

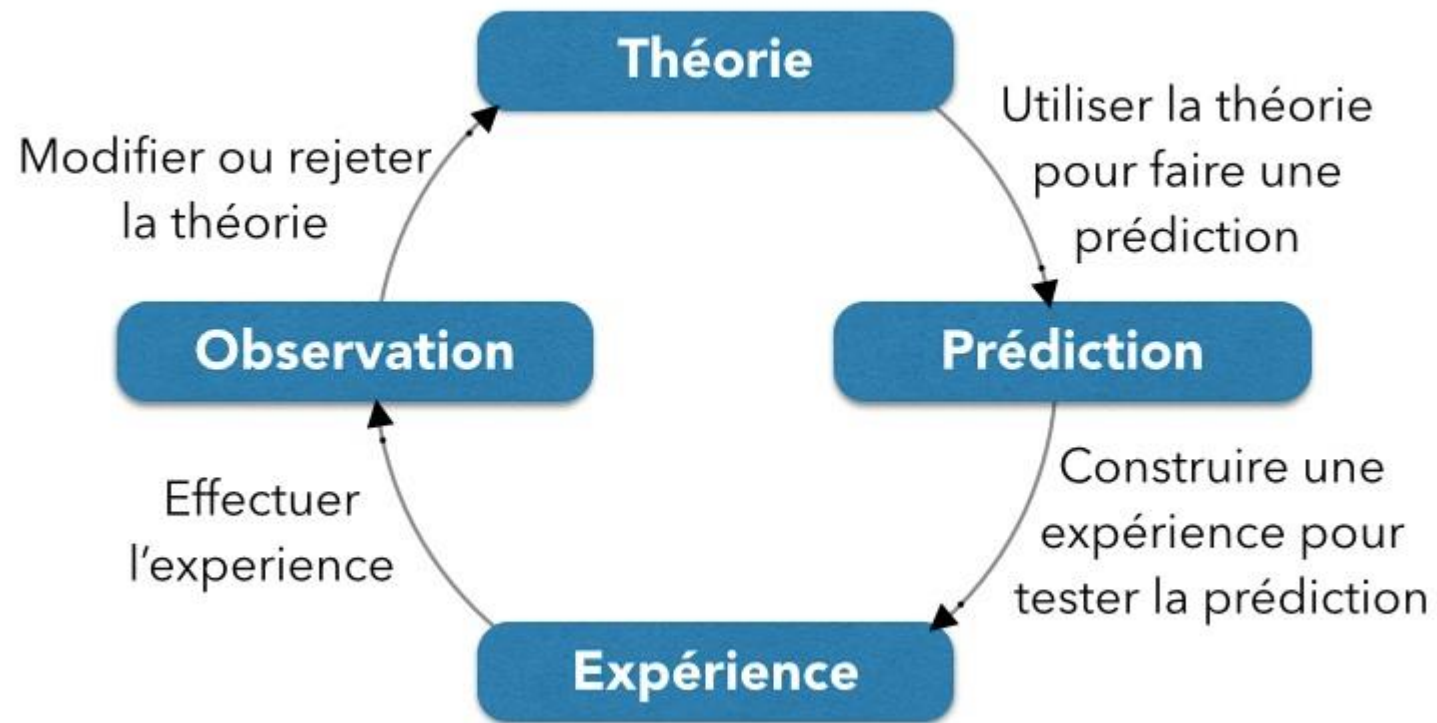
# Combien de répets dans ma manipe?

