

Chapitre 4 – Comparaison de deux populations

STT-1920 Méthodes statistiques

Département de mathématiques et de statistique



UNIVERSITÉ
LAVAL

Introduction et objectifs

Où en sommes-nous ?

Au chapitre 3c, on a testé si **un** paramètre prend une valeur précise :

$H_0 : \mu = 120 \text{ mmHg}$, $H_0 : p = 0,10$, etc.

Au chapitre 4, on dispose de **deux échantillons** et on veut savoir si le paramètre prend la **même valeur dans deux populations** :

- La concentration moyenne de zinc est-elle la même dans deux rivières ?
- Un antihypertenseur diminue-t-il la pression artérielle des patients ?
- Le taux de guérison d'un nouveau traitement dépasse-t-il celui du traitement standard ?
- La variabilité du poids des fruits est-elle la même dans deux vergers ?

Rappel du chapitre 3c : les étapes d'un test

1) H_0 et H_1 ; 2) seuil α et loi de la statistique sous H_0 ; 3) statistique observée; 4) décision (région critique ou valeur-p); 5) **conclusion en mots**.

Objectifs du chapitre

Dans de nombreuses applications réelles (agronomie, biologie, santé, écologie), on souhaite **comparer deux traitements, deux groupes ou deux conditions expérimentales**.

Objectifs principaux

1. **Estimer** la différence ou le rapport de deux paramètres associés à deux populations ($\mu_1 - \mu_2$, σ_1^2/σ_2^2 , $p_1 - p_2$).
2. **Calculer** l'erreur-type associée et construire un **intervalle de confiance**.
3. **Tester** une hypothèse sur ces deux paramètres (ex. : $H_0 : \mu_1 = \mu_2$ ou $H_0 : \sigma_1^2 = \sigma_2^2$) à l'aide d'un **test d'hypothèses**.
4. **Choisir** le bon test selon le plan d'échantillonnage et **lire** une sortie R.

Les 5 situations étudiées

Paramètres	Hypothèse	Conditions expérimentales
Deux variances	$H_0 : \sigma_1^2 = \sigma_2^2$	Échantillons indépendants, populations normales
Deux moyennes	$H_0 : \mu_1 = \mu_2$	1. Indépendants, normales, $\sigma_1^2 = \sigma_2^2$ (variance combinée) 2. Indépendants, normales, $\sigma_1^2 \neq \sigma_2^2$ (Welch-Satterthwaite) 3. Échantillons appariés (mesures dépendantes deux à deux)
Deux proportions	$H_0 : p_1 = p_2$	Indépendants, n_1, n_2 grands (approximation normale)

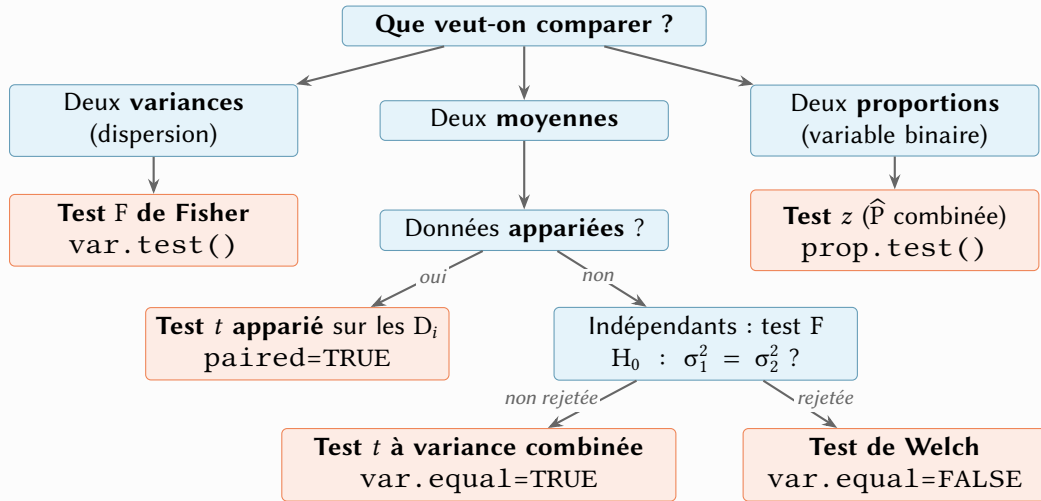
Les 5 situations étudiées

Paramètres	Hypothèse	Conditions expérimentales
Deux variances	$H_0 : \sigma_1^2 = \sigma_2^2$	Échantillons indépendants, populations normales
Deux moyennes	$H_0 : \mu_1 = \mu_2$	1. Indépendants, normales, $\sigma_1^2 = \sigma_2^2$ (variance combinée) 2. Indépendants, normales, $\sigma_1^2 \neq \sigma_2^2$ (Welch-Satterthwaite) 3. Échantillons appariés (mesures dépendantes deux à deux)
Deux proportions	$H_0 : p_1 = p_2$	Indépendants, n_1, n_2 grands (approximation normale)

Ordre d'analyse logique pour deux moyennes

Lorsque les échantillons sont indépendants, on commence **toujours** par tester l'égalité des variances (σ_1^2 vs σ_2^2) pour choisir le test de comparaison des moyennes approprié!

Feuille de route : quel test utiliser ?



Exemple 1 : Le zinc dans les rivières

On cherche à diversifier l'alimentation en eau potable d'une municipalité.

- Deux rivières sont candidates, mais les biologistes soupçonnent que la **rivière 1** a une concentration moyenne en zinc supérieure à celle de la **rivière 2**.
- On suppose que les concentrations de zinc suivent des lois normales.

Mesures prélevées (en $\mu\text{g/l}$) à des sites choisis au hasard dans chaque rivière :

Site	1	2	3	4	5	6	7	$\bar{x}$	s
Rivière 1 ($n_1 = 7$)	6,51	5,43	4,32	5,73	3,90	4,11	5,00	5,00	0,954
Rivière 2 ($n_2 = 6$)	6,03	4,92	3,60	5,13	2,58	4,08	–	4,39	1,226

Exemple 1 : Le zinc dans les rivières

On cherche à diversifier l'alimentation en eau potable d'une municipalité.

- Deux rivières sont candidates, mais les biologistes soupçonnent que la **rivière 1** a une concentration moyenne en zinc supérieure à celle de la **rivière 2**.
- On suppose que les concentrations de zinc suivent des lois normales.

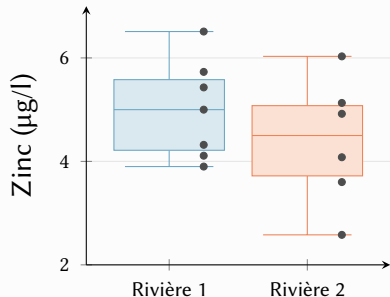
Mesures prélevées (en $\mu\text{g/l}$) à des sites choisis au hasard dans chaque rivière :

Site	1	2	3	4	5	6	7	$\bar{x}$	s
Rivière 1 ($n_1 = 7$)	6,51	5,43	4,32	5,73	3,90	4,11	5,00	5,00	0,954
Rivière 2 ($n_2 = 6$)	6,03	4,92	3,60	5,13	2,58	4,08	–	4,39	1,226

Question 1 : au seuil $\alpha = 10\%$, les variances sont-elles égales ?

Question 2 : au seuil $\alpha = 5\%$, la rivière 1 contient-elle plus de zinc en moyenne ?

Exemple 1 : Toujours regarder les données d'abord



- Les **diagrammes en boîtes juxtaposés** permettent de comparer d'un coup d'œil le **centre** et la **dispersion**.
- Ici : médianes proches, boîtes qui se chevauchent beaucoup; dispersion un peu plus grande pour la rivière 2.
- Avec $n_1 = 7$ et $n_2 = 6$, difficile de conclure « à l'œil » : il faut un **test**.

QUIZ!

Dans l'exemple du zinc :

- Les deux échantillons sont-ils indépendants ou appariés?
- Quelles hypothèses pour la question 2? Test unilatéral ou bilatéral?

QUIZ!

