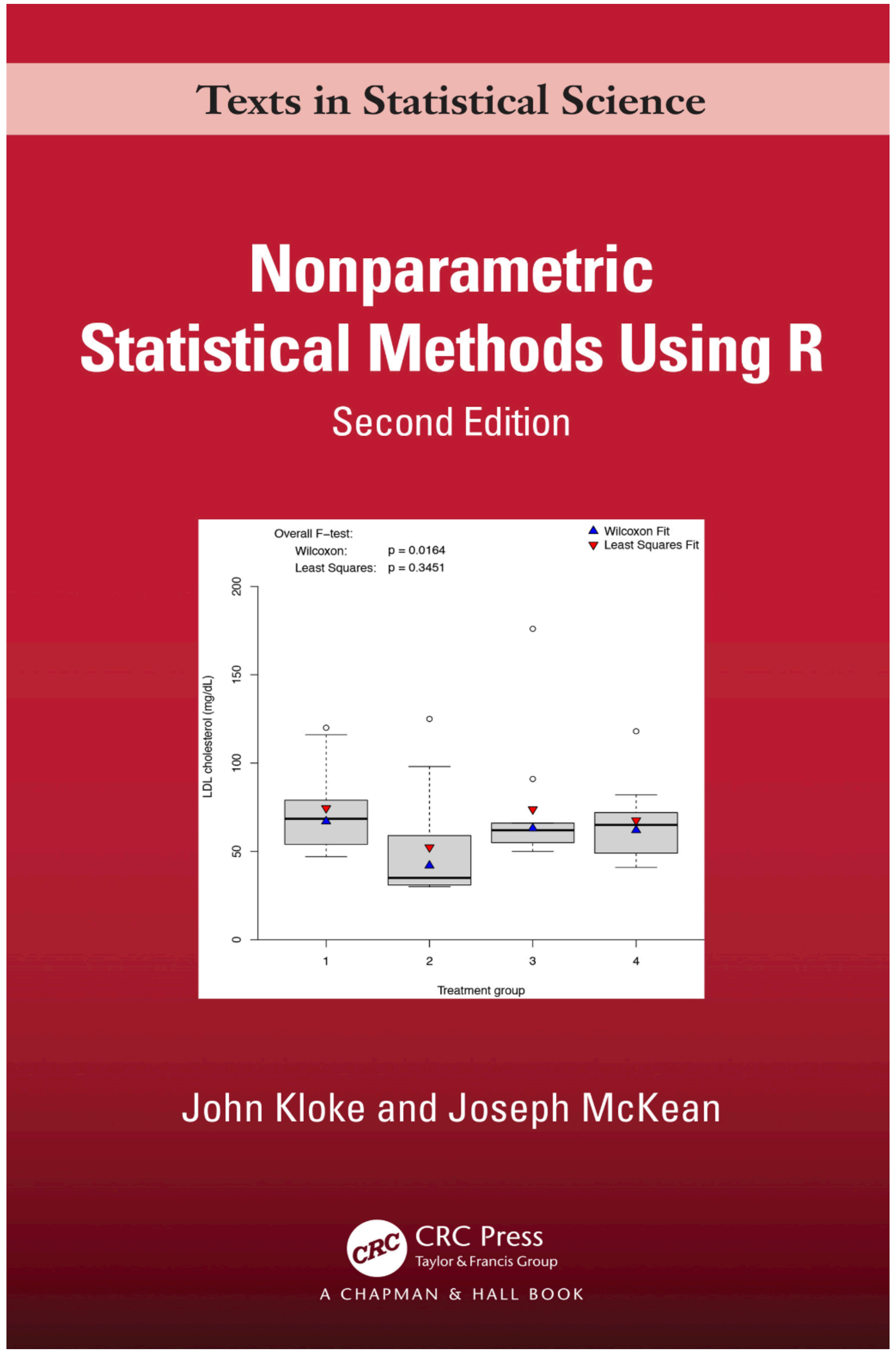
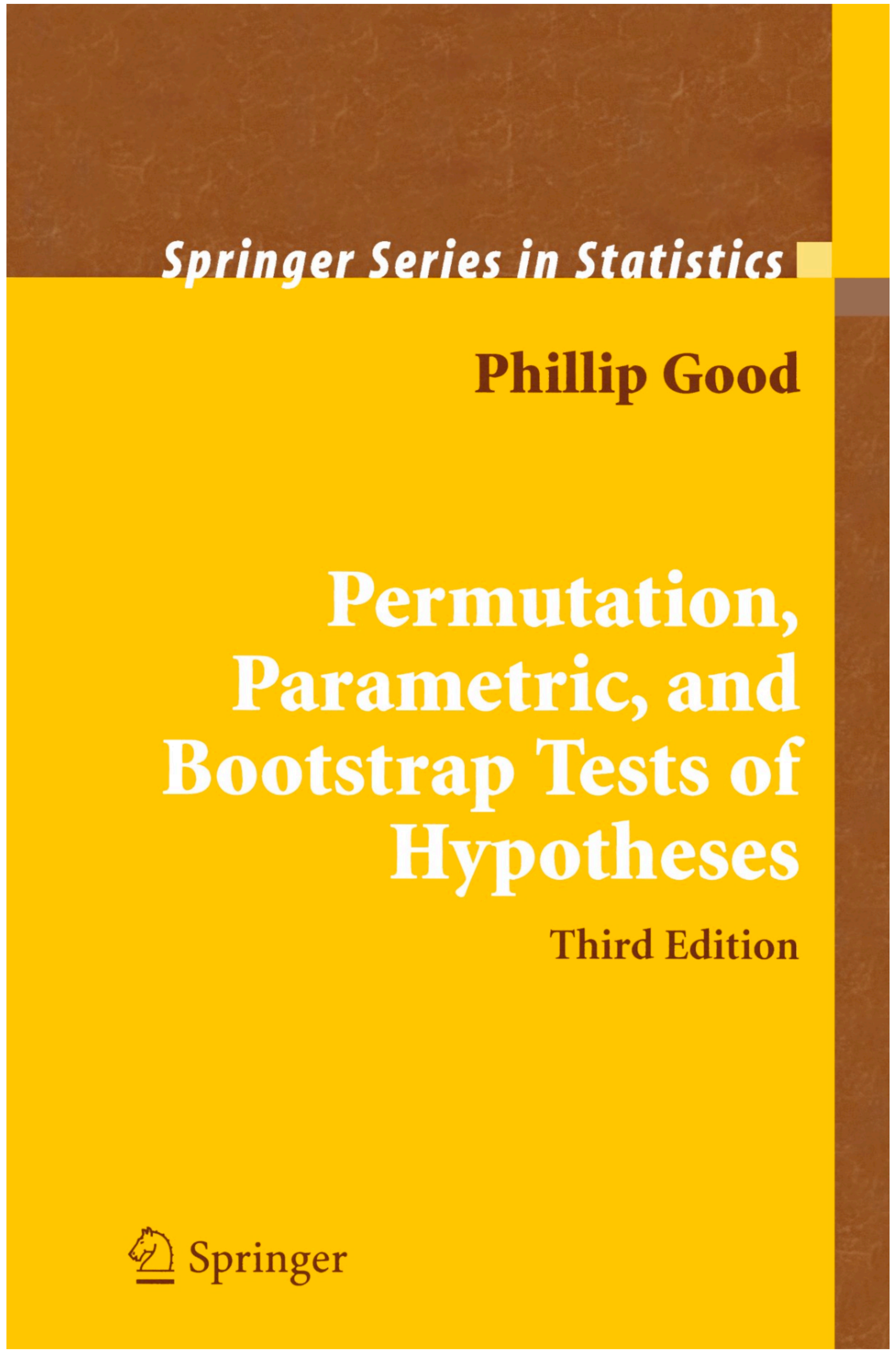
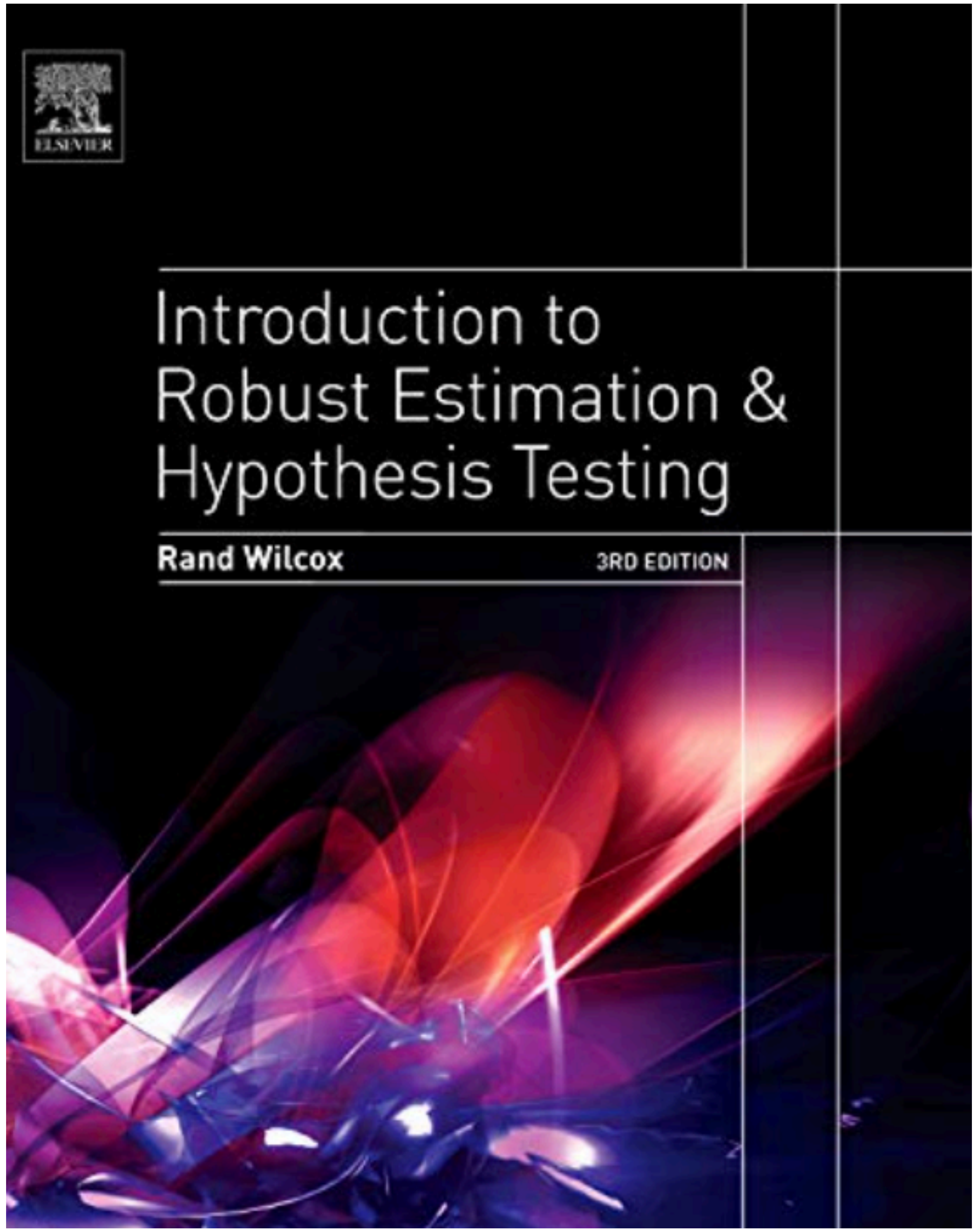
Tests non paramétriques

