

Chapitre 3 – Estimations et tests d'hypothèse (3^e partie)

STT-1920 Méthodes statistiques

Département de mathématiques et de statistique



UNIVERSITÉ
LAVAL

Des hypothèses de recherche aux hypothèses statistiques

Où en sommes-nous ?

- Au **chapitre 3a**, nous avons **estimé** des paramètres ($\bar{x}$, s^2 , $\hat{p}$) et mesuré leur précision (erreur-type).
- Au **chapitre 3b**, nous avons construit des **intervalles de confiance**.
- Souvent, le chercheur doit plutôt **prendre une décision** face à une affirmation :
 - la concentration moyenne de mercure dépasse-t-elle la norme ?
 - la nouvelle diète augmente-t-elle la masse des souris ?

Le test d'hypothèses

Une procédure qui permet de décider, à partir d'un échantillon, si les données fournissent une **preuve suffisante** contre une affirmation. C'est l'outil de base des chapitres 4 à 7 (comparaisons de groupes, ANOVA, régression, khi-deux).

Des questions de recherche

Une question de recherche porte souvent sur un **paramètre** d'une population :

- **Toxicologie** : la concentration moyenne de mercure dans les dorés d'un réservoir est **inférieure** à la ligne directrice de 0,5 µg/g.
- **Santé animale** : la prévalence d'une infection dans un troupeau **dépasse** 10%.
- **Pharmacologie** : l'écart-type de la dose de principe actif des comprimés est **supérieur** à 2 mg.
- **Physiologie** : la pression systolique moyenne d'une population **diffère** de 130 mm Hg.

Chaque énoncé porte sur un paramètre (μ , p , σ) et le compare à une **valeur de référence**.

Hypothèses statistiques : H_0 et H_1

Hypothèse statistique

Énoncé sur la loi de probabilité d'une variable aléatoire, le plus souvent sur la valeur d'un paramètre (ex. : $\mu = 130$, $p > 0,10$, $\sigma^2 \neq 4$).

Un test confronte deux hypothèses, par exemple $H_0 : \mu = 25$ contre $H_1 : \mu > 25$.

- H_0 : l'**hypothèse nulle** (« rien de nouveau », « pas d'effet »). Elle est **supposée vraie** tant que les données n'apportent pas une preuve suffisante du contraire.
- H_1 : l'**hypothèse alternative**. Elle incarne l'effet ou le changement que le chercheur veut **démontrer**.

Quelques règles pour formuler H_0 et H_1

- H_0 et H_1 sont **mutuellement exclusives**.
- H_0 contient toujours l'**égalité** : $H_0 : \mu = \mu_0$, où μ_0 est la valeur de référence.
- H_1 peut être :
 - **bilatérale** : $H_1 : \mu \neq \mu_0$;
 - **unilatérale** à droite ($H_1 : \mu > \mu_0$) ou à gauche ($H_1 : \mu < \mu_0$).
- C'est l'**hypothèse de recherche** (ce qu'on veut démontrer) qui dicte H_1 , et elle doit être fixée **avant** de voir les données.

QUIZ!

Formulez H_0 et H_1 :

- Un agronome veut démontrer qu'un nouvel engrais porte le rendement moyen du blé **au-delà** de 3,5 t/ha.
- Une biologiste veut vérifier si la proportion de tiques porteuses de la bactérie de Lyme dans un parc **dépasse** 20%.
- Un médicament **modifie**-t-il la fréquence cardiaque moyenne au repos (72 battements/min)?
- Sur 10 parcelles fertilisées, l'agronome obtient $\bar{x} = 3,62$ t/ha. Peut-il conclure que l'engrais est efficace?

La logique d'un test

Exemple 7 : une diète enrichie chez la souris

Dans une lignée de souris de laboratoire, la masse à 8 semaines suit une loi $\mathcal{N}(\mu, \sigma^2 = 2^2)$; sous la diète standard, $\mu = 25$ g. Grâce aux nombreuses données historiques, on considère $\sigma = 2$ g comme **connu**.

On nourrit $n = 16$ souris avec une **diète enrichie** et on obtient $\bar{x} = 26,1$ g.

Question : la diète enrichie augmente-t-elle la masse moyenne? (seuil $\alpha = 5\%$)

$$H_0 : \mu = 25 \quad \text{contre} \quad H_1 : \mu > 25$$

Le dilemme : $26,1 > 25$, mais $\bar{X}$ varie d'un échantillon à l'autre. L'écart de 1,1 g est-il **trop grand pour être dû au hasard** si H_0 est vraie?

Les étapes d'un test d'hypothèses

1. **Hypothèses** : formuler H_0 et H_1 et choisir le **seuil** α .
2. **Statistique de test** : choisir une statistique dont on connaît la **loi si H_0 est vraie**.
3. **Règle de décision** : déterminer la **région de rejet** (ou utiliser la valeur-p).
4. **Calcul** : calculer la statistique observée à partir de l'échantillon.
5. **Décision et conclusion** : rejeter ou ne pas rejeter H_0 , puis conclure **dans les termes de l'étude**.

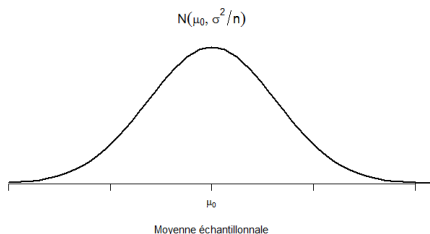
Seulement deux décisions possibles

« On rejette H_0 » ou « on ne rejette pas H_0 ». On ne dit jamais « on accepte H_0 ».