Dans l'exemple du zinc :

- Les deux échantillons sont-ils indépendants ou appariés?
- **Indépendants** : sites différents dans deux rivières différentes, aucune correspondance naturelle entre le site 1 de la rivière 1 et le site 1 de la rivière 2 (d'ailleurs $n_1 \neq n_2$).
- Quelles hypothèses pour la question 2? Test unilatéral ou bilatéral?
- $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 > \mu_2$: **unilatéral à droite** (les biologistes soupçonnent une concentration *supérieure* dans la rivière 1).

Comparaison de deux variances (Test de Fisher)

Situation et distribution d'échantillonnage

Soient deux échantillons aléatoires indépendants :

$$X_{11}, \dots, X_{1n_1} \stackrel{iid}{\sim} \mathcal{N}(\mu_1, \sigma_1^2) \quad \text{et} \quad X_{21}, \dots, X_{2n_2} \stackrel{iid}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$$

- On sait que $\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi_{n_1-1}^2$ et $\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$.
- Le rapport de deux v.a. khi-deux indépendantes divisées par leurs degrés de liberté respectifs suit une **loi de Fisher-Snedecor**.

Situation et distribution d'échantillonnage

Soient deux échantillons aléatoires indépendants :

$$X_{11}, \dots, X_{1n_1} \stackrel{iid}{\sim} \mathcal{N}(\mu_1, \sigma_1^2) \quad \text{et} \quad X_{21}, \dots, X_{2n_2} \stackrel{iid}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$$

- On sait que $\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi_{n_1-1}^2$ et $\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$.
- Le rapport de deux v.a. khi-deux indépendantes divisées par leurs degrés de liberté respectifs suit une **loi de Fisher-Snedecor**.

Statistique fondamentale

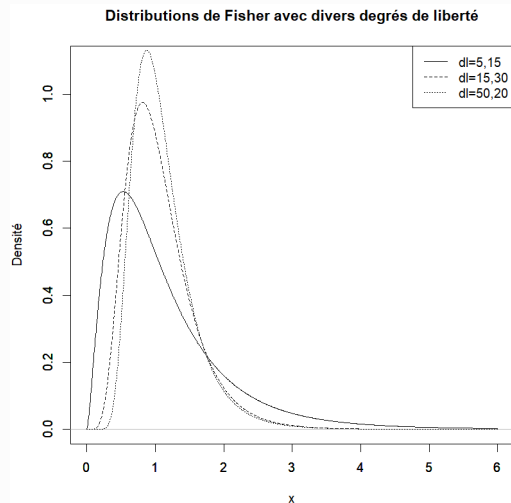
$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim \mathcal{F}(n_1 - 1, n_2 - 1)$$

Propriétés de la loi de Fisher $\mathcal{F}(v_1, v_2)$

- Variable toujours **strictement positive** ($F > 0$).
- Distribution **asymétrique à droite**.
- Dépend de deux paramètres : v_1 degrés de liberté au numérateur, v_2 au dénominateur (**l'ordre compte!**).
- Propriété d'inversion des quantiles :

$$F_{1-\alpha, v_1, v_2} = \frac{1}{F_{\alpha, v_2, v_1}}$$

- Dans R : `qf(0.95, 6, 5)`
donne $F_{0,05;6,5} = 4,95$.



Intervalle de confiance pour σ_1^2/σ_2^2

En partant de $P\left(F_{1-\alpha/2} \leq \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq F_{\alpha/2}\right) = 1 - \alpha$, on isole le rapport $\frac{\sigma_1^2}{\sigma_2^2}$:

Intervalle de confiance à $1 - \alpha$ pour σ_1^2/σ_2^2

$$\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha/2, n_1-1, n_2-1}}, \quad \frac{S_1^2}{S_2^2} F_{\alpha/2, n_2-1, n_1-1} \right]$$

Intervalle de confiance pour σ_1^2/σ_2^2

En partant de $P\left(F_{1-\alpha/2} \leq \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq F_{\alpha/2}\right) = 1 - \alpha$, on isole le rapport $\frac{\sigma_1^2}{\sigma_2^2}$:

Intervalle de confiance à $1 - \alpha$ pour σ_1^2/σ_2^2

$$\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{\alpha/2, n_1-1, n_2-1}}, \quad \frac{S_1^2}{S_2^2} F_{\alpha/2, n_2-1, n_1-1} \right]$$

- Attention à l'**inversion des degrés de liberté** dans la borne supérieure!
- Si l'intervalle contient la valeur **1**, on ne peut pas rejeter l'hypothèse d'égalité des variances ($\sigma_1^2 = \sigma_2^2$).

Test de Fisher pour l'égalité des variances

- **Hypothèses** : $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$ (bilatéral)
- **Statistique de test sous H_0** : $F_{obs} = s_1^2 / s_2^2$, à comparer à la loi $\mathcal{F}(n_1 - 1, n_2 - 1)$
- **Règle de décision au seuil α** :

$$\text{Rejeter } H_0 \iff F_{obs} < \frac{1}{F_{\alpha/2, n_2-1, n_1-1}} \quad \text{ou} \quad F_{obs} > F_{\alpha/2, n_1-1, n_2-1}$$

- **Unilatéral** ($H_1 : \sigma_1^2 > \sigma_2^2$) : rejeter H_0 si $F_{obs} > F_{\alpha, n_1-1, n_2-1}$.
- **Valeur-p** (rappel 3c) : rejeter H_0 si valeur-p $< \alpha$.

Test de Fisher pour l'égalité des variances

- **Hypothèses** : $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$ (bilatéral)
- **Statistique de test sous H_0** : $F_{obs} = s_1^2/s_2^2$, à comparer à la loi $\mathcal{F}(n_1 - 1, n_2 - 1)$
- **Règle de décision au seuil α** :

$$\text{Rejeter } H_0 \iff F_{obs} < \frac{1}{F_{\alpha/2, n_2-1, n_1-1}} \quad \text{ou} \quad F_{obs} > F_{\alpha/2, n_1-1, n_2-1}$$

- **Unilatéral** ($H_1 : \sigma_1^2 > \sigma_2^2$) : rejeter H_0 si $F_{obs} > F_{\alpha, n_1-1, n_2-1}$.
- **Valeur-p** (rappel 3c) : rejeter H_0 si valeur-p $< \alpha$.

Astuce pratique (convention numérateur plus grand)

En plaçant la plus grande variance au numérateur ($s_{\max}^2/s_{\min}^2$), on a $F_{obs} > 1$ et, en bilatéral, on compare simplement à $F_{\alpha/2, v_{\max}, v_{\min}}$.

Exemple 1 (suite) : Test de Fisher

Données : $n_1 = 7$, $s_1 = 0,954 \implies s_1^2 = 0,910$; $n_2 = 6$, $s_2 = 1,226 \implies s_2^2 = 1,503$.

Seuil : $\alpha = 0,10 \implies \alpha/2 = 0,05$. Degrés de liberté : $v_1 = 6$, $v_2 = 5$.

1. **Hypothèses** : $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$.

2. **Statistique observée** : $F_{obs} = \frac{s_1^2}{s_2^2} = \frac{0,910}{1,503} = 0,606$

Exemple 1 (suite) : Test de Fisher

Données : $n_1 = 7$, $s_1 = 0,954 \implies s_1^2 = 0,910$; $n_2 = 6$, $s_2 = 1,226 \implies s_2^2 = 1,503$.

Seuil : $\alpha = 0,10 \implies \alpha/2 = 0,05$. Degrés de liberté : $v_1 = 6$, $v_2 = 5$.

1. **Hypothèses** : $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$.