Club stats iEES-Paris

Jerome Mathieu

16-11-2022

# La méthode scientifique



# Les différentes étapes

Bibliographie

Question

Test statistique

Plan d'échantillonnage

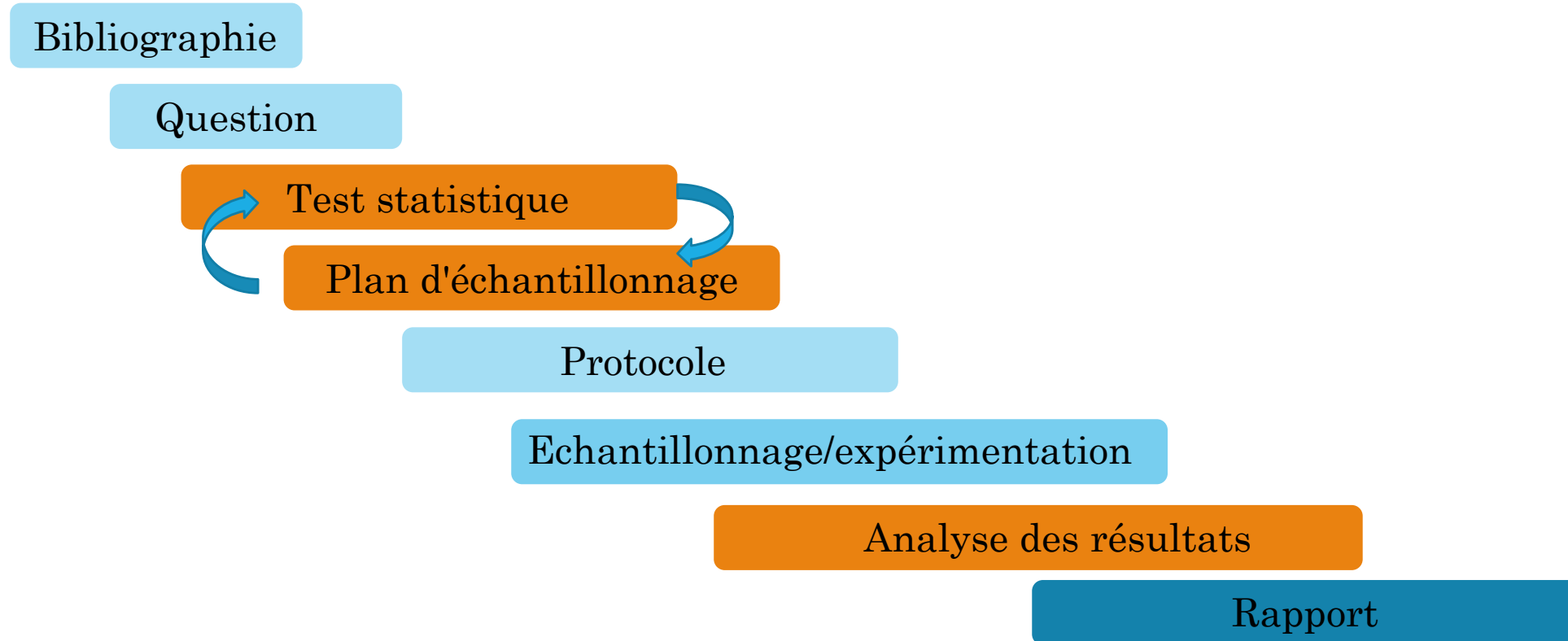
Protocole

Echantillonnage/expérimentation

Analyse des résultats

Rapport

# Je te tiens par la barbichette



# Plan d'échantillonnage

Bibliographie

Question

Test statistique

Plan d'échantillonnage

Protocole

Echantillonnage/expérimentation

Analyse des résultats

Rapport

# Définir un Plan expérimental ou d'échantillonnage

- Plan =
  - Facteur(s) manipulé(s) ou gradient(s) échantillonné(s)
  - Nombre de niveaux des facteurs ou nombre de valeurs du gradient
  - Valeurs des niveaux de facteur ou du gradient
  - Combinaisons manquantes
  - Hiérarchisation
  - Organisation spatiale
  - Localisation géographique
  - **Nombre de répétitions par modalité unique de facteur**
- Souvent représenté sous forme de tableau

		X1 →			
		A	B	C	D
X2 ↓	oui	3	5	...	$n_{ij}$
	non	10	...	...	...

nb de répétitions

# Designs expérimentaux : Design of Experiments (DoE)

Task dans R

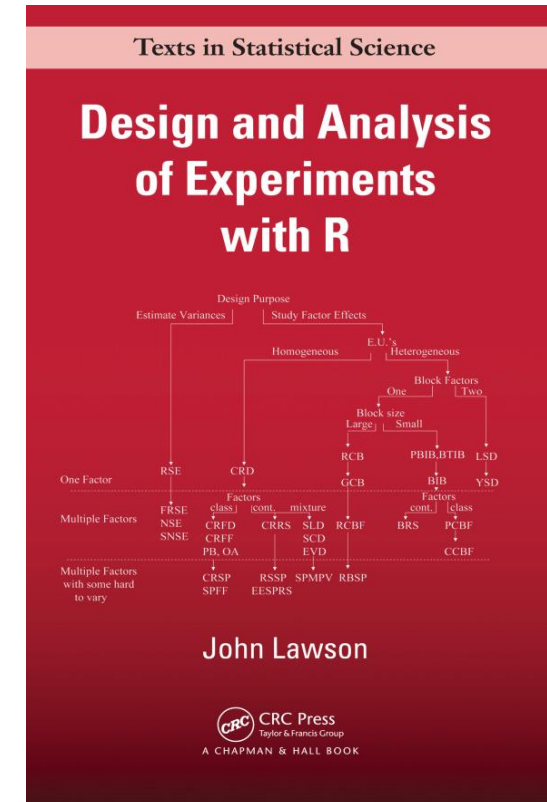
<https://cran.r-project.org/web/views/ExperimentalDesign.html>

Bouquin en ligne

[https://bookdown.org/gerhard\\_krennrich/doe\\_and\\_optimization/](https://bookdown.org/gerhard_krennrich/doe_and_optimization/)

Cours en ligne

<https://www.datacamp.com/courses/experimental-design-in-r>



# Agricolae – a free statistical toolbox for agricultural experiments

### 1) Design of experiments:

design.crd (Completely Randomized Design), rcdb (Randomized Complete Block Design), lsd (Latin Square Design), bib (Balanced Incomplete Block Design), lattice (Lattice Design), alpha (Alpha Design), cyclic (Cyclic Design), graeco (Graeco-Latin Square Design).