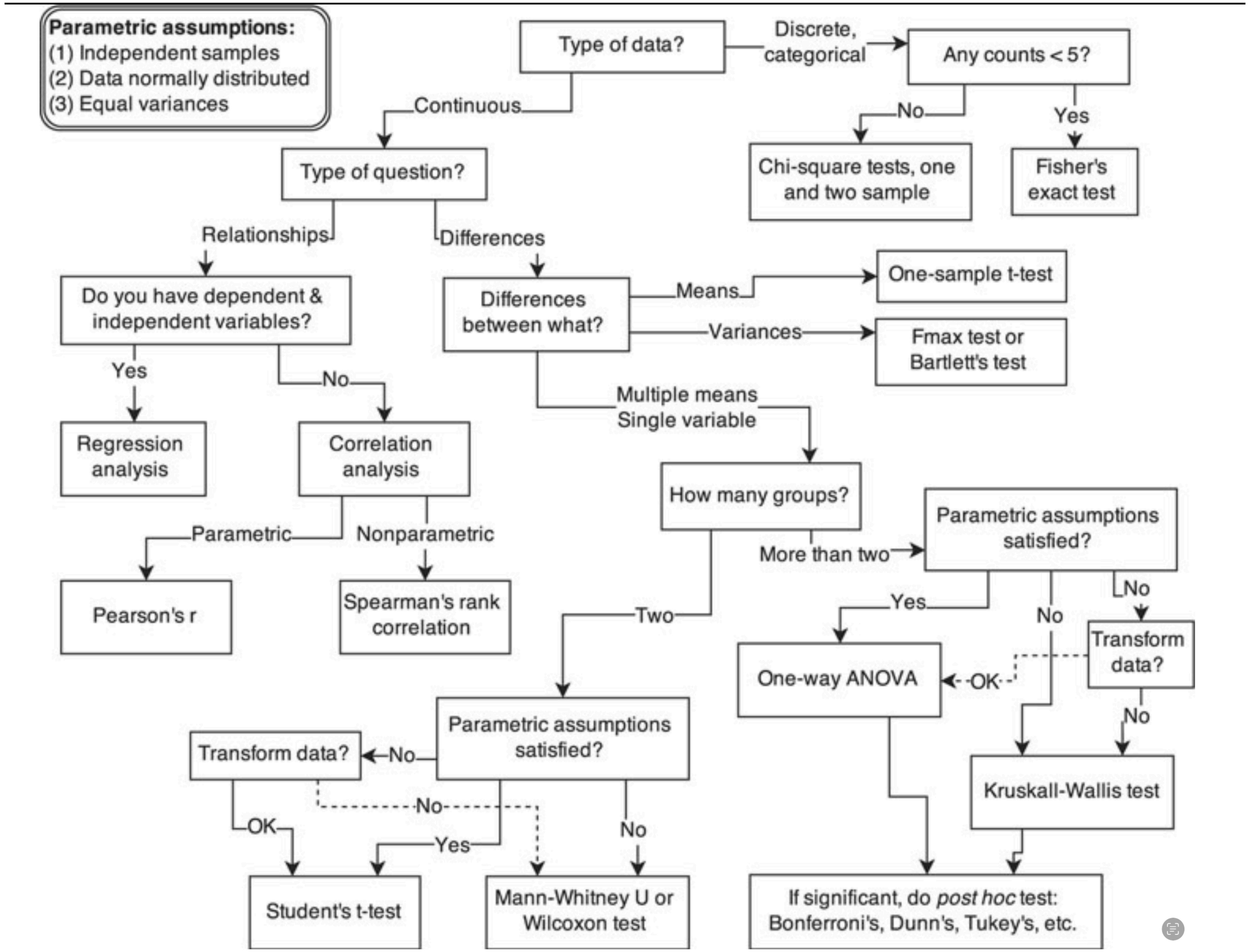
GALHARRET J-M,
MC Science des données
ONIRIS VetAGroBio


Texts in Statistical Science
Nonparametric Statistical Methods Using R
Second Edition
Overall F-test:
Wilcoxon: p = 0.0164
Least Squares: p = 0.3451
Wilcoxon Fit
Least Squares Fit
LDL cholesterol (mg/dL)
Treatment group
John Kloke and Joseph McKean
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK


Springer Series in Statistics
Phillip Good
Permutation, Parametric, and Bootstrap Tests of Hypotheses
Third Edition
Springer


ELSEVIER
Introduction to Robust Estimation & Hypothesis Testing
Rand Wilcox
3RD EDITION

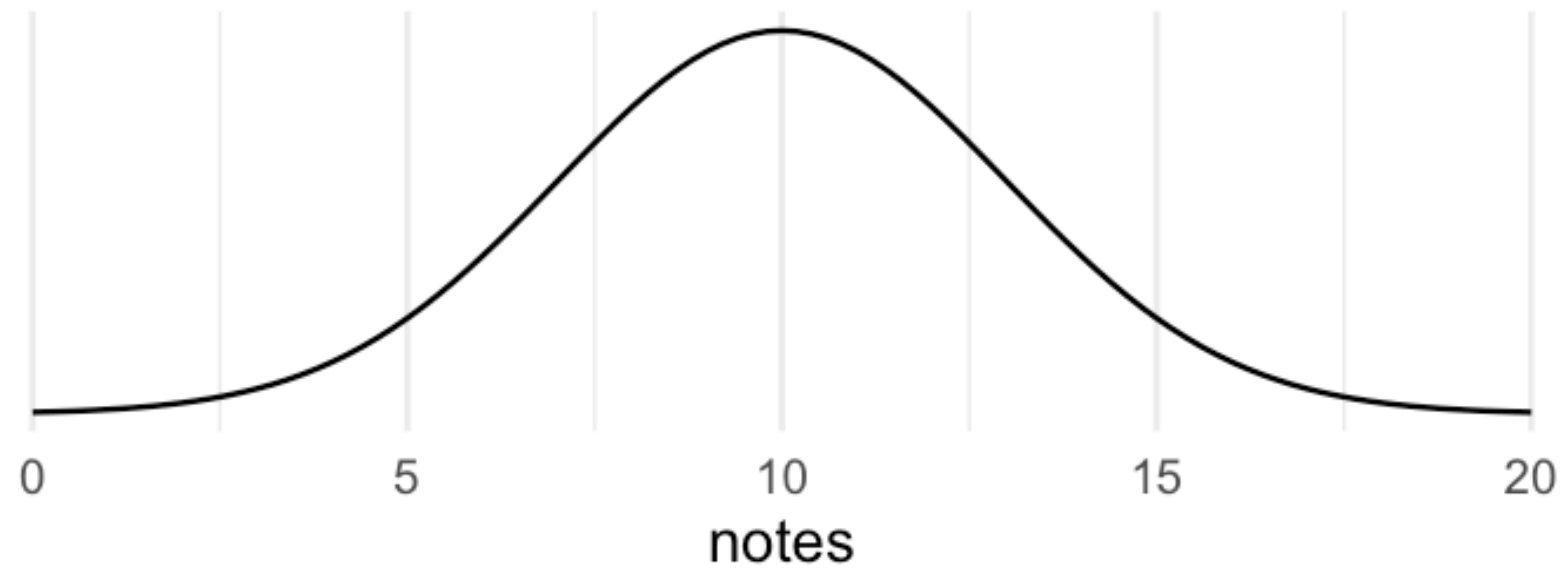
Tests paramétriques versus tests non paramétriques