2. **Statistique observée** : $F_{obs} = \frac{s_1^2}{s_2^2} = \frac{0,910}{1,503} = 0,606$

3. **Valeurs critiques** : table : $F_{0,05;6,5} = 4,95$ et $F_{0,05;5,6} = 4,39$, d'où

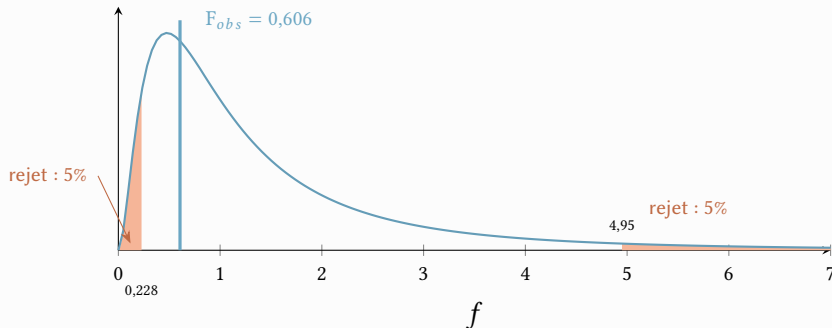
$$\text{Borne inférieure} = \frac{1}{4,39} = 0,228, \quad \text{Borne supérieure} = 4,95$$

4. **Décision** : Puisque $0,228 \leq 0,606 \leq 4,95$, on **ne rejette pas H_0** (valeur- $p = 0,557 > 0,10$).

5. **Conclusion** : les données ne montrent pas de différence de variabilité entre les deux rivières; on peut utiliser le test t avec **variance combinée**.

Exemple 1 (suite) : La décision en image

Loi $F(6, 5)$ sous H_0 , $\alpha = 0,10$



- Les zones orange forment la **région critique** : chacune a une aire de $\alpha/2 = 0,05$.
- F_{obs} tombe dans la zone de non-rejet : **on ne rejette pas H_0** .

Exemple 1 (suite) : Sortie R de `var.test()`

Les données sont dans un tableau `rivieres` (colonnes `Zinc` et `Riviere`, facteur à 2 niveaux).

```
> var.test(Zinc ~ Riviere, alternative='two.sided', conf.level=.9, data=rivieres)

      F test to compare two variances

data:  Zinc by Riviere
F = 0.60577, num df = 6, denom df = 5, p-value = 0.557
alternative hypothesis: true ratio of variances is not equal to 1
90 percent confidence interval:
 0.1223707 2.6577408
sample estimates:
ratio of variances
 0.6057703
```

- $F = 0.60577$, $\text{num df} = 6$, $\text{denom df} = 5$: c'est notre F_{obs} et ses degrés de liberté.
- $p\text{-value} = 0.557 > \alpha = 0,10 \Rightarrow$ on **ne rejette pas** H_0 .
- IC à 90 % pour σ_1^2/σ_2^2 : $[0,12; 2,66]$ **contient 1** : même conclusion.

QUIZ!

- Vrai ou faux? « Puisqu'on ne rejette pas H_0 , on a *démontré* que les deux rivières ont la même variance. »
- Un collègue place s_2^2 au numérateur : $F_{obs} = 1,503/0,910 = 1,652$. Sa conclusion change-t-elle?

QUIZ!

- Vrai ou faux? « Puisqu'on ne rejette pas H_0 , on a *démontré* que les deux rivières ont la même variance. »
- **Faux.** Ne pas rejeter H_0 n'est pas une preuve de H_0 (rappel du chapitre 3c). Avec $n_1 = 7$ et $n_2 = 6$, le test F a une **faible puissance** : l'IC à 90 % va de 0,12 à 2,66!
- Un collègue place s_2^2 au numérateur : $F_{obs} = 1,503/0,910 = 1,652$. Sa conclusion change-t-elle?
- **Non.** Il compare à $F_{0,05;5,6} = 4,39$: $1,652 < 4,39$, on ne rejette pas H_0 . Même valeur-p (0,557).

Attention !

Le test F est très sensible à la non-normalité

Contrairement aux tests t (assez robustes), le test de Fisher peut donner des conclusions trompeuses si les populations ne sont **pas normales**, même avec de grands échantillons.

- Vérifier la normalité de **chaque** échantillon (histogramme, diagramme quantile-quantile, chapitre 2).
- Toujours accompagner le test d'un graphique (diagrammes en boîtes juxtaposés).
- Une grande différence de dispersion peut être un **résultat biologique** en soi (ex. : plus grande variabilité de la réponse à un traitement).

Comparaison de deux moyennes (Échantillons indépendants)

Cas 1 : Variances égales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

Puisque les deux populations ont la même variance inconnue σ^2 , on combine l'information des deux échantillons pour obtenir la **variance combinée** (pooled variance) :

Variance échantillonnale combinée S_p^2

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Cas 1 : Variances égales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

Puisque les deux populations ont la même variance inconnue σ^2 , on combine l'information des deux échantillons pour obtenir la **variance combinée** (pooled variance) :

Variance échantillonnale combinée S_p^2

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- S_p^2 est une moyenne pondérée de S_1^2 et S_2^2 par leurs degrés de liberté.
- Elle possède $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ degrés de liberté.
- L'erreur-type estimée de la différence des moyennes est :

$$\widehat{\text{ET}}(\bar{X}_1 - \bar{X}_2) = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Intervalle de confiance pour $\mu_1 - \mu_2$ (variances égales)

Sous l'hypothèse de normalité et de variances égales :

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Intervalle de confiance de niveau $1 - \alpha$

$$\left[(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2, \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

La marge d'erreur dépend de la variance combinée S_p^2 et de la somme des inverses des tailles d'échantillons. Si l'intervalle **contient 0**, on ne peut pas conclure que les moyennes diffèrent (test bilatéral au seuil α).

Test d'hypothèses pour $\mu_1 - \mu_2$ (variances égales)

Pour tester $H_0 : \mu_1 = \mu_2$ (c'est-à-dire $\mu_1 - \mu_2 = 0$) :

Statistique observée sous H_0

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Règles de décision :

- Bilatéral ($H_1 : \mu_1 \neq \mu_2$) : rejeter H_0 si $|t_{obs}| > t_{n_1+n_2-2, \alpha/2}$.
- Unilatéral à droite ($H_1 : \mu_1 > \mu_2$) : rejeter H_0 si $t_{obs} > t_{n_1+n_2-2, \alpha}$.
- Unilatéral à gauche ($H_1 : \mu_1 < \mu_2$) : rejeter H_0 si $t_{obs} < -t_{n_1+n_2-2, \alpha}$.

Ce test est souvent appelé **test t à deux échantillons** (*two-sample t -test*).

Exemple 1 (suite) : Comparaison des moyennes

$H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 > \mu_2$ (question 2, $\alpha = 5\%$).

Données : $n_1 = 7$, $\bar{x}_1 = 5,00$, $s_1^2 = 0,910$; $n_2 = 6$, $\bar{x}_2 = 4,39$, $s_2^2 = 1,503$.

1. **Variance combinée** (le test F n'a pas rejeté l'égalité) :

$$s_p^2 = \frac{6(0,910) + 5(1,503)}{7 + 6 - 2} = \frac{5,460 + 7,515}{11} = 1,180 \implies s_p = 1,086$$

2. **Erreur-type** : $\widehat{ET} = 1,086 \sqrt{\frac{1}{7} + \frac{1}{6}} = 1,086 \times 0,5563 = 0,604$.

3. **Statistique observée** : $t_{obs} = \frac{5,00 - 4,39}{0,604} = \frac{0,61}{0,604} = 1,01$

Exemple 1 (suite) : Comparaison des moyennes

$H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 > \mu_2$ (question 2, $\alpha = 5\%$).

Données : $n_1 = 7$, $\bar{x}_1 = 5,00$, $s_1^2 = 0,910$; $n_2 = 6$, $\bar{x}_2 = 4,39$, $s_2^2 = 1,503$.

1. **Variance combinée** (le test F n'a pas rejeté l'égalité) :

$$s_p^2 = \frac{6(0,910) + 5(1,503)}{7 + 6 - 2} = \frac{5,460 + 7,515}{11} = 1,180 \implies s_p = 1,086$$

2. **Erreur-type** : $\widehat{ET} = 1,086 \sqrt{\frac{1}{7} + \frac{1}{6}} = 1,086 \times 0,5563 = 0,604$.