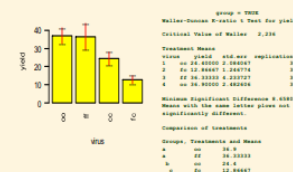
**alpha design (0,1) - Serie I**  
**Parameters Alpha design**

treatments : 12  
Block size : 3  
Blocks : 4  
Replication: 3

$\{, \}$	$\{, \}$	$\{, \}$
$\{1, \}$ j	a	e
$\{2, \}$ i	f	h
$\{3, \}$ e	k	i
$\{4, \}$ b	g	d

## 2) Comparison of treatments:

parametric and non-parametric methods: LSD.test (Minimum Significant Difference), HSD.test (Tukey), waller.test (Waller-Duncan), kruskal (Kruskal-Wallis), friedman (Friedman), durbin.test (Test of Durbin), PBIB.test (Analysis of partially balanced blocks, applicable to the designs lattices and alpha).



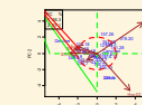
### 3) Analysis of genetic designs:

Carolina I, II and III with the function `carolina()` and line by tester with the function `lineXtester()`.

[illegible]

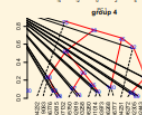
#### 4) Stability of genotypes:

stability.par(), stability.nonpar()  
and AMMI().



### 5) Consensus of dendrograms:

(e.g. for molecular marker analysis)  
consensus



### 6) Biodiversity analysis:

index.bio

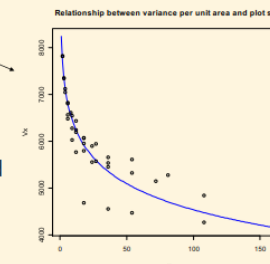
Method: Shannon The index: 3.973060  
95 percent confidence interval: 3.841353 ;  
4.60517

### 7) Optimizing plot size:

index.smith

### 8) Descriptive statistics:

graph.freq (graphic of frequencies), poligon.freq (polygon), ojiva.freq (point), and table.freq



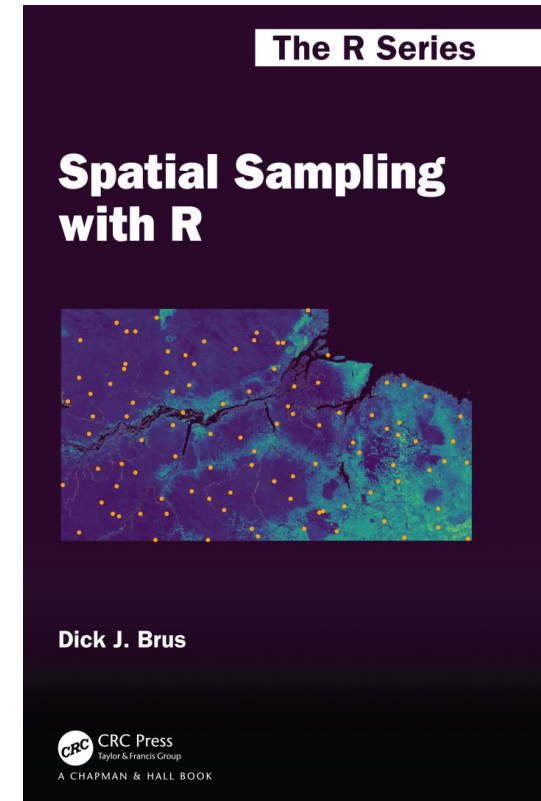
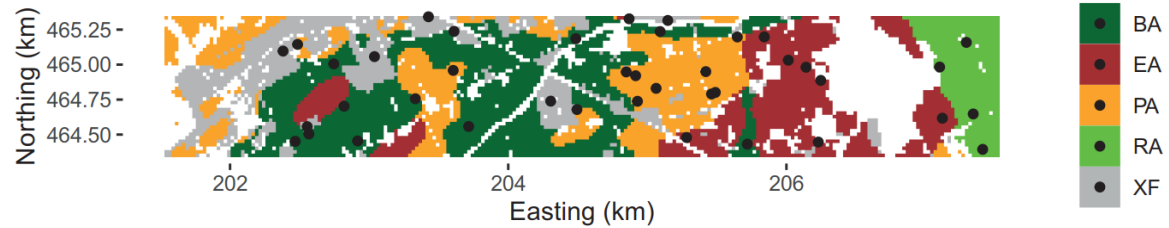
9) Various: AUDPC, helper

functions



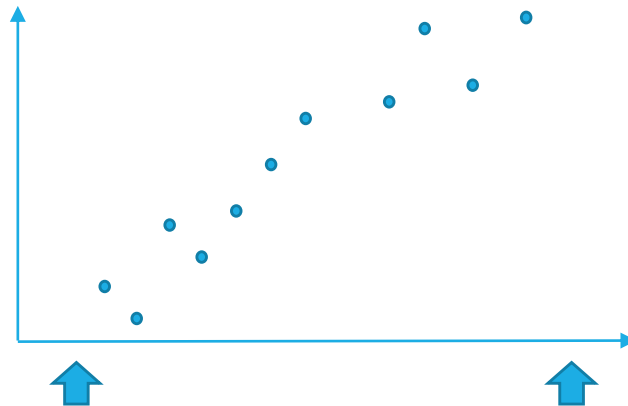
# Plans spatiaux

- Cartographie
- estimer une valeur globale (ex stock C)