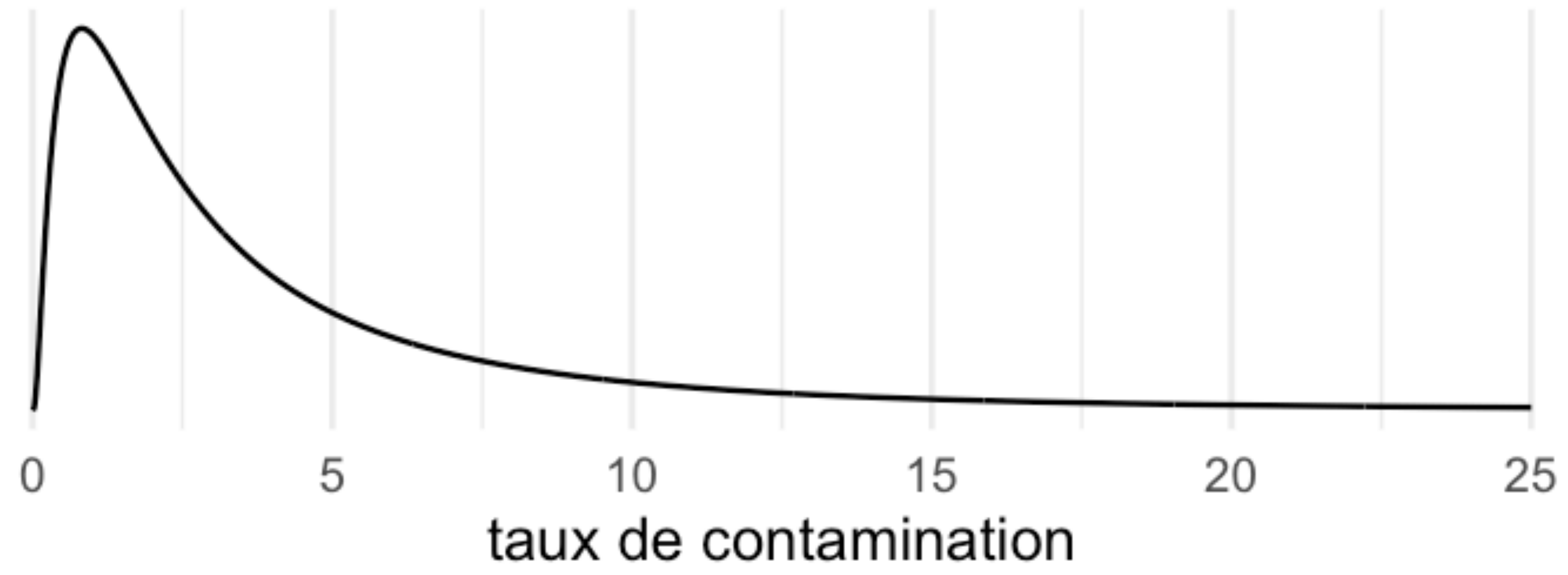
Distribution de probabilité
Fonction de répartition

Variables aléatoires continues

- Soit X une variable aléatoire mesurée sur une échelle continue. Par exemple un taux de contamination entre 0 % et 100%, une note entre 0 et 20.
- Question : comment caractériser cette variable ?
- Cas simple on connait la distribution de probabilité de cette variable



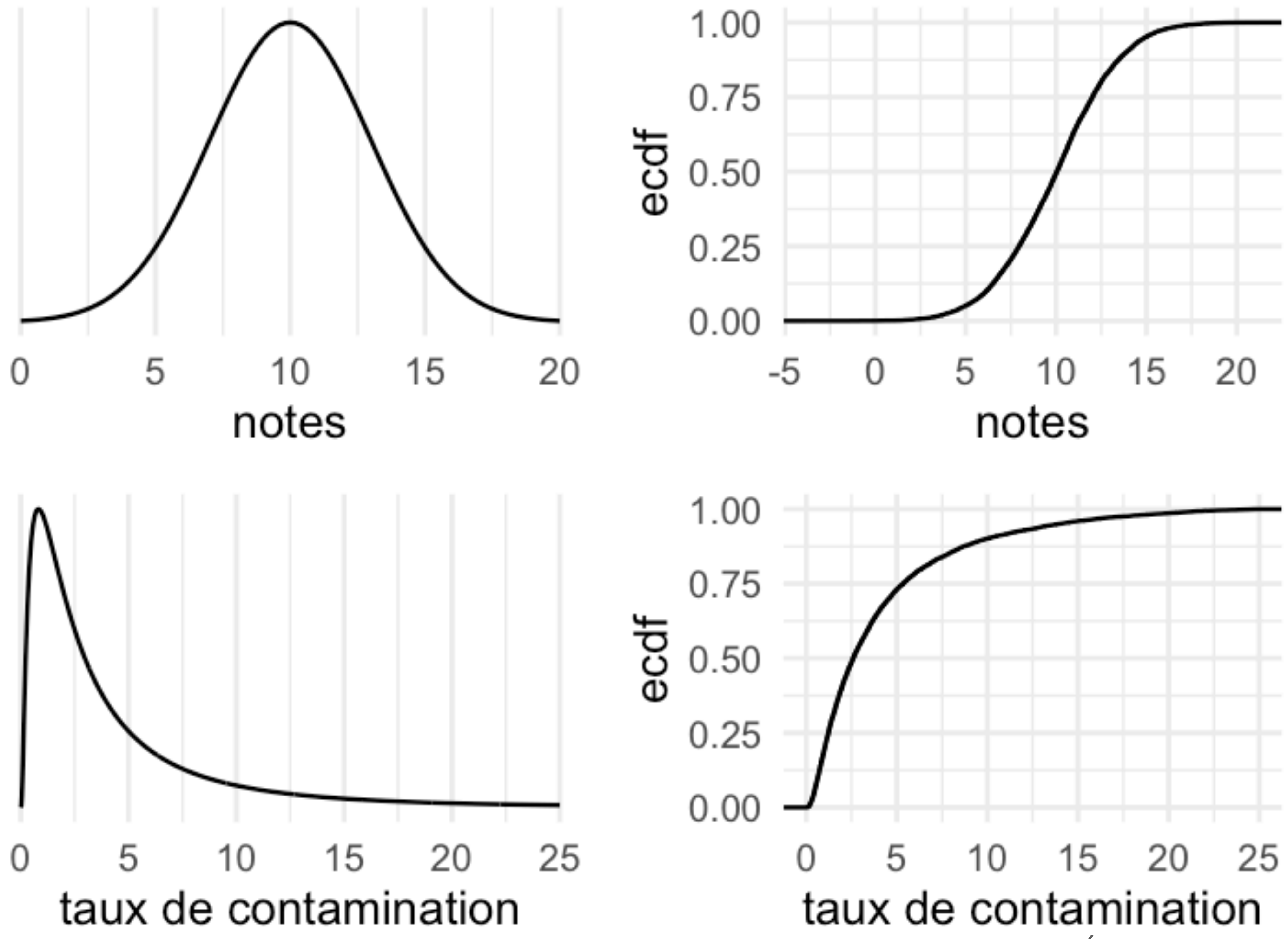
Mode, médiane, moyenne faciles à trouver
Même l'écart type est facile à trouver