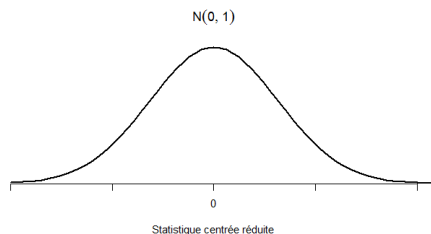
Statistique de test et loi sous H_0

Si H_0 est vraie ($\mu = \mu_0$), alors $\bar{X} \sim \mathcal{N}(\mu_0, \sigma^2/n)$ et on standardise :

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad \text{sous } H_0.$$



Loi de $\bar{X}$ sous H_0

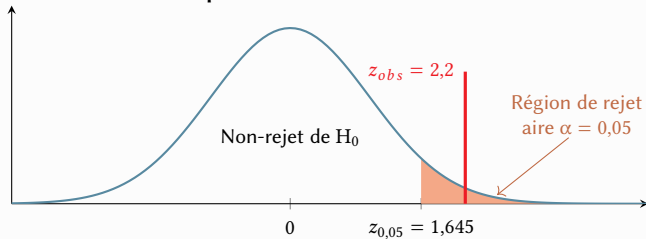


Loi de Z_0 sous H_0

Exemple 7 : $z_{obs} = \frac{26,1 - 25}{2/\sqrt{16}} = \frac{1,1}{0,5} = 2,2$: la moyenne observée est à 2,2 erreurs-types de μ_0 .

Règle de décision et région de rejet

On rejette $H_0 : \mu = 25$ en faveur de $H_1 : \mu > 25$ si $\bar{x}$ est **trop grand**, c'est-à-dire si z_{obs} tombe dans les $\alpha = 5\%$ les plus extrêmes **à droite** de la loi $\mathcal{N}(0, 1)$.



Règle : rejeter H_0 si $z_{obs} > z_{\alpha} = 1,645$. Sur l'échelle de $\bar{x}$: si $\bar{x} > 25 + 1,645 \times 0,5 = 25,82$ g.

Exemple 7 (suite) : décision et conclusion

1. **Hypothèses** : $H_0 : \mu = 25$ contre $H_1 : \mu > 25$, avec $\alpha = 0,05$.
2. **Statistique** : $Z_0 = \frac{\bar{X}-25}{2/\sqrt{16}} \sim \mathcal{N}(0, 1)$ sous H_0 .
3. **Règle de décision** : rejeter H_0 si $z_{obs} > 1,645$.
4. **Calcul** : $z_{obs} = 2,2$.
5. **Décision** : $2,2 > 1,645$, donc on **rejette** H_0 .

Conclusion dans les termes de l'étude

Au seuil de 5%, les données démontrent que la diète enrichie **augmente** la masse moyenne des souris de 8 semaines.

Erreurs de type I et II

Deux erreurs possibles

La décision est prise à partir d'un échantillon : on peut se tromper !

Décision	Réalité (inconnue!)	
	H_0 vraie	H_0 fausse
Rejeter H_0	Erreur de type I	Bonne décision
Ne pas rejeter H_0	Bonne décision	Erreur de type II

- $\alpha = P(\text{erreur de type I}) = P(\text{rejeter } H_0 \mid H_0 \text{ vraie})$: le **seuil** du test.
- $\beta = P(\text{erreur de type II}) = P(\text{ne pas rejeter } H_0 \mid H_0 \text{ fausse})$.
- $1 - \beta = P(\text{rejeter } H_0 \mid H_0 \text{ fausse})$: la **puissance** du test.

Deux analogies

Un procès

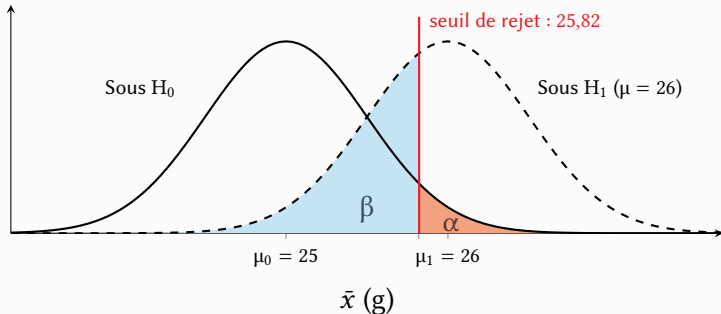
- H_0 : l'accusé est innocent ; H_1 : il est coupable.
- On condamne seulement si la preuve est **hors de tout doute raisonnable**.
- Type I : condamner un innocent.
Type II : acquitter un coupable.
- Acquitter $\neq$ prouver l'innocence!

Un test diagnostique

- H_0 : le patient n'a pas la maladie ; H_1 : il l'a.
- Type I : **faux positif**. Type II : **faux négatif**.
- $1 - \alpha$ correspond à la **spécificité**, la puissance $1 - \beta$ à la **sensibilité**.

Visualiser α et β (exemple 7)

Si en réalité la diète fait passer la moyenne à $\mu_1 = 26$ g, $\bar{X} \sim \mathcal{N}(26, 0,5^2)$ au lieu de $\mathcal{N}(25, 0,5^2)$.



$\alpha = 0,05$ (orange); $\beta = P(\bar{X} \leq 25,82 \mid \mu = 26) = 0,361$ (bleu); puissance = $1 - \beta = 0,639$.

