Chapitre 3

Tests statistiques
Tests de comparaisons de moyenne

Cédric Wolf [a]

[a]Université de Rennes I, Unité Mixte de Recherche « ECOBIO »

Définition du H0

Le **H0**, aussi appelée **hypothèse neutre** ou **hypothèse nulle** est l'hypothèse qui sera testée

Le H0 est bien entendu déterminé en fonction de l'objectif de l'étude, et dépend du test qui sera effectué

H0 traduit le fait que la différence observée entre la valeur réalisée de la statistique associée au test et sa valeur théorique est due <u>uniquement</u> au hasard

Définition du H0

Le **H0**, aussi appelée **hypothèse neutre** ou **hypothèse nulle** est l'hypothèse qui sera testée

Le H0 est bien entendu déterminé en fonction de l'objectif de l'étude, et dépend du test qui sera effectué

H0 traduit le fait que la différence observée entre la valeur réalisée de la statistique associée au test et sa valeur théorique est due uniquement au hasard

Le plus souvent, l'hypothèse H0 sera donc du type “pas de différence entre X_{pop} et $X_{théo}$ “, “Rien à signaler” ; “pas d'effet”, etc...

Exemples :

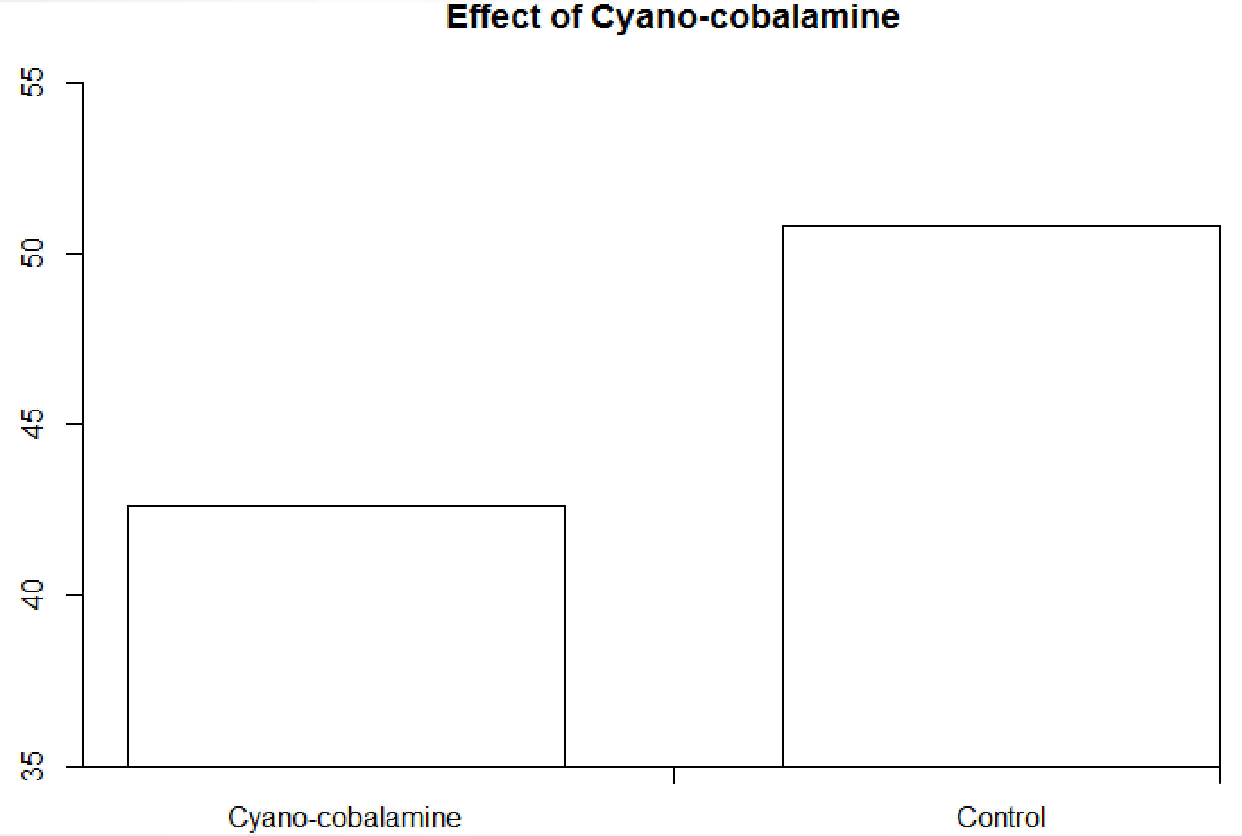
- H0 : Pas de différence entre la taille des hommes et celle des femmes
- H0 : $X_{pop} = X_{théo}$
- H0 : Le traitement n'a pas d'effet
- H0 : La population étudiée correspond à la population de référence
- H0 : la distribution des données ne diffère pas de celle d'une loi Normale

On peut lui associer l'hypothèse inverse **H1**, qui traduit donc le fait que la différence observée entre la valeur réalisée de la statistique associée au test et sa valeur théorique n'est pas due que au hasard (et qu'il y a donc une différence réelle due à autre chose).

Définition du H0 : exemple

Exemple : Effet de la cyano-cobalamine sur le rendement du riz : moyenne de 30 rendements pour chaque modalité

```
> Cyano
 [1] 43 98 45  3 29 46 11 41 43  5 19 57 13 48  3 65 49 61 69 10 64 38 29 18 13 71 81 81 42 84
> Control
 [1] 31 37 41 92 19 58 14 93 46 76 31 29 81  4 88 36 94 96 58 14 82 47 66 27 82 62 73  6 18 24
```



H0 : la pollution n'a pas d'effet sur le rendement

⇔

H0 : le rendement moyen est identique pour les parcelles polluées et pour les parcelles non polluées

Interprétation : pvalue, seuil de signification

La **pvalue** p est définie comme la probabilité d'observer la valeur $X_{rés}$ (ou “pire”) si H0 est vraie. Elle dépend de l'effectif, du test utilisé, ….
Plus la pvalue est petite, moins H0 est probable

Le **seuil de signification α** , ou **seuil de risque**, est le risque que l'on est prêt à prendre de rejeter une hypothèse H0 vraie. Le plus souvent (=par défaut), il est fixé de 5% ($\alpha=0.05$).

Ainsi :

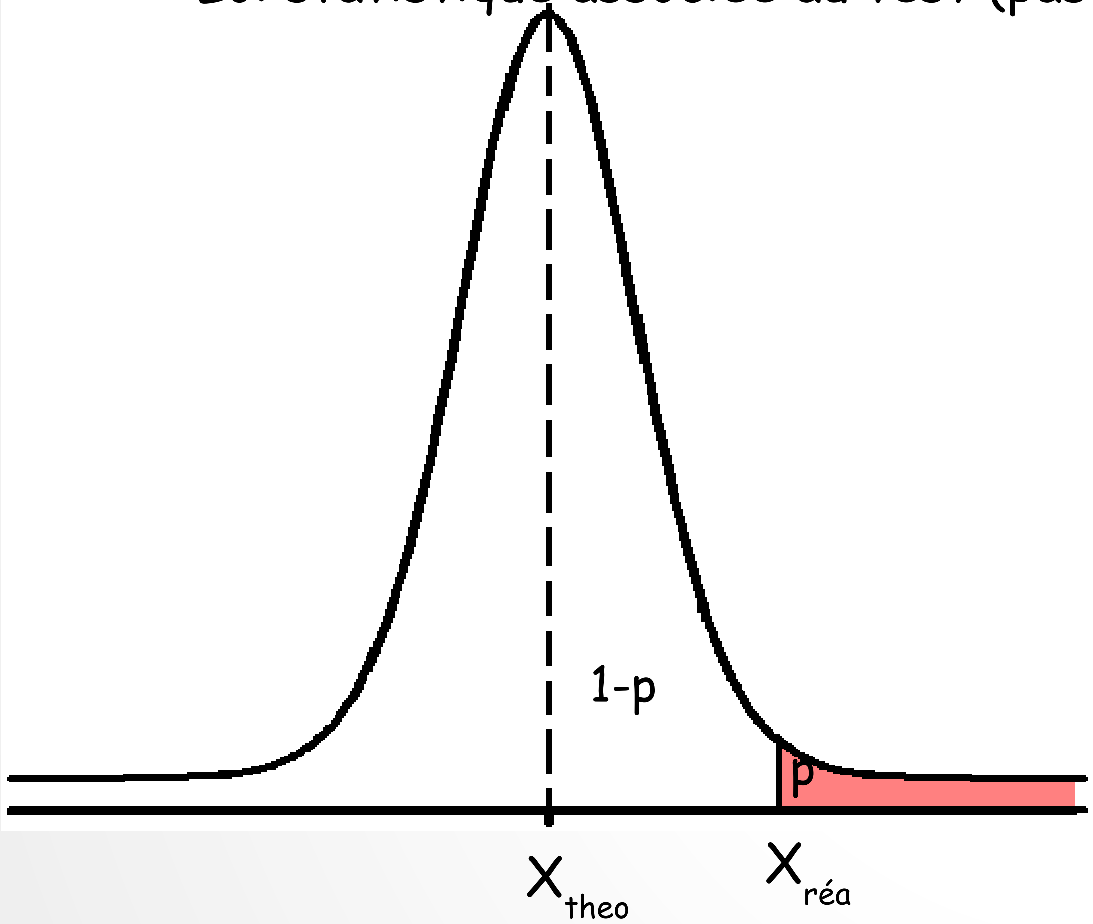
Si $p < \alpha$: On parle de **test significatif**. On rejette H0 (car improbable) et on valide H1.
(différence significative entre ce qu'on observe et ce qu'on devrait observer si H0 était vraie)

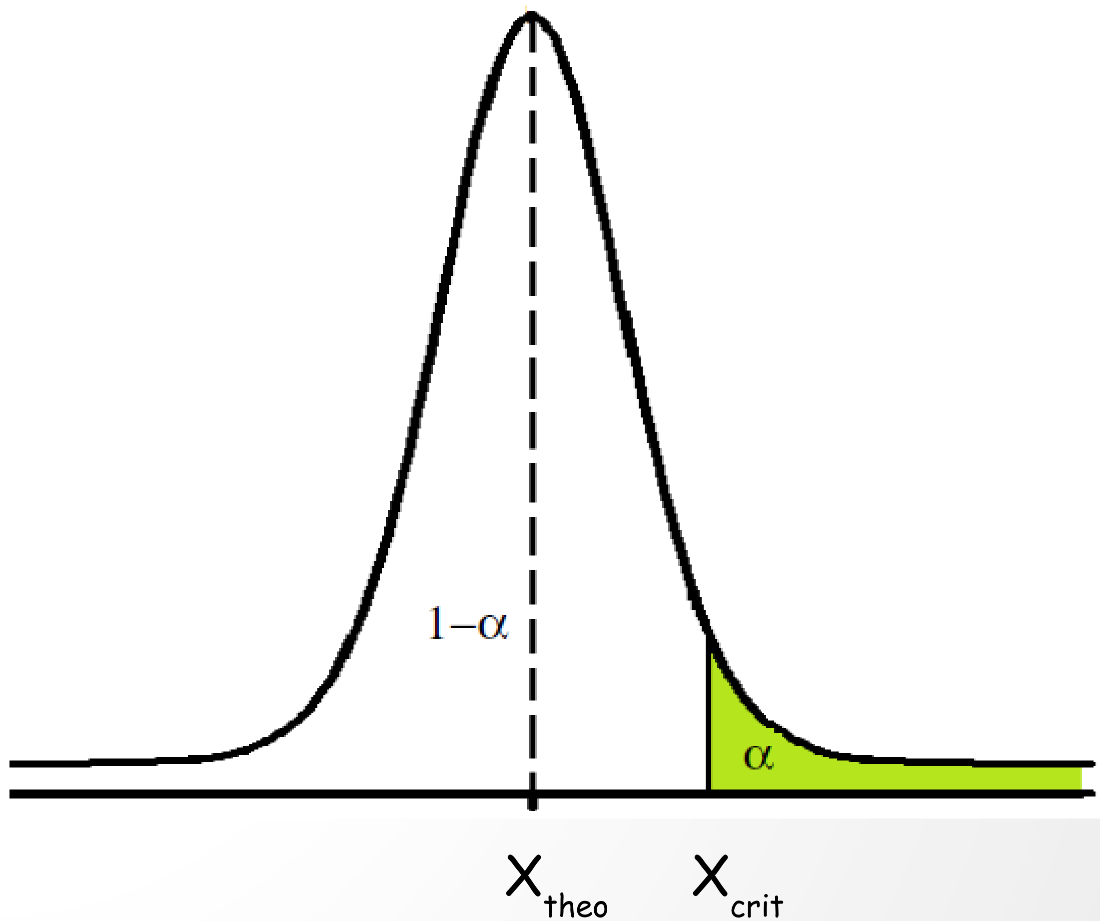
Si $p > \alpha$: **Test non significatif**. On ne rejette pas H0.