Mode facile à trouver.
Médiane, moyenne difficiles à trouver

Autre représentation

Fonction de répartition

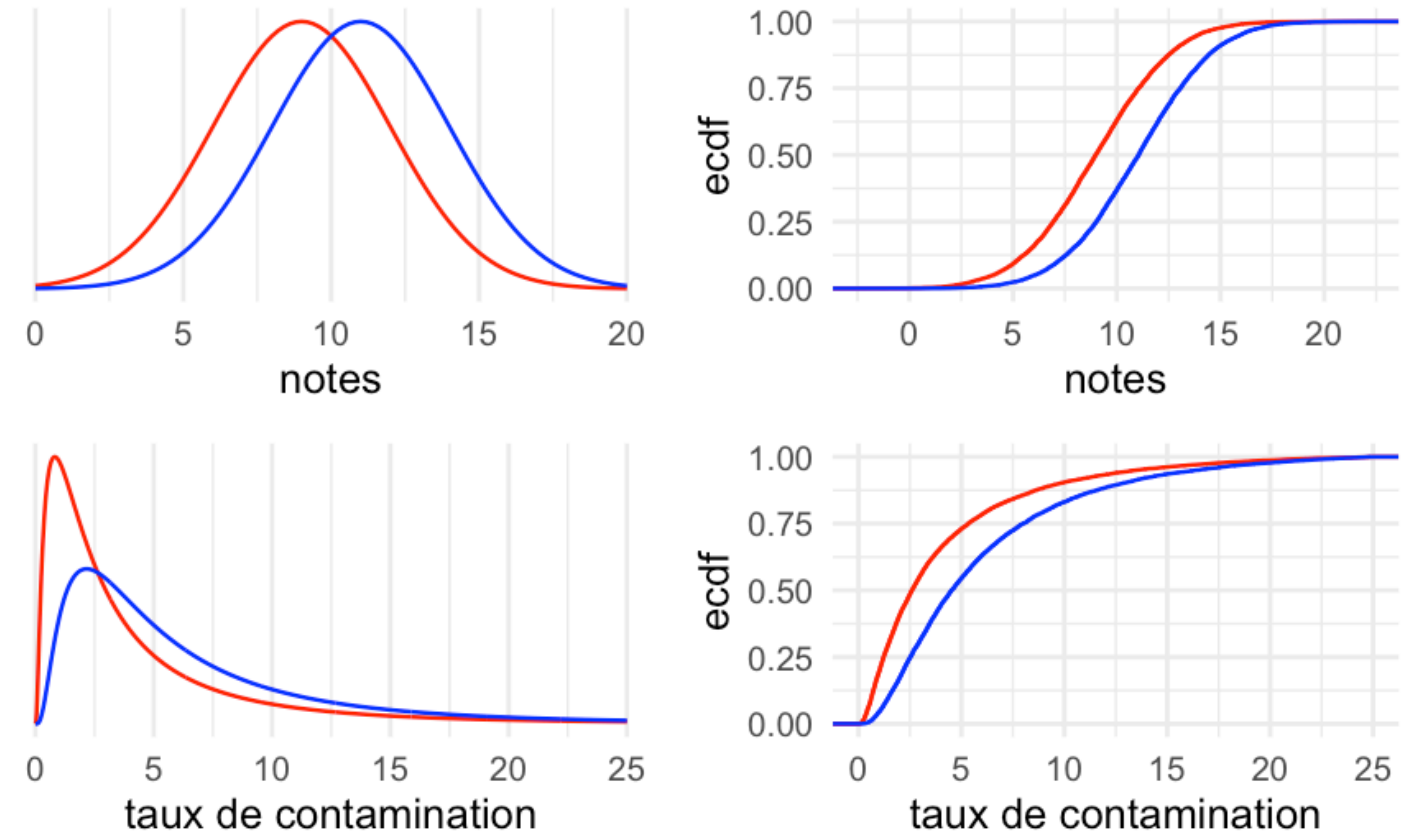


La fonction de répartition d'une variable continue X est définie par :

$$F(x) = \mathbb{P}(X < x)$$

Question : quels indices de la distribution sont visibles sur la fonction de répartition ?

Comparaison de variables aléatoires



Dans les deux situations les deux variables aléatoires ont le même écart type et une différence de moyenne égale à 2

Les tests statistiques

Quelques notions

- Un test statistique a pour objectif de choisir entre deux hypothèses l'une dite nulle H0 contre le contraire H1 de celle-ci (appelée hypothèse alternative).
- Cette décision est basée sur un échantillon d'observations et donc compte tenu de la fluctuation d'échantillonnage on est obligé de se fixer un risque d'erreur α appelé niveau de significativité du test

$$\alpha = P(\text{Rejet de } H_0 \,|\, H_0 \text{ est vraie})$$

- Dans la procédure des tests on construit un paramètre appelé statistique du test dont on connait la distribution sous H0.
- On rejette H0 si la p-value du test est inférieur à α.

		Réalité (jamais connue)	
		H0 est vraie	H0 est fausse
Décision à l'issue du test	Non rejet de H0	$1-\alpha$ VRAI NEGATIF	β FAUX NEGATIF
	Rejet de HO	α FAUX POSITIF	$1-\beta$ VRAI POSITIF

Puissance d'un test statistique

- Bonne nouvelle pour le statisticien : la puissance de la plupart des tests (la proportion de vrais positifs) tend vers 1 lorsque le nombre d'observation tend vers l'infini !!!! C'est à dire que lorsque n tend vers l'infini la p-value tend vers 0
- Pourquoi c'est une mauvaise nouvelle pour le praticien ????

  - Cela signifie qu'un écart aussi infime soit il pourra toujours être détecté à supposer que l'on prenne un nombre d'observations suffisamment grand.
  - Cohen[1] (1988) a introduit la taille d'effet pour quantifier cet écart (indépendamment de la taille de l'échantillon).

[1]Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Lawrence Erlbaum Associates.

Comment savoir si un test est bon ???

- Un test statistique est basé sur des conditions d'application : que se passe-t-il lorsque ces conditions ne sont plus vérifiées ?
  - Pour comparer, par un test de Student, une variable sur deux populations indépendantes on suppose que les distributions de cette variable sur ces deux populations sont normales et qu'elles ont la même variance.
  - Sous ces deux conditions on sait que le niveau du test est bien égal à 5% et que sa puissance tend vers 1 lorsque l'échantillon à un nombre infini de valeurs.
  - Que se passerait-il sur ces deux critères si l'une (ou les deux) conditions ne sont plus vérifiées ?
- En pratique pour vérifier ces deux conditions on utilise des simulations de Monte Carlo.

Principe des simulations de Monte Carlo

La plus simple : si on considère des variables aléatoires $X_1, \ldots, X_n$ iid ayant la même loi de moyenne M et d'écart type S. Alors les estimateurs de M et de S par la méthode de Monte Carlo sont

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

$$S_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_n)^2$$

Cette est une conséquence de la loi des grands nombres.

On va appliquer ce principe pour estimer :

- Le niveau empirique du test (nombre moyen de fois que l'on rejette H0 alors qu'il est vrai)
- Sa puissance empirique (nombre de fois où l'on rejette H0 alors qu'il est faux).