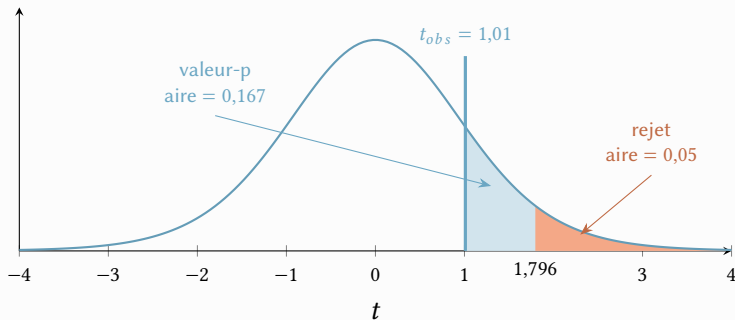
3. **Statistique observée** : $t_{obs} = \frac{5,00 - 4,39}{0,604} = \frac{0,61}{0,604} = 1,01$

4. **Décision** : avec $v = 7 + 6 - 2 = 11$, $t_{11;0,05} = 1,796$. Puisque $t_{obs} = 1,01 < 1,796$, on **ne rejette pas H_0** (valeur-p = 0,167).

5. **Conclusion** : les données ne permettent pas de conclure que la concentration moyenne de zinc est plus élevée dans la rivière 1.

Exemple 1 (suite) : Région critique et valeur-p

Loi t_{11} sous H_0 (test unilatéral à droite, $\alpha = 0,05$)



- **Région critique** : $t > 1,796$ (zone orange).
- **Valeur-p** : $P(T_{11} > 1,01) = 0,167$ (zone bleue, qui *contient* la zone orange) $> 0,05$.

Exemple 1 (suite) : Sortie R de `t.test()`

```
> t.test(Zinc~Riviere, alternative='greater', conf.level=.95, var.equal=TRUE,
+ data=rivieres)

      Two Sample t-test

data:  Zinc by Riviere
t = 1.0095, df = 11, p-value = 0.1672
alternative hypothesis: true difference in means between group 1 and group 2 is greater than 0
95 percent confidence interval:
 -0.475208      Inf
sample estimates:
mean in group 1 mean in group 2
          5.00          4.39
```

- `alternative='greater'` : $H_1 : \mu_1 > \mu_2$; `var.equal=TRUE` : variance combinée.
- `t = 1.0095`, `df = 11` : notre t_{obs} (à l'arrondi près) et $v = n_1 + n_2 - 2$.
- `p-value = 0.1672 > 0,05` : on **ne rejette pas** H_0 .
- L'IC unilatéral $[-0,475; \infty[$ **contient 0** : même conclusion.

Exemple 1 (suite) : Test bilatéral et intervalle de confiance

À vous : construire l'IC à 95 % pour $\mu_1 - \mu_2$ (utiliser $t_{11;0,025} = 2,201$). Vérifier ensuite avec R.

Exemple 1 (suite) : Test bilatéral et intervalle de confiance

À vous : construire l'IC à 95 % pour $\mu_1 - \mu_2$ (utiliser $t_{11;0,025} = 2,201$). Vérifier ensuite avec R.

```
> t.test(Zinc~Riviere, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=rivieres)

Two Sample t-test

data:  Zinc by Riviere
t = 1.0095, df = 11, p-value = 0.3344
alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
95 percent confidence interval:
 -0.7199999  1.9399999
sample estimates:
mean in group 1 mean in group 2
      5.00      4.39
```

- Test bilatéral : $p\text{-value} = 0.3344 = 2 \times 0,1672$ (la valeur-p bilatérale est le **double** de l'unilatérale).
- L'IC à 95 % pour $\mu_1 - \mu_2$ contient 0 : pas de différence significative au seuil de 5 %.

QUIZ!

Un conseiller municipal lit le rapport et conclut : « Les deux rivières ont la même concentration moyenne de zinc. On peut choisir l'une ou l'autre. »

- Êtes-vous d'accord ?
- Que faudrait-il faire pour trancher ?

QUIZ!

Un conseiller municipal lit le rapport et conclut : « Les deux rivières ont la même concentration moyenne de zinc. On peut choisir l'une ou l'autre. »

- Êtes-vous d'accord ?
- **Non.** « Ne pas rejeter H_0 » $\neq$ « H_0 est vraie ». L'IC à 95 % pour $\mu_1 - \mu_2$ va d'environ $-0,72$ à $1,94$ $\mu\text{g/l}$: une différence de près de 2 $\mu\text{g/l}$ reste plausible.
- Que faudrait-il faire pour trancher ?
- Augmenter les tailles d'échantillons pour gagner en **puissance** (rappel 3c) et réduire la largeur de l'IC.

Cas 2 : Variances inégales ($\sigma_1^2 \neq \sigma_2^2$)

Si le test de Fisher a rejeté l'égalité des variances, on ne peut plus utiliser S_p^2 . On utilise alors l'approximation de **Welch-Satterthwaite** (**test de Welch**).

Statistique de Welch

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$$

Cas 2 : Variances inégales ($\sigma_1^2 \neq \sigma_2^2$)

Si le test de Fisher a rejeté l'égalité des variances, on ne peut plus utiliser S_p^2 . On utilise alors l'approximation de **Welch-Satterthwaite** (**test de Welch**).

Statistique de Welch

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$$

Les degrés de liberté ajustés v sont donnés par la formule de Welch :

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$$

En général v n'est pas entier : avec une table, on **arrondit vers le bas** (plus prudent). Dans R : `t.test(..., var.equal = FALSE)` (par défaut).

Cas 2 : Règles de décision et intervalle de confiance

Sous $H_0 : \mu_1 = \mu_2$, on calcule $t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$ et :

- Bilatéral : rejeter H_0 si $|t_{obs}| > t_{v,\alpha/2}$.
- Unilatéral à droite / à gauche : rejeter si $t_{obs} > t_{v,\alpha}$ / $t_{obs} < -t_{v,\alpha}$.

IC approximatif de niveau $1 - \alpha$ pour $\mu_1 - \mu_2$

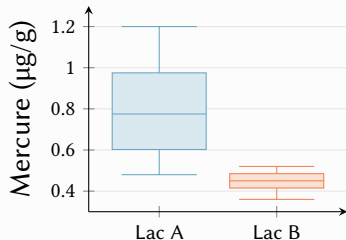
$$\left[(\bar{x}_1 - \bar{x}_2) \pm t_{v,\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right]$$

Remarque : le test de Welch reste valide même si les variances sont égales (il est alors à peine moins puissant). Sur l'exemple du zinc, il donne $t_{obs} = 0,99$, $v = 9,42$, valeur-p = 0,174 : même conclusion.

Exemple 2 : Mercure dans les dorés de deux lacs

Des biologistes comparent la concentration de mercure (en $\mu\text{g/g}$) dans la chair de dorés jaunes pêchés dans un **réservoir hydroélectrique** récent (lac A) et dans un **lac naturel** voisin (lac B). La concentration moyenne diffère-t-elle ($\alpha = 5\%$) ?

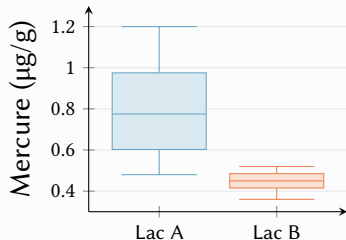
Mesures										$\bar{x}$	s	
Lac A ($n_1 = 8$)	0,62	0,95	0,48	1,20	0,81	0,55	1,05	0,74		0,800	0,2524	
Lac B ($n_2 = 10$)	0,41	0,46	0,38	0,52	0,44	0,49	0,36	0,47	0,43	0,50	0,446	0,0521



Exemple 2 : Mercure dans les dorés de deux lacs

Des biologistes comparent la concentration de mercure (en $\mu\text{g/g}$) dans la chair de dorés jaunes pêchés dans un **réservoir hydroélectrique** récent (lac A) et dans un **lac naturel** voisin (lac B). La concentration moyenne diffère-t-elle ($\alpha = 5\%$) ?

