Outils Mathématiques

Caroline Verhoeven

Médecine factuelle

le 22 février 2018

Table des matières

Les erreurs

Données = erreurs

Les erreurs

Données = erreurs

Nous considérons 3 types d'erreurs :

- les erreurs aléatoires :
- les erreurs systématiques :
- les fraudes

Les erreurs

Données = erreurs

Nous considérons 3 types d'erreurs :

- les erreurs aléatoires :
  d'une mesure à l'autre, la valeur réelle peut être surévaluée ou sous-évaluée.
- les erreurs systématiques :
- les fraudes

Les erreurs

Données = erreurs

Nous considérons 3 types d'erreurs :

- les erreurs aléatoires :
  d'une mesure à l'autre, la valeur réelle peut être surévaluée ou sous-évaluée.
- les erreurs systématiques :
  toujours une surévaluation (ou sous-évaluation) de la valeur réelle.
- les fraudes

Les erreurs

Données = erreurs

Nous considérons 3 types d'erreurs :

- les erreurs aléatoires :
  d'une mesure à l'autre, la valeur réelle peut être surévaluée ou sous-évaluée.
- les erreurs systématiques :
  toujours une surévaluation (ou sous-évaluation) de la valeur réelle.
- les fraudes

Les erreurs de mesure

Les erreurs de mesure : dans toutes les sciences dans laquelle il y a des mesures

Les variabilités biologiques

Les variabilités biologiques : dans les sciences biomédicales

Les variabilités biologiques

Les variabilités biologiques : dans les sciences biomédicales

- Variabilités inter-individuelles
- Variabilités intra-individuelles

Variabilités inter-individuelles

Nous sommes tous uniques

Exemple 1

Les analyses de sang : les normes sont établies de manière à ce que 95% de la population saine se trouve dans la norme.

Variabilités inter-individuelles

Nous sommes tous uniques

Exemple 1

Les analyses de sang : les normes sont établies de manière à ce que 95% de la population saine se trouve dans la norme.

- Par variable testée : une personne saine a une probabilité de 0,95 d'être dans la norme.

Variabilités inter-individuelles

Nous sommes tous uniques

Exemple 1

Les analyses de sang : les normes sont établies de manière à ce que 95% de la population saine se trouve dans la norme.

- Par variable testée : une personne saine a une probabilité de 0,95 d'être dans la norme.
- Supposons 15 variables indépendantes : la probabilité qu'une personne personne saine soit dans la norme pour les 15 :

$$(0{,}95)^{15} = 0{,}463$$

Variabilités inter-individuelles

Nous sommes tous uniques

Exemple 1

Les analyses de sang : les normes sont établies de manière à ce que 95% de la population saine se trouve dans la norme.

- Par variable testée : une personne saine a une probabilité de 0,95 d'être dans la norme.
- Supposons 15 variables indépendantes : la probabilité qu'une personne personne saine soit dans la norme pour les 15 :

$$(0{,}95)^{15} = 0{,}463$$

C'est normal d'être anormal

Variabilités intra-individuelles

Nous changeons avec le temps

Variabilités intra-individuelles

Nous changeons avec le temps

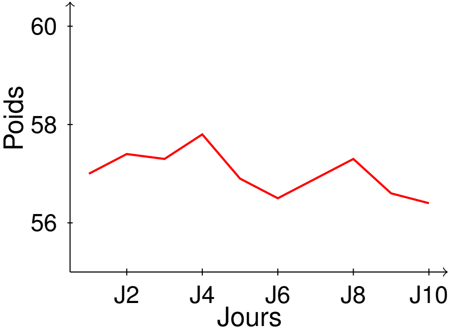
Exemple 2

Variabilités intra-individuelles

Nous changeons avec le temps

Exemple 2

Mon poids jour après jour ...



On a besoin de la biostat ...

En sciences biomédicales on cumule les erreurs aléatoires :

On a besoin de la biostat ...

En sciences biomédicales on cumule les erreurs aléatoires :


Scanned by CamScanner


Scanned by CamScanner

On a besoin de la biostat ...

En sciences biomédicales on cumule les erreurs aléatoires :

erreurs de mesures + variabilités biologiques

La biostat aide à gérer les erreurs aléatoires

Ce que je veux faire

Exemple 3

Je veux mesurer le diamètre moyen (μ) de tous les globules rouge d'un patient.