Estimation des p-values sous H0

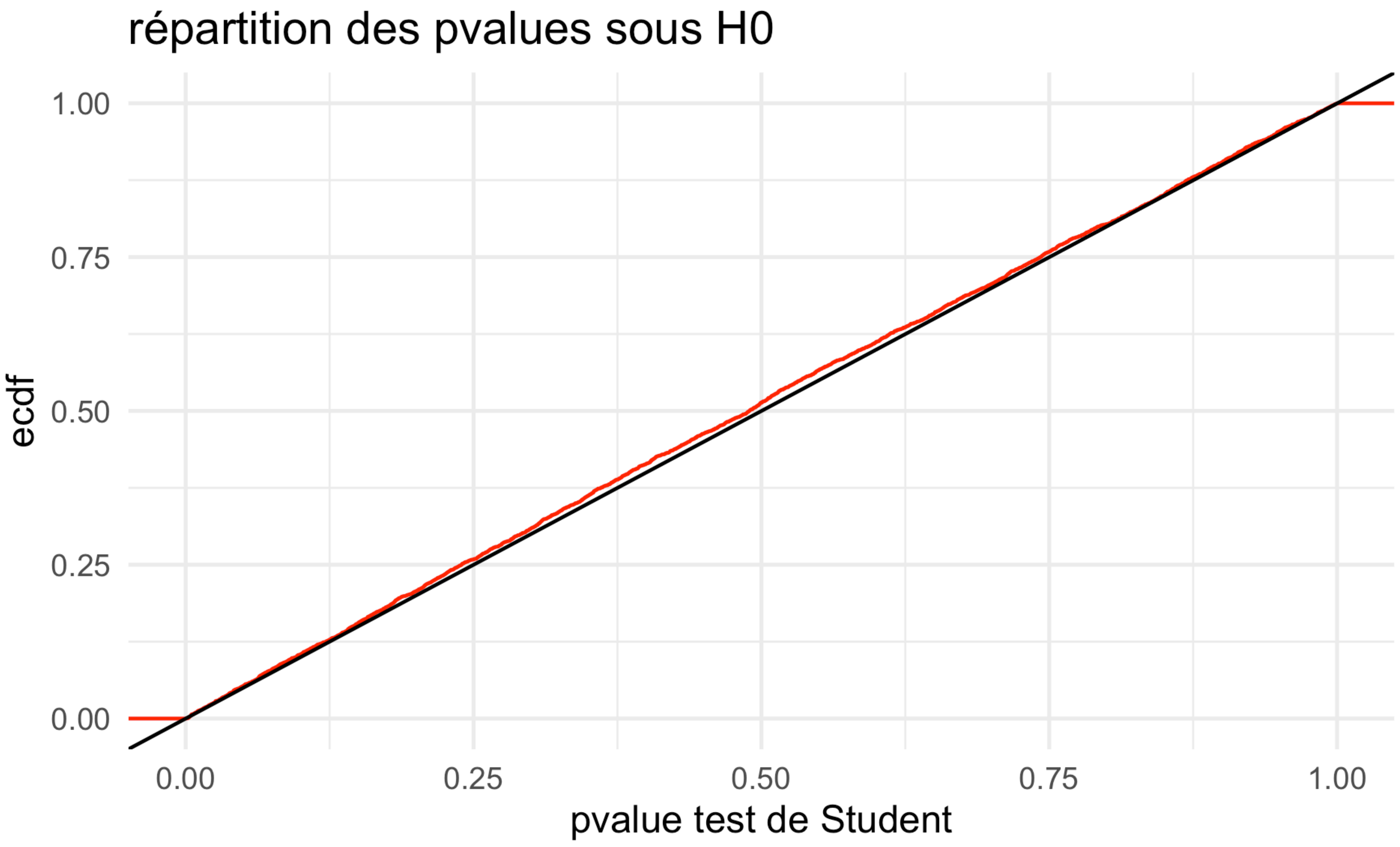


Schéma de simulations :

$X_1 \sim \mathcal{N}(10,2), \quad X_2 \sim \mathcal{N}(10,2)$

On prend considère B=5000 échantillons de tailles respectives n1=10 et n2=20.

Confirmation par les simulations de ce que l'on savait déjà :

Les p-values sont distribuées uniformément sous H0

```
n1=10
n2=20
B=5000
pval=NULL
for(i in 1:B){
  X1=rnorm(n1,10,2)
  X2=rnorm(n2,10,2)
  pval=c(pval,t.test(X1,X2,var.equal = T)$p.value)
}
```

Calcul de rangs

- Une partie des tests non paramétriques est basée sur des calculs de rang.
- On considère n observations $x_1, \ldots, x_n$, les rangs de ces observations sont notés $R(x_1), \ldots, R(x_n)$ et définis par

$$R(x_i) = \#\{j = 1,\ldots n \ : \ x_j \leq x_i\}$$

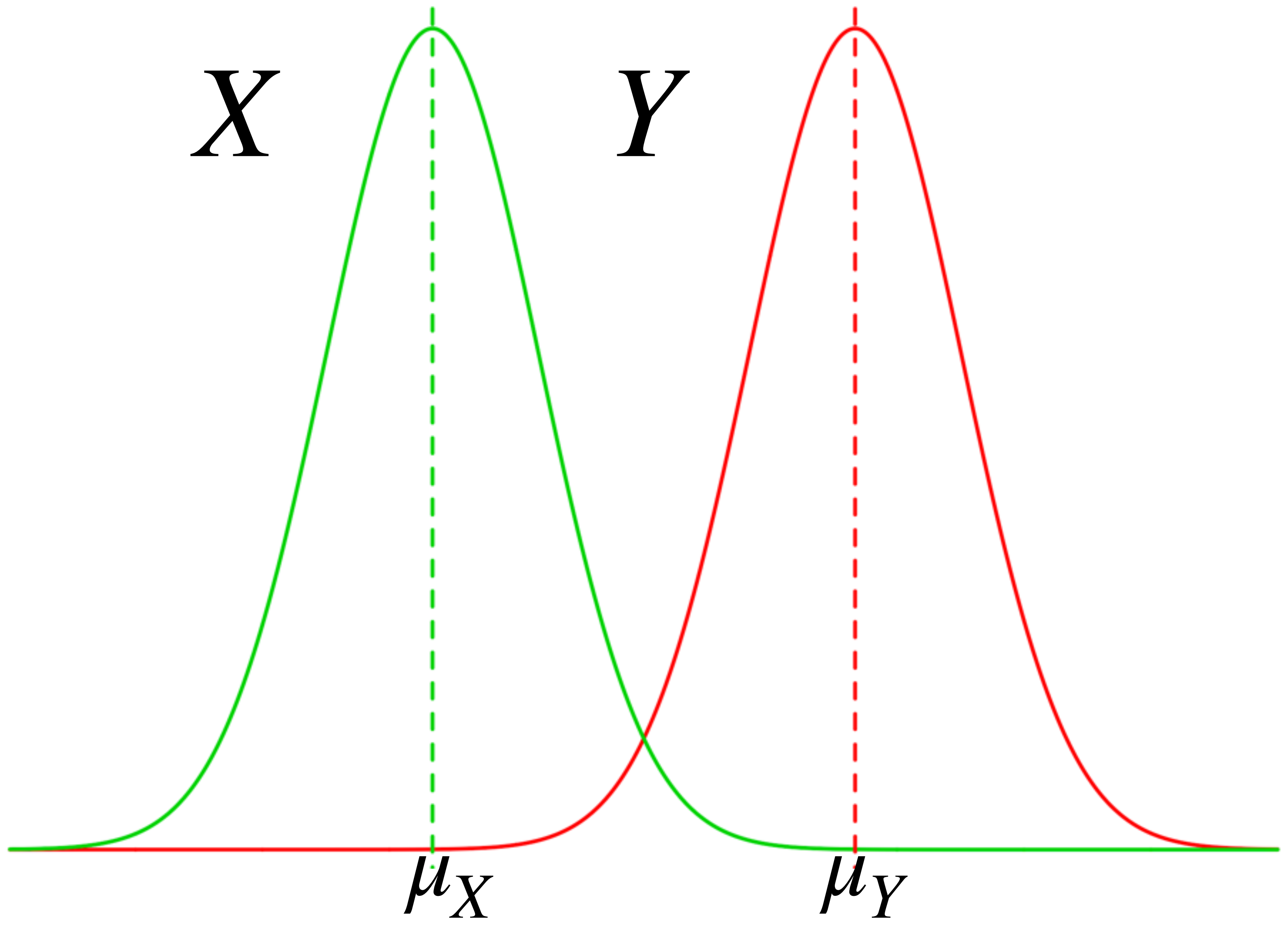
(ie le nombre d'observations inférieures ou égales à x_i.

- En cas d'ex-aequo on attribue à chaque la valeur le rang qu'ils auraient obtenus si ils n'avaient pas été ex-aequo.
- Dans R on utilise la fonction rank

```
x<-c(12,10,8,15,17)
rank(x)
[1] 3 2 1 4 5
x<-c(12,10,12,15,17,10,10)
rank(x)
[1] 4.5 2.0 4.5 6.0 7.0 2.0 2.0
```

Comparaison d’indices de position
(deux groupes indépendants)

Les tests paramétriques



$$H_0 : \mu_X = \mu_Y$$

- Soit $X_1, \ldots, X_{n_1}$ un échantillon aléatoire distribué à partir d'une fonction de répartition F normale de paramètres μ_X, σ_X .
- Soit $Y_1, \ldots, Y_{n_2}$ un échantillon aléatoire distribué à partir d'une fonction de répartition G normale de paramètres μ_Y, σ_Y.

- Ces deux variables ont la même variance : on utilise le **test de Student**.
- Ces deux variables n'ont pas la même variance on utilise le **test de Welch** (correction du test de Student).

