Des différences F/H significatives ?
Introduction aux tests statistiques !

Tiffany Cherchi

Mars 2022
Women's day ... is not enough

Pré-requis : Vocabulaire en statistiques

Test statistique : comparaison de deux proportions

Test d'indépendance : de deux caractéristiques

Conclusions

Un peu de vocabulaire statistique

Population

↓

Ensemble exhaustif

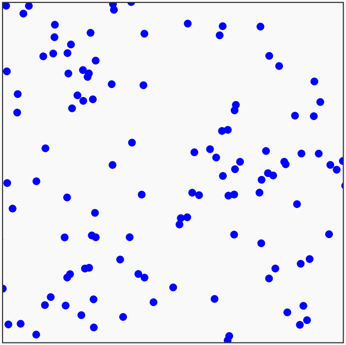
↓

Paramètre théorique θ^*

↓

μ^*, p^*, σ^*

Population



Sample

Un peu de vocabulaire statistique

Population $\xrightarrow{\text{Tirage}}$ Échantillon

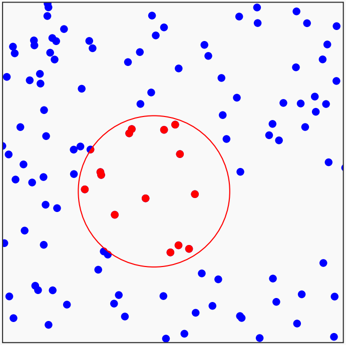
Population

Ensemble exhaustif

Paramètre théorique θ^*

μ^*, p^*, σ^*

Population



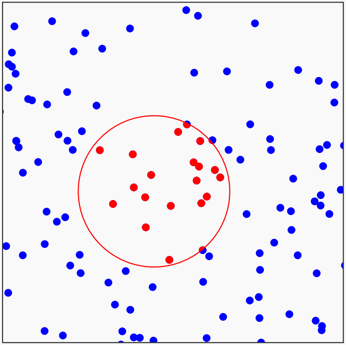
Échantillon (n)

Un peu de vocabulaire statistique

Population $\xrightarrow{\text{Tirage}}$ Échantillon

Population	Échantillon
Ensemble exhaustif	Sous-ensemble représentatif
Paramètre théorique θ^*	Paramètre *estimé* $\hat{\theta}$
μ^*, p^*, σ^*	$\hat{\mu}$, $\hat{p}$, $\hat{\sigma}$

Population



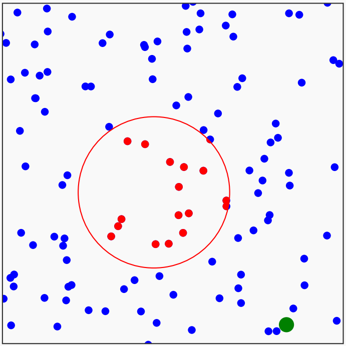
Échantillon (n)

Un peu de vocabulaire statistique

Population $\xrightarrow{\text{Tirage}}$ Échantillon

Population	Échantillon
Ensemble exhaustif	Sous-ensemble représentatif
Paramètre théorique θ^*	Paramètre *estimé* $\hat{\theta}$
μ^*, p^*, σ^*	$\hat{\mu}$, $\hat{p}$, $\hat{\sigma}$

Population

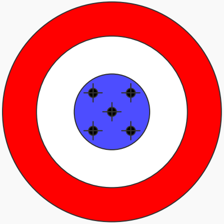


Échantillon (n)

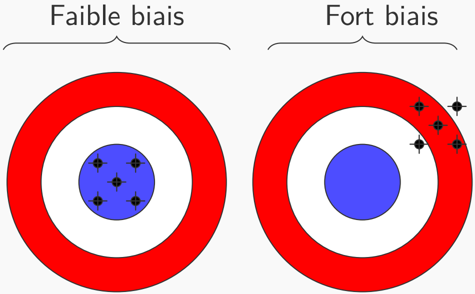
"Moi je connais quelqu'un qui.."