Cela ne signifie par que H0 est vraie, simplement qu'elle est possible, mais d'autres hypothèses non compatibles peuvent l'être également !

Interprétation : pvalue, seuil de signification

Loi statistique associée au test (pas Normale mais souvent proche)





p : p_value observée

X_{ech} : Valeur observée

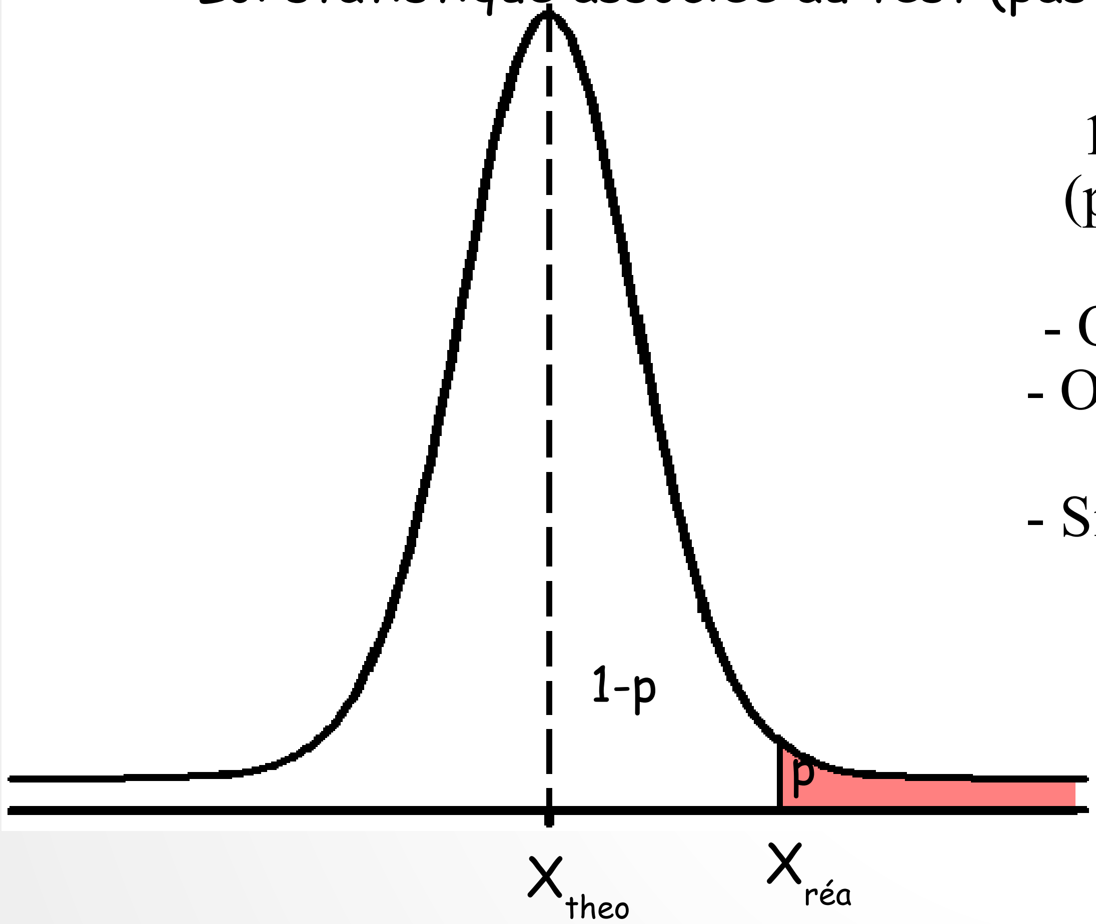
$X_{théo}$: Valeur théorique (0 le plus souvent)

α : Risque toléré

X_{crit} : Valeur limite autorisée au risque α

Interprétation : pvalue, seuil de signification

Loi statistique associée au test (pas Normale mais souvent proche)



1ère approche :
(privilégiée avec ordinateur)

- On calcule p
- On compare p avec le seuil de risque α souhaité

- Si p< α , le test est significatif

p : p_value observée

$X_{réa}$: Valeur observée

$X_{théo}$: Valeur théorique (0 le plus souvent)

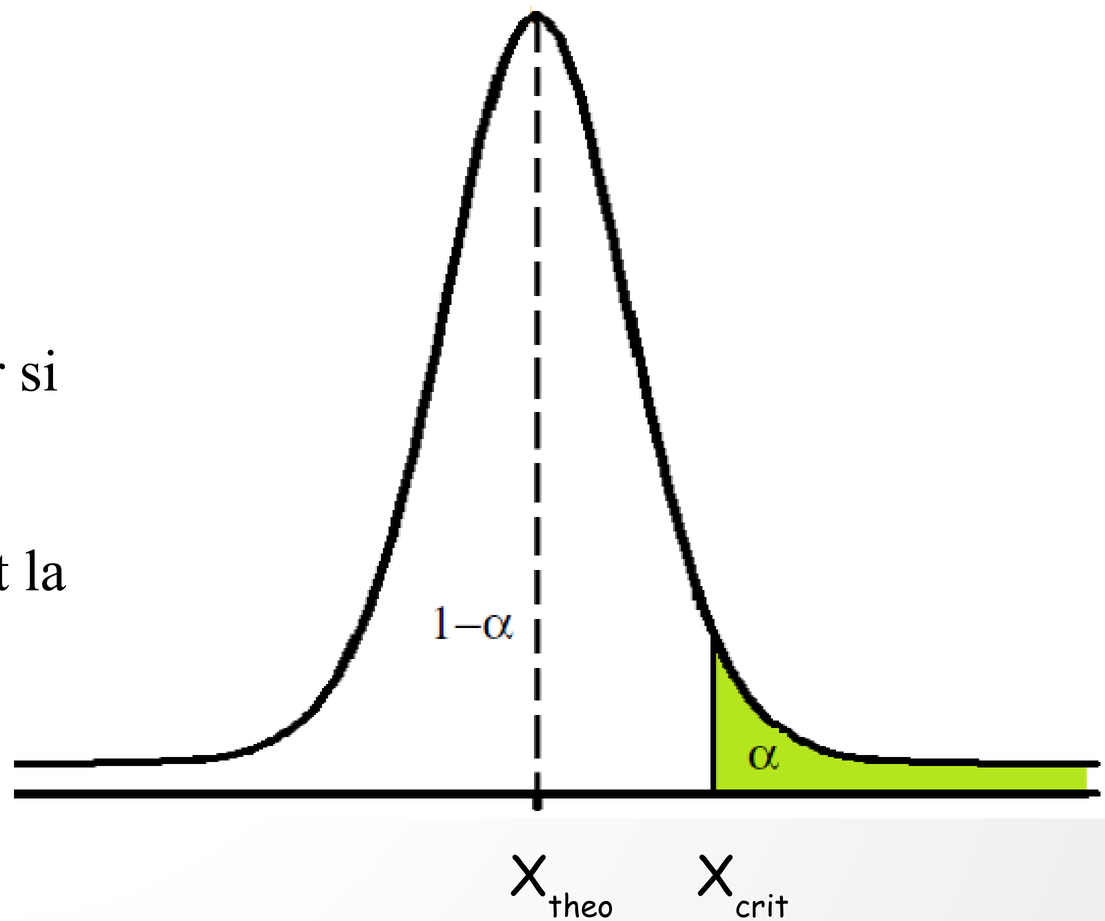
Interprétation : pvalue, seuil de signification

Loi statistique associée au test (pas Normale mais souvent proche)

2ème approche :
(privilégiée « à la main »)

- On fixe α
- On en déduit X_{crit} = valeur à ne pas dépasser si H0 est vraie (tables...)

- Si $X_{réa}$ dépasse X_{crit}, le test est significatif (et la pvalue est donc inférieur à α)



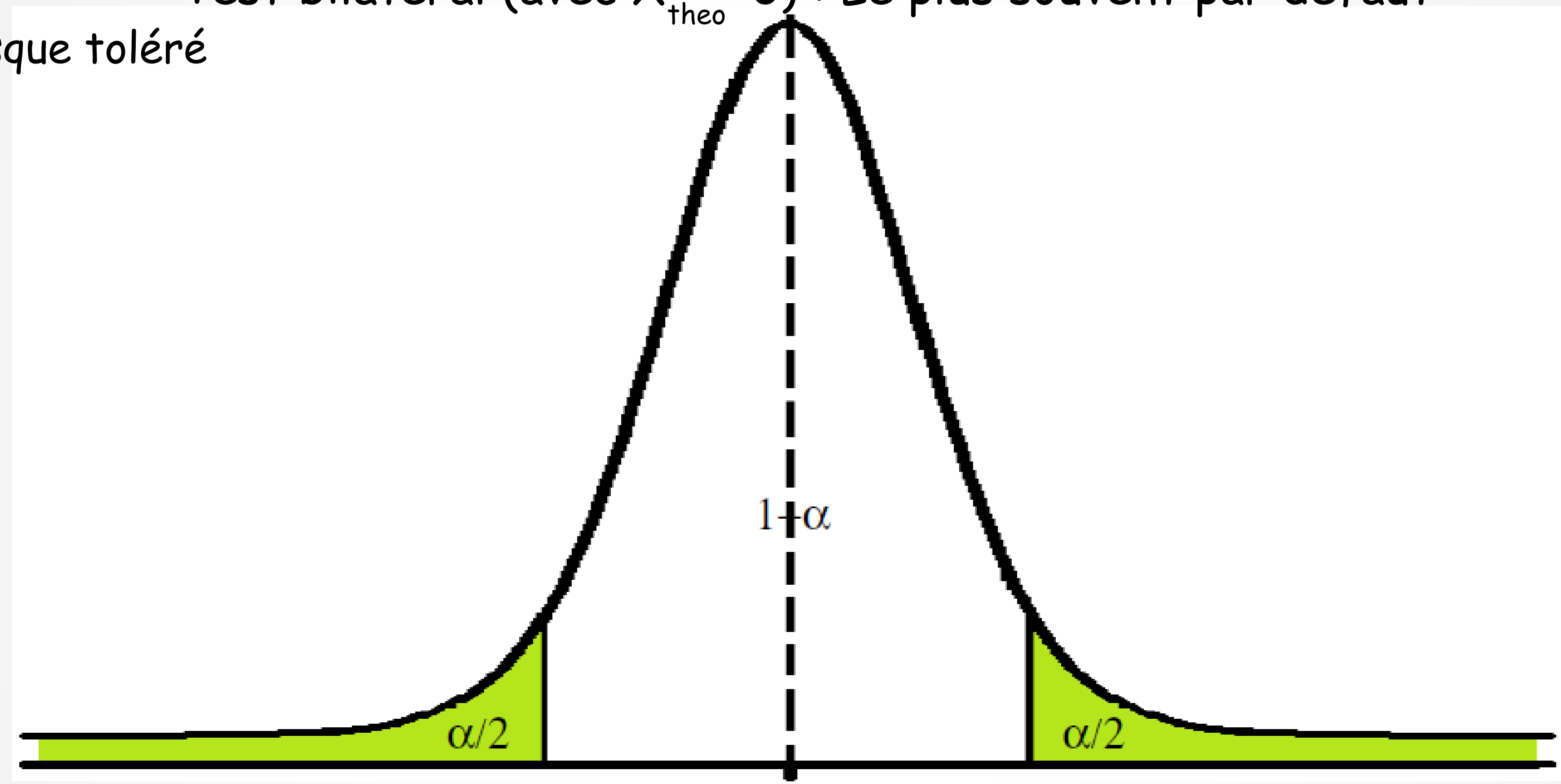
α : Risque toléré

X_{crit} : Valeur limite autorisée au risque α

Test unilatéral ou bilatéral ?

