Poisson, etc.).

H0 : la distribution suit la loi théorique de référence.
H1 : la distribution ne suit pas cette loi.

Statistique de test χ^2

Pour k classes, avec effectifs observés O_i et effectifs théoriques $E_i = n \cdot p_i$:

$$\chi^2_{obs} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Sous H0, cette statistique suit une loi du χ^2 à $\nu = k - 1 - r$ degrés de liberté, où r est le nombre de paramètres estimés à partir des données.

- **Condition de validité :** tous les $E_i \geq 5$ (sinon, regrouper des classes).
- **Zone de rejet :** $\chi^2_{obs} > \chi^2_{\nu;\,\alpha}$ (valeur critique dans la table).

Exemple 11 — Répartition des pannes sur la semaine

Un contremaître dénombre les pannes sur 5 jours sur une ligne de production : Lundi 18, Mardi 22, Mercredi 20, Jeudi 25, Vendredi 15 (total = 100 pannes). Les pannes sont-elles réparties uniformément ? (α = 5 %)

Ho : distribution uniforme → $E_i = 100/5 = 20$ pour chaque jour.

Jour	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
Lundi	18	20	−2	$4/20 = 0{,}20$
Mardi	22	20	+2	$4/20 = 0{,}20$
Mercredi	20	20	0	$0/20 = 0$
Jeudi	25	20	+5	$25/20 = 1{,}25$
Vendredi	15	20	−5	$25/20 = 1{,}25$
Total	**100**	**100**	0	**2,90**

$$\chi^2_{obs} = 2{,}90$$

$\nu = k - 1 = 4$; valeur critique : $\chi^2_{4;\,0{,}05} = 9{,}488$

$\chi^2_{obs} = 2{,}90 < 9{,}488 \longrightarrow$ hors zone de rejet.

On ne rejette pas Ho au seuil 5 %.

Conclusion : Il n'y a pas de preuve statistique contre une répartition uniforme des pannes sur la semaine.

Exemple 12 — Contrôle de conformité d'une distribution

Un gestionnaire de stock modélise le nombre de commandes urgentes par jour selon une loi de Poisson de paramètre $\lambda = 2$. Sur 100 jours, il observe :

Commandes urgentes	0	1	2	3	4	≥5
Observé O_i	14	25	30	18	9	4
Théorique E_i	13,5	27,1	27,1	18,0	9,0	5,3

(Effectifs théoriques calculés avec $P(X = k) = e^{-2} \cdot 2^k/k!$, $n = 100$)

$$\chi^2_{obs} = \frac{(14-13{,}5)^2}{13{,}5} + \frac{(25-27{,}1)^2}{27{,}1} + \frac{(30-27{,}1)^2}{27{,}1} + \frac{(18-18)^2}{18} + \frac{(9-9)^2}{9} + \frac{(4-5{,}3)^2}{5{,}3} \approx 0{,}019 + 0{,}163 + 0{,}310 + 0 + 0 + 0{,}319 = 0{,}$$

$\nu = 6 - 1 - 1 = 4$ (λ estimé depuis les données) ; $\chi^2_{4;\,0{,}05} = 9{,}488$.

$\chi^2_{obs} = 0{,}81 < 9{,}488$

Le modèle de Poisson($\lambda = 2$) est validé au seuil 5 %.

7. Table des valeurs de la loi normale standard

La table donne $\Phi(z) = \mathbb{P}(Z \leq z)$ pour $Z \sim \mathcal{N}(0, 1)$. Rappel : $\mathbb{P}(Z \leq -z) = 1 - \Phi(z)$.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	**0,9505**	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	**0,9756**	0,9761	0,9767
2,0	**0,9772**	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,8	**0,9974**	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

Valeurs clés à mémoriser : $\Phi(1{,}645) = 0{,}95$ | $\Phi(1{,}960) = 0{,}975$ | $\Phi(2{,}326) = 0{,}990$ | $\Phi(2{,}576) = 0{,}995$