# Remarque sur la distribution des co variables

# Si $X$ numérique



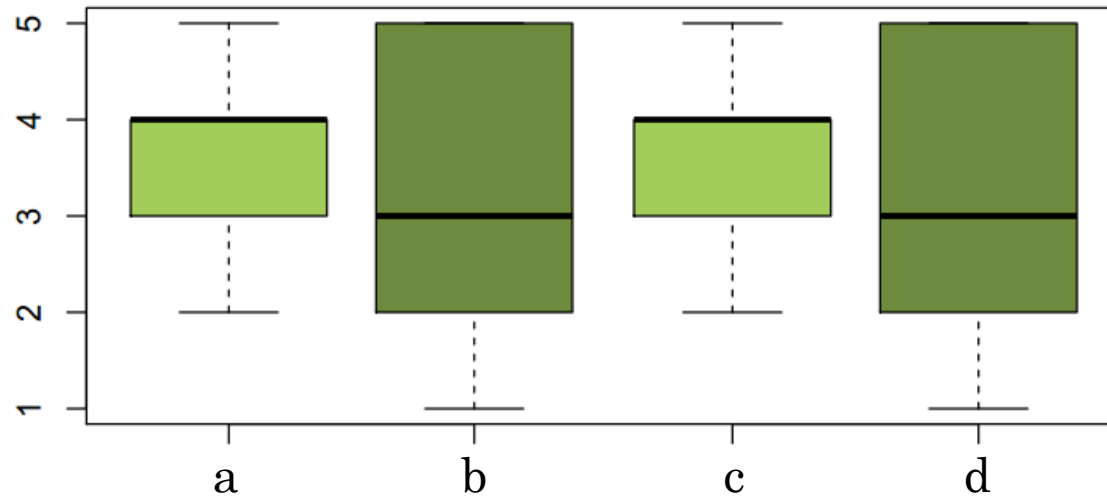
Couvrir les extrêmes de  $X$

Couvrir  $X$  de manière homogène

Minimum : 10 mesures

$X$  ne suit pas de loi particulière

# si X catégorique



Même nb d'observations par catégories : plan équilibré  
Minimun : 10 observations / catégorie  
Si possible Pas de trous : pas de catégorie non échantillonnée

Questions qu'on peut se poser sur un test  
(ou une manipe)

# Ex.

- J'ai/ il a une  $p \text{ val} > 0.05 \rightarrow$  pourquoi?
- J'ai une  $p \text{ val} < 0.05 \rightarrow$  est ce que l'effet est fort?
- Combien de répét il me faut pour avoir  $p < 0.05$ ?
- Quelles sont les chances de détecter un effet avec ma manipe? (puissance)

# Avant de pouvoir répondre, quelques def (désolé!)

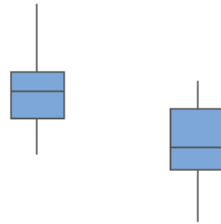
- Risques I et II :  $\alpha$  et  $\beta$

Unknown to you, the Null Hypothesis is:	As a result of sampling, the Null Hypothesis is:	
	Rejected You believe there has been an environmental impact	Retained You believe there has been no environmental impact
True	Type I error With probability $\alpha$	Correct decision
False	Correct decision	Type II error With probability $\beta$

- **$\alpha$  risque de "surdétecter" des effets** : rejeter  $H_0$  alors que  $H_0$  vraie
- **$\beta$  risque de "rater" un effet** : Ne pas rejeter  $H_0$  alors que  $H_0$  fausse
- **Puissance : capacité de "détecter" un effet** : rejeter  $H_0$  alors que  $H_0$  fausse :  $1-\beta$

# $\alpha$ vs $\beta$

- Un nouveau variant du covid arrive. Quel est son effet?
- Manipe : 2 traitements virus ou pas



- si  $\alpha$  élevé
  - → **risque de "surdétecter" des effets élevé**: rejet  $H_0$  alors que  $H_0$  vraie : létalité sur évaluée → trop de précautions
  - → COUTS : grosses dépenses de santé publique mais pas de pertes humaine
- si  $\beta$  élevé (puissance  $(1-\beta)$  faible )
  - → **risque de "rater" l'effet du virus élevé**: on ne rejete pas assez  $H_0$  alors que  $H_0$  fausse : létalité sous évaluée → pas assez de précautions
  - → COUTS : peu de dépenses de santé publique mais grosses pertes humaines potentielles
- → **vous choisissez quoi entre  $\alpha$  et  $\beta$  ? Quelles valeurs??**



# The five eighty convention

- $\alpha = 0.05$
- $1 - \beta = 0.8$

- Mais controversé :

*Functional  
Ecology* 2003  
17, 707–709

## Forum

JULIAN DI STEFANO

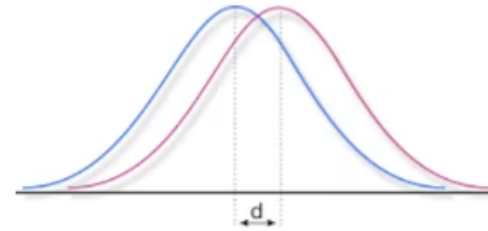
How much power is enough? Against the development of an arbitrary convention for statistical power calculations

- **Alternative** : estimer les couts acceptables de  $\alpha$  et  $\beta$  et fixer le ratio  $\alpha/\beta$

# Quizz :qu'est ce qu'un "effet"?

- **ex d'un test t :**
  - effet =
    - p value?
    - différence de moyenne?