- microcytose : anémies ferriprives et des thalassémies
- macrocytose : carences en folates ou vitamine B12, alcoolisme

Ce que je veux faire

Exemple 3

Je veux mesurer le diamètre moyen (μ) de tous les globules rouge d'un patient.

- microcytose : anémies ferriprives et des thalassémies
- macrocytose : carences en folates ou vitamine B12, alcoolisme

Ce que je veux faire

Exemple 3

Je veux mesurer le diamètre moyen (μ) de tous les globules rouge d'un patient.

- microcytose : anémies ferriprives et des thalassémies
- macrocytose : carences en folates ou vitamine B12, alcoolisme

Problèmes

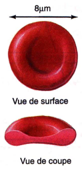
- Entre $\pm$ 4 et 6 millions de globules rouges par mm^3 de sang

Ce que je veux faire

Exemple 3

Je veux mesurer le diamètre moyen (μ) de tous les globules rouge d'un patient.

- microcytose : anémies ferriprives et des thalassémies
- macrocytose : carences en folates ou vitamine B12, alcoolisme



Problèmes

- Entre $\pm$ 4 et 6 millions de globules rouges par mm^3 de sang
- Le patient risque de ne pas être d'accord

Ce que je fais

Exemple 3

Je mesure qu’un petit nombre de globules rouges

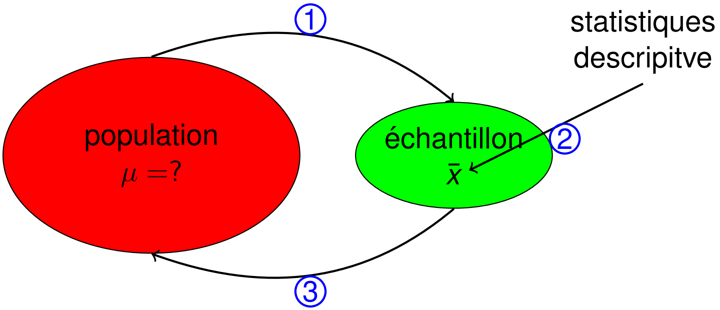
Ce que je fais

Exemple 3

Je mesure qu'un petit nombre de globules rouges

- Je mesure par exemple 20 globules rouges
- Je calcule le diamètre moyen ($\overline{x}$) de ces 20 globules
- J'essaie de conclure si le diamètre moyen est normal

Inférence statistique



- μ : moyenne pour la population
- $\overline{x}$: moyenne pour l'échantillon

Pas simple l'inférence

Exemple 3

- J'obtiens le diamètre moyen de 20 globules rouges

Pas simple l'inférence

Exemple 3

- J'obtiens le diamètre moyen de 20 globules rouges
- Je prend un autre échantillon se 20 globules rouges du même patient

Pas simple l'inférence

Exemple 3

- J'obtiens le diamètre moyen de 20 globules rouges
- Je prend un autre échantillon se 20 globules rouges du même patient
  $\Rightarrow$ j'obtiens une autre valeur pour la moyenne

Pas simple l'inférence

Exemple 3

- J'obtiens le diamètre moyen de 20 globules rouges
- Je prend un autre échantillon se 20 globules rouges du même patient
  $\Rightarrow$ j'obtiens une autre valeur pour la moyenne
- moyennes de 10 échantillons de 20 sujets d'une même population :

$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$	$\overline{x}_4$	$\overline{x}_5$	$\overline{x}_6$	$\overline{x}_7$	$\overline{x}_8$	$\overline{x}_9$	$\overline{x}_{10}$
7,29	7,31	7,15	6,79	6,96	7,12	7,10	7,27	7,37	7,18

Pas simple l'inférence

Exemple 3

- J'obtiens le diamètre moyen de 20 globules rouges
- Je prend un autre échantillon se 20 globules rouges du même patient
  $\Rightarrow$ j'obtiens une autre valeur pour la moyenne
- moyennes de 10 échantillons de 20 sujets d'une même population :

$\overline{x}_1$	$\overline{x}_2$	$\overline{x}_3$	$\overline{x}_4$	$\overline{x}_5$	$\overline{x}_6$	$\overline{x}_7$	$\overline{x}_8$	$\overline{x}_9$	$\overline{x}_{10}$
7,29	7,31	7,15	6,79	6,96	7,12	7,10	7,27	7,37	7,18

Conclusion :
La variabilité des individus fait varier les moyennes des échantillons