Test bilatéral (avec $X_{theo}=0$) : Le plus souvent par défaut

α : Risque toléré



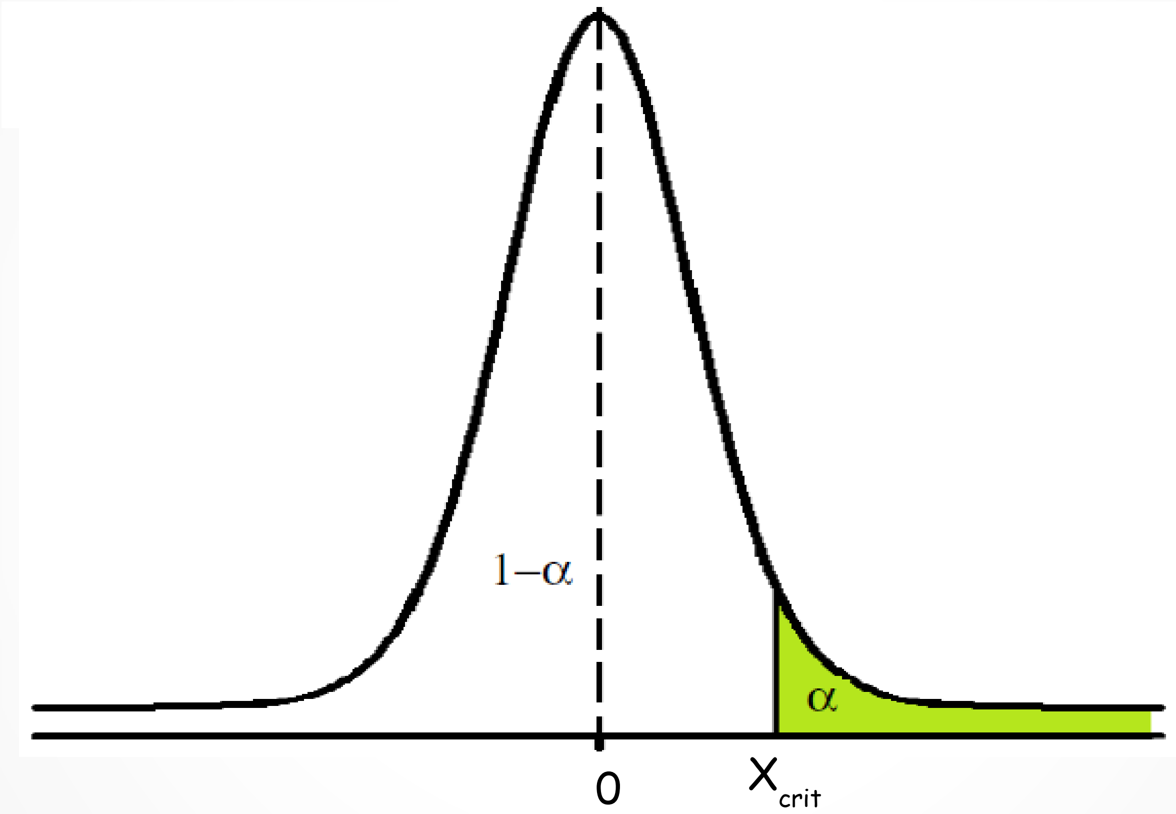
$-X_{crit}$ 0 X_{crit}

$\lvert X_{réa}\rvert \geq \lvert X_{crit}\rvert$	$\lvert X_{réa}\rvert \leq \lvert X_{crit}\rvert$	$\lvert X_{réa}\rvert \geq \lvert X_{crit}\rvert$
Test significatif	Test non significatif	Test significatif
H0 rejetée ; H1 validée : $X_{théo} \neq 0$	H0 non rejetée	H0 rejetée ; H1 validée : $X_{théo} \neq 0$

Test unilatéral ou bilatéral ?

Test unilatéral droit (avec X_{theo}=0)

α : Risque toléré



$X_{réa} \leq X_{crit}$

Test non significatif

H0 non rejetée

$X_{réa} > X_{crit}$

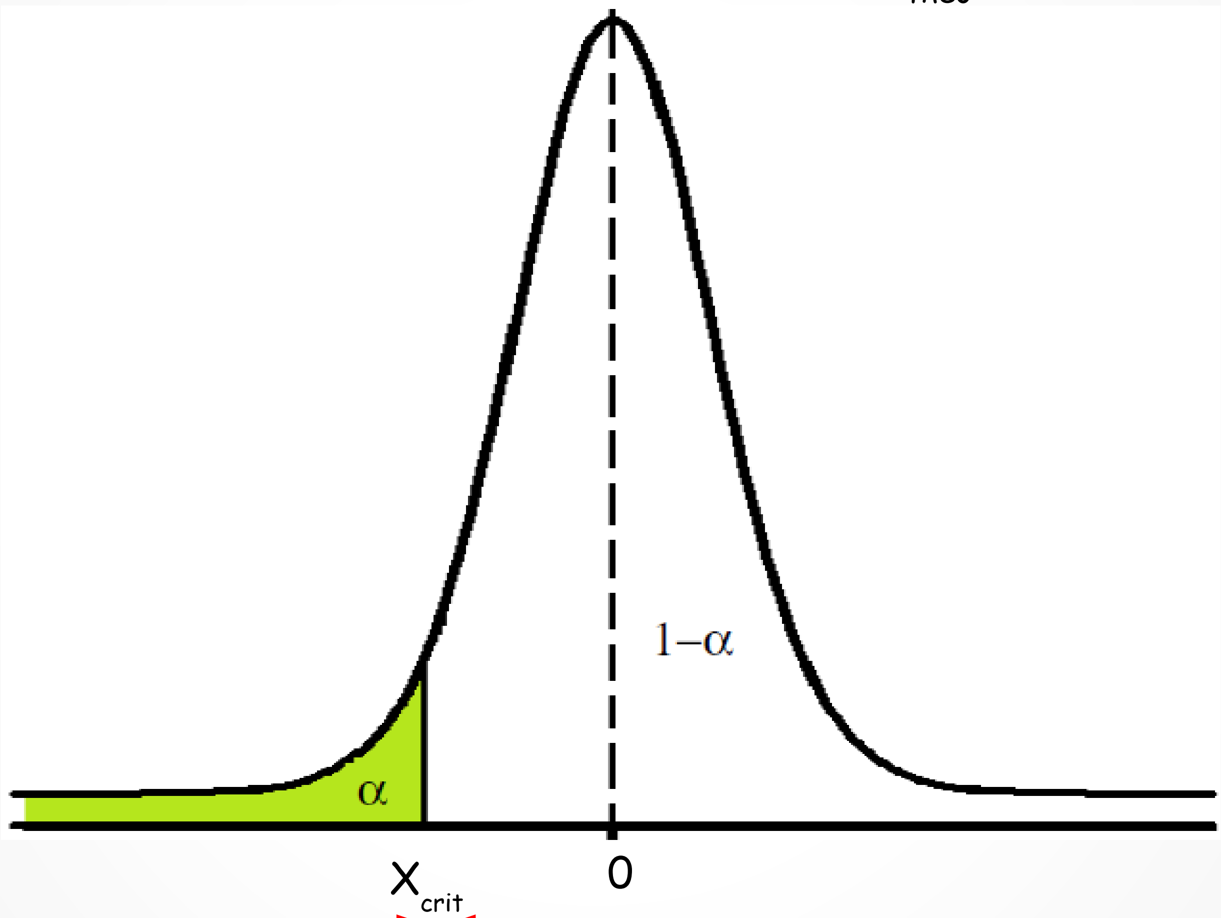
Test significatif

H0 rejetée ; H1 validée : $X_{théo} > 0$

Test unilatéral ou bilatéral ?

Test unilatéral gauche (avec $X_{theo}=0$)

α : Risque toléré



$X_{réa} < X_{crit}$

Test significatif

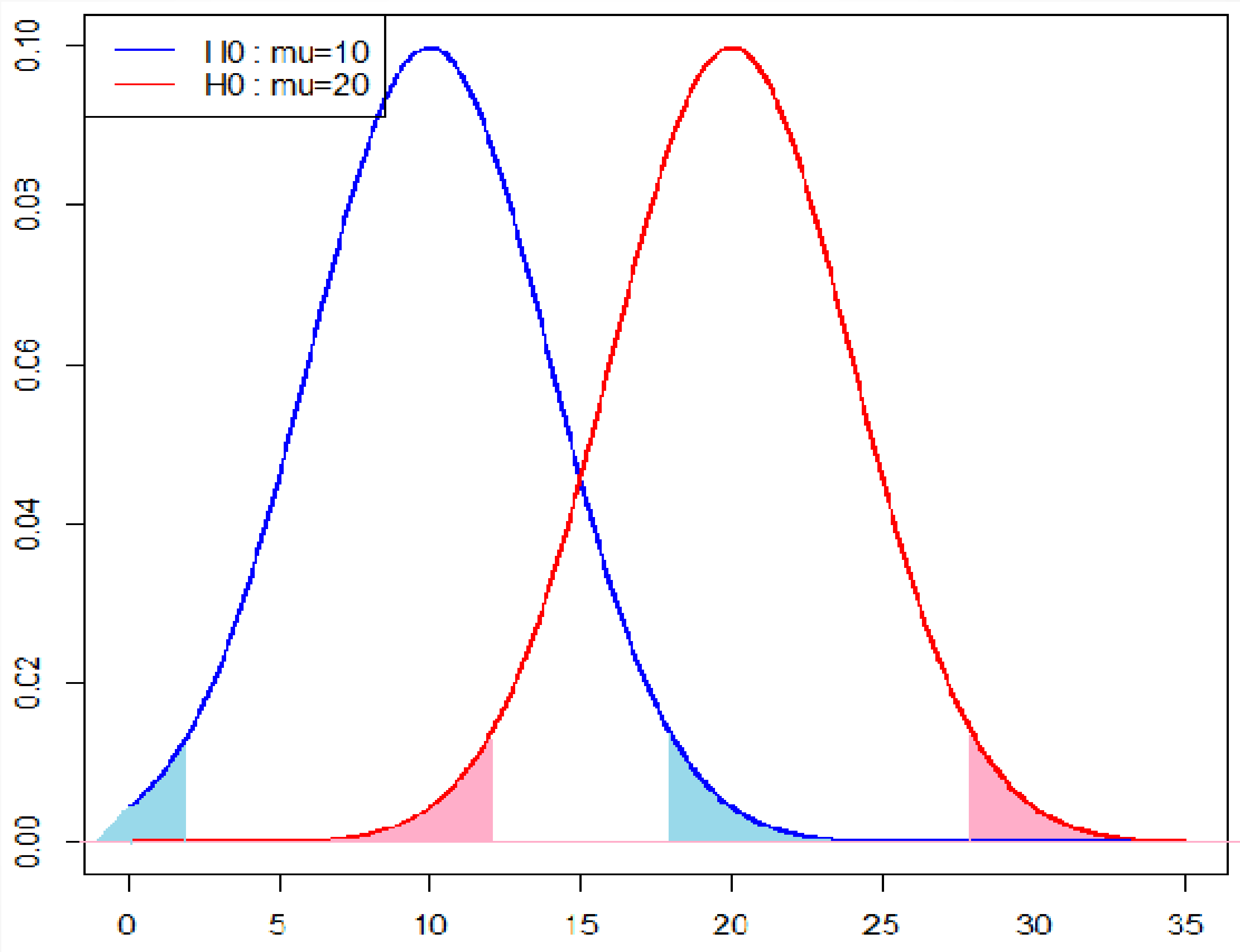
H0 rejetée ; H1 validée : $X_{théo} < 0$

$X_{réa} \geq X_{crit}$

Test non significatif

H0 non rejetée

Test non significatif ≠ H0 vraie !



$x_{\text{réa}} = 15$ est compatible avec les deux hypothèses !