# Effect size



- Effect size is a quantitative measure of the *strength of a phenomenon*.
- Effect size emphasizes the **size** of the difference or relationship
- Examples:
  - the correlation between two variables (specifically  $r^2$ )
    - $r=.1$  weak,  $r=.5$  moderate,  $r=.7$  strong,  $r=.9$  very strong
  - the regression coefficient in a regression ( $B_0, B_1, B_2$ )
    - Relative to model and field
  - the mean differences in t tests (use Cohen's D)
    - $d = .2$  is small;  $r = .5$  is medium;  $r = .8$  is large
  - The mean differences in ANOVA (use eta)
    - .01 is small, .06 medium, .14 large

- calcul différent selon la comparaison que l'on veut faire

# Puissance moyenne des études en éco comportementale (effect size moyen)

**Table 2**  
z scores and power to detect a medium effect for 10 biological journals

Journal	z score		Power (medium)	
	(First test)	(Last test)	(First test)	(Last test)
American Naturalist	2.67 ± 0.19 (33)	2.33 ± 1.88 (32)	45.0 ± 5.2 (33)	42.0 ± 4.8 (32)
Animal Behaviour	2.33 ± 0.08 (184)	1.88 ± 0.08 (176)	46.8 ± 2.2 (187)	37.6 ± 2.0 (179)
Behavioral Ecology	2.13 ± 0.13 (68)	1.66 ± 0.13 (67)	51.2 ± 3.6 (68)	41.0 ± 3.3 (67)
Behavioural Ecology and Sociobiology	2.25 ± 0.11 (99)	1.75 ± 0.11 (98)	52.9 ± 2.9 (102)	42.2 ± 2.7 (99)
Behavioural Processes	2.38 ± 0.17 (39)	1.98 ± 0.17 (38)	40.0 ± 4.8 (39)	33.5 ± 4.4 (38)
Behaviour	2.45 ± 0.13 (67)	2.05 ± 0.14 (63)	44.6 ± 3.6 (69)	40.2 ± 3.4 (63)
Ethology, Ecology, and Evolution	2.15 ± 0.23 (22)	2.13 ± 0.24 (21)	58.0 ± 6.3 (22)	44.0 ± 5.9 (21)
Ethology	2.24 ± 0.13 (73)	2.03 ± 0.13 (66)	43.0 ± 3.5 (73)	35.5 ± 3.3 (69)
Journal of Animal Ecology	2.28 ± 0.15 (49)	1.73 ± 0.16 (45)	49.3 ± 4.2 (50)	40.3 ± 4.0 (45)
Journal of Insect Behaviour	2.23 ± 0.15 (50)	1.85 ± 0.16 (48)	42.2 ± 4.0 (54)	43.1 ± 3.8 (52)

Mean ± SE. Sample sizes are in parentheses.

# Qu'est ce qui compte? p val ou Effect size?

- Différence de 1°C, p val = 0.001
- Différences de 5°C, pval = 0.1

# Différence entre p value et taille d'effet

- p value reflète plutôt l'effort d'échantillonnage
- Effect size reflète plutôt le phénomène observé