Quelques remarques importantes

- L'erreur de type I est habituellement jugée la plus grave : on fixe α **petit** (0,05 ou 0,01) **avant** l'analyse.
- À n fixe, diminuer α **augmente** β (la ligne rouge se déplace vers la droite).
- On contrôle donc α directement, et β surtout par la **taille d'échantillon**.
- β est souvent grand avec de petits échantillons : « ne pas rejeter H_0 » signifie seulement que la preuve est **insuffisante**, pas que H_0 est vraie.

L'absence de preuve n'est pas la preuve de l'absence

Un test non significatif ne démontre pas l'absence d'effet.

QUIZ!

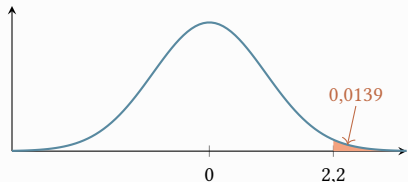
- Dans l'exemple 7, décrivez en mots une erreur de type I et une erreur de type II.
- On dépiste une maladie grave mais traitable (H_0 : patient sain). Quelle erreur est la plus grave? Que faut-il privilégier?
- Vrai ou faux : si on ne rejette pas H_0 , la probabilité que H_0 soit vraie est $1 - \alpha$.

La valeur-p

La valeur-p (seuil observé)

Valeur-p

Probabilité, si H_0 est vraie, d'obtenir une statistique de test au moins aussi extrême (dans le sens de H_1) que celle observée. C'est aussi le plus petit α pour lequel on rejetterait H_0 avec ces données.



Exemple 7 : valeur-p
 $= P(Z \geq 2,2) = 0,0139$.

Règle universelle

Rejeter H_0 au seuil $\alpha \iff$ valeur-p
 $\leq \alpha$.

Ici $0,0139 \leq 0,05$: on rejette H_0 .

Calcul de la valeur-p selon H_1

Soit z_{obs} la valeur observée de Z_0 (même principe pour t_{obs} avec la loi t_{n-1}) :

Alternative	Valeur-p	Rejet de H_0 au seuil α si
$H_1 : \mu > \mu_0$ (droite)	$P(Z \geq z_{obs})$	$z_{obs} > z_\alpha$
$H_1 : \mu < \mu_0$ (gauche)	$P(Z \leq z_{obs})$	$z_{obs} < -z_\alpha$
$H_1 : \mu \neq \mu_0$ (bilatérale)	$2 P(Z \geq z_{obs})$	$ z_{obs} > z_{\alpha/2}$

Dans R : `1 - pnorm(2.2)` donne 0,0139.

Pour un test t bilatéral : `2 * (1 - pt(abs(t_obs), df))`.

Les logiciels donnent toujours la valeur-p : c'est ainsi que les résultats sont rapportés dans les articles scientifiques et dans les chapitres 4 à 7.

Interpréter une valeur-p

La valeur-p mesure la **compatibilité des données avec H_0** :

- valeur-p **petite** : les données seraient **surprenantes** si H_0 était vraie $\Rightarrow$ preuve contre H_0 ;
- valeur-p **grande** : les données sont compatibles avec H_0 (ce qui ne prouve pas H_0).

Erreurs d'interprétation courantes

- **Faux** : « La valeur-p est la probabilité que H_0 soit vraie. »
- **Faux** : « $1 - \text{valeur-p}$ est la probabilité que H_1 soit vraie. »
- **Faux** : « valeur-p $> 0,05$, donc il n'y a pas d'effet. »
- **Faux** : « Une très petite valeur-p signifie un effet très important. »
- **Faux** : « $p = 0,049$ et $p = 0,051$ sont des résultats fondamentalement différents. »

Exemple 8 : la valeur-p n'est pas $P(H_0 \text{ vraie})$

Un laboratoire teste **1000 composés** pour détecter une activité antibactérienne. En réalité, seuls **100** sont actifs. Chaque test utilise $\alpha = 0,05$ et a une puissance de 0,80.

	Inactif (H_0 vraie)	Actif (H_0 fausse)	Total
Test significatif	$900 \times 0,05 = 45$	$100 \times 0,80 = 80$	125
Test non significatif	855	20	875
Total	900	100	1000

Parmi les 125 résultats « significatifs », $45/125 = 36\%$ sont des **faux positifs**, bien loin de 5%! La probabilité que H_0 soit vraie dépend aussi de la **proportion d'hypothèses vraies au départ**, ce que la valeur-p ignore.

QUIZ!

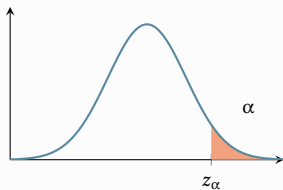
On mesure la croissance (cm) de 20 plantules et on teste $H_0 : \mu = 5$ contre $H_1 : \mu < 5$. La valeur-p est 0,09. Vrai, faux ou information insuffisante ?

- La moyenne échantillonnale est inférieure à 5.
- On rejette H_0 au seuil de 5%.
- On rejette H_0 au seuil de 10%.
- On rejetterait $H_0 : \mu = 4$ contre $H_1 : \mu < 4$ au seuil de 5%.
- On rejetterait $H_0 : \mu = 6$ contre $H_1 : \mu < 6$ au seuil de 5%.
- Avec 40 plantules, la même moyenne et le même écart-type, la valeur-p serait plus grande que 0,09.