Puissance d'un test

4 situations sont possibles :

	H0 Vraie	H0 Fausse
H0 rejetée (test significatif)	Erreur de première espèce proba α	OK ! proba $1-\beta$
H0 non rejetée (test non significatif)	OK ! proba $1-\alpha$	Erreur de seconde espèce proba β

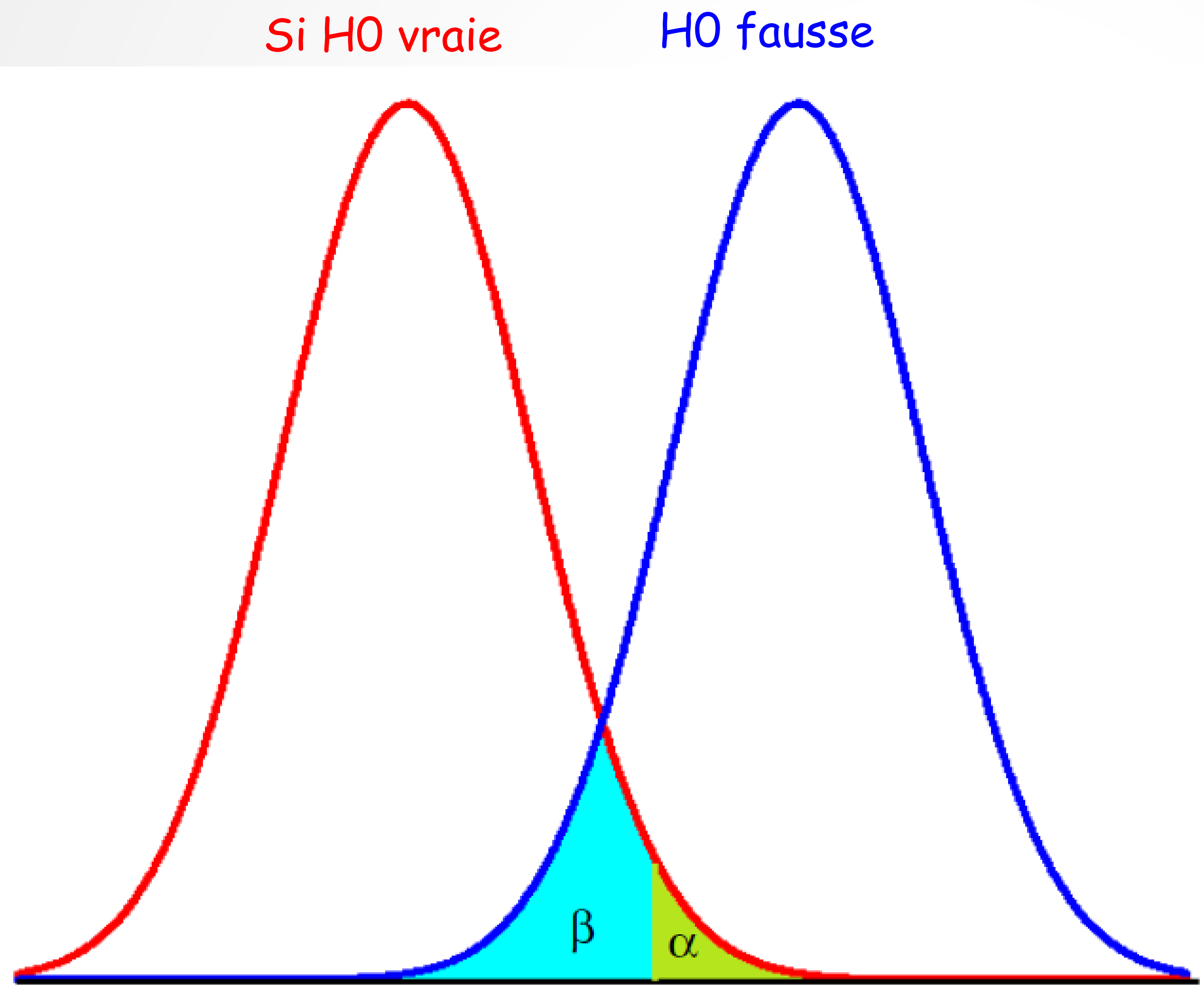
α est le seuil de signification choisit par l'utilisateur : maîtrisé !

β est inconnu ! $1-\beta$ est la **puissance** du test : Probabilité de rejeter H0 si elle est effectivement fausse

La puissance d'un test paramétrique est toujours meilleure que celle du test non paramétrique correspondant (pour les mêmes données).

La puissance d'un test est difficile à déterminer. On peut parfois l'estimer a posteriori (=une fois les données connues), mais on peut essayer de l'améliorer préventivement

Puissance d'un test

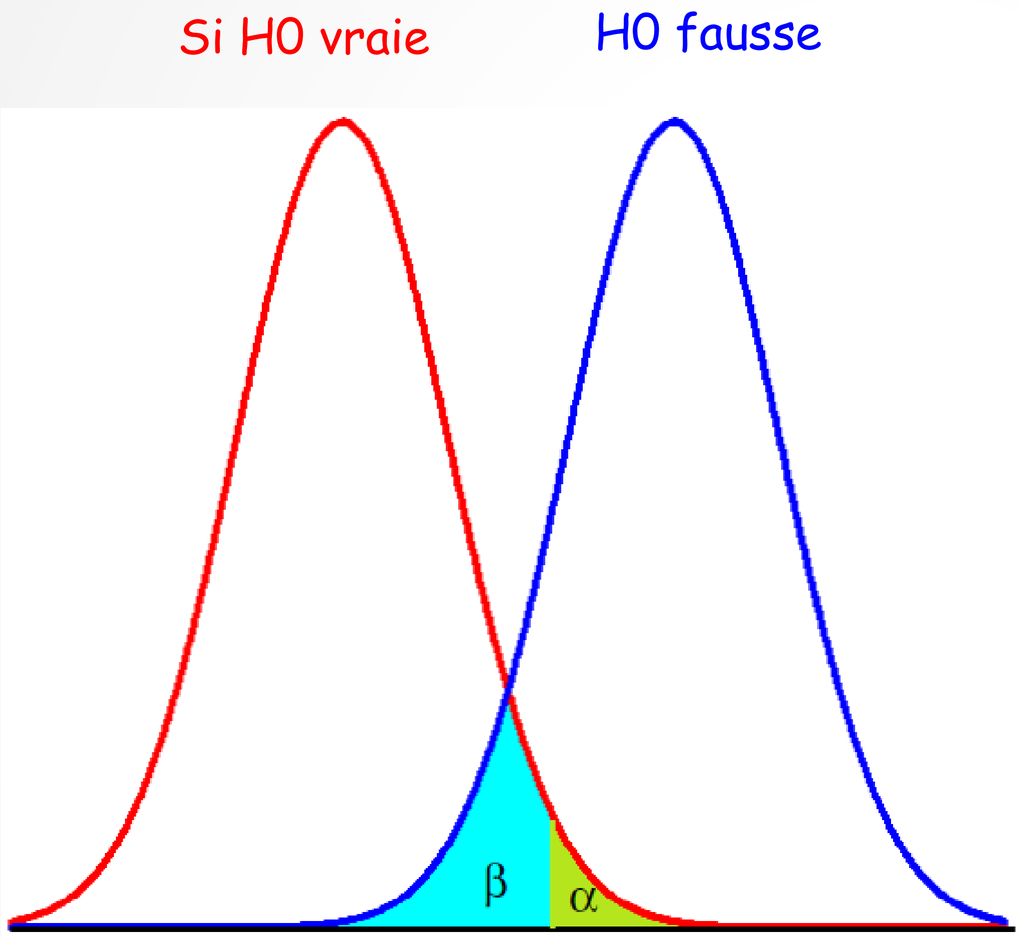


α : Risque toléré de rejeter H0 si vraie

β : Risque non maîtrisé d'accepter H0 si fausse

Puissance = 1-β
Comment l'améliorer ?

Puissance d'un test



Puissance = 1-β
Comment l'améliorer ?
Bilan :

- Favoriser les effets
- Diminuer la variance
- Augmenter α
- Augmenter l'effectif (cela diminue l'étalement)
- Privilégier test paramétrique (si possible!)

α : Risque toléré de rejeter H0 si vraie

β : Risque non maîtrisé d'accepter H0 si fausse

Test statistique ou Intervalle de confiance ?

Lien entre une valeur estimée X_{ech} de X_{pop} et une valeur théorique $X_{théo}$?
(avec un risque d'erreur de 5%)

Le test statistique associé (s'il existe !) donne l'intervalle dans lequel X_{ech} a 95% de chance de se trouver si H0 est vraie (H0 : $X_{pop} = X_{théo}$)

L'intervalle de confiance $IC_{95\%}$ donne un intervalle autour de X_{ech} dans lequel X_{pop} à 95% de chance de se trouver.

Les deux sont liés mais n'ont pas la même finalité... complémentaires !

- L'intervalle de confiance à l'avantage de vous indiquer la plage des valeurs théoriques possibles d'après la valeur échantillonée.
- Le test permet de tester la crédibilité d'une valeur théorique

Rappel : la pvalue associée au test n'est pas la probabilité que H0 soit vraie, mais la probabilité d'avoir observé X_{ech} <u>si</u> H0 est vraie.

Exemple : Test de Shapiro-Wilk

Le **test de Shapiro-Wilk** est un test d'ajustement à la loi Normale destiné à tester si la distribution des données observées est compatible avec l'hypothèse que la distribution théorique dont sont issues les données suit une loi Normale (de même moyenne et variance)

Conditions d'application : Variable quantitative continue, au moins 30 données idéalement (mais applicable sinon).

H0 : La distribution des données suit une loi Normale (hypothèse de normalité des données)

H1 : La distribution des données ne suit pas une loi Normale

Conclusion ? (avec un seuil à 95%, communément retenu)

pvalue < 0.05 : H0 rejetée, les données ne suivent pas une loi Normale
pvalue > 0.05 : H0 non rejetée, l'hypothèse de normalité est retenue (Rappel : mais pas prouvée !)

Le test de Shapiro n'est pas infaillible.
En particulier, il est “valide” un peu trop facilement la normalité si N est petit
Or c'est justement dans ce cas là que l'hypothèse de normalité est importante (le théorème centrale limite assure que pour N grand ça l'est moins)

Cf fonction Shapi(N) qui permet de tester le taux de rejet par le test de shapiro pour un échantillon de taille N

→ Intérêt de compléter avec une approche visuelle

Test de comparaison à une moyenne théorique

Cadre : On souhaite comparer une moyenne observée $\overline{X}$ a une moyenne théorique μ.

H0 : La moyenne μ_{pop} de la population (estimée par $\overline{X}$) est égale à $\mu \Leftrightarrow \mu_{pop} = \mu$

On a vu que pourvu que l'hypothèse de <u>normalité des données</u> soit respectée :

$$\boxed{IC_p = [\overline{X} - t_{N-1;\alpha}\frac{s}{\sqrt{N}} ; \overline{X} + t_{N-1;\alpha}\frac{s}{\sqrt{N}}]}$$

Donc SI H0 est vraie, l'écart entre $\overline{X}$ et μ devrait être inférieur à $t_{N-1;\alpha}\frac{s}{\sqrt{N}}$ avec une probabilité p, et donc :

$$\frac{|\overline{X}-\mu|}{s/\sqrt{N}} \leq t_{N-1;\alpha}$$

On calculera donc la statistique $\boxed{t = \frac{|\overline{X}-\mu|}{s/\sqrt{N}}}$ que l'on comparera à $t_{N-1;\alpha}$

$t > t_{N-1;\alpha} \Leftrightarrow pvalue < \alpha$: test significatif (seuil α) : H0 rejetée

$t < t_{N-1;\alpha} \Leftrightarrow pvalue > \alpha$: test non significatif (seuil α) : H0 non rejetée

Test de comparaison à une moyenne théorique : exemple

On a mesuré 25 pins sylvestres par la méthode trigonométrique et obtenus les résultats suivants (X_i en m) :

Ind	1	2	3	4	5	6	7	8	9	10	11	12	13
X	21.1	25.2	28.7	19.1	25.6	17.7	17.4	23.8	22.1	16.4	22.1	16.1	22.6
Ind.	14	15	16	17	18	19	20	21	22	23	24	25	
X	27.2	19.5	26.4	20.8	27.9	28.6	25.3	18.8	30.2	21.2	18.8	25.9	

L'échantillon mesuré correspond-il à une population type de taille moyenne 25 m ?