# En fait tout est lié

Type de test

Risque  $\alpha$

Risque  $\beta$

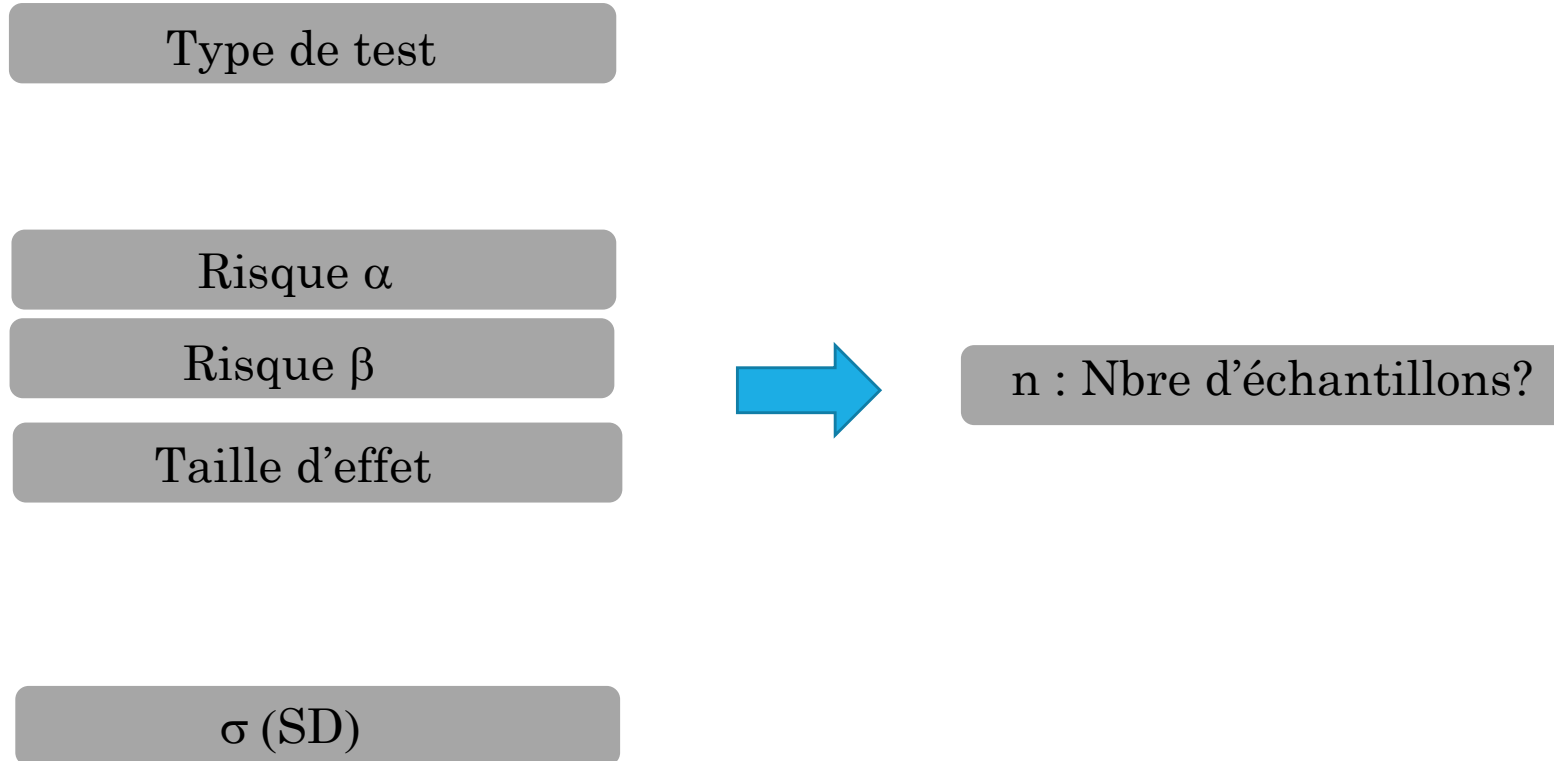
Taille d'effet

n : Nbre d'échantillons?

$\sigma$  (SD)

Si connaît tout les éléments sauf 1, on peut le déduire des autres éléments

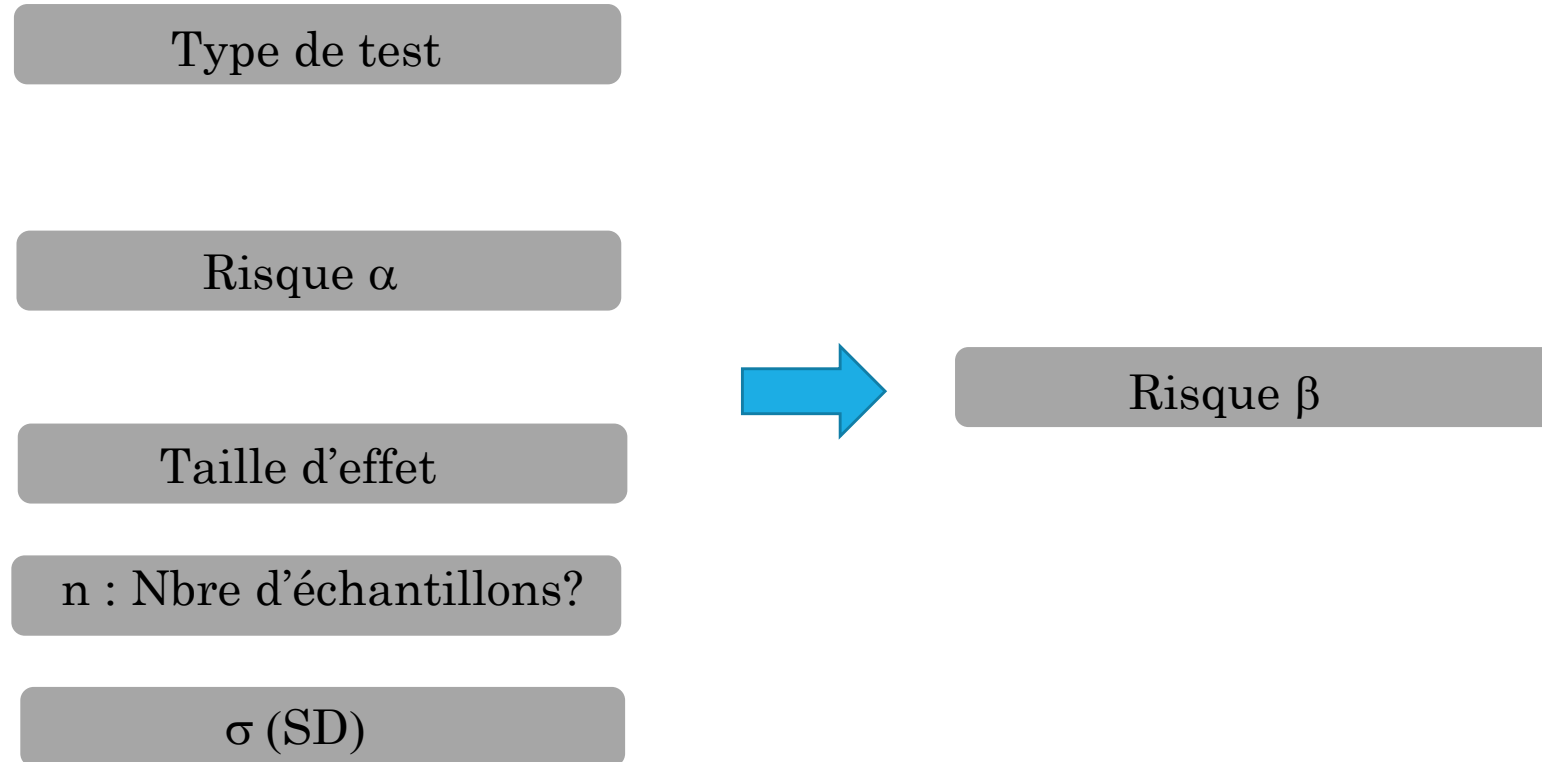
# n pour une manipe?