	Mesures										$\bar{x}$	s
Lac A ($n_1 = 8$)	0,62	0,95	0,48	1,20	0,81	0,55	1,05	0,74			0,800	0,2524
Lac B ($n_2 = 10$)	0,41	0,46	0,38	0,52	0,44	0,49	0,36	0,47	0,43	0,50	0,446	0,0521



- Échantillons **indépendants** (poissons différents, lacs différents).
- Les boîtes suggèrent une dispersion **beaucoup plus grande** dans le lac A.
- Étape 1 : test F ($\alpha = 10\%$). Étape 2 : test sur les moyennes.

Exemple 2 (suite) : Étape 1, test de Fisher

$$s_1^2 = 0,2524^2 = 0,06371 \quad \text{et} \quad s_2^2 = 0,0521^2 = 0,002716.$$

1. $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$, $\alpha = 0,10$.

2. $F_{obs} = \frac{0,06371}{0,002716} = 23,46$ (plus grande variance au numérateur).

Exemple 2 (suite) : Étape 1, test de Fisher

$$s_1^2 = 0,2524^2 = 0,06371 \quad \text{et} \quad s_2^2 = 0,0521^2 = 0,002716.$$

1. $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$, $\alpha = 0,10$.

2. $F_{obs} = \frac{0,06371}{0,002716} = 23,46$ (plus grande variance au numérateur).

3. Valeur critique : $F_{0,05;7,9} = 3,29$.

4. Puisque $23,46 > 3,29$, on **rejette H_0** (valeur-p $< 0,0001$).

5. **Conclusion** : la variabilité de la concentration de mercure est plus grande dans le réservoir. On **ne peut pas** combiner les variances $\Rightarrow$ **test de Welch**.

Exemple 2 (suite) : Étape 2, test de Welch

$H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$, $\alpha = 0,05$.

1. **Erreur-type** : $\frac{s_1^2}{n_1} = \frac{0,06371}{8} = 0,007964$, $\frac{s_2^2}{n_2} = \frac{0,002716}{10} = 0,000272$

$$\widehat{ET} = \sqrt{0,007964 + 0,000272} = \mathbf{0,0908}$$

2. **Degrés de liberté** : $\nu = \frac{(0,008236)^2}{\frac{(0,007964)^2}{7} + \frac{(0,000272)^2}{9}} = 7,48 \rightsquigarrow 7$

3. **Statistique** : $t_{obs} = \frac{0,800 - 0,446}{0,0908} = \mathbf{3,90}$

Exemple 2 (suite) : Étape 2, test de Welch

$H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$, $\alpha = 0,05$.

1. **Erreur-type** : $\frac{s_1^2}{n_1} = \frac{0,06371}{8} = 0,007964$, $\frac{s_2^2}{n_2} = \frac{0,002716}{10} = 0,000272$

$$\widehat{ET} = \sqrt{0,007964 + 0,000272} = 0,0908$$

2. **Degrés de liberté** : $v = \frac{(0,008236)^2}{\frac{(0,007964)^2}{7} + \frac{(0,000272)^2}{9}} = 7,48 \rightsquigarrow 7$

3. **Statistique** : $t_{obs} = \frac{0,800 - 0,446}{0,0908} = 3,90$

4. **Décision** : $|t_{obs}| = 3,90 > t_{7;0,025} = 2,365$: on **rejette H_0** (valeur-p = 0,005).

5. **Conclusion** : la concentration moyenne de mercure est significativement plus élevée dans le réservoir (lac A) que dans le lac naturel.

Exemple 2 (suite) : Sortie R du test de Welch

```
> A <- c(0.62, 0.95, 0.48, 1.20, 0.81, 0.55, 1.05, 0.74)
> B <- c(0.41, 0.46, 0.38, 0.52, 0.44, 0.49, 0.36, 0.47, 0.43, 0.50)
> t.test(A, B)           # var.equal = FALSE par défaut : Welch
```

Welch Two Sample t-test

data: A and B

t = 3.9008, df = 7.4787, p-value = 0.005177

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.1421602 0.5658398

sample estimates:

mean of x mean of y

0.800 0.446

- $df = 7.4787$: les degrés de liberté de Welch, non entiers.
- **À vous** : calculer l'IC à 95 % à la main avec $v = 7$ et le comparer à celui de R.

QUIZ!

- Dans l'exemple 2, on a obtenu $v = 7,48$. Entre quelles bornes les degrés de liberté de Welch se situent-ils toujours?
- Qu'aurait-on obtenu avec le test t à variance combinée (à tort)?

QUIZ!

- Dans l'exemple 2, on a obtenu $v = 7,48$. Entre quelles bornes les degrés de liberté de Welch se situent-ils toujours?
- $\min(n_1 - 1, n_2 - 1) \leq v \leq n_1 + n_2 - 2$, soit ici entre 7 et 16. Plus les variances s_i^2/n_i sont déséquilibrées, plus v est proche de la borne inférieure.
- Qu'aurait-on obtenu avec le test t à variance combinée (à tort)?
- $t_{obs} = 4,35$ avec 16 degrés de liberté (valeur- $p = 0,0005$) : un résultat **trop optimiste**, car l'erreur-type est mal estimée quand les variances diffèrent autant (surtout si $n_1 \neq n_2$).

Données appariées (Échantillons dépendants)

Deux scénarios d'échantillonnage

On veut comparer deux engrais (A et B) sur le rendement du maïs, avec 40 parcelles en tout.

Scénario 1 : échantillons indépendants

20 fermes reçoivent l'engrais A, 20 **autres** fermes reçoivent l'engrais B (une parcelle par ferme). On note 40 observations indépendantes.

Deux scénarios d'échantillonnage

On veut comparer deux engrais (A et B) sur le rendement du maïs, avec 40 parcelles en tout.

Scénario 1 : échantillons indépendants

20 fermes reçoivent l'engrais A, 20 **autres** fermes reçoivent l'engrais B (une parcelle par ferme). On note **40 observations indépendantes**.

Scénario 2 : échantillons appariés

20 fermes : chacune est divisée en deux parcelles voisines, qui reçoivent A et B (attribution **au hasard**). On note **20 paires** d'observations.

Quel scénario permet de détecter plus facilement une différence entre les engrais ?

Notion d'appariement : pourquoi appairer ?

- **Scénario 1** : les fermes diffèrent beaucoup (sol, climat, drainage). Une différence de rendement est-elle due aux **engrais** ou aux **fermes** ? La variance de $\bar{X}_1 - \bar{X}_2$ contient toute la **variabilité entre fermes** : elle risque d'être trop grande pour conclure.
- **Scénario 2** : chaque ferme sert de **son propre témoin**. En comparant A et B *dans* la même ferme, on élimine la variabilité entre fermes. Une différence significative sera due aux engrais, pourvu que l'attribution de A et B aux deux parcelles soit **aléatoire**.
- **Problème** : dans le scénario 2, X_{1j} et X_{2j} ne sont **pas indépendants** (même ferme). On ne peut pas utiliser les tests pour échantillons indépendants.

Notion de données appariées

Définition

Deux échantillons sont dits **appariés** s'il existe une correspondance naturelle deux à deux entre les observations des deux groupes.

Exemples typiques :

- **Mesures répétées sur le même sujet** : pression artérielle avant et après la prise d'un médicament (X_{1i} et X_{2i}).
- **Paires naturelles** : les deux yeux d'un patient (un traité, l'autre témoin); les deux moitiés d'une feuille; deux parcelles d'une même ferme.
- **Jumeaux** : l'un reçoit le traitement A, l'autre le placebo.

Notion de données appariées

Définition

Deux échantillons sont dits **appariés** s'il existe une correspondance naturelle deux à deux entre les observations des deux groupes.

Exemples typiques :

- **Mesures répétées sur le même sujet** : pression artérielle avant et après la prise d'un médicament (X_{1i} et X_{2i}).
- **Paires naturelles** : les deux yeux d'un patient (un traité, l'autre témoin); les deux moitiés d'une feuille; deux parcelles d'une même ferme.
- **Jumeaux** : l'un reçoit le traitement A, l'autre le placebo.

Attention : Ne jamais appliquer un test à deux échantillons indépendants sur des données appariées (violation majeure de l'indépendance)!