La distribution d'échantillonnage

- Calculer la moyenne de tous les échantillons possibles de N sujets venant de la même population
  (en théorie)

La distribution d'échantillonnage

- Calculer la moyenne de tous les échantillons possibles de N sujets venant de la même population
  (en théorie)
- Faire un histogramme

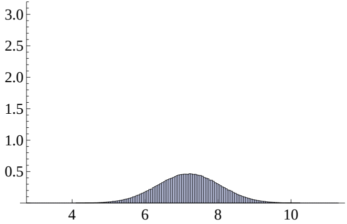
La distribution d’échantillonnage

- Calculer la moyenne de tous les échantillons possibles de N sujets venant de la même population
  (en théorie)
- Faire un histogramme
- Distribution d’échantillonnage

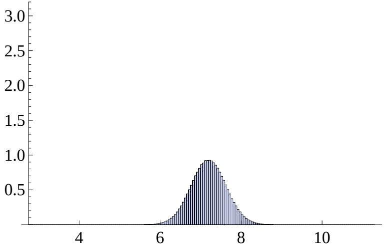
Théorème centrale limite I

Distribution des valeurs individuelles gaussienne

Distribution des individus (X)
$\mu = 7{,}2$, $\sigma = 0{,}86$



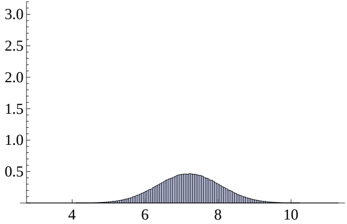
Distribution d'échantillonnage ($\overline{X}$)
N=4



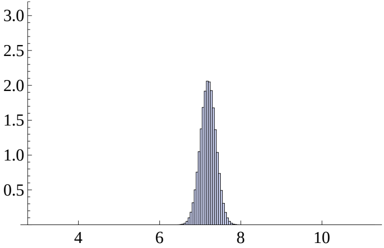
Théorème centrale limite I

Distribution des valeurs individuelles gaussienne

Distribution des individus (X)
$\mu = 7{,}2$, $\sigma = 0{,}86$



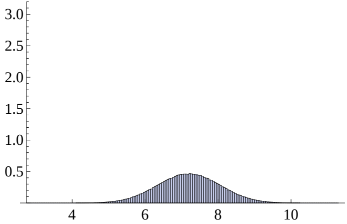
Distribution d'échantillonnage ($\overline{X}$)
N=20



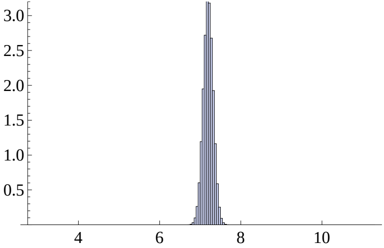
Théorème centrale limite I

Distribution des valeurs individuelles gaussienne

Distribution des individus (X)
$\mu = 7{,}2$, $\sigma = 0{,}86$



Distribution d'échantillonnage ($\overline{X}$)
N=50

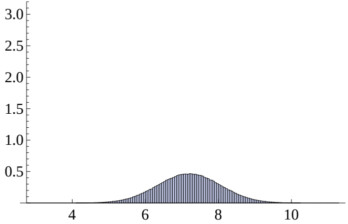


Théorème centrale limite I

Distribution des valeurs individuelles gaussienne

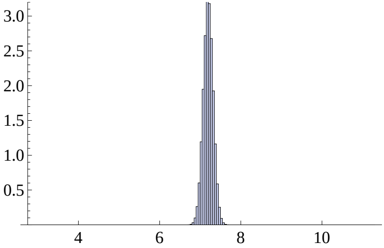
Distribution des individus (X)

$\mu = 7{,}2$, $\sigma = 0{,}86$



Distribution d'échantillonnage ($\overline{X}$)

N=50



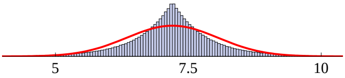
Conlusions :

- $\mu_{\overline{X}} = \mu$
- $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{N}}$ erreur standard de la moyenne (ESM)

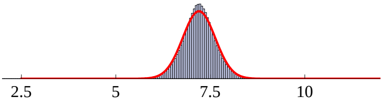
Théorème centrale limite II

Distribution des valeurs individuelles non gaussienne

Distribution des individus (X)
$\mu =$ 7,2, $\sigma =$ 0,86



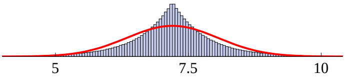
Distribution d'échantillonnage ($\overline{X}$)
N=4