Qualités d'un estimateur

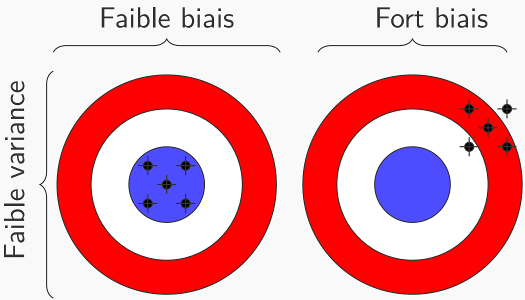
Faible biais



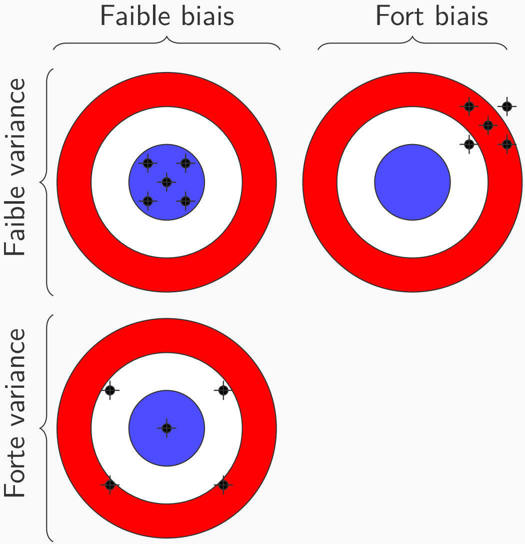
Qualités d'un estimateur



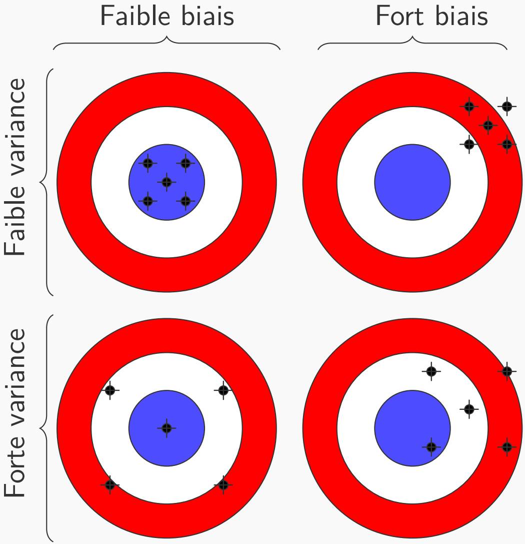
Qualités d'un estimateur



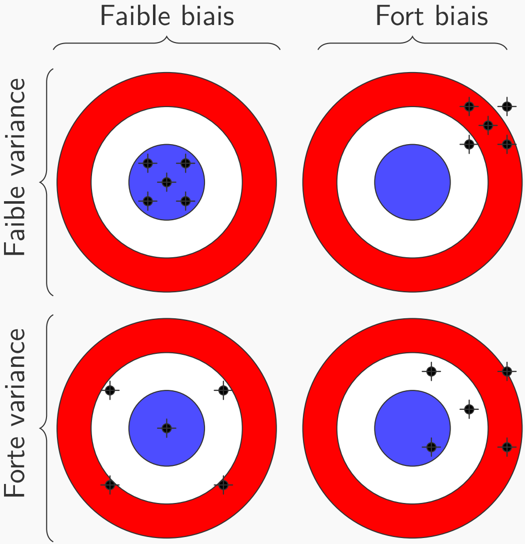
Qualités d'un estimateur



Qualités d'un estimateur



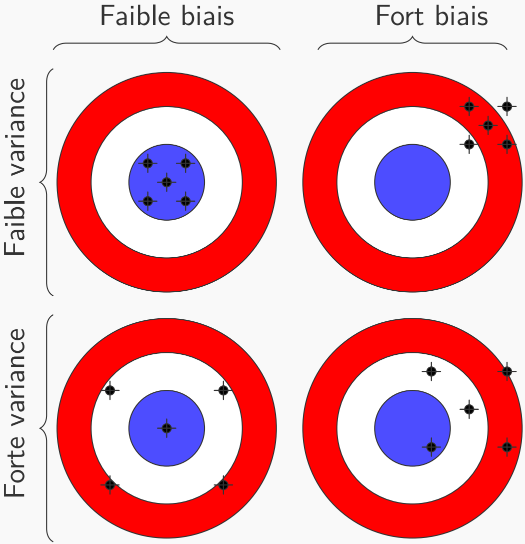
Qualités d'un estimateur



Des estimateurs sans biais :

- moyenne : $\hat{\mu} = \frac{1}{n} \sum_1^n x_i$

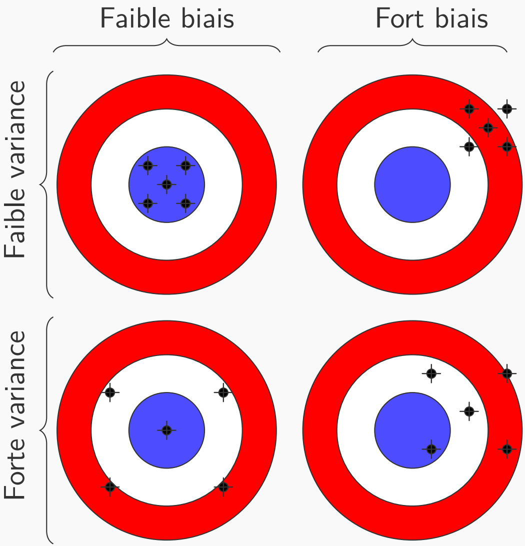
Qualités d'un estimateur



Des estimateurs sans biais :

- moyenne : $\hat{\mu} = \frac{1}{n} \sum_1^n x_i$
- proportion : $\hat{p} = \frac{\#succes}{n}$