Méthode d'analyse : réduction à un échantillon

On calcule pour chaque paire i la **différence individuelle** $D_i = X_{1i} - X_{2i}$, $i = 1, \dots, n$. Les D_i , eux, sont **indépendants** d'une paire à l'autre.

Si les $D_i \sim \mathcal{N}(\mu_D, \sigma_D^2)$, le problème se ramène exactement à un **test t à un échantillon** (chapitre 3c) sur la moyenne des différences $\mu_D = \mu_1 - \mu_2$!

Test t apparié

- Moyenne des différences : $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$
- Écart-type des différences : $S_D = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2}$
- Statistique de test sous $H_0 : \mu_D = 0$: $t_{obs} = \frac{\bar{D}}{S_D / \sqrt{n}} \sim t_{n-1}$

Test apparié : règles de décision et IC

Pour tester $H_0 : \mu_1 = \mu_2 \iff \mu_D = 0$ au seuil α :

- $H_1 : \mu_D \neq 0$: rejeter H_0 si $|t_{obs}| > t_{n-1, \alpha/2}$
- $H_1 : \mu_D > 0$: rejeter H_0 si $t_{obs} > t_{n-1, \alpha}$
- $H_1 : \mu_D < 0$: rejeter H_0 si $t_{obs} < -t_{n-1, \alpha}$

Intervalle de confiance de niveau $1 - \alpha$ pour μ_D

$$\left[\bar{D} \pm t_{n-1, \alpha/2} \frac{S_D}{\sqrt{n}} \right]$$

Attention : n est le nombre de **paires**, pas le nombre total de mesures.

Exemple 3 : Un antihypertenseur

On mesure la pression artérielle systolique (mmHg) de 10 patients hypertendus **avant** et **après** 8 semaines de traitement. Le médicament diminue-t-il la pression moyenne ($\alpha = 5\%$)?

Patient	1	2	3	4	5	6	7	8	9	10	Moy.
Avant (x_{1i})	152	138	165	147	171	143	158	136	161	149	152,0
Après (x_{2i})	141	133	154	140	160	139	146	131	155	138	143,7

Exemple 3 : Un antihypertenseur

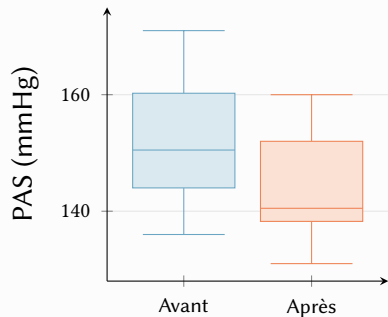
On mesure la pression artérielle systolique (mmHg) de 10 patients hypertendus **avant** et **après** 8 semaines de traitement. Le médicament diminue-t-il la pression moyenne ($\alpha = 5\%$) ?

Patient	1	2	3	4	5	6	7	8	9	10	Moy.
Avant (x_{1i})	152	138	165	147	171	143	158	136	161	149	152,0
Après (x_{2i})	141	133	154	140	160	139	146	131	155	138	143,7
$d_i = x_{1i} - x_{2i}$	11	5	11	7	11	4	12	5	6	11	8,3

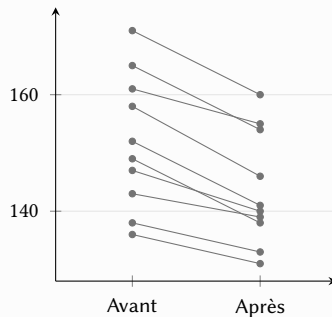
- Données **appariées** : les deux mesures proviennent du **même patient**.
- $H_0 : \mu_D = 0$ contre $H_1 : \mu_D > 0$ (la pression *baisse* : avant > après).
- Écarts-types : $s_{\text{avant}} = 11,6$, $s_{\text{après}} = 9,8$, mais $s_D = 3,16$ seulement !

Exemple 3 (suite) : L'appariement en image

Vue « indépendante »



Vue « appariée »



- À gauche, les boîtes se chevauchent : la **variabilité entre patients** masque l'effet.
- À droite, **les 10 segments descendent** : chaque patient baisse, de 4 à 12 mmHg.

Exemple 3 (suite) : Solution

$n = 10$ paires, $\bar{d} = 8,3$ mmHg, $s_D = 3,164$ mmHg.

1. **Erreur-type** : $s_D / \sqrt{n} = 3,164 / \sqrt{10} = 1,001$.

2. **Statistique observée** : $t_{obs} = \frac{\bar{d}}{s_D / \sqrt{n}} = \frac{8,3}{1,001} = 8,30$

Exemple 3 (suite) : Solution

$n = 10$ paires, $\bar{d} = 8,3$ mmHg, $s_D = 3,164$ mmHg.

1. **Erreur-type** : $s_D / \sqrt{n} = 3,164 / \sqrt{10} = 1,001$.

2. **Statistique observée** : $t_{obs} = \frac{\bar{d}}{s_D / \sqrt{n}} = \frac{8,3}{1,001} = 8,30$

3. **Décision** : $t_{9;0,05} = 1,833$. Puisque $8,30 > 1,833$, on **rejette H_0** (valeur- $p < 0,0001$).

4. **Conclusion** : le traitement diminue significativement la pression artérielle systolique moyenne des patients.

5. **À vous** : calculer l'IC à 95 % pour μ_D (utiliser $t_{9;0,025} = 2,262$) et l'interpréter.

Dans R : `t.test(avant, apres, paired = TRUE, alternative = "greater")` donne `t = 8.2954`, `df = 9`, `p-value = 8.274e-06`.

QUIZ!

Par erreur, vous analysez l'exemple 3 comme s'il s'agissait d'échantillons *indépendants* avec variances égales ($\alpha = 5\%$, $H_1 : \mu_1 > \mu_2$).

- Quelle est votre règle de rejet de H_0 ?
- On obtient $s_p = 10,72$ et $t_{obs} = 8,3 / (10,72 \sqrt{2/10}) = 1,73$. Conclusion ?

QUIZ!

Par erreur, vous analysez l'exemple 3 comme s'il s'agissait d'échantillons *indépendants* avec variances égales ($\alpha = 5\%$, $H_1 : \mu_1 > \mu_2$).

- Quelle est votre règle de rejet de H_0 ?
- $v = 10 + 10 - 2 = 18$: rejeter H_0 si $t_{obs} > t_{18;0,05} = 1,734$.
- On obtient $s_p = 10,72$ et $t_{obs} = 8,3/(10,72\sqrt{2/10}) = 1,73$. Conclusion ?
- $1,73 < 1,734$: on **ne rejetterait pas** H_0 (valeur-p = 0,050)! L'erreur-type (4,80 au lieu de 1,00) inclut la variabilité entre patients. **Ignorer l'appariement fait perdre beaucoup de puissance.**

Lecture d'une sortie R : zinc au fond et en surface

On mesure la concentration de zinc **au fond** et **en surface** de l'eau à 6 sites d'un lac (deux mesures **au même site**). L'eau du fond est-elle plus chargée en zinc ($\alpha = 1\%$)?

```
> with(Fond, (t.test(Fond, Surface, alternative='greater', conf.level=.99, paired=TRUE)))

Paired t-test

data:  Fond and Surface
t = 3.6902, df = 5, p-value = 0.007071
alternative hypothesis: true difference in means is greater than 0
99 percent confidence interval:
 0.08050277      Inf
sample estimates:
mean of the differences
      0.9133333
```

- `paired=TRUE` : test t sur les 6 différences $d_i = \text{fond}_i - \text{surface}_i$; $df = 5 = n - 1$.
- $\bar{d} = 0,913$; $t_{obs} = 3,69$; $p\text{-value} = 0.007 < 0,01 \Rightarrow$ on **rejette** H_0 .
- Au seuil de 1 %, la concentration moyenne est plus élevée au fond qu'en surface.