8. Extrait de la table de Student

La table donne $t_{\nu;\,\alpha/2}$ tel que $\mathbb{P}(|T| > t_{\nu;\,\alpha/2}) = \alpha$ pour une loi de Student à ν degrés de liberté.

ddl ν	$t_{0,10}$ (IC 80 %)	$t_{0,05}$ (IC 90 %)	$t_{0,025}$ (IC 95 %)	$t_{0,01}$ (IC 98 %)	$t_{0,005}$ (IC 99 %)
5	1,476	2,015	**2,571**	3,365	4,032
8	1,397	1,860	**2,306**	2,896	3,355
10	1,372	1,812	**2,228**	2,764	3,169
11	1,363	1,796	**2,201**	2,718	3,106
14	1,345	1,761	**2,145**	2,624	2,977
19	1,328	1,729	**2,093**	2,539	2,861
24	1,318	1,711	**2,064**	2,492	2,797
29	1,311	1,699	**2,045**	2,462	2,756
49	1,299	1,677	**2,010**	2,403	2,680
∞	1,282	1,645	**1,960**	2,326	2,576

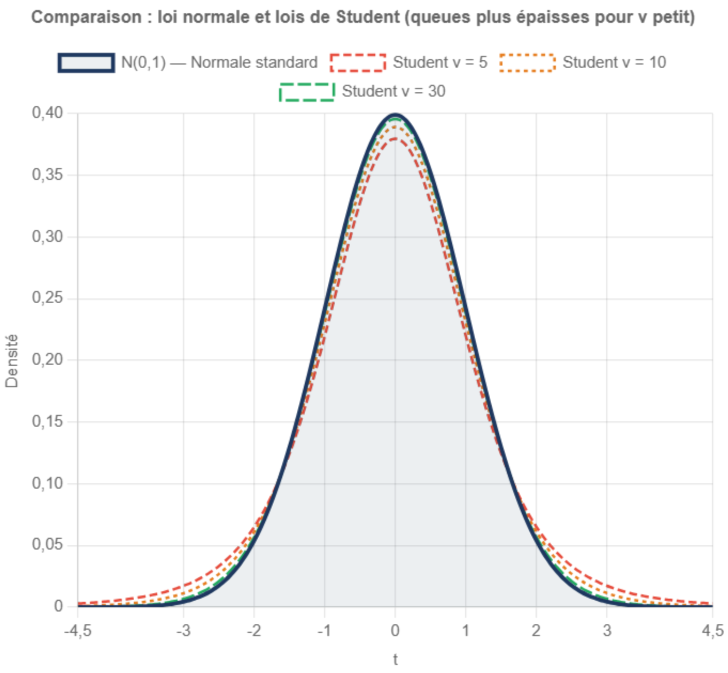

Fig. 2 — Densités : loi normale $\mathcal{N}(0,1)$ et lois de Student pour différents ddl

9. Exemple industriel complet — Décision d'acceptation de lot

Contexte : Un responsable qualité dans une fonderie reçoit un lot de $n = 100$ pièces coulées. La spécification indique une masse nominale de $\mu_0 = 850$ g avec un écart-type historique de $\sigma = 12$ g. Il mesure la masse de toutes les pièces et obtient $\bar{x} = 847{,}2$ g.

Répondre aux questions suivantes au seuil $\alpha = 5\,\%$ puis $\alpha = 1\,\%$:

1. Construire un IC à 95 % et à 99 % pour la masse moyenne.
2. La masse moyenne est-elle conforme à la spécification ?
3. Calculer la p-valeur et l'interpréter.