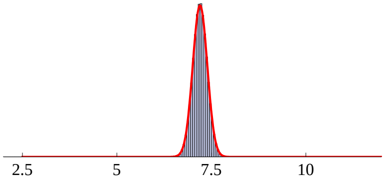
Théorème centrale limite II

Distribution des valeurs individuelles non gaussienne

Distribution des individus (X)
$\mu = 7{,}2$, $\sigma = 0{,}86$



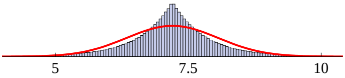
Distribution d'échantillonnage ($\overline{X}$)
N=20



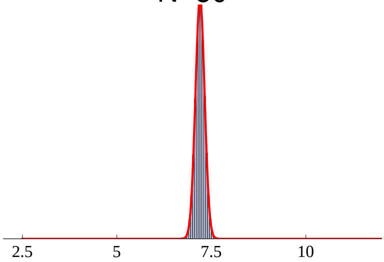
Théorème centrale limite II

Distribution des valeurs individuelles non gaussienne

Distribution des individus (X)
$\mu = 7{,}2$, $\sigma = 0{,}86$



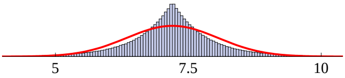
Distribution d'échantillonnage ($\overline{X}$)
N=50



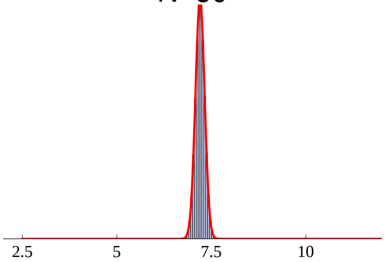
Théorème centrale limite II

Distribution des valeurs individuelles non gaussienne

Distribution des individus (X)
$\mu = 7{,}2$, $\sigma = 0{,}86$



Distribution d'échantillonnage ($\overline{X}$)
N=50



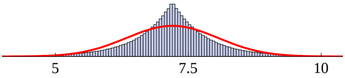
Conlusions :

- N grand $\Rightarrow$ Distribution d'échantillonnage va vers une gaussienne

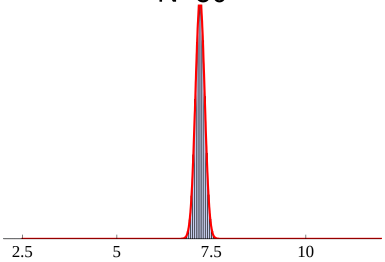
Théorème centrale limite II

Distribution des valeurs individuelles non gaussienne

Distribution des individus (X)
$\mu = 7{,}2,\ \sigma = 0{,}86$



Distribution d'échantillonnage ($\overline{X}$)
N=50



Conlusions :

- N grand $\Rightarrow$ Distribution d'échantillonnage va vers une gaussienne
- $\mu_{\overline{X}} = \mu$
- $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{N}}$ erreur standard de la moyenne (ESM)

Exemple

Exemple 4

QI moyen de la population : $\mu_0 = 100$,
écart-type pour la popluation : $\sigma_0 = 15$.

Un savant fou développe un médicament pouvant augmenter le QI. 30 personnes prennent ce médicament, ensuite on teste leur QI.
QI moyen de ces 30 personnes : $\overline{x} = 140$.

Ce médicament augmente-t-il en moyenne le QI ?

Test d'hypothèse : principe

Principe : Confronter deux hypothèse contradictoire

Exemple 4

Test d'hypothèse : principe

Principe : Confronter deux hypothèse contradictoire

Exemple 4

- Formulation de l'hypothèse nulle et alternative :

Test d'hypothèse : principe

Principe : Confronter deux hypothèse contradictoire

Exemple 4

- Formulation de l'hypothèse nulle et alternative :
  - H_0 : $\mu = 100$
  - H_a : $\mu > 100$

Test d'hypothèse : principe

Principe : Confronter deux hypothèse contradictoire

Exemple 4

- Formulation de l'hypothèse nulle et alternative :
  - $H_0 : \mu = 100$
  - $H_a : \mu > 100$
- Les données donnent $\overline{x} > \mu_0$
- A la fin, on décide de rejeter l'hypothèse nulle ou non

Test d'hypothèse : principe

Principe : Confronter deux hypothèse contradictoire

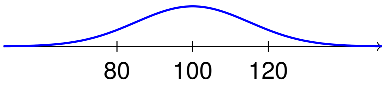
Exemple 4

- Formulation de l'hypothèse nulle et alternative :
  - $H_0 : \mu = 100$
  - $H_a : \mu > 100$
- Les données donnent $\overline{x} > \mu_0$

- A la fin, on décide de rejeter l'hypothèse nulle ou non
- Quand doit-on rejeter H_0 ou non ?