Si connaît tout les éléments sauf 1, on peut le déduire des autres éléments



# Puissance d'une manipe?



Si connaît tout les éléments sauf 1, on peut le déduire des autres éléments

# Déterminer $n$ le nombre de répétitions

- Dans le cas  $Y$  univarié uniquement : **analyse de puissance**

# Analyse de puissance

Risques  $\alpha$  et  $\beta$

Type de test

Taille d'effet recherché

A décider en 1er



Analyse de puissance

n : Nbre d'échantillons?

# Comment faire?

- si modèle linéaire à effets fixes unique → formules analytiques → analyse de puissance (cf dia suivantes)
- si tests non paramétriques ou modèle mixte, ou GLM(M) → simulations
- si approche bayésienne -> n'a pas vraiment de sens

# modèles simples fixes

- <https://cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html>

# 1° On fixe les seuils de risque

- $\alpha$  = proba de "surdétecter un effet" : faux positifs
- $\beta$  = proba de rater un effet significatif : faux négatifs
- **Puissance** =  $1 - \beta$

## 2° On identifie le type de test à faire

- one-sample proportion test
- two-sample proportion test
- two-sample proportion test (unequal sample sizes)
- two-sample, one-sample and paired t-tests
- two-sample t-tests (unequal sample sizes)
- one-way balanced ANOVA (X catégorique)
- correlation test
- chi-squared test (goodness of fit and association)
- test for the general linear model ( X continu)

# Les fonctions R pour l'analyse de puissance dans R

```
library(pwr)
```

- **pwr.p.test** : one-sample proportion test
- **pwr.2p.test** : two-sample proportion test
- **pwr.2p2n.test** : two-sample proportion test (unequal sample sizes)
- **pwr.t.test** : two-sample, one-sample and paired t-tests
- **pwr.t2n.test** : two-sample t-tests (unequal sample sizes)
- **pwr.anova.test** : one-way balanced ANOVA
- **pwr.r.test** : correlation test
- **pwr.chisq.test** : chi-squared test (goodness of fit and association)
- **pwr.f2.test** : test for the general linear model



### 3° On Détermine la taille de l'effet recherché

Dépend du test à faire :

Test	small	medium	large
tests for proportions ( $p$ )	0.2	0.5	0.8
tests for means ( $t$ )	0.2	0.5	0.8
chi-square tests ( $\chi^2$ )	0.1	0.3	0.5
correlation test ( $r$ )	0.1	0.3	0.5
anova ( $anov$ )	0.1	0.25	0.4
general linear model ( $f^2$ )	0.02	0.15	0.35

cf fonction dans pwr

4° On identifie le nombre de paramètres du test et on déduit "n"

- **Si X qualitatif :**
  - on fixe  $k$  : le nombre de niveaux
  - $\rightarrow$  on déduit  $n$  : le nombre d'échantillons par niveau
- **Si X continu ou si plusieurs variables :**
  - on fixe  $u$  = nbre de coefficients = ddl du numérateur  
= nbre de variables
  - $\rightarrow$  on obtient  $v$  et on déduit  $n$  par la formule :  $n = v + u + 1$

## Exemple avec un effet moyen (.25) pour 1 facteur à 3 niveaux

```
> library(pwr)
> pwr.anova.test(k = 3, f = 0.25, sig.level = 0.05, power = 0.9)

> Balanced one-way analysis of variance power calculation
```

```
      k = 3
      n = 52.3966
      f = 0.25
sig.level = 0.05
power = 0.8
```

NOTE: n is number in each group

→ **n = 53 échantillons par niveau (53x3 = 159 en tout)**

# Exemple d'une régression simple avec un effet large ( $f^2 = 0.5$ )

```
> library(pwr)
```

```
> pwr.f2.test(u = 1, f2 = .5, sig.level = 0.001, power = 0.8)
```

```
> Multiple regression power calculation
```

```
u = 1
```

```
v = 37.67315
```

```
f2 = 0.5
```