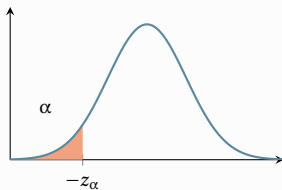
Tests unilatéraux et bilatéraux

Trois formes de région de rejet (seuil α)

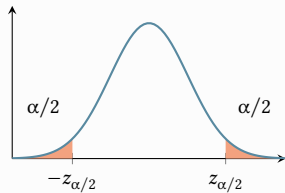
$$H_1 : \mu > \mu_0$$



$$H_1 : \mu < \mu_0$$



$$H_1 : \mu \neq \mu_0$$



- Au seuil $\alpha = 5\%$: $z_{0,05} = 1,645$ (unilatéral) mais $z_{0,025} = 1,96$ (bilatéral).
- Le test bilatéral répartit le risque α **des deux côtés** : il faut un écart plus grand pour rejeter H_0 dans une direction donnée.

Unilatéral ou bilatéral : comment choisir ?

- Le choix découle de la **question de recherche**, et il est fait **avant** de regarder les données.
- En cas de doute, ou si un effet dans les deux sens serait intéressant (ex. : un médicament peut augmenter *ou* diminuer la fréquence cardiaque), on utilise un test **bilatéral**.
- Un test unilatéral se justifie quand seul un sens intéresse (ex. : dépassement d'une norme environnementale).
- Si la statistique est dans le sens de H_1 : $\text{valeur-p}_{\text{unilatérale}} = \frac{1}{2} \text{valeur-p}_{\text{bilatérale}}$.

Attention !

Choisir un test unilatéral *après* avoir vu le sens des données pour « diviser la valeur-p par 2 » est une faute : le vrai risque d'erreur de type I devient 2α .

QUIZ!

Un test t bilatéral sur μ donne $t_{obs} = 1,85$ et une valeur-p de 0,08. Vrai ou faux?

- Pour $H_1 : \mu > \mu_0$, la valeur-p serait 0,04.
- Pour $H_1 : \mu < \mu_0$, la valeur-p serait 0,96.
- Au seuil de 5%, le test bilatéral rejette H_0 .

La puissance d'un test

Calcul de la puissance (exemple 7)

- Puissance = $1 - \beta = P(\text{rejeter } H_0 \mid H_1 \text{ vraie})$.
- $H_1 : \mu > 25$ contient une infinité de valeurs : on calcule la puissance pour une valeur **jugée importante** μ_1 (ex. : $\mu_1 = 26$ g, un gain de 1 g).

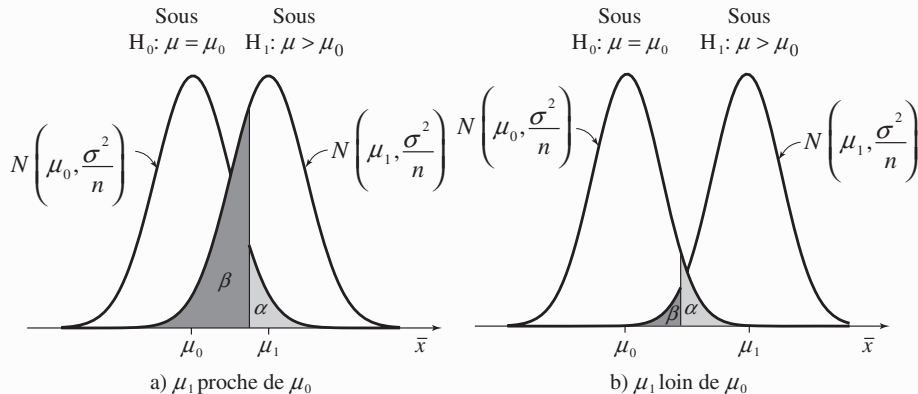
On rejette H_0 si $\bar{X} > 25,82$. Si $\mu = 26$:

$$1 - \beta = P(\bar{X} > 25,82 \mid \mu = 26) = P\left(Z > \frac{25,82 - 26}{0,5}\right) = P(Z > -0,36) = \mathbf{0,639}.$$

Interprétation : si la diète augmente vraiment la masse moyenne de 1 g, le test avec 16 souris ne le détectera que dans environ 64% des expériences.

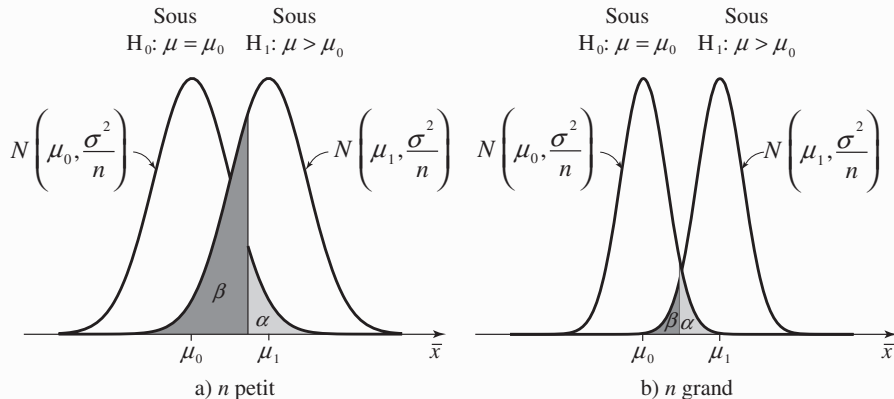
Pour $\mu_1 = 25,5$: puissance = 0,26 ; pour $\mu_1 = 27$: puissance = 0,99.