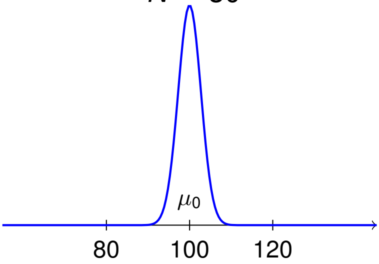
Exemple

Distribution des individus si H_0 vraie



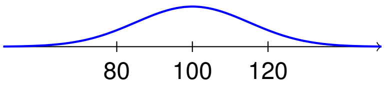
Distribution d'échantillonnage si H_0

$N = 30$



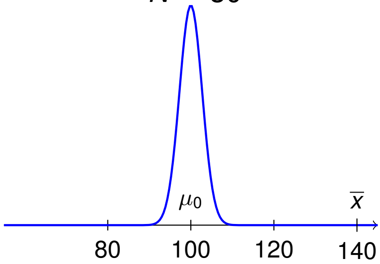
Exemple

Distribution des individus si H_0 vraie



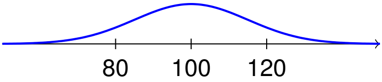
Distribution d'échantillonnage si H_0

$N = 30$



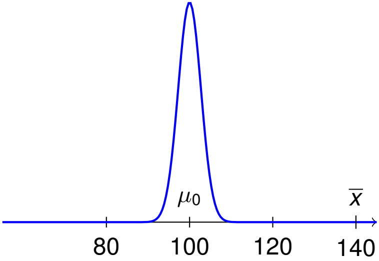
Exemple

Distribution des individus si H_0 vraie



Distribution d'échantillonnage si H_0

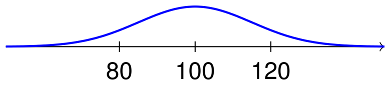
$N = 30$



- *p*-valeur : probabilité qu'on obtienne un échantillon de moyenne ≥ 140 (ici $p = 0{,}00003$)

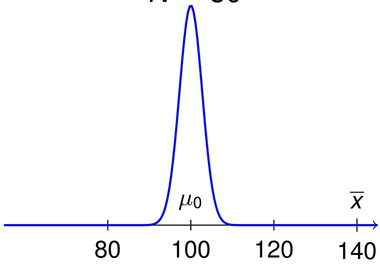
Exemple

Distribution des individus si H_0 vraie



Distribution d'échantillonnage si H_0

$N = 30$

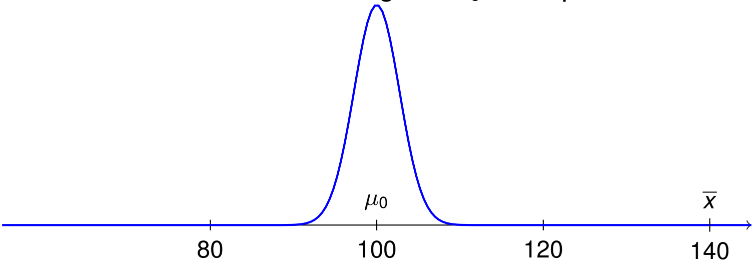


- p-valeur : probabilité qu'on obtienne un échantillon de moyenne ≥ 140 (ici $p = 0{,}00003$)
- On rejette H_0 si p est suffisamment petit

Le risque de 1ère espèce

Que veut dire "*p* suffisamment petit" ?

Distribution d'échantillonnage si H_0 vraie pour $N = 30$

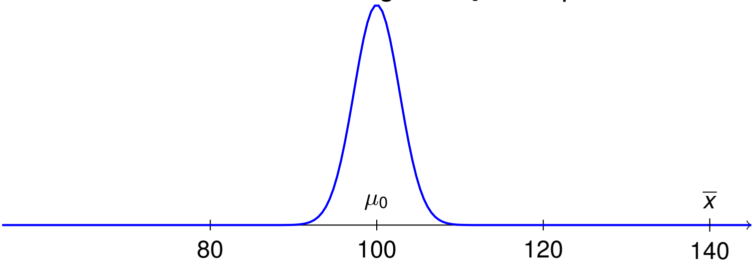


Le risque de 1ère espèce

Que veut dire "p suffisamment petit" ?