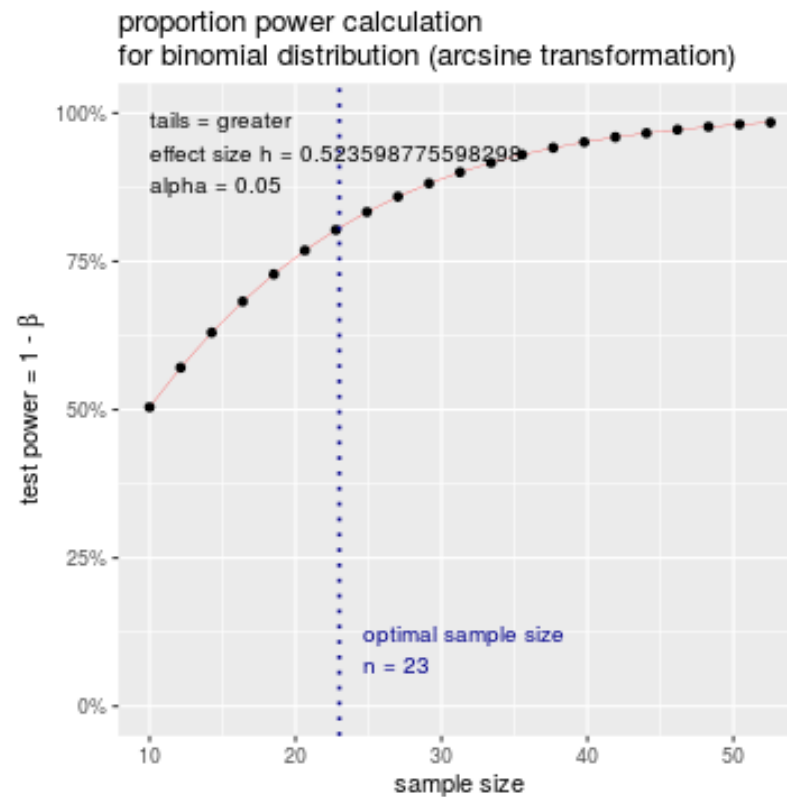
```
sig.level = 0.001
```

```
power = 0.8
```

→  $n = v + u + 1 = 38 + 1 + 1 : \underline{40 \text{ échantillons}}$

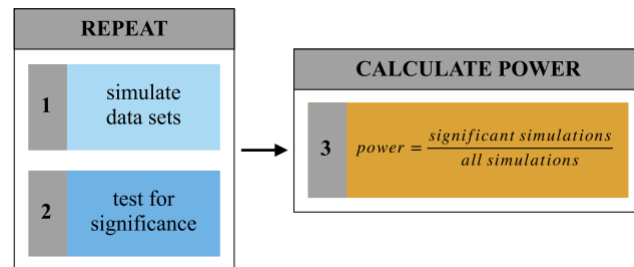
# courbes de puissance

```
p.out <- pwr.p.test(h = ES.h(p1 = 0.75, p2 = 0.50),  
  sig.level = 0.05,  
  power = 0.80,  
  alternative = "greater")  
  
plot(p.out)
```



# Cas plus complexes

# On fait des simuls



## Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R

Leah Kumle<sup>1</sup> • Melissa L.-H. Vö<sup>1</sup> • Dejan Draschkow<sup>2</sup>

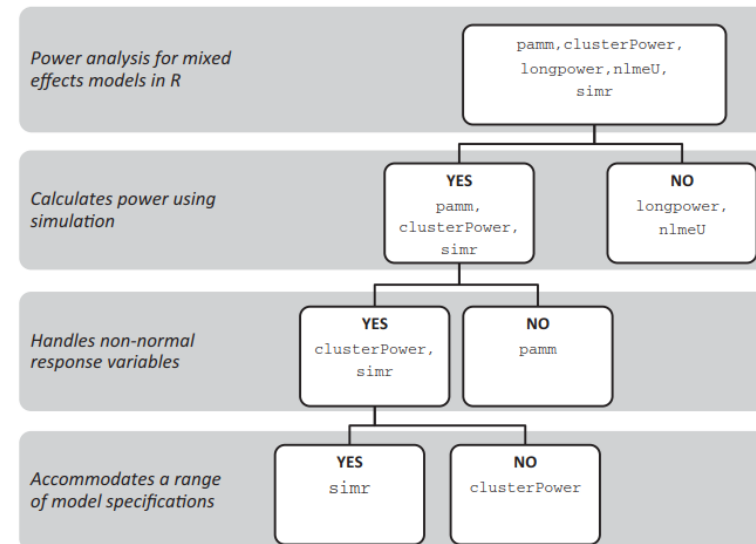
Behavior Research Methods (2021) 53:2528–2543  
<https://doi.org/10.3758/s13428-021-01546-0>

*Methods in Ecology and Evolution* 2015, 6, 133–142

doi: 10.1111/2041-210X.12306

## Power analysis for generalized linear mixed models in ecology and evolution

Paul C. D. Johnson<sup>1,2\*</sup>, Sarah J. E. Barry<sup>2</sup>, Heather M. Ferguson<sup>1</sup> and Pie Müller<sup>3,4</sup>



*Methods in Ecology and Evolution* 2016, 7, 493–498

doi: 10.1111/2041-210X

### APPLICATION

**SIMR: an R package for power analysis of generalized linear mixed models by simulation**

Peter Green\* and Catriona J. MacLeod

Peut on montrer qu'il n'y a pas d'effet?



# TOST

- démarche :
- utiliser effets sizes non intéressants pour prouver que  $H_0$  vraie

<https://cran.r-project.org/web/packages/TOSTER/index.html>

**Equivalence Tests: A Practical Primer for  
t Tests, Correlations, and Meta-Analyses**

Daniël Lakens<sup>1</sup>

Social Psychological and  
Personality Science  
2017, Vol. 8(4) 355-362