Déterminant 1 : l'écart entre μ_1 et μ_0



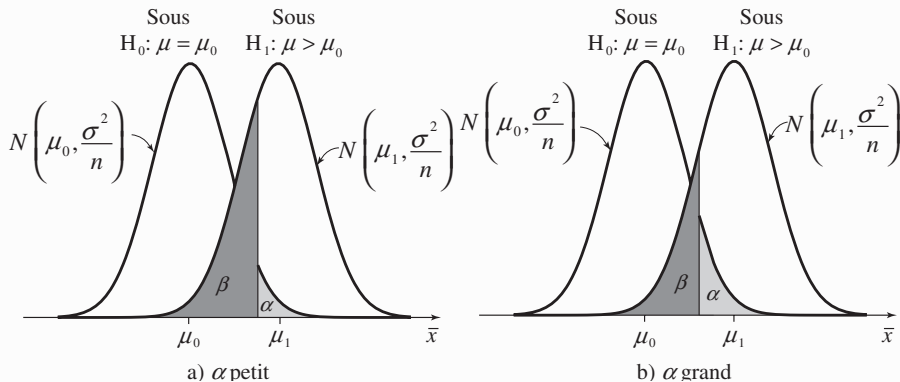
Plus la vraie valeur μ_1 est **éloignée** de μ_0 , plus β est petit et plus la **puissance est grande** : les grands effets sont plus faciles à détecter.

Déterminant 2 : la taille d'échantillon n



Quand n augmente, les deux lois de $\bar{X}$ se **resserrent** ($\sigma/\sqrt{n}$ diminue) : elles se chevauchent moins et la **puissance augmente**.

Déterminant 3 : le seuil α



Un α plus grand déplace la frontière de rejet vers μ_0 : la **puissance augmente**, mais au prix de plus d'erreurs de type I.

En résumé : comment augmenter la puissance ?

Exemple 7 ($\mu_0 = 25$, $\mu_1 = 26$, $\sigma = 2$) :

α	n	Puissance	α	n	Puissance
0,01	16	0,37	0,05	16	0,64
0,05	16	0,64	0,05	25	0,80
0,10	16	0,76	0,05	36	0,91

1. On ne contrôle pas μ_1 : on demande au chercheur la **plus petite différence importante** à détecter.
2. Augmenter α est rarement acceptable (plus de faux positifs).
3. En pratique, le levier principal est la **taille d'échantillon** n , à planifier **avant** l'étude.

Taille d'échantillon pour une puissance donnée

Pour un test unilatéral sur μ (σ connue), seuil α , puissance $1 - \beta$ pour détecter un écart $\delta = |\mu_1 - \mu_0|$:

Taille d'échantillon

$$n \geq \left(\frac{(z_\alpha + z_\beta) \sigma}{\delta} \right)^2 \quad (\text{bilatéral : remplacer } z_\alpha \text{ par } z_{\alpha/2})$$

Exemple 7 : combien de souris pour une puissance de 80% si $\mu_1 = 26$?

$$n \geq \left(\frac{(1,645 + 0,842) \times 2}{1} \right)^2 = 24,7 \quad \Rightarrow \quad \mathbf{n = 25 \text{ souris.}}$$

Pour σ inconnue, les calculs font intervenir une loi de Student « décentrée » ; R le fait avec `power.t.test()`.

QUIZ!

Vrai ou faux?

- Passer de $\alpha = 5\%$ à $\alpha = 1\%$ augmente la puissance du test.
- Il est plus facile de détecter une augmentation de 2 g que de 0,5 g.
- Avec seulement 5 souris par expérience, un résultat non significatif montre que la diète est inefficace.
- Un comité d'éthique animale peut exiger un calcul de puissance avant une expérience.

Tests sur une moyenne μ

Test t sur μ : loi normale, σ inconnue

En pratique, σ est inconnue : on la remplace par S (comme pour l'IC du chapitre 3b).

Statistique de test

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \quad \text{sous } H_0 : \mu = \mu_0 \quad \text{valeur observée : } t_{obs}$$

Alternative	Rejeter H_0 si	Valeur-p
$H_1 : \mu \neq \mu_0$	$ t_{obs} > t_{n-1, \alpha/2}$	$2P(t_{n-1} \geq t_{obs})$
$H_1 : \mu > \mu_0$	$t_{obs} > t_{n-1, \alpha}$	$P(t_{n-1} \geq t_{obs})$
$H_1 : \mu < \mu_0$	$t_{obs} < -t_{n-1, \alpha}$	$P(t_{n-1} \leq t_{obs})$

Ce sont les mêmes règles qu'au chapitre 4, avec d'autres degrés de liberté.

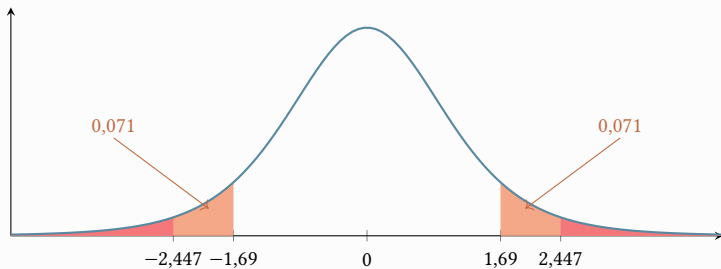
Exemple 1 (retour) : la pression artérielle

Mesures ($n = 7$) : 141, 130, 135, 138, 137, 122, 135; $\bar{x} = 134$, $s = 6,27$ mm Hg. Une étude affirme que la pression systolique moyenne de cette population est de 130 mm Hg. Nos données contredisent-elles cette valeur ($\alpha = 5\%$) ?