À retenir — Formules essentielles

Théorème central limite

$$\bar{X} \sim \mathcal{N}\left(\mu,\ \frac{\sigma^2}{n}\right) \quad (n \geq 30)$$

IC pour μ (σ connue)

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

IC pour μ (σ inconnue, Student)

$$\bar{x} \pm t_{n-1;\,\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

IC pour p

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Z-test — Moyenne

$$Z_{obs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

T-test — Moyenne (σ inconnue)

$$T_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Test sur proportion

$$Z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

χ^2 d'adéquation

$$\chi^2_{obs} = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Niveau	$z_{\alpha/2}$	Usage typique
90 %	1,645	Contrôle courant, sondages rapides
95 %	1,960	Standard industriel, études de marché
99 %	2,576	Sécurité, pharmacologie, nucléaire

11

MATHÉMATIQUES

Statistique inférentielle

BTS | Mathématiques | Durée : 40 min | /20

Nom : _____ Prénom : _____ Date : _____

Valeurs utiles : $z_{\alpha/2} = 1{,}645$ (90 %), 1,960 (95 %), 2,576 (99 %). $\Phi(2{,}50) = 0{,}9938$; $\Phi(1{,}43) = 0{,}9236$.
Table de Student : $t_{14;\,0{,}025} = 2{,}145$.

Exercice 1 — Erreur standard et estimation (3 pts)

Une chaîne produit des axes dont l'écart-type historique du diamètre est $\sigma = 0{,}08$ mm. On prélève un échantillon de $n = 64$ axes.

a. Calculer l'erreur standard $\sigma/\sqrt{n}$ de la moyenne d'échantillon. *(1,5 pt)*
b. Par quel facteur faut-il multiplier la taille n pour diviser cette erreur standard par 2 ? *(1,5 pt)*

Exercice 2 — Intervalle de confiance pour une moyenne (4 pts)

Sur $n = 100$ pièces coulées, un responsable qualité mesure une masse moyenne $\bar{x} = 848{,}5$ g. L'écart-type de production, connu, est $\sigma = 12$ g.

a. Calculer l'erreur standard $\sigma/\sqrt{n}$. *(1 pt)*
b. Construire l'intervalle de confiance à 95 % de la masse moyenne. *(2 pts)*
c. La masse nominale est 850 g. Cette valeur appartient-elle à l'intervalle ? Que peut-on en conclure ? *(1 pt)*

Exercice 3 — Intervalle de confiance pour une proportion (4 pts)

Sur $n = 400$ clients interrogés, 240 se déclarent satisfaits.

a. Calculer la fréquence empirique $\hat{p}$ et vérifier les conditions de validité $n\hat{p} \geq 5$ et $n(1-\hat{p}) \geq 5$. *(1,5 pt)*
b. Construire l'intervalle de confiance à 95 % du taux de satisfaction. *(2,5 pts)*

Exercice 4 — Test sur la moyenne, Z-test (5 pts)

Le diamètre nominal d'un roulement est $\mu_0 = 25{,}000$ mm, avec $\sigma = 0{,}03$ mm (connu). Sur $n = 36$ pièces, le technicien mesure $\bar{x} = 25{,}0125$ mm. On teste l'existence d'une dérive au seuil $\alpha = 5\,\%$.

a. Énoncer les hypothèses H_0 et H_1 (test bilatéral). *(1 pt)*
b. Calculer la statistique de test Z_{obs}. *(2 pts)*
c. Conclure (zone de rejet : $|Z_{obs}| > 1{,}96$). *(1 pt)*
d. Calculer la p-valeur du test (sachant $\Phi(2{,}50) = 0{,}9938$) et l'interpréter. *(1 pt)*

Exercice 5 — Test du χ^2 d'adéquation (4 pts)

Un contremaître relève le nombre de pannes par jour ouvré sur une semaine (total = 100 pannes). On veut tester une répartition uniforme au seuil $\alpha = 5\,\%$. Valeur critique : $\chi^2_{4;\,0{,}05} = 9{,}488$.

Jour	Lun	Mar	Mer	Jeu	Ven
Effectif observé O_i	16	24	20	22	18

a. Donner l'effectif théorique E_i attendu sous l'hypothèse d'une répartition uniforme. *(1 pt)*
b. Calculer la statistique χ^2_{obs}. *(2 pts)*
c. Conclure. *(1 pt)*