Quelques remarques importantes

- C'est le **contexte** (le plan d'expérience) qui détermine si les données sont appariées ou non, pas les nombres eux-mêmes.
- Si on analyse des données appariées **sans tenir compte de l'appariement**, on risque de diminuer grandement la significativité des données (valeur-p surestimée), comme dans le quiz précédent.
- L'appariement est avantageux quand les **unités expérimentales sont hétérogènes** (patients, fermes, sites) : on retire la variabilité entre unités.
- Pour des mesures « avant/après », prudence : sans groupe témoin, un autre facteur (saison, effet placebo) peut expliquer le changement.

QUIZ!

Appariés ou indépendants ?

1. On mesure la glycémie de 15 patients avec deux glucomètres différents (chaque patient est mesuré avec les deux appareils).
2. On compare la masse de 12 souris nourries au régime A à celle de 12 autres souris nourries au régime B.
3. On compte les stomates sur la face supérieure et la face inférieure de 20 feuilles.
4. On compare l'abondance d'une espèce d'insecte dans 8 champs biologiques et 8 champs conventionnels.

QUIZ!

Appariés ou indépendants ?

1. On mesure la glycémie de 15 patients avec deux glucomètres différents (chaque patient est mesuré avec les deux appareils).
⇒ **Appariés** (même patient).
2. On compare la masse de 12 souris nourries au régime A à celle de 12 autres souris nourries au régime B.
⇒ **Indépendants** (souris différentes, aucune correspondance).
3. On compte les stomates sur la face supérieure et la face inférieure de 20 feuilles.
⇒ **Appariés** (les deux faces d'une même feuille).
4. On compare l'abondance d'une espèce d'insecte dans 8 champs biologiques et 8 champs conventionnels.
⇒ **Indépendants**, sauf si les champs sont formés en paires voisines (même sol).

Comparaison de deux proportions

Comparaison de deux proportions

Soient deux échantillons indépendants de grandes tailles : $Y_1 \sim \text{Binomiale}(n_1, p_1)$ et $Y_2 \sim \text{Binomiale}(n_2, p_2)$.

Estimations ponctuelles : $\hat{p}_1 = Y_1/n_1$ et $\hat{p}_2 = Y_2/n_2$.

Test d'hypothèses d'égalité des proportions

On estime la proportion commune sous $H_0 : p_1 = p_2$ par la **proportion combinée** $\hat{P} = \frac{Y_1 + Y_2}{n_1 + n_2}$.

Statistique de test (normale centrée réduite) :

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{P}(1 - \hat{P}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim \mathcal{N}(0, 1)$$

Règles de décision, conditions et intervalle de confiance

- Bilatéral : rejeter H_0 si $|z_{obs}| > z_{\alpha/2}$; unilatéral : $z_{obs} > z_{\alpha}$ ou $z_{obs} < -z_{\alpha}$.
- **Conditions** (approximation normale) : $n_i \hat{P} \geq 5$ et $n_i(1 - \hat{P}) \geq 5$ pour $i = 1, 2$.

Intervalle de confiance de niveau $1 - \alpha$ pour $p_1 - p_2$

$$\left[(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \right]$$

Règles de décision, conditions et intervalle de confiance

- Bilatéral : rejeter H_0 si $|z_{obs}| > z_{\alpha/2}$; unilatéral : $z_{obs} > z_{\alpha}$ ou $z_{obs} < -z_{\alpha}$.
- **Conditions** (approximation normale) : $n_i \hat{P} \geq 5$ et $n_i(1 - \hat{P}) \geq 5$ pour $i = 1, 2$.

Intervalle de confiance de niveau $1 - \alpha$ pour $p_1 - p_2$

$$\left[(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \right]$$

Attention : l'erreur-type n'est pas la même!

- Pour le **test**, on se place sous H_0 ($p_1 = p_2$) : on utilise $\hat{P}$ combinée.
- Pour l'**IC**, on ne suppose pas $p_1 = p_2$: on utilise $\hat{p}_1$ et $\hat{p}_2$ séparément.

Exemple 4 : Taux de guérison de deux traitements

Un essai clinique randomisé compare un **nouvel antibiotique** au **traitement standard** contre une infection urinaire. Guérison après 14 jours :

	Patients (n_i)	Guéris (y_i)	$\hat{p}_i$
Nouvel antibiotique (1)	120	90	0,750
Traitement standard (2)	110	66	0,600

Au seuil $\alpha = 5\%$, les taux de guérison diffèrent-ils ?

Exemple 4 : Taux de guérison de deux traitements

Un essai clinique randomisé compare un **nouvel antibiotique** au **traitement standard** contre une infection urinaire. Guérison après 14 jours :

	Patients (n_i)	Guéris (y_i)	$\hat{p}_i$
Nouvel antibiotique (1)	120	90	0,750
Traitement standard (2)	110	66	0,600

Au seuil $\alpha = 5\%$, les taux de guérison diffèrent-ils ?

- Deux groupes de patients **différents** (attribution au hasard) : échantillons indépendants.
- $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.
- $\hat{P} = \frac{90 + 66}{120 + 110} = \frac{156}{230} = 0,678$; conditions : $n_1\hat{P} = 81,4$, $n_1(1 - \hat{P}) = 38,6$, $n_2\hat{P} = 74,6$, $n_2(1 - \hat{P}) = 35,4$, tous ≥ 5 .

Exemple 4 (suite) : Solution

1. **Erreur-type sous H_0 :**

$$\sqrt{0,678 \times 0,322 \left(\frac{1}{120} + \frac{1}{110} \right)} = \sqrt{0,2182 \times 0,01742} = 0,0617$$

2. **Statistique observée :** $z_{obs} = \frac{0,750 - 0,600}{0,0617} = 2,43$

3. **Décision :** $|z_{obs}| = 2,43 > z_{0,025} = 1,96$: on **rejette H_0** . Valeur-p
 $= 2 P(Z > 2,43) = 0,015$.

Exemple 4 (suite) : Solution

1. **Erreur-type sous H_0 :**

$$\sqrt{0,678 \times 0,322 \left(\frac{1}{120} + \frac{1}{110} \right)} = \sqrt{0,2182 \times 0,01742} = 0,0617$$

2. **Statistique observée :** $z_{obs} = \frac{0,750 - 0,600}{0,0617} = 2,43$

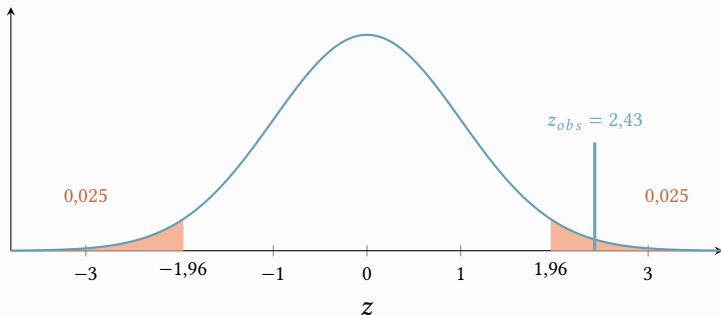
3. **Décision :** $|z_{obs}| = 2,43 > z_{0,025} = 1,96$: on **rejette H_0** . Valeur-p
 $= 2 P(Z > 2,43) = 0,015$.

4. **IC à 95 % pour $p_1 - p_2$:** $\widehat{ET} = \sqrt{\frac{0,75 \times 0,25}{120} + \frac{0,60 \times 0,40}{110}} = 0,0612$
 $0,150 \pm 1,96 \times 0,0612 = 0,150 \pm 0,120 = [0,030 ; 0,270]$

5. **Conclusion :** le nouvel antibiotique a un taux de guérison significativement plus élevé : de 3 à 27 points de pourcentage de plus que le traitement standard (confiance 95 %).

Exemple 4 (suite) : La décision en image

Loi $\mathcal{N}(0, 1)$ sous H_0 (test bilatéral, $\alpha = 0,05$)



- z_{obs} tombe dans la **région critique** : on rejette H_0 .
- Cohérent avec l'IC à 95 % $[0,030 ; 0,270]$, qui **ne contient pas 0**.