H0 : La moyenne de taille pour la population est de 25m

N=25

Test de Shapiro : pvalue = 0.36 donc hypothèse de normalité validée

$$\overline{X}=\frac{1}{N}\sum_{i=1}^{N}X_i=22.7$$

$$s^2=\frac{N}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N}X_i^2-\overline{X}^2\right)=17.50 \quad \text{donc} \quad s=4.18$$

$$t=\frac{|\overline{X}-\mu|}{s/\sqrt{N}}=\frac{|22.7-25|}{4.18/\sqrt{25}}=2.75$$

à comparer avec $t_{24;\alpha}$

$t>t_{24;0.05}$ Donc on rejette H0 au risque de 5 % ; l'échantillon mesuré ne provient pas d'une population de taille moyenne 25m, p<0.05

Rem : L'ordinateur donnera pvalue = 0.012

Test de comparaison de 2 moyennes : Cadre

La problématique est ici est de comparer les moyennes obtenues pour une même variable quantitative sur deux échantillons différents.

On notera $\overline{X}_1$ et $\overline{X}_2$ les moyennes calculées respectivement sur les deux échantillons, de taille N_1 et N_2

Il existe plusieurs tests de comparaison de moyennes : le choix se fera en fonction des données.
Dans certain cas, on sera amené à comparer en réalité les <u>médianes</u> et non pas les moyennes

H0 : La moyenne ne dépend pas de l'échantillon $\Leftrightarrow$ La moyenne des populations dont sont issus les échantillons sont identiques $\Leftrightarrow$ $\mu_1 = \mu_2$

H1 : Les deux échantillons proviennent de populations différentes $\Leftrightarrow$ $\mu_1 \neq \mu_2$

(à formuler selon la problématique !)

Test de comparaison de 2 moyennes : Cadre

Cadre : On a donc à comparer 2 moyennes $\overline{X}_1$ et $\overline{X}_2$ sur 2 échantillons de taille N_1 et N_2.

H0 : Les moyennes sont les mêmes pour les populations dont sont issus les échantillons
$\Leftrightarrow \mu_1 - \mu_2 = 0$ (si μ_1 et μ_2 sont les moyennes des populations 1 et 2)

Idée : On est donc ramenée à étudier la différence entre une moyenne observée $\overline{M}=\overline{X}_1 - \overline{X}_2$ et une moyenne théorique $\mu=0$.

Or on a vu que pourvu que l'hypothèse de normalité des données soit respectée :

$$\boxed{IC_p = [\overline{M} - t_{N-1;\alpha} es_M ; \overline{M} + t_{N-1;\alpha} es_M]}$$ (es_M = erreur standard associée à M)

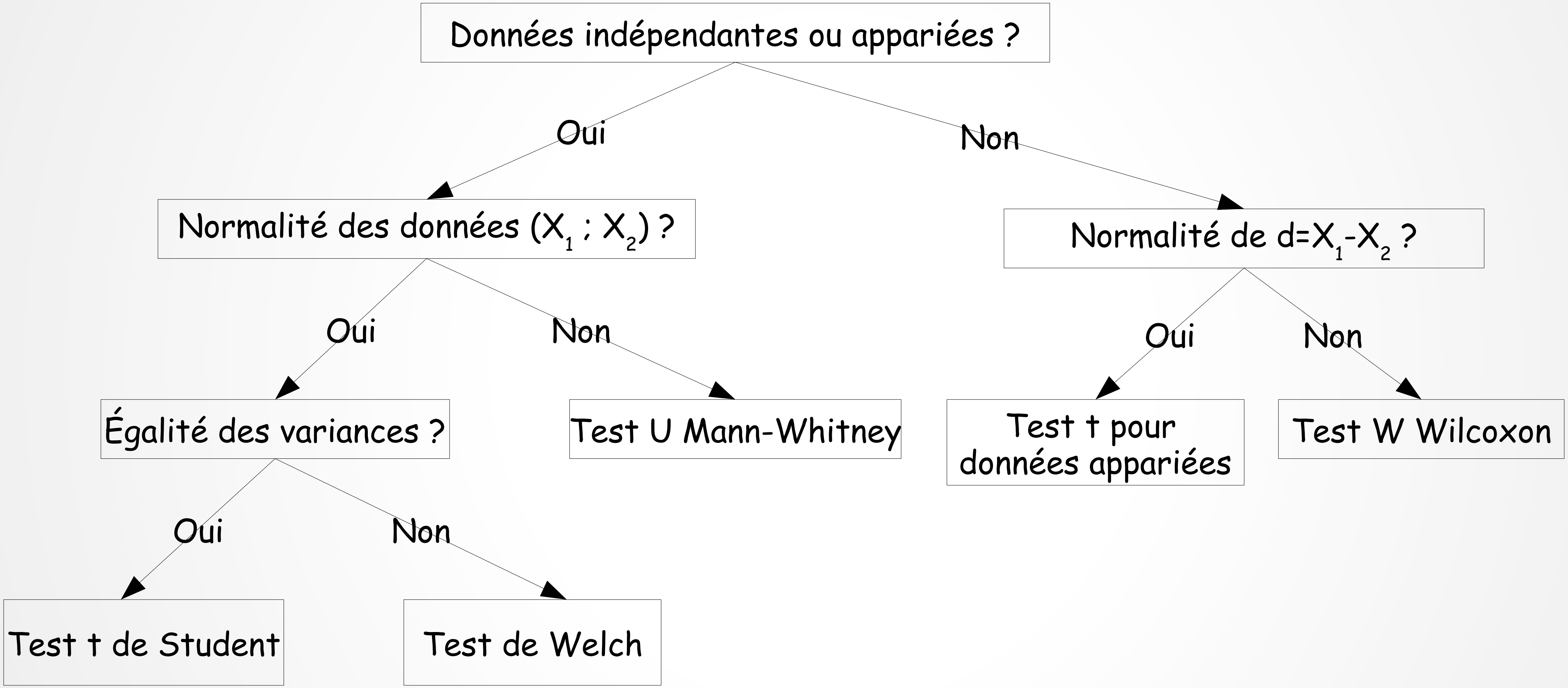
Donc SI H0 est vraie, l'écart entre $\overline{M}$ et μ devrait être inférieur à $t_{N-1;\alpha} es_M$ avec une probabilité p, et donc :

$$\frac{|\overline{M}-\mu|}{es_M} \leq t_{N-1;\alpha} \Leftrightarrow \frac{|\overline{X}_1 - \overline{X}_2|}{es_M} \leq t_{N-1;\alpha}$$

d'où l'idée de la statistique associée $Z = \dfrac{|\overline{X}_1 - \overline{X}_2|}{es_M}$ (la valeur à calculer) que l'on comparera

Avec $t_{DL;\alpha}$. DL et es_M seront adaptée selon les données

Test de comparaison de 2 moyennes : Quel test ?



Test de normalité : Test de Shapiro
Test d'égalité de variances (homoscédasticité) : Test de Fisher

Les test t et test de Welch sont des tests paramétriques
Les test de Mann-Whitney et de Wilcoxon sont des tests non paramétriques (moins puissants)

Données indépendantes ou appariées ?

Les données (= les valeurs obtenues dans les 2 échantillons) sont **indépendantes** si elles ne sont pas appariées. Elles sont **appariées** si les individus de chaque échantillon peuvent être associés par paire (avec une bonne raison !)
C'est donc que les échantillons sont liés (voire identiques)
Ex : Même individu avant et après traitement.

Rem : Ce n'est possible que lorsque $N_1 = N_2 (= N)$ (mais réciproque fausse !)

Le fait que les données soient appariées implique qu'il y a également de l'information dans la comparaison des données 2 à 2 et pas seulement à travers les moyennes.

Cela permet également de définir une nouvelle variable aléatoire $d = X_1 - X_2$, qui est donc évaluée pour les N individus de l'échantillon ; l'idée sera donc de comparer la moyenne $\bar{d}$ avec la moyenne 0.

Rem : Lorsque les données sont fournis par un tiers (ex : le prof !), ce n'est pas toujours évident de savoir si les données sont appariées ou pas ; il faut faire avec les informations disponibles....
En l'absence d'indications vous permettant d'apparier les données (= de faire des paires), on considèrera les données comme indépendantes.
Ce sera plus facile lorsque vous travaillerez sur vos propres donneés !

Normalité des données ?

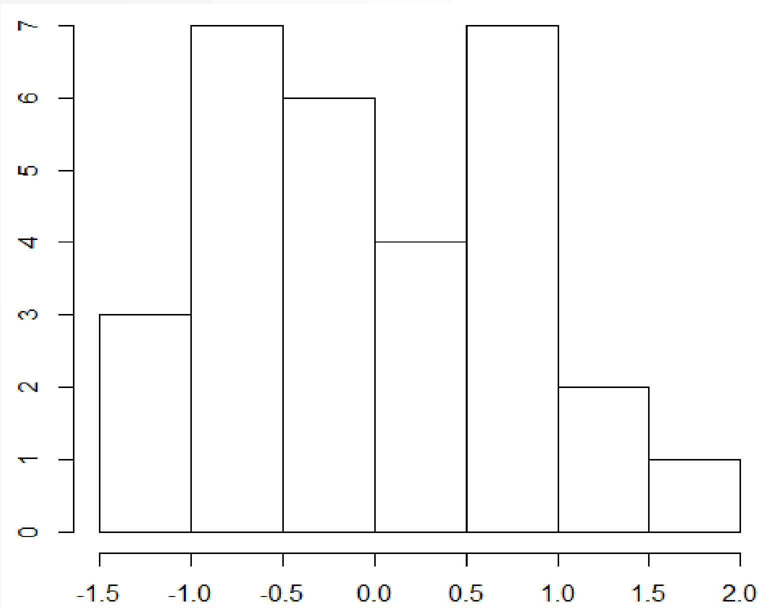
L'hypothèse de normalité des données sera validée par un **test de Shapiro-Wilk** .

Rappel : Conclusion ? (avec un seuil à 95%, communément retenu)

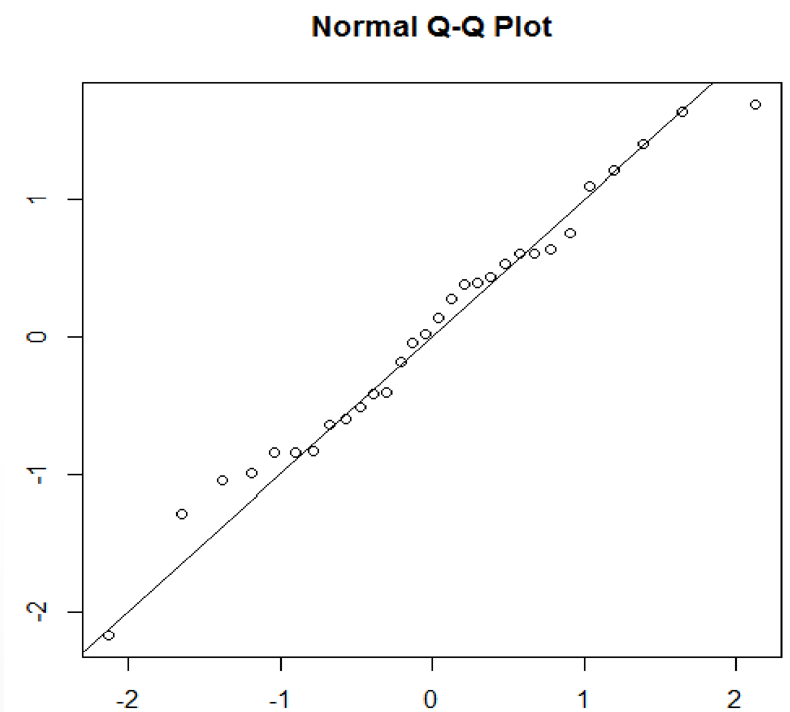
pvalue < 0.05 : H0 rejetée, les données ne suivent pas une loi Normale

pvalue > 0.05 : H0 non rejetée, l'hypothèse de normalité est retenue (rappel : pas prouvée !)

Le test de Shapiro-Wilk étant particulièrement peu puissant pour des effectifs petits, on complètera l'analyse de normalité pour des effectifs <30 par l'examen de la représentation graphique des données et d'un QQplot :



→ Ok si pas trop loin d'une “cloche”



→ Ok si points à peu près alignés sur la droite

Normalité des données ?

La normalité des données est à vérifier :

- Pour chaque échantillon dans le cas de données indépendantes
- Pour $d=X_1-X_2$ dans le cas de données appariées

Si un jeu de données n'a pas une distribution Normale, on a 2 choix :

- Solution 1 : "Transformer" les données pour revenir dans le cadre Normal

Note : la transformation est a effectuer sur les 2 échantillons dans le cas des données indépendantes, et sur d dans le cas de données appariées

- Solution 2 : Utiliser un test non-paramétrique (**test de Mann-Whitney** si les données sont indépendantes ; **test de Wilcoxon** si les données sont appariées)

Transformation des données

La transformation consiste à transformer les données (toutes les données de tous les échantillons) en appliquant une même fonction en espérant aboutir à des données transformées distribuées suivant une loi Normale.