Etude d’un exemple sur R : on charge les data neuropathy du package coin :

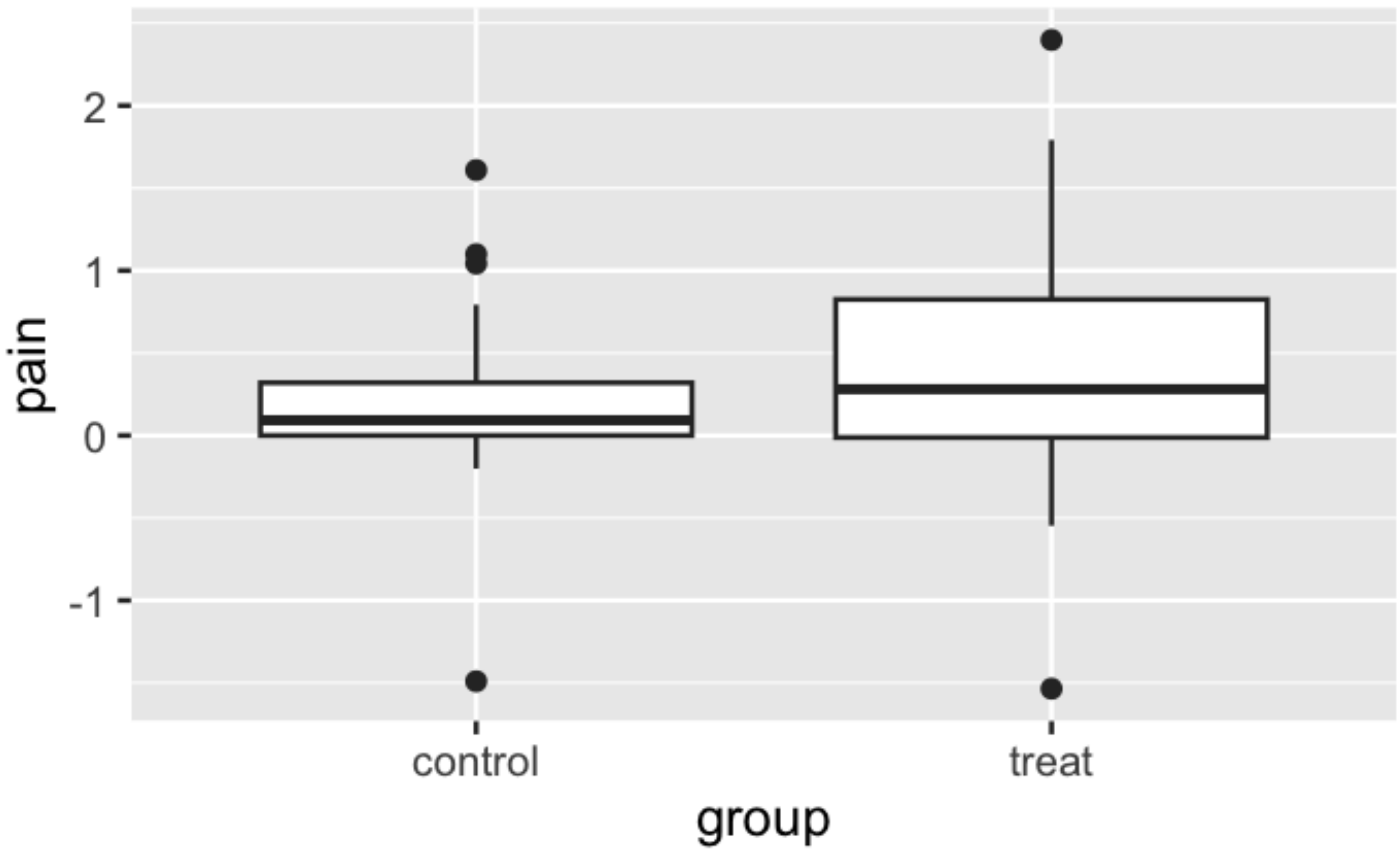
```
x=neuropathy$pain[neuropathy$group=="control"]
y=neuropathy$pain[neuropathy$group==« treat"]
```

Version test de Student :

```
t.test(x,y,var.equal = T)

 Two Sample t-test

data:  x and y
t = -1.3279, df = 56, p-value = 0.1896
```

Version test de Welch :

```
t.test(x,y)

Welch Two Sample t-test

data:  x and y
t = -1.3094, df = 46.051, p-value = 0.1969
```



Test de Yuen[1]

Moyennes « trimmées »

- Une alternative robuste au test de Student est basée sur le calcul de moyennes triées c'est à dire calculées sur un échantillon dont on a retiré des p/2 valeurs minimales et maximales. En général, on prend p=10%.

Sur R :

```
library(robnptests)
trimmed_test(x,y)

data:  x and y
trimmed t = -1.3775, df = 34, p-value = 0.1578
```

[1]Yuen, K.K. (1974) The two-sample trimmed t for unequal population variances. Biometrika, 61, 165-170.

Introduction aux tests non paramétriques

- Ils s'utilisent couramment dans les cas suivants :
  - Présence de valeur(s) extrême(s) dans l'un au moins des échantillons.
  - Tailles d'échantillons faibles ($n_A < 10$ ou $n_B < 10$) $\rightsquigarrow$ on verra !!!
  - Très forte asymétrie des distributions (effet plancher ou effet plafond).
  - Les variables X_A, X_B sont ordinales (c'est à dire nominales mais pour lesquelles on peut classer les modalités).
- Attention : conditions d'utilisation ?

Test de Mann-Whitney/ Wilcoxon

- Soit $X_1, \ldots, X_{n_1}$ un échantillon aléatoire distribué à partir d'une fonction de répartition F.
- Soit $Y_1, \ldots, Y_{n_2}$ un échantillon aléatoire distribué à partir d'une fonction de répartition G.

Test de Wilcoxon :

On suppose que les fonctions F et G vérifient $G(x) = F(x - \Delta)$.
Δ peut être vu comme une différence entre les moyennes ou entre les médianes [shift].

Ainsi, on suppose que les variables X et Y sont mesurées sur la même échelle.
On teste l'hypothèse $H_0 : \Delta = 0$ versus $H_1 : \Delta \neq 0$

- La statistique de ce test est $T = \sum_{j=1}^{n_2} R(Y_j)$.
- Sous H0 cette statistique est connue et on rejettera H0 pour des petites valeurs de T

Remarque : L'estimateur du shift est donné par Hodges-Lehmann $\Delta = med\{(Y_j - X_i)_{ij}\}$.

Autre expression de la statistique de test

Mann Whitney :

Dans les deux cas précédents on peut utiliser une statistique de test équivalente à T. On considère les $n_1 n_2$ différences $Y_j - X_i$ et on définit $T^+ = \#\{(i,j) : Y_j - X_i > 0\}$ le nombre de différences positives.

On a $T^+ = T - \dfrac{n_2(n_2+1)}{2}$ et cette statistique s'appelle la statistique de Mann-Whitney.

Version asymptotique :

La statistique T de Wilcoxon peut être standardisée : $z = \dfrac{T - n_2(n_1+n_2+1)/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1/12}}$

La p-value est alors calculée en utilisant la loi normale centrée réduite .

Définition du score de Wilcoxon

- La statistique $z = \dfrac{T - n_2(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1/12}}$ peut aussi s'écrire sous la forme $z = \sum_{i=1}^{n_2} a(R(Y_i))$ où $a(R(Y_i)) = \dfrac{\sqrt{12}}{n_1 + n_2 + 1}(R(Y_i) - \dfrac{1}{2})$.
- Ce score étant appelé score de Wilcoxon il est utilisé dans la régression basée sur les rangs.
- On peut généraliser ce score et ainsi obtenir d'autres estimateurs et d'autres tests pour le paramètre de shift Δ.

La fonction wilcox.test de R calcule la statistique de Mann-Whitney.

Pour la version classique

```r
wilcox.test(x,y)
```

```
Wilcoxon rank sum test with continuity correction

data:  pain by group
W = 357, p-value = 0.3301
alternative hypothesis: true location shift is not equal to 0
```

Pour la version asymptotique

```r
library(npsm)
rank.test(x,y)
```

```
statistic =  0.9816897 , p-value =  0.3262528
```

Test de Fligner-Policello

- Soit $X_1, \ldots, X_{n_1}$ un échantillon aléatoire distribué à partir d'une fonction de répartition F.
- Soit $Y_1, \ldots, Y_{n_2}$ un échantillon aléatoire distribué à partir d'une fonction de répartition G.

Soient $\tilde{\mu}_X, \tilde{\mu}_Y$ les médianes respectives de F et de G

On teste l'hypothèse $H_0 : \tilde{\mu}_X = \tilde{\mu}_Y$ versus $H_1 : \tilde{\mu}_X \neq \tilde{\mu}_Y$

- La statistique de ce test est une modification de la statistique de test de Wilcoxon.
- La fonction fp.test permet de réaliser ce test sur R.

```
library(npsm)
fp.test(x,y)
```