Exemple 4 (suite) : Sortie R de `prop.test()`

```
> prop.test(x = c(90, 66), n = c(120, 110), correct = FALSE)
```

2-sample test for equality of proportions without continuity correction

```
data:  c(90, 66) out of c(120, 110)
X-squared = 5.9174, df = 1, p-value = 0.01499
alternative hypothesis: two.sided
95 percent confidence interval:
 0.03006817 0.26993183
sample estimates:
prop 1 prop 2
 0.75    0.60
```

- R affiche $X^2 = z_{obs}^2 : 2,433^2 = 5,917$ (un khi-deux à 1 d.l., voir chapitre 7). Même valeur-p : 0,015.
- `correct = FALSE` : sans correction de continuité, pour retrouver le test z du cours (par défaut, R applique une correction).
- L'IC est exactement celui calculé à la main.

QUIZ!

Une petite étude pilote avait obtenu les *mêmes proportions* : 9 guéris sur 12 (nouvel antibiotique) et 6 sur 10 (standard).

- La différence observée (0,15) est la même. La conclusion est-elle la même?
- Le test est-il seulement valide ici?

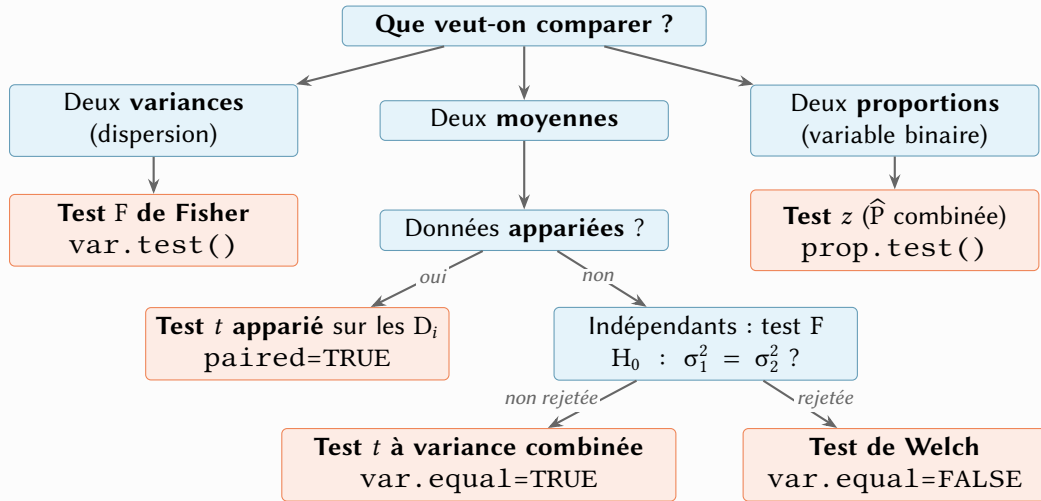
QUIZ!

Une petite étude pilote avait obtenu les *mêmes proportions* : 9 guéris sur 12 (nouvel antibiotique) et 6 sur 10 (standard).

- La différence observée (0,15) est la même. La conclusion est-elle la même?
- **Non.** $\hat{P} = 15/22 = 0,682$, erreur-type = 0,199, $z_{obs} = 0,75$ (valeur-p = 0,45) : on ne rejette pas H_0 . La même différence n'est pas significative avec de petits échantillons.
- Le test est-il seulement valide ici?
- Douteux : $n_2(1 - \hat{P}) = 10 \times 0,318 = 3,2 < 5$. R avertit d'ailleurs :
Chi-squared approximation may be incorrect.

Synthèse : choisir le bon test

Quel test utiliser ?



QUIZ!

Quel test utiliser ?

1. La variabilité du calibre des pommes diffère-t-elle entre deux vergers ?
2. Proportion de tiques porteuses de la bactérie de la maladie de Lyme dans deux régions ($n = 250$ et 300 tiques).
3. Masse de souris de deux lignées ($n = 12$ et 15); le test F rejette l'égalité des variances.
4. Fréquence cardiaque au repos de 20 athlètes avant et après un programme d'entraînement.
5. Longueur de truites de deux lacs ($n = 25$ chacun); le test F ne rejette pas l'égalité des variances.

QUIZ!

Quel test utiliser ?

1. La variabilité du calibre des pommes diffère-t-elle entre deux vergers ? $\Rightarrow$ **Test F de Fisher** (`var.test`).
2. Proportion de tiques porteuses de la bactérie de la maladie de Lyme dans deux régions ($n = 250$ et 300 tiques). $\Rightarrow$ **Test z pour deux proportions** (`prop.test`).
3. Masse de souris de deux lignées ($n = 12$ et 15); le test F rejette l'égalité des variances. $\Rightarrow$ **Test de Welch** (`t.test`, `var.equal = FALSE`).
4. Fréquence cardiaque au repos de 20 athlètes avant et après un programme d'entraînement. $\Rightarrow$ **Test t apparié** (`paired = TRUE`).
5. Longueur de truites de deux lacs ($n = 25$ chacun); le test F ne rejette pas l'égalité des variances. $\Rightarrow$ **Test t à variance combinée** (`var.equal = TRUE`).

Tableau récapitulatif du Chapitre 4

Problème	Statistique de test	Loi sous H_0
$H_0 : \sigma_1^2 = \sigma_2^2$ (indép.)	$F_{obs} = \frac{S_1^2}{S_2^2}$	$\mathcal{F}(n_1 - 1, n_2 - 1)$
$H_0 : \mu_1 = \mu_2$ (σ^2 égales)	$t_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{1/n_1 + 1/n_2}}$	$t_{n_1 + n_2 - 2}$
$H_0 : \mu_1 = \mu_2$ (σ^2 inégales)	$t_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$	t_v (Welch)
$H_0 : \mu_1 = \mu_2$ (appariés)	$t_{obs} = \frac{\bar{D}}{S_D / \sqrt{n}}$	t_{n-1}
$H_0 : p_1 = p_2$ (grandes tailles)	$z_{obs} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{P}(1-\hat{P})(1/n_1 + 1/n_2)}}$	$\mathcal{N}(0, 1)$

Tous ces tests sont disponibles sous R avec `var.test()`, `t.test()` et `prop.test()`.

Réponses aux exemples

1. **Zinc.** Test F : $0,228 \leq F_{obs} = 0,606 \leq 4,95$ (valeur-p = 0,557), on ne rejette pas l'égalité des variances. Test t ($H_1 : \mu_1 > \mu_2$) : $t_{obs} = 1,01 < 1,796$ (valeur-p = 0,167), on ne rejette pas H_0 : la concentration moyenne de la rivière 1 n'excède pas celle de la rivière 2 de façon significative. IC à 95 % pour $\mu_1 - \mu_2$: $0,61 \pm 2,201 \times 0,604 = [-0,72 ; 1,94]$ $\mu\text{g/l}$.
2. **Mercure.** $F_{obs} = 23,46 > 3,29$: variances inégales. Welch : $t_{obs} = 3,90$, $v = 7,48$, $|t_{obs}| > 2,365$ (valeur-p = 0,005), on rejette H_0 . IC à 95 % avec $v = 7$: $0,354 \pm 2,365 \times 0,0908 = [0,139 ; 0,569]$ $\mu\text{g/g}$ (R, avec $v = 7,48$: $[0,142 ; 0,566]$).
3. **Antihypertenseur.** $t_{obs} = 8,30 > 1,833$ (valeur-p < 0,0001), on rejette H_0 . IC à 95 % pour μ_D : $8,3 \pm 2,262 \times 1,001 = [6,04 ; 10,56]$ mmHg : baisse moyenne entre 6 et 10,6 mmHg. Analyse (erronée) en échantillons indépendants : $t_{obs} = 1,73 < 1,734$, valeur-p = 0,050.
4. **Guérison.** $z_{obs} = 2,43 > 1,96$ (valeur-p = 0,015), on rejette H_0 ; IC à 95 % pour $p_1 - p_2$: $[0,030 ; 0,270]$. Étude pilote : $z_{obs} = 0,75$ (valeur-p = 0,45), conditions non respectées.
5. **Zinc fond/surface** (sortie R) : $t_{obs} = 3,69$, 5 d.l., valeur-p = 0,007 < 0,01 : on rejette H_0 .