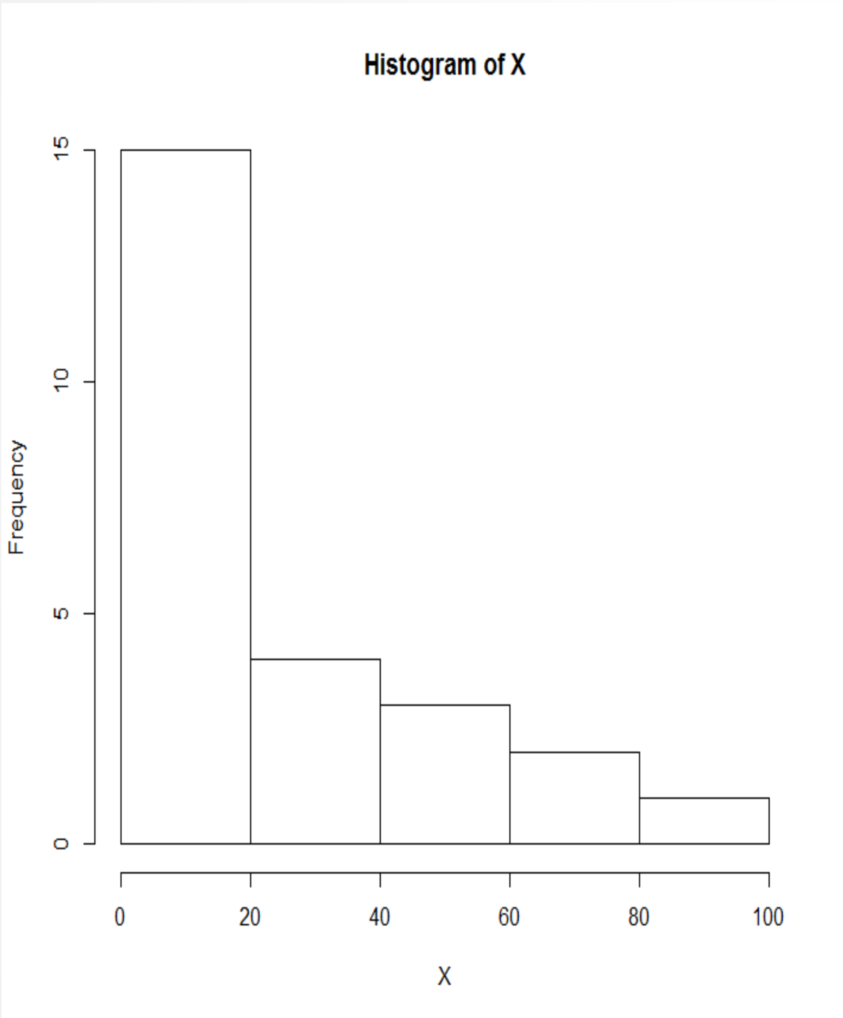
Le procédé est parfaitement licite (et les conclusions sur les données transformées s'appliquent aux données d'origine) à partir du moment ou la fonction utilisée est inversible (ce qui signifie que l'on peut retrouver la valeur d'origine par une fonction inverse)

Les plus couramment utilisées sont : log(X), log(1+X), sqrt(X) ou Arcsin(X) dans le cas de proportions notamment

Exemple de transformation des données

X=c(2,3,69,6,19,1,9,19,95,4,4,21,47,2,29,41,7,2,73,5,23,32,11,55,0)

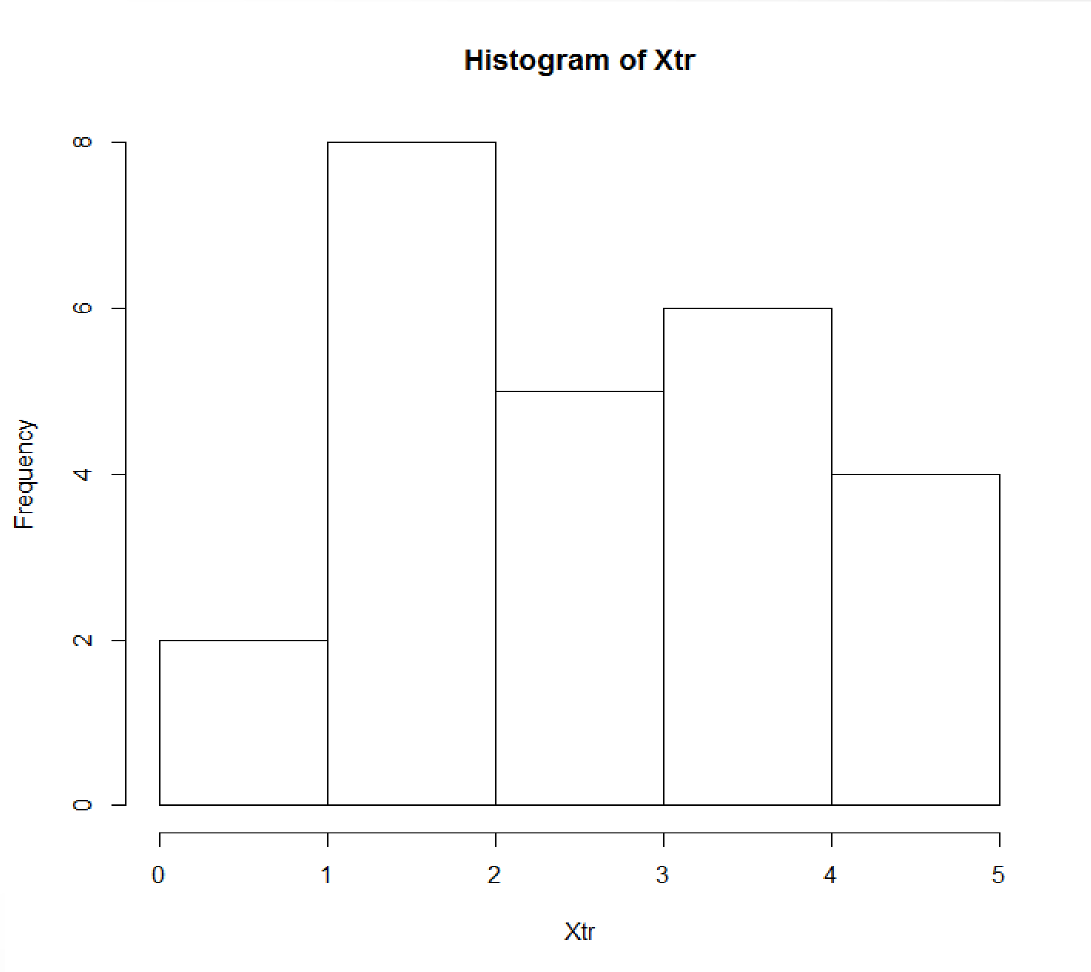
X



Test de Shapiro : pvalue=0.0004

→ Pas normalité

Xtr=log(1+X)



Test de Shapiro : pvalue=0.496

→ Normalité

Test t pour données appariées

Les données étant appariées, on définit une nouvelle variable aléatoire $d=X_1-X_2$, qui est donc évaluée pour les N individus de l'échantillon.

L'idée est donc de comparer la moyenne $\overline{d}$ avec la moyenne théorique 0.

Ainsi, si les données de la variable d sont Normales (test de Shapiro), on utilisera le **test t de Student pour données appariées** :

$$\boxed{t=\frac{|\overline{d}|}{s_d/\sqrt{N}}} \quad \text{à comparer avec} \quad t_{N-1;\alpha}$$

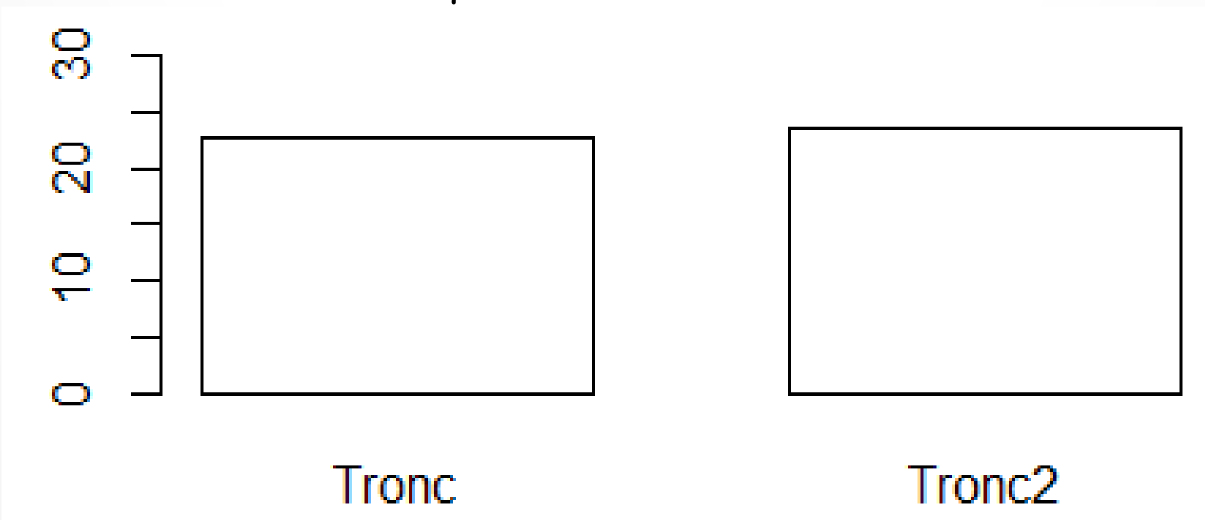
Sinon, on utilisera le **test de Wilcoxon**, qui est un test non-paramétrique (donc moins puissant ; moins de chance de rejter H0 si fausse)

Test t pour données appariées : exemple

On coupe les pins et on les mesure au sol. On obtient les résultats suivant :

Ind.	1	2	3	4	5	6	7	8	9	10	11	12	13
X1	21.1	25.2	28.7	19.1	25.6	17.7	17.4	23.8	22.1	16.4	22.1	16.1	22.6
X2	24.5	24.2	30.8	19.7	28.4	19.0	18.0	23.3	21.2	17.8	23.3	15.8	22.8
Ind.	14	15	16	17	18	19	20	21	22	23	24	25	
X1	27.2	19.5	26.4	20.8	27.9	28.6	25.3	18.8	30.2	21.2	18.8	25.9	
X2	28.4	21.0	27.7	21.4	28.1	29.0	25.5	18.4	31.2	21.8	19.4	27.1	

Les deux méthodes de mesures sont-elles équivalentes ?



Test t pour données appariées : exemple

d=X1-X2

```
> d
 [1] -3.4  1.0 -2.1 -0.6 -2.8 -1.3 -0.6  0.5  0.9 -1.4 -1.2  0.3 -0.2 -1.2 -1.5
[16] -1.3 -0.6 -0.2 -0.4 -0.2  0.4 -1.0 -0.6 -0.6 -1.2
```

Test de Shapiro-Wilk sur d : pvalue = 0.36 ; l'hypothèse de normalité est validée

$$\bar{d}=\frac{1}{N}\sum_{i=1}^{N} d_i=-0.77$$

$$s_d^2=\frac{N}{N-1}\left(\frac{1}{N}\sum_{i=1}^{N} d_i^2-\overline{d}^2\right)=1.09$$

donc $s_d=1.045$

$$t=\frac{|\bar{d}|}{s_d/\sqrt{N}}=\frac{0.77}{1.045/\sqrt{25}}=3.69$$

à comparer avec $t_{24\,;\alpha}$

$t>t_{24\,;0.01}$ Donc on rejette H0 au risque de 1 % ; les deux méthodes ne sont pas équivalentes, p<0.01

Rem : L'ordinateur donnera pvalue = 0.001

Égalité des variances, homoscédasticité

L'hypothèse d'homoscédasticité ou d'égalité des variances entre deux échantillons se vérifie avec le **test de Fisher** (ou **test de Fisher-Snedecor**)

H0 : Les variances sont égales (homoscédacticité)

Supposons que la variance est estimée la plus grande est celle de l'échantillon 1, s_1^2
Ce test consiste à comparer le rapport des variance $F = s_1^2 / s_2^2$, sensé suivre la loi de Fisher-Snedecor.