- Cela dépend du risque qu'on accepte de prendre

Distribution d'échantillonnage si H_0 vraie pour $N = 30$

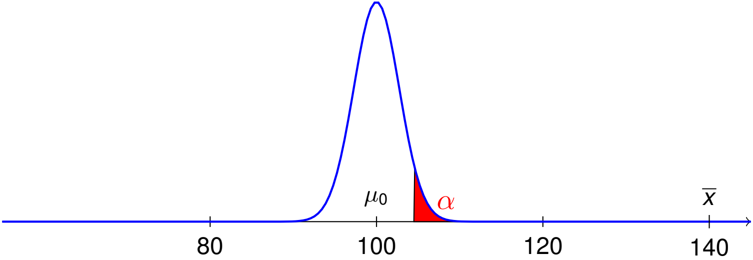


Le risque de 1ère espèce

Que veut dire "*p* suffisamment petit" ?

- Cela dépend du risque qu'on accepte de prendre
- On détermine ce risque α avant de commencer le test
- α est le risque de 1ère espèce

Distribution d'échantillonnage si H_0 vraie pour $N = 30$



Choix standard pour α

- On choisit souvent $\alpha = 0{,}05$

Choix standard pour α

- On choisit souvent $\alpha = 0{,}05$
- Si $p < 0{,}05$: le résultat est significatif
- Si $p < 0{,}01$: le résultat est hautement significatif
- Si $p < 0{,}001$: le résultat est très hautement significatif

Statistique vs clinique

Notre savant fou conclue que son traitement augmente le QI. Est-ce une augmentation suffisante ?

Statistiquement significatif $\neq$ cliniquement significatif

La différence réelle peut être petite, même si p est très petit !

Erreurs possible

	Réalité	
	H_0 vraie	H_0 fausse
NRH_0	☺	Erreur de 2ème espèce ☹
RH_0	Erreur de 1ère espèce ☹	☺

Erreur de 2ème espèce

- Erreur de 2ème espèce : Ne pas rejeter H_0 alors H_0 fausse

Erreur de 2ème espèce

- Erreur de 2ème espèce : Ne pas rejeter H_0 alors H_0 fausse
- β : risque de faire une erreur de 2ème espèce.
- $1 - \beta$: puissance du test.

Erreur de 2ème espèce

- Erreur de 2ème espèce : Ne pas rejeter H_0 alors H_0 fausse
- β : risque de faire une erreur de 2ème espèce.
- $1 - \beta$: puissance du test.
- On ne sait calculer β que si on connaît la taille de l'effet.

Erreur de 2ème espèce : Exemple I

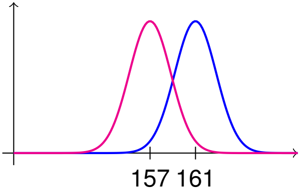
Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille

Erreur de 2ème espèce : Exemple I

Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille

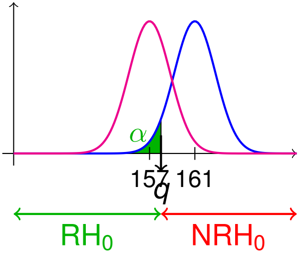


- $\mu_0 = 161$cm, $\mu_a = 157$cm $\Rightarrow \beta = 0{,}293$

Erreur de 2ème espèce : Exemple I

Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille

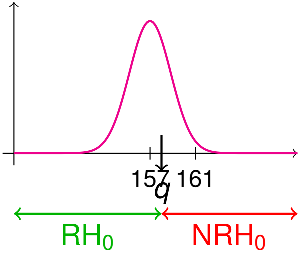


- $\mu_0 = 161$cm, $\mu_a = 157$cm $\Rightarrow \beta = 0{,}293$

Erreur de 2ème espèce : Exemple I

Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille

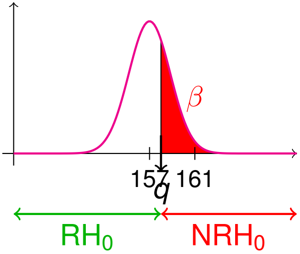


- $\mu_0 = 161$cm, $\mu_a = 157$cm $\Rightarrow \beta = 0{,}293$
- $\mu_0 = 161$cm, $\mu_a = 155$cm $\Rightarrow \beta = 0{,}050$