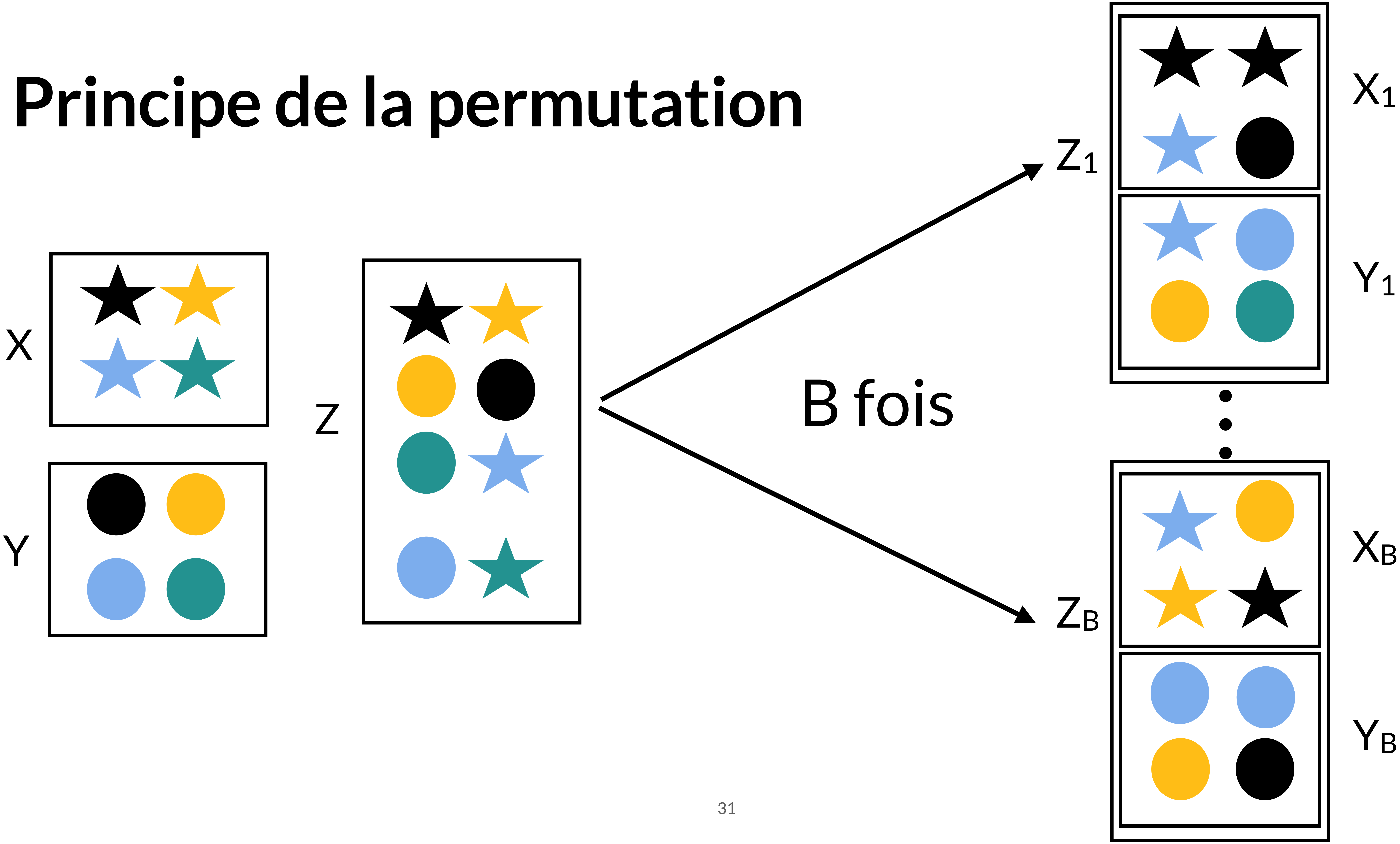
```
statistic =  0.9124277 , p-value =  0.3653668
```

Tests basés sur le ré-échantillonnage

Test par permutation

- Principe : on mélange les échantillons, on ré-échantillonne (bootstrap) et sur chaque échantillon on calcule une statistique de test
- On peut prendre comme statistique de test :
  - $D = |\mu_X - \mu_Y|.$
  - La statistique du test de Student.
  - La statistique du test de Welch.

Principe de la permutation



```r
permutation_D_two_groups <- function(x, y, n_b = 5000){
  n1=length(x)
  n2=length(y)
  D=abs(mean(x)-mean(y))
  D_b <- numeric(n_b)
  Z <- c(x,y)

  for (i in 1:n_b) {
    # Échantillonnage bootstrap avec remplacement
    Zb<- sample(Z,n1+n2 ,replace = TRUE)

    # Diviser l'échantillon bootstrap en deux groupes
    xb <- Z[1:n1]
    yb <- Z[(n1 + 1):(n1+n2)]

    # Calculer la différence de moyennes pour le bootstrap
    D_b[i] <- abs(mean(xb) - mean(yb))
  }
  # Calculer la p-value
  # On considère un test bilatéral
  p_value_b <- mean(D_b>D)

  return(p_value_b)
}
```

On trouve p=0.1844 pour l'exemple précédent.
Sur R package coin avec la fonction oneway_test

Test de Brunner-Munzel

Principe du test

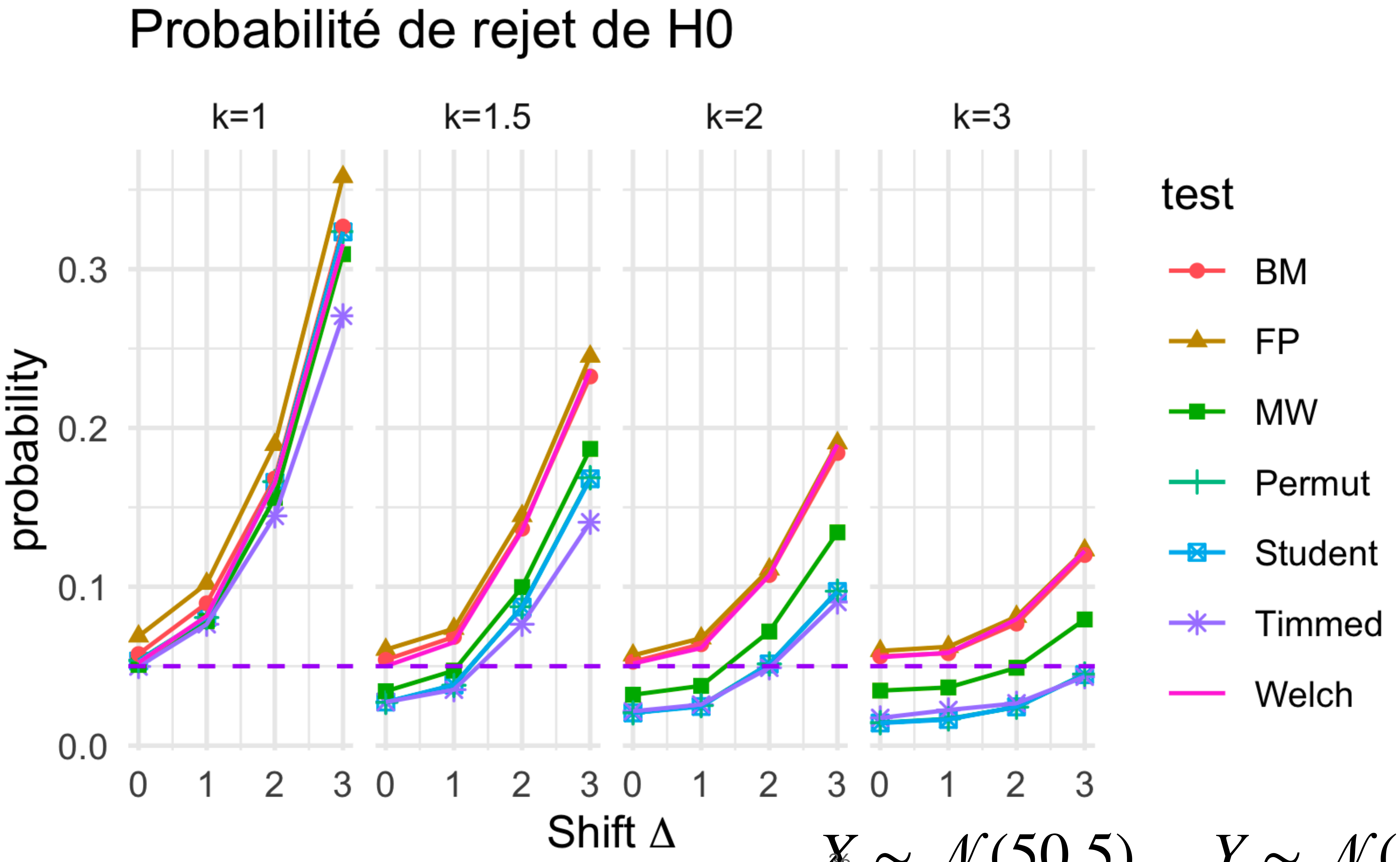
- Il s'agit d'un test non paramétrique basé sur un principe similaire au test de Mann-Whitney et sur le principe de la permutation.
- Soit S(x,y) la fonction qui vaut 1 si x<y, 1/2 si x=y et 0 sinon.
- L'hypothèse d'égalité des fonctions de répartitions F et G est équivalente à l'hypothèse $H_0 : P(X < Y) + 0.5P(X = Y) = 0.5$ (versus $\neq 0.5$).
- On construit la statistique de test $U^* = \dfrac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} S(X_i, Y_j)$.
- L'hypothèse H0 est alors vraie lorsque $U^* \simeq 0.5$.