“A la main”, une table permet de connaitre la valeur de F à ne pas dépasser en fonction de N_1 et N_2

Plus généralement (test fait sur ordinateur), on concluera à partir de la pvalue :

pvalue <0.05 : H0 rejetée, l'hypothèse d'homoscédasticité est rejetée
On effectuera un test de Welch

pvalue > 0.05 : Hypothèse d'homoscédasticité non contredite (donc conservée)
On effectuera un test t de Student

Table de Fisher

TABLE 1 $\alpha = .05$

Degrés de liberté pour le dénominateur	Degrés de liberté pour le numérateur														
	1	2	3	4	5	6	7	8	9	10	15	20	25	30	40
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.73	1.69	1.63
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.69	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.60	1.55	1.50
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.56	1.52	1.46
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.53	1.48	1.42
1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.68	1.58	1.52	1.47	1.41

Égalité des variances, homoscédasticité

On a aussi la possibilité d'essayer de transformer les données pour favoriser l'homoscédasticité, afin d'effectuer le test t de Student au lieu du test de Welch, et gagner ainsi (un peu) en puissance

Exemple :

X=c(2,3,69,6,19,1,9,19,95,4,4,21,47,2,29,41,7,2,73,5,23,32,11,55,0) → $S^2_X=689$; $N_1=25$

Y=c(110,0,33,7,0,30,77,6,36,52,150,40,1,14,52) → $S^2_y=1890$; $N_2=15$

D'où $F=\frac{\sigma_Y^2}{\sigma_X^2}=\frac{1890}{689}=2.74$ > 2.28 donc H0 rejetée : hétéroscédasticité

En transformant : $x \rightarrow \sqrt{1+x}$

Xtr=c(1.73, 2, 8.37, 2.65, 4.47, 1.41, 3.16, 4.47, 9.8, 2.24, 2.24, 4.69, 6.93, 1.73, 5.48, 6.48, 2.83, 1.73, 8.6, 2.45, 4.9, 5.74, 3.46, 7.48, 1) → $S^2_{Xtr}=6.42$

Ytr=c(10.54, 1, 5.83, 2.83, 1, 5.57, 8.83, 2.65, 6.08, 7.28, 12.29, 6.4, 1.41, 3.87, 7.28) → $S^2_{Ytr}=11.80$

D'où $F=\frac{\sigma_{Ytr}^2}{\sigma_{Xtr}^2}=\frac{11.80}{6.42}=1.84$ < 2.28
donc H0 non rejetée : homoscédasticité validée

Test t de Student

Rappel : Statistique $Z=\frac{|\overline{X}_1-\overline{X}_2|}{es_M}$ à comparer avec $t_{DL;\alpha}$

S'utilise lorsqu'on a vérifié les hypothèses :
- d'indépendance des données
- de normalité des données (pour chaque échantillon)
- d'homoscédasticité.

Les données étant indépendantes, la variance de la différence M est la somme des variances : $es_M^2=\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}$

Les variances étant "égales", on estimera la variance commune par : $\boxed{s^2=\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{(N_1-1)+(N_2-1)}}$

Ainsi, le **test t de Student** consistera à calculer la statistique $\boxed{t=\frac{|\overline{X}_1-\overline{X}_2|}{s\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}}}$

et à comparer le résultat avec $t_{DL;\alpha}$ pour $DL=N_1+N_2-2$

Test t de Student : Exemple

```
> Cyano
 [1] 43 98 45  3 29 46 11 41 43  5 19 57 13 48  3 65 49 61 69 10 64 38 29 18 13 71 81 81 42 84
> Control
 [1] 31 37 41 92 19 58 14 93 46 76 31 29 81  4 88 36 94 96 58 14 82 47 66 27 82 62 73  6 18 24
```

H0 : le rendement moyen est identique pour les parcelles polluées et pour les parcelles non polluées

Test de Shapiro sur Cyano : pvalue = 0.27 → Normalité des données validée pour Cyano

Test de Shapiro sur Control : pvalue = 0.06 → Normalité des données validée pour Control

Test de Fisher : pvalue = 0.60 → Homoscédasticité validée

Cyano : $N_1=30$; $\overline{X}_1=42.6$; $s^2_1 = 715.5$

Control : $N_2=30$; $\overline{X}_2=50.8$; $s^2_2 = 872.4$

Variance : $$s^2=\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{(N_1-1)+(N_2-1)}=\frac{29\,x\,715.5+29\,x\,872.4}{29+29}=793.9$$

D'où : $$t=\frac{|\overline{X}_1-\overline{X}_2|}{s\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}}=\frac{|42.6-50.8|}{793.9\,s\sqrt{\frac{1}{29}+\frac{1}{29}}}=1.13$$ à comparer avec $t_{N1+N2-1;\alpha}=t_{58;\alpha}$

Table de la loi de Student

ν \ α	0,05	0,01	0,001
1	12,61	63,66	636,6
2	4,303	9,925	31,60
3	3,182	5,841	12,94
4	2,776	4,604	8,610
5	2,571	4,032	6,859
6	2,447	3,707	5,959
7	2,365	3,499	5,405
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
11	2,201	3,106	4,437
12	2,179	3,055	4,318
13	2,160	3,012	4,221
14	2,145	2,977	4,140
15	2,131	2,947	4,073
16	2,120	2,921	4,015
17	2,110	2,898	3,965
18	2,101	2,878	3,922
19	2,093	2,861	3,883
20	2,086	2,845	3,850
21	2,080	2,831	3,819
22	2,074	2,819	3,792
23	2,069	2,807	3,767
24	2,064	2,797	3,745
25	2,060	2,787	3,725
26	2,056	2,779	3,707
27	2,052	2,771	3,690
28	2,048	2,763	3,674
29	2,045	2,756	3,659
30	2,042	2,750	3,646
35	2,030	2,724	3,591
40	2,021	2,704	3,551
45	2,014	2,690	3,520
50	2,008	2,678	3,496
55	2,004	2,669	3,476
60	2,000	2,660	3,460
70	1,994	2,648	3,435
80	1,990	2,639	3,415
90	1,986	2,631	3,402
100	1,984	2,626	3,389
120	1,980	2,617	3,373
200	1,972	2,601	3,339
500	1,965	2,586	3,310
∞	1,960	2,576	3,291

$t = 1.13 < t_{58\,;\alpha}$ quel que soit le seuil retenu

Le test est donc non significatif ; on ne rejette pas H0 ; on ne met pas en évidence de différence de rendement entre les deux échantillons

Rem : L'ordinateur donnera pvalue = 0.26

Test de Welch

Rappel : Statistique $Z=\frac{|\overline{X}_1-\overline{X}_2|}{es_M}$ à comparer avec $t_{DL;\alpha}$

S'utilise lorsqu'on a vérifié les hypothèses :

- d'indépendance des données
- de normalité des données (pour chaque échantillon)
- mais pas d'homoscédasticité.

Les données étant indépendantes, la variance de la différence M est la somme des variances : $es_M^2=\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}$

Ainsi, le **test de Welch** consistera à calculer la statistique

$$\boxed{t=\frac{|\overline{X}_1-\overline{X}_2|}{\sqrt{\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}}}}$$

et à comparer le résultat avec $t_{DL;\alpha}$

Cependant un biais est à corriger, qui amène à considérer

$$DL=\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{1}{N_1-1}\left(\frac{s_1^2}{N_1}\right)^2+\frac{1}{N_2-1}\left(\frac{s_2^2}{N_2}\right)^2}$$

Test de Welch : exemple

On a mesuré 25 pins sylvestres par la méthode trigonométrique et obtenus les résultats suivants (en m) :

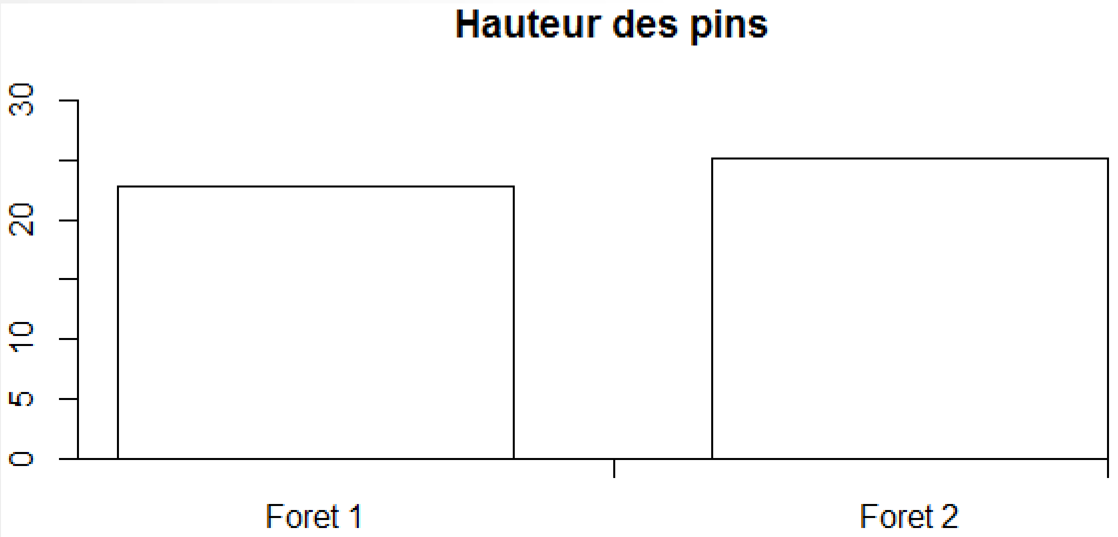
X_1=c(21.1,25.2,28.7,19.1,25.6,17.7,17.4,23.8,22.1,16.4,22.1,16.1,22.6,27.2,19.5,26.4,20.8,27.9,28.6, 25.3,18.8,30.2,21.2,18.8,25.9)

Rappel : N_1=25 $\overline{X_1}=22.7$ $s_1^2=17.50$

pvalue du Test de Shapiro = 0.36 donc hypothèse de normalité validée

On a mesuré par la même méthode 35 pins d'une autre forêt :

X_2=c(37.7,34.1,22.9,20,15.4,26.6,19,24.1,13.8,24.7,18.8,26.6,26.4,26.6,24.7,24.8,13.2,29.6,14,31.7, 32.7,33.6,20.4,29.6,27.1,20.3,26.5,24.5,28.3,28,23.6,29.5,37.9,20.7,27)



H0 : les hauteurs des pins dans les deux forêts sont indentiques

Test de Welch : exemple

X_1=c(21.1,25.2,28.7,19.1,25.6,17.7,17.4,23.8,22.1,16.4,22.1,16.1,22.6,27.2,19.5,26.4,20.8,27.9,28.6, 25.3,18.8,30.2,21.2,18.8,25.9)

X_2=c(37.7,34.1,22.9,20,15.4,26.6,19,24.1,13.8,24.7,18.8,26.6,26.4,26.6,24.7,24.8,13.2,29.6,14,31.7, 32.7,33.6,20.4,29.6,27.1,20.3,26.5,24.5,28.3,28,23.6,29.5,37.9,20.7,27)

Échantillons indépendants

pvalue du Test de Shapiro pour X_1 = 0.36 donc hypothèse de normalité validée

pvalue du Test de Shapiro pour X_2 = 0.56 donc hypothèse de normalité validée

pvalue du Test Fisher = 0.04 donc hypothèse d'homoscédasticité rejetée

On a : $N_1=25$ $\quad \overline{X_1}=22.7 \quad s_1^2=17.50$

$N_2=35$ $\quad \overline{X_2}=25.3 \quad s_2^2=39.3$

$$t=\frac{|\overline{X_1}-\overline{X_2}|}{\sqrt{\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}}}=\frac{|22.7-25.3|}{\sqrt{\frac{17.5}{25}+\frac{39.3}{35}}}=1.92$$

à comparer avec $t_{DL,\alpha}$

$$DL=\frac{\left(\frac{s_1^2}{N_1}+\frac{s_2^2}{N_2}\right)^2}{\frac{1}{N_1-1}\left(\frac{s_1^2}{N_1}\right)^2+\frac{1}{N_2-1}\left(\frac{s_2^2}{N_2}\right)^2}=...=57.79$$