1. **Hypothèses** : $H_0 : \mu = 130$ contre $H_1 : \mu \neq 130$ (bilatéral).
2. **Règle** : rejeter H_0 si $|t_{obs}| > t_{6, 0,025} = 2,447$.
3. **Calcul** : $t_{obs} = \frac{134 - 130}{6,27 / \sqrt{7}} = \frac{4}{2,37} = 1,69$.
4. **Décision** : $|1,69| < 2,447$: on **ne rejette pas** H_0 (valeur-p = 0,14).
5. **Conclusion** : au seuil de 5%, les données ne permettent pas de conclure que la pression moyenne diffère de 130 mm Hg.

Exemple 1 (retour) : la valeur-p en image

Loi t_6 (loi de T_0 si H_0 est vraie)



- Valeur-p = $P(|t_6| \geq 1,69) = 2 \times 0,071 = 0,14$ (aire orange totale) $> 0,05$.
- Région de rejet au seuil de 5% (en rouge) : $|t| > 2,447$; $t_{obs} = 1,69$ n'y tombe pas.

Exemple 1 (retour) : avec R

```
> x <- c(141, 130, 135, 138, 137, 122, 135)
> t.test(x, mu = 130)
```

One Sample t-test

```
data:  x
t = 1.6874, df = 6, p-value = 0.1425
alternative hypothesis: true mean is not equal to 130
95 percent confidence interval:
 128.1997 139.8003
sample estimates:
mean of x
      134
```

Options: alternative = "greater" ($H_1 : \mu > \mu_0$) ou "less";
conf.level = 0.99. Sans mu = , R teste $H_0 : \mu = 0$ (rarement pertinent!).

Ne pas rejeter H_0 ne veut pas dire $\mu = 130$!

- Avec seulement $n = 7$ personnes, l'erreur-type (2,37) est grande : le test a **peu de puissance**.
- L'IC à 95% [128,20 ; 139,80] contient 130, mais aussi 135 ou 139 : beaucoup de valeurs de μ restent **plausibles**.
- Conclusion correcte : « les données sont **compatibles** avec $\mu = 130$ », et non « on a montré que $\mu = 130$ ».

Bonne pratique

Rapporter à la fois la **valeur-p** et l'**intervalle de confiance** : l'IC montre l'ampleur plausible de l'effet et la précision obtenue.

Cas n grand, loi inconnue : exemple 3 (mercure)

Si n est grand ($n \geq 30$), par le TLC : $Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \approx \mathcal{N}(0, 1)$ sous H_0 (niveau approximatif).

Exemple 3 : $n = 50$ dorés, $\bar{x} = 0,42 \mu\text{g/g}$, $s = 0,18 \mu\text{g/g}$. La concentration moyenne est-elle **inférieure** à la ligne directrice de $0,5 \mu\text{g/g}$ ($\alpha = 1\%$) ?

1. $H_0 : \mu = 0,5$ contre $H_1 : \mu < 0,5$.
2. Rejeter H_0 si $z_{obs} < -z_{0,01} = -2,326$.
3. $z_{obs} = \frac{0,42 - 0,5}{0,18/\sqrt{50}} = \frac{-0,08}{0,0255} = -3,14$.
4. $-3,14 < -2,326$: on **rejette H_0** (valeur- $p = P(Z \leq -3,14) = 0,0008$).
5. Au seuil de 1%, la concentration moyenne est significativement inférieure à $0,5 \mu\text{g/g}$.

Exemple 9 : nitrates dans l'eau de puits

On mesure la concentration d'azote des nitrates (mg/L) dans 12 puits d'une zone agricole. La norme de qualité de l'eau potable est de 10 mg/L.

9,8 12,4 10,9 13,1 8,7 11,5 12,8 10,2 9,5 13,6 11,9 10,6

On suppose les concentrations normales. $\bar{x} = 11,25$ mg/L et $s = 1,55$ mg/L.

1. Au seuil de 5%, la concentration moyenne **dépasse**-t-elle la norme? Suivez les 5 étapes.
2. Donnez un encadrement de la valeur-p à l'aide de la table de Student, puis la commande R qui la calcule.
3. Pourquoi un test unilatéral est-il justifié ici?

(Réponses à la fin du chapitre.)

Tests sur une proportion et sur une variance

Test sur une proportion

$Y \sim \text{Binomiale}(n, p)$, $\hat{p} = y/n$. On teste $H_0 : p = p_0$.

Statistique de test (approximation normale)

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{p_0(1 - p_0)/n}} \approx \mathcal{N}(0, 1) \quad \text{sous } H_0, \quad \text{si } np_0 \geq 5 \text{ et } n(1 - p_0) \geq 5.$$

- Mêmes règles de décision et valeurs-p que pour Z_0 sur μ (unilatéral ou bilatéral).
- **Attention** : l'erreur-type utilise p_0 (la valeur supposée sous H_0), alors que l'IC du chapitre 3b utilise $\hat{p}$.
- Si n est petit, on utilise le **test binomial exact** (`binom.test()` dans R), qui calcule la valeur-p directement avec la loi binomiale.

Exemple 4 (retour) : infection chez des veaux

Sur 80 veaux, 12 sont infectés ($\hat{p} = 0,15$). Le vétérinaire intervient si la prévalence **dépasse 10%**. Au seuil de 5%?