Sur R :

```
library(brunnermunzel)
brunnermunzel.test(x,y)
 Brunner-Munzel Test

data:  x and y
Brunner-Munzel Test Statistic = 0.94424, df = 41.597, p-value = 0.3505
```

Comparaison des différents tests

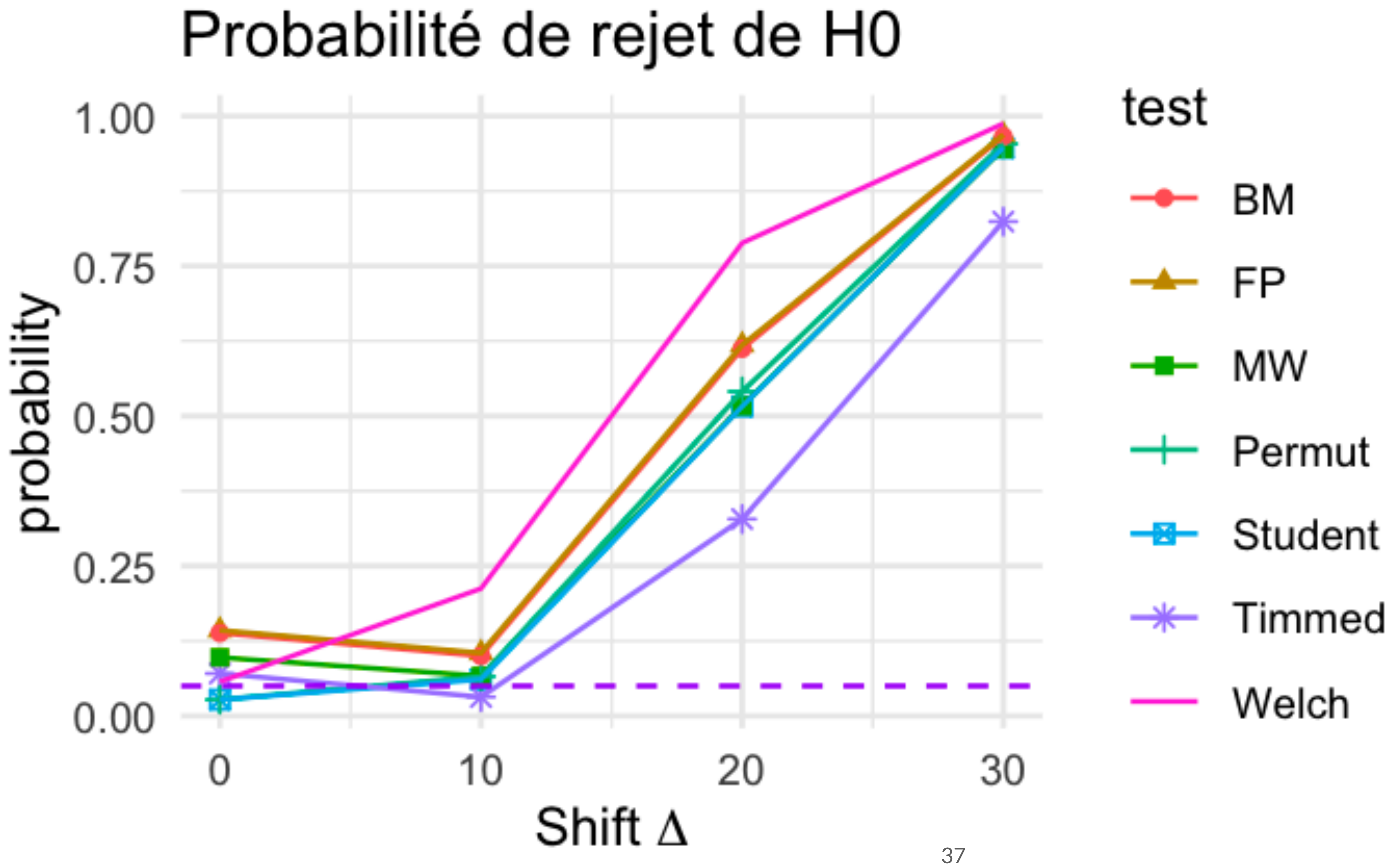
Comparaison 1 : Loi normales



$$X \sim \mathcal{N}(50,5), \quad Y \sim \mathcal{N}(50+\Delta,5k)$$

Comparaison 2 : Lois non normales

$$X \sim \mathscr{LN}(50,10), \quad Y \sim \mathscr{LN}(50+\Delta,30)$$



Comparaison d’échelles
(deux groupes indépendants)

Test de Levene

Test paramétrique

- Soit $X_1, \ldots, X_{n_1}$ un échantillon aléatoire distribué à partir d'une fonction de répartition F normale de paramètres μ_X, σ_X.
- Soit $Y_1, \ldots, Y_{n_2}$ un échantillon aléatoire distribué à partir d'une fonction de répartition G normale de paramètres μ_Y, σ_Y.

On teste $H_0 : \sigma_X = \sigma_Y$ versus $H_1 : \sigma_X \neq \sigma_Y$

```
library(car)
leveneTest(c(x,y),c(rep(0,length(x)),rep(1,length(y))))
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1  4.4733 0.03889 *
      56
```

Test de Fligner-Killeen

- Soit $X_1, \ldots, X_{n_1}$ un échantillon aléatoire distribué à partir d'une densité de probabilité $\frac{1}{\sigma_X} f\left(\frac{t-\mu_X}{\sigma_X}\right)$ où f est symétrique par rapport à 0.
- Soit $Y_1, \ldots, Y_{n_2}$ un échantillon aléatoire distribué à partir d'une densité de probabilité $\frac{1}{\sigma_Y} f\left(\frac{t-\mu_Y}{\sigma_Y}\right)$.

On teste $H_0 : \sigma_X = \sigma_Y$ versus $H_1 : \sigma_X \neq \sigma_Y$.

```
library(npsm)
fk.test(x,y)
```

```
statistic =  2.193009 , p-value =  0.02830674
```

Comparaison de k groupes

Anova à un facteur sur groupes indépendants

Principe de la régression basée sur les rangs

On considère le modèle de régression usuel $Y = \mathbb{X}\beta + \varepsilon$.

Il existe plusieurs approches pour déterminer les coefficients β.

- Régression par la méthode des moindres carrés : minimiser la fonction de dispersion f définie par la norme euclidienne : on cherche β tel que $f(\beta) = ||Y - \mathbb{X} \, . \, \beta||^2 = \sum_{i=1}^{n} \varepsilon_i^2$ soit minimale.
- Régression basée sur les rangs : la fonction de dispersion à minimiser est basée sur un calcul de rang : $f(\beta) = \sum_{i=1}^{n} a(R(\varepsilon_i))\varepsilon_i$ où $a(R(\varepsilon_i)) = \frac{\sqrt{12}}{n+1}(R(\varepsilon_i) - \frac{1}{2})$ est le score de Wilcoxon associé aux résidus.
    - Contrairement au problème des moindres carrés il n'y a pas de solution explicite pour β, on utilise des méthodes de descente de gradient pour obtenir une estimation de β.

ANOVA sur les rangs

Un modèle d'ANOVA est un modèle de régression dans lequel la matrice de design a une forme particulière, on peut par exemple l'écrire :

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \, .$$

On considère que les résidus sont i.i.d. selon une même densité f (c'est à dire selon une même fonction de répartition F). Si f est la loi normale on re trouve le modèle usuel.

⇝ On peut alors considérer les α_j comme des paramètres de shifts entre les distributions sur les différents groupes.

Sur R :

oneway.rfit() du package Rfit (même syntaxe que celle de aov())

ANOVA à plusieurs facteurs

- L'ANOVA sur les rangs se généralise au cas où on a plusieurs facteurs.
- Le modèle s'écrit de la même façon que le modèle usuel et on utilise la fonction raov().
- On peut tester des effets principaux, des effets d'interaction
- On peut aussi tester des contrastes (voir TP).

Tests sur Echantillons appariés

Deux échantillons appariés

- On considère des échantillons appariés, le problème de leur comparaison revient à comparer l'échantillon des différence $D_1, \ldots, D_n$ par rapport à 0 : c'est à dire que l'on considère que

$D_i = \Delta + \varepsilon_i$ où les erreurs ε_i sont i.i.d. selon une densité f symétrique par rapport à 0.

On teste alors $H_0 : \Delta = 0$ versus $H_1 : \Delta \neq 0$.

- On a plusieurs tests :
  - Test des signes : signtest_pvalue() dans npsm (par exemple)
  - Test des signes de Wilcoxon : wilcox.test(x)
  - Test de permutations.

RM Anova

- Pour 1 seul facteur le test de Friedman friedman.test sur R
- Pas possible d'utiliser les fonctions raov()
- On peut toujours utiliser de la permutation … (voir TP).
- La version paramétrique reste la plus indiquée.