Table de la loi de Student

ν \ α	0,05	0,01	0,001
1	12,61	63,66	636,6
2	4,303	9,925	31,60
3	3,182	5,841	12,94
4	2,776	4,604	8,610
5	2,571	4,032	6,859
6	2,447	3,707	5,959
7	2,365	3,499	5,405
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
11	2,201	3,106	4,437
12	2,179	3,055	4,318
13	2,160	3,012	4,221
14	2,145	2,977	4,140
15	2,131	2,947	4,073
16	2,120	2,921	4,015
17	2,110	2,898	3,965
18	2,101	2,878	3,922
19	2,093	2,861	3,883
20	2,086	2,845	3,850
21	2,080	2,831	3,819
22	2,074	2,819	3,792
23	2,069	2,807	3,767
24	2,064	2,797	3,745
25	2,060	2,787	3,725
26	2,056	2,779	3,707
27	2,052	2,771	3,690
28	2,048	2,763	3,674
29	2,045	2,756	3,659
30	2,042	2,750	3,646
35	2,030	2,724	3,591
40	2,021	2,704	3,551
45	2,014	2,690	3,520
50	2,008	2,678	3,496
55	2,004	2,669	3,476
60	2,000	2,660	3,460
70	1,994	2,648	3,435
80	1,990	2,639	3,415
90	1,986	2,631	3,402
100	1,984	2,626	3,389
120	1,980	2,617	3,373
200	1,972	2,601	3,339
500	1,965	2,586	3,310
∞	1,960	2,576	3,291

$t = 1.92 < t_{58\,;\alpha}$ quel que soit le seuil retenu

Le test est donc non significatif ; on ne rejette pas H0 ; on ne met pas en évidence de différence de taille pour les pins des deux forêts

Rem : L'ordinateur donnera pvalue = 0.006

Test de Mann-Whitney

S'utilise pour des données indépendantes pour lesquelles l'hypothèse de normalité n'est pas validée. Ce test est aussi connu sous le nom de Mann-Whitney-Wilcoxon (à ne pas confondre avec le test de Wilcoxon pour données appariées).

Ce test est basé sur les rangs et compare en réalité plutôt les médianes que les moyennes
H0 : les échantillons sont identiquement positionnés.

Principe :

- On classe ensemble les valeurs provenant des deux échantillons, et on associe à chaque valeur son rang (=classement)
- En cas d'ex-aequo, chaque valeur se voit attribuer le rang moyen
- On calcule ensuite la somme des rangs des valeurs des échantillons : $U_i = \sum_{r \in E_i} r$

(U est connu sous le nom de statistique de Wilcoxon)

- On estime ensuite la statistique de Mann-Whitney : $W_i = U_i - \frac{N_i(N_i+1)}{2}$

Rem : Il suffit de calculer cette valeur pour 1 seul des 2 échantillons.

- On compare ensuite cette valeur à la valeur théorique $\frac{N_1 N_2}{2}$ si H0 est vraie

Test de Mann-Whitney

Pourquoi la valeur attendue si H0 est vraie est-elle $\frac{N_1 N_2}{2}$ *?*

La somme de tous les rangs vaut : $1+2+\cdots+(N_1+N_2)=\frac{(N_1+N_2)(N_1+N_2+1)}{2}$

Si H0 est vraie, la somme des rangs de l'échantillon i représentera donc une propotion de ce total, soit une valeur théorique $\tilde{U}_i=\frac{N_1}{N_1+N_2}\frac{(N_1+N_2)(N_1+N_2+1)}{2}=\frac{N_i(N_1+N_2+1)}{2}$

La valeur théorique de la satistique W est donc :

$$\tilde{W}_1=\tilde{U}_1-\frac{N_1(N_1+1)}{2}=\frac{N_1}{2}\left((N_1+N_2+1)-(N_1+1)\right)=\frac{N_1 N_2}{2}$$

$$\tilde{W}_2=\tilde{U}_2-\frac{N_2(N_2+1)}{2}=\frac{N_2}{2}\left((N_1+N_2+1)-(N_2+1)\right)=\frac{N_1 N_2}{2}$$

Test de Mann-Whitney

n_2	α	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	3	3
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	.05	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	.01	--	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	.05	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
	.01	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	.05	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	.01	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	.05	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	.01	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	.05	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	.01	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	.05	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	.01	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	.05	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	.01	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	.05	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	.01	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	.05	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	.01	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	.05	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	.01	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	.05	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	.01	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	.05	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	.01	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	.05	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
	.01	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

$N_1 = 10$

$N_2 = 7$

$\alpha = 0.05$

→ Le plus petit des deux W_i ne doit être <u>inférieur</u> à 14

Il s'agit bien ici des valeurs “plancher” autorisées et non pas de valeurs “plafonds” comme c'est le cas le plus souvent

Test de Mann-Whitney

Comment comparer W_i et $\frac{N_1 N_2}{2}$ *?*

- A la main ?

Il existe des tables donnant pour le plus petit des deux W_i la valeur à ne pas dépasser, en fonction de N_1, N_2 et du seuil de risque α choisi

Si N_1,et N_2 sont grands (pas dans la table !), on montre que la statistique Wi se rapproche d'une loi normale de moyenne $\tilde{W}=\frac{N_1 N_2}{2}$ et de variance $\tilde{\sigma}_W^2=\frac{N_1 N_2(N_1+N_2+1)}{12}$

On peut calculer la statistique $t_W=\frac{|\tilde{W}-W_i|}{\tilde{\sigma}_W}$ et la comparer avec les seuils de la loi NCR :

$\alpha=0.05$ ⟶ t_W ne doit pas dépasser 1.96

$\alpha=0.01$ ⟶ t_W ne doit pas dépasser 2.6

$\alpha=0.001$ ⟶ t_W ne doit pas dépasser 3.3

- Avec un ordinateur ? Il vous renvoie directement la pvalue, à comparer avec α

H0 rejetée si $\text{pvalue}<\alpha$

Test de Mann-Whitney

n_2	α	n_1																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	--	0	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	.01	--	0	0	0	0	0	0	0	0	1	1	1	2	2	2	2	3	3
4	.05	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	14
	.01	--	--	0	0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	.05	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	.01	--	--	0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	.05	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	.01	--	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	.05	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	.01	--	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	.05	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	.01	--	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	.05	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
	.01	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	.05	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	.01	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	.05	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	.01	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	48
12	.05	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	.01	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	.05	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	.01	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	.05	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	.01	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	.05	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	.01	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	.05	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	.01	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	.05	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	.01	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	.05	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	.01	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	.05	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	.01	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	.05	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
	.01	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105

$N_1 = 6$

$N_2 = 9$

Test de Mann-Whitney : exemple

E1=c(2,3,5,5,7,8) → $N_1 = 6$

E2=c(2,4,5,6,6,7,8,9,10) → $N_2 = 9$

Classement : 2 2 3 4 5 5 5 6 6 7 7 8 8 9 10

rang : 1.5 3 6 6 10.5 12.5 → $U_1 = 39.5$

1.5 3 4 6 8.5 8.5 10.5 12.5 14 15 → $U_2 = 80.5$

Donc : $W_1 = 39.5 - 6 * 7 /2 = 18.5$

$W_2 = 80.5 - 9*10/2 = 35.5$

Regardons dans la table pour N1=6 et N2=9

Le minimum des deux Wi est supérieur au min « autorisé » au risque α= 0.05 (et donc a fortiori celle pour α=0.01)

On ne peut donc pas rejeter H0 et on ne montre pas de différence entre les deux échantillons

Rem : L'ordinateur donnera pvalue = 0.31

Test de Wilcoxon

S'utilise pour des données appariées pour lesquelles l'hypothèse de normalité n'est pas validée. Ce test est aussi connu sous le nom de test des rangs signés de Wilcoxon

Ce test est basé sur les rangs et compare en réalité plutôt les médianes que les moyennes
H0 : les échantillons sont identiquement positionnés.

Principe :

- On calcule les valeurs de $d=X_1-X_2$ pour chaque paire (N paires)

- On classe les valeurs de |d| et on associe à chaque valeur son rang (=classement). S'il y a des 0, on n'en tient pas compte.
- En cas d'ex-aequo, chaque valeur se voit attribuer le rang moyen
- On calcule ensuite la somme des rangs respectivements des valeurs positives et négatives : $V_+=\sum_{d_i>0} r$ $\quad V_-=\sum_{d_i<0} r$

Rem : Il suffit de calculer cette valeur pour 1 seul des 2 échantillons.
- On compare ensuite cette valeur à la valeur théorique $\frac{N(N+1)}{4}$ si H0 est vraie

Test de Wilcoxon

Pourquoi la valeur attendue si H0 est vraie est-elle $\frac{N(N+1)}{4}$ *?*

La somme de tous les rangs vaut : $1+2+\cdots+N=\frac{N(N+1)}{2}$

Si H0 est vraie, les valeurs de d positives et négatives sont équitablement réparties, on s'attend donc a ce que de $V_{+}=V_{-}=\frac{N(N+1)}{4}$

Test de Wilcoxon

Comment comparer V_{+ou-} *et* $\frac{N(N+1)}{4}$ *?*

- A la main ?

 Il existe des tables donnant pour le plus petit des deux V (=V_+ ou V_-) la valeur à ne pas dépasser, en fonction de N et du seuil de risque α choisi

 Si N est grand (pas dans la table !), on montre que la statistique V se rapproche d'une loi normale de moyenne $\tilde{V}=\frac{N(N+1)}{4}$ et de variance $\tilde{\sigma}_V^2=\frac{N(N+1)(2N+1)}{24}$

On peut calculer la statistique $t_V=\frac{|\tilde{V}-V|}{\tilde{\sigma}_V}$ et la comparer avec les seuils de la loi NCR :

α=0.05 → t_V ne doit pas dépasser 1.96

α=0.01 → t_V ne doit pas dépasser 2.6

α=0.001 → t_V ne doit pas dépasser 3.3

- Avec un ordinateur ? Il vous renvoie directement la pvalue, à comparer avec α

 H0 rejetée si pvalue<α

Test de Wilcoxon

Test bilatéral		
n	risque 5%	risque 1%
6	0	
7	2	
8	3	0
9	5	1
10	8	3
11	10	5
12	13	9
13	17	9
14	21	12
15	25	15
16	29	19
17	34	23
18	40	27
19	46	32
20	52	37
21	59	43
22	66	49
23	73	55
24	81	61
25	89	68

Il s'agit bien ici “également des valeurs “plancher” autorisées et non pas de valeurs “plafonds” comme c'est le cas le plus souvent

Test de Wilcoxon : exemple

E1=c(5,9,3,8,12,14,4)
E2=c(11,1,6,7,13,10,2) → $N_1 = 7$

d=E1 - E2	-6	8	-3	1	-1	4	2
|d|	6	8	3	1	1	4	2
Rangs :	6	7	4	1.5	1.5	5	3
signes :	-	+	-	+	-	+	+

Donc :

$V_- = 6+4+1.5 = 11.5$ $V_+ = 7+1.5+5+3 = 16.5$

Test bilatéral		
n	risque 5%	risque 1%
6	0	
7	2	
8	3	0
9	5	1
10	8	3
11	10	5
12	13	9
13	17	9
14	21	12
15	25	15
16	29	19
17	34	23
18	40	27
19	46	32
20	52	37
21	59	43
22	66	49
23	73	55
24	81	61
25	89	68

Regardons dans la table pour N=7
Au risque de 5 %, la plus petite des 2 valeurs ne doit pas être inférieure à 2

On ne rejette donc pas H0 et on ne met pas en évidence de différence entre les 2 échantillons)

Rem : L'ordinateur donnera pvalue = 0.67

Généralisation a plus de 2 moyennes ?

Comment comparer k moyennes ? (k>2)

Il existe des tests (paramétriques et non paramétriques) dédiés pour comparer simultanément k moyennes (cf an prochain, conditions d'applications plus complexes) : Anova (test paramétrique), Test de Kruskal-Wallis (non paramétrique), anovar, Friedman, ...

Toutefois, on peut déjà avec les tests à notre disposition comparer 2 à 2 les moyennes (ce qui fait $p=k(k+1)/2$ comparaisons) avec les tests à notre disposition.

Dans ce cas, il faudra adapter le seuil de risque, en appliquant la correction de Bonferonni : le seuil de risque sera à diviser par p, pour que globalement le risque reste maitrisé