1. $H_0 : p = 0,10$ contre $H_1 : p > 0,10$. Conditions : $80 \times 0,10 = 8 \geq 5$ et $72 \geq 5$.
2. Rejeter H_0 si $z_{obs} > z_{0,05} = 1,645$.
3.
$$z_{obs} = \frac{0,15 - 0,10}{\sqrt{0,10 \times 0,90/80}} = \frac{0,05}{0,0335} = 1,49.$$
4. $1,49 < 1,645$: on **ne rejette pas H_0** (valeur-p = $P(Z \geq 1,49) = 0,068$).
5. Au seuil de 5%, les données ne démontrent pas que la prévalence dépasse 10%.

Remarque : $\hat{p} = 15\%$ est pourtant supérieur à 10%... mais avec 80 veaux, un tel écart peut facilement survenir par hasard.

Exemple 4 (retour) : avec R

```
> prop.test(12, 80, p = 0.10, alternative = "greater",  
+           correct = FALSE)  
X-squared = 2.2222, df = 1, p-value = 0.06802  
> binom.test(12, 80, p = 0.10, alternative = "greater")  
number of successes = 12, number of trials = 80,  
p-value = 0.1004
```

- `prop.test(..., correct = FALSE)` reproduit le calcul à la main ($z_{obs}^2 = 1,49^2 = 2,22$).
- `binom.test()` donne la valeur-p **exacte** : $P(Y \geq 12 \mid n = 80, p = 0,10) = 0,100$.
- Avec $np_0 = 8$, l'approximation normale est grossière : on préfère le test exact. La conclusion est la même.

QUIZ!

Exemple 2 (sondage) : sur $n = 1028$ Québécois, 216 citent l'environnement ($\hat{p} = 0,2101$). Un groupe affirme que 25% des Québécois ont cette opinion.

- Formulez H_0 et H_1 pour vérifier cette affirmation.
- Calculez z_{obs} et la valeur-p. Conclusion au seuil de 5%?
- Retrouvez la conclusion à l'aide de l'IC à 95% du chapitre 3b : $[0,185 ; 0,235]$.

Test sur σ^2 : population normale

Pour contrôler la **variabilité** (uniformité d'un produit, précision d'un appareil), on teste $H_0 : \sigma^2 = \sigma_0^2$.

Statistique de test

$$U_0 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2 \quad \text{sous } H_0 \quad \text{valeur observée : } u_{obs}$$

Alternative	Rejeter H_0 si	Valeur-p
$H_1 : \sigma^2 > \sigma_0^2$	$u_{obs} > \chi_{n-1, \alpha}^2$	$P(\chi_{n-1}^2 \geq u_{obs})$
$H_1 : \sigma^2 < \sigma_0^2$	$u_{obs} < \chi_{n-1, 1-\alpha}^2$	$P(\chi_{n-1}^2 \leq u_{obs})$
$H_1 : \sigma^2 \neq \sigma_0^2$	$u_{obs} < \chi_{n-1, 1-\alpha/2}^2$ ou $u_{obs} > \chi_{n-1, \alpha/2}^2$	2× la plus petite queue

Comme pour l'IC, ce test est **très sensible à la normalité**.

Exemple 10 : uniformité de comprimés

Un fabricant exige que l'écart-type de la quantité de principe actif par comprimé ne dépasse pas **2 mg** ($\sigma_0^2 = 4 \text{ mg}^2$). Sur $n = 20$ comprimés d'un nouveau lot, $s = 2,6 \text{ mg}$. La variabilité est-elle trop grande ($\alpha = 5\%$)?

1. $H_0 : \sigma^2 = 4$ contre $H_1 : \sigma^2 > 4$.
2. Rejeter H_0 si $u_{obs} > \chi_{19;0,05}^2 = 30,144$.
3. $u_{obs} = \frac{19 \times 2,6^2}{4} = \frac{19 \times 6,76}{4} = 32,11$.
4. $32,11 > 30,144$: on **rejette** H_0 (valeur-p = $P(\chi_{19}^2 \geq 32,11) = 0,030$).
5. Au seuil de 5%, l'écart-type du lot dépasse significativement 2 mg : le lot ne respecte pas l'exigence d'uniformité.

Dans R : `u <- 19 * 2.6^2 / 4 ; 1 - pchisq(u, df = 19).`

Bien interpréter un test : IC et importance pratique

IC et test bilatéral : deux faces d'une même médaille

Équivalence

On rejette $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au seuil $\alpha \iff$ l'IC de niveau $1 - \alpha$ pour θ **ne contient pas** θ_0 .

Pourquoi ? Pour $\mu : |t_{obs}| \leq t_{n-1, \alpha/2} \iff |\bar{x} - \mu_0| \leq t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \iff \mu_0 \in \bar{x} \pm ME$.

Exemples :

- Pression : l'IC à 95% $[128,20 ; 139,80]$ contient 130 $\Rightarrow$ on ne rejette pas $H_0 : \mu = 130$ (valeur-p = 0,14).
- Mercure : l'IC à 95% $[0,370 ; 0,470]$ exclut 0,5 $\Rightarrow$ on rejette $H_0 : \mu = 0,5$ en bilatéral à 5%.

L'IC donne **plus d'information** : il indique toutes les valeurs de θ_0 qu'on ne rejetterait pas. Cette équivalence vaut aussi pour σ^2 (et, approximativement, pour p).

QUIZ!

Dans un lac, l'IC à 99% pour la concentration moyenne de chlorophylle *a* est $[2,26 ; 3,41]$ $\mu\text{g/L}$. Tests bilatéraux : vrai, faux ou information insuffisante ?

- Au seuil de 1%, on rejette $H_0 : \mu = 3$.
- Au seuil de 1%, on rejette $H_0 : \mu = 4$.
- Au seuil de 5%, on rejette $H_0 : \mu = 4$.
- Au seuil de 5%, on rejette $H_0 : \mu = 2,5$.