Erreur de 2ème espèce : Exemple I

Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille

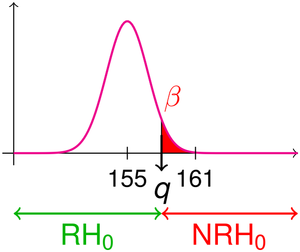


- $\mu_0 = 161$cm, $\mu_a = 157$cm $\Rightarrow \beta = 0{,}293$
- $\mu_0 = 161$cm, $\mu_a = 155$cm $\Rightarrow \beta = 0{,}050$
- $\mu_0 = 161$cm, $\mu_a = 153$cm $\Rightarrow \beta = 0{,}003$

Erreur de 2ème espèce : Exemple I

Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille

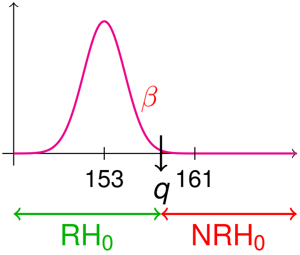


- $\mu_0 = 161$cm, $\mu_a = 157$cm $\Rightarrow \beta = 0{,}293$
- $\mu_0 = 161$cm, $\mu_a = 155$cm $\Rightarrow \beta = 0{,}050$
- $\mu_0 = 161$cm, $\mu_a = 153$cm $\Rightarrow \beta = 0{,}003$

Erreur de 2ème espèce : Exemple I

Exemple 5

En 2002, 2 chercheurs français ont étudié l'impact de l'entraînement intensif sur les jeunes gymnaste féminines de haut niveau.
Ces jeunes gymnastes de haut niveau sont-elles en moyennes plus petites que la moyenne des filles de 15 ans ?
En moyenne les filles de 15 ans ont une taille $\mu_0 = 161$cm, $\sigma = 10$cm
On regarde 30 gymnastes de 15 ans.
Déterminons β pour de valeurs différentes supposition de taille



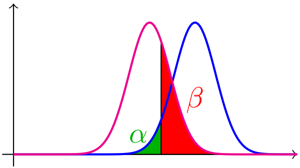
- $\mu_0 = 161$cm, $\mu_a = 157$cm $\Rightarrow \beta = 0{,}293$
- $\mu_0 = 161$cm, $\mu_a = 155$cm $\Rightarrow \beta = 0{,}050$
- $\mu_0 = 161$cm, $\mu_a = 153$cm $\Rightarrow \beta = 0{,}003$

Erreur de 2ème espèce : Exemple II

On peut diminuer β en agrandissant N

Erreur de 2ème espèce : Exemple II

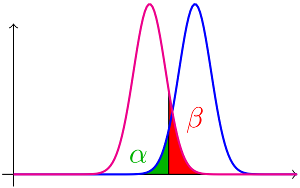
On peut diminuer β en agrandissant N



- $\mu_0 = 161\text{cm}$, $\mu_a = 157\text{cm}$
- $N = 30 \Rightarrow \beta = 0{,}293$

Erreur de 2ème espèce : Exemple II

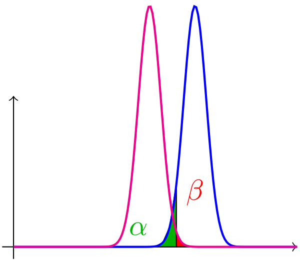
On peut diminuer β en agrandissant N



- $\mu_0 = 161\text{cm}$, $\mu_a = 157\text{cm}$
- $N = 30 \Rightarrow \beta = 0{,}293$
- $N = 50 \Rightarrow \beta = 0{,}118$

Erreur de 2ème espèce : Exemple II

On peut diminuer β en agrandissant N



- $\mu_0 = 161\text{cm}$, $\mu_a = 157\text{cm}$
- $N = 30 \Rightarrow \beta = 0{,}293$
- $N = 50 \Rightarrow \beta = 0{,}118$
- $N = 100 \Rightarrow \beta = 0{,}009$

Réunir statistique et clinique

- Déterminer ce qui est cliniquement significatif
- Choisir N tel que que $1-\beta$ soit grand si c'est cliniquement significatif