« Statistiquement significatif » $\neq$ « important »

Une étude sur $n = 5000$ adultes mesure l'effet d'une diète sur la pression systolique : baisse moyenne de $\bar{d} = 0,8$ mm Hg, $s = 12$ mm Hg.

$$H_0 : \mu_D = 0 : t_{obs} = \frac{0,8}{12/\sqrt{5000}} = 4,71, \text{ valeur-}p = 2,5 \times 10^{-6}. \text{ Très significatif!}$$

Mais l'IC à 95% pour la baisse moyenne est $[0,47 ; 1,13]$ mm Hg : un effet **cliniquement négligeable**.

À retenir

- Avec un très grand n , presque toute différence devient significative.
- Avec un petit n , un effet important peut ne pas être significatif (faible puissance).
- L'importance pratique se juge avec l'**estimation** et son **IC**, selon l'expertise du domaine.

QUIZ!

Vrai ou faux?

- Une valeur- p de 0,0001 indique que l'effet du traitement est très grand.
- Deux études testent le même traitement. L'étude A ($n = 20$) donne $p = 0,20$ et l'étude B ($n = 2000$) donne $p = 0,001$. L'étude B a nécessairement trouvé un effet plus grand.
- Pour juger de l'importance d'un effet, on doit regarder l'estimation et son intervalle de confiance.

Synthèse

Tableau récapitulatif : tests à un échantillon

Hypothèse nulle	Statistique	Loi sous H_0	R
$\mu = \mu_0$ (normale, σ connue)	$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	$\mathcal{N}(0, 1)$	<code>pnorm()</code>
$\mu = \mu_0$ (normale, σ inconnue)	$t_{obs} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	t_{n-1}	<code>t.test(x, mu =)</code>
$\mu = \mu_0$ (n grand)	$z_{obs} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	$\approx \mathcal{N}(0, 1)$	<code>t.test(x, mu =)</code>
$p = p_0$ ($np_0, n(1 - p_0) \geq 5$)	$z_{obs} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$\approx \mathcal{N}(0, 1)$	<code>prop.test()</code>
$\sigma^2 = \sigma_0^2$ (normale)	$u_{obs} = \frac{(n-1)s^2}{\sigma_0^2}$	χ_{n-1}^2	<code>pchisq()</code>

Dans tous les cas : **rejeter H_0 si valeur- $p \leq \alpha$** . Au **chapitre 4**, on applique la même démarche pour **comparer deux populations**.

Réponses aux exemples

- **Quiz (H_0/H_1)** : $H_0 : \mu = 3,5$ vs $H_1 : \mu > 3,5$; $H_0 : p = 0,20$ vs $H_1 : p > 0,20$; $H_0 : \mu = 72$ vs $H_1 : \mu \neq 72$. Avec $\bar{x} = 3,62$ seulement, on ne peut rien conclure : il faut connaître la variabilité (erreur-type) pour juger si l'écart de 0,12 dépasse le hasard.
- **Exemple 7** : $z_{obs} = 2,2 > 1,645$, rejet de H_0 ; valeur-p = 0,0139; puissance à $\mu_1 = 26$: 0,639; $n = 25$ pour 80%.
- **Quiz (erreurs)** : type I : conclure que la diète augmente la masse alors qu'elle ne le fait pas; type II : ne pas détecter une augmentation réelle. Pour une maladie grave, le faux négatif (type II) est le plus grave : on privilégie la puissance (sensibilité). Faux : $1 - \alpha$ n'est pas $P(H_0 \text{ vraie})$.
- **Quiz (valeur-p, plantules)** : V; F; V; F (avec $\mu_0 = 4$, la valeur-p est plus grande que 0,09); info insuffisante; F (erreur-type plus petite, donc valeur-p plus petite).
- **Quiz (unilatéral)** : V; V; F ($0,08 > 0,05$).
- **Quiz (puissance)** : F; V; F (puissance très faible); V.

Réponses aux exemples (suite)

- **Exemple 9 (nitrates)** : $H_0 : \mu = 10$ vs $H_1 : \mu > 10$; rejet si $t_{obs} > t_{11;0,05} = 1,796$; $t_{obs} = (11,25 - 10)/(1,551/\sqrt{12}) = 2,79$: rejet de H_0 . Valeur-p entre 0,005 et 0,01 ($t_{11;0,01} = 2,718$, $t_{11;0,005} = 3,106$); exacte : 0,0088 avec `t.test(x, mu = 10, alternative = "greater")`. Unilatéral : seul un dépassement de la norme pose problème.
- **Exemple 4 (veaux)** : $z_{obs} = 1,49$, valeur-p = 0,068 (exacte : 0,100); non-rejet.
- **Quiz (sondage)** : $H_0 : p = 0,25$ vs $H_1 : p \neq 0,25$; $z_{obs} = (0,2101 - 0,25)/\sqrt{0,25 \times 0,75/1028} = -2,95$; valeur-p = 0,003 : rejet. L'IC $[0,185 ; 0,235]$ exclut 0,25 : même conclusion.
- **Exemple 10 (comprimés)** : $u_{obs} = 32,11 > 30,144$; valeur-p = 0,030 : rejet.
- **Quiz (IC et test)** : F; V; V (l'IC à 95% est inclus dans celui à 99%); info insuffisante (l'IC à 95% est plus étroit, mais on ne sait pas de combien).
- **Quiz (importance pratique)** : F; F (la valeur-p dépend aussi de n); V.