Qualités d'un estimateur

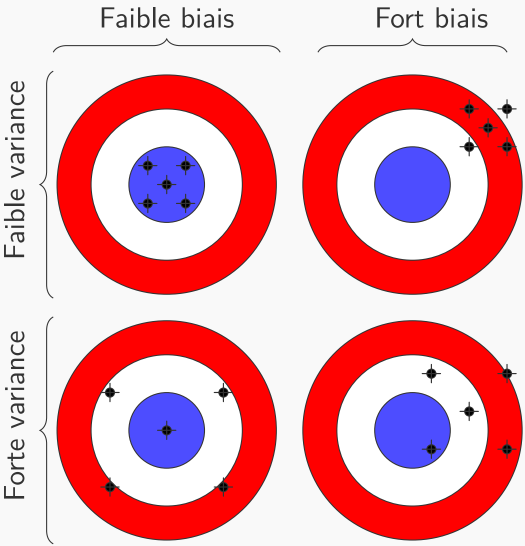


Des estimateurs sans biais :

- moyenne : $\hat{\mu} = \frac{1}{n}\sum_1^n x_i$
- proportion : $\hat{p} = \frac{\#succes}{n}$
- variance (corrigée) :

$$\hat{\sigma}^2 = \frac{\sum_1^n (x_i - \hat{\mu})^2}{n-1}$$

Qualités d'un estimateur



Des estimateurs sans biais :

- moyenne : $\hat{\mu} = \frac{1}{n}\sum_1^n x_i$
- proportion : $\hat{p} = \frac{\#succes}{n}$
- variance (corrigée) :

$$\hat{\sigma}^2 = \frac{\sum_1^n (x_i - \hat{\mu})^2}{n-1}$$

Un estimateur biaisé

- variance :

$$\hat{s}^2 = \frac{\sum_1^n (x_i - \hat{\mu})^2}{n}$$

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0, 1]$,

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0, 1]$,
- la variance (de l'estimateur) est : $\hat{\sigma}^2 = \frac{p \times (1-p)}{n}$,

Données binaires : estimer une proportion

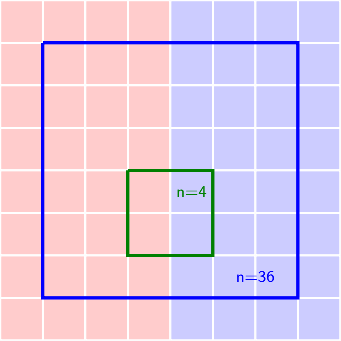
Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0, 1]$,
- la variance (de l'estimateur) est : $\hat{\sigma}^2 = \frac{p \times (1-p)}{n}$,

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0, 1]$,
- la variance (de l'estimateur) est : $\hat{\sigma}^2 = \frac{p \times (1-p)}{n}$,



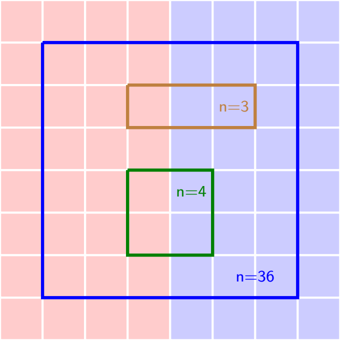
$\hat{p} = \frac{1}{2}$, $\hat{\sigma} = \sqrt{\frac{0.5 \times (1-0.5)}{4}} = 0.25.$

$\hat{p} = \frac{1}{2}$, $\hat{\sigma} = \sqrt{\frac{0.5 \times (1-0.5)}{36}} \approx 0,08.$

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0, 1]$,
- la variance (de l'estimateur) est : $\hat{\sigma}^2 = \frac{p \times (1-p)}{n}$,



$\hat{p} = \frac{1}{2}$, $\hat{\sigma} = \sqrt{\frac{0.5 \times (1-0.5)}{4}} = 0.25.$

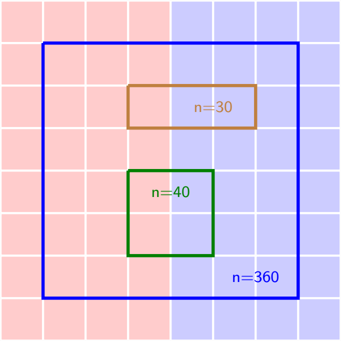
$\hat{p} = \frac{1}{2}$, $\hat{\sigma} = \sqrt{\frac{0.5 \times (1-0.5)}{36}} \approx 0,08.$

$\hat{p} = \frac{1}{3}$, $\hat{\sigma} = \sqrt{\frac{0.33 \times (1-0.33)}{3}} \approx 0,27.$

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0,1]$,
- la variance (de l'estimateur) est : $\hat{\sigma}^2 = \frac{p\times(1-p)}{n}$,



$\hat{p} = \frac{1}{2}$, $\hat{\sigma} \approx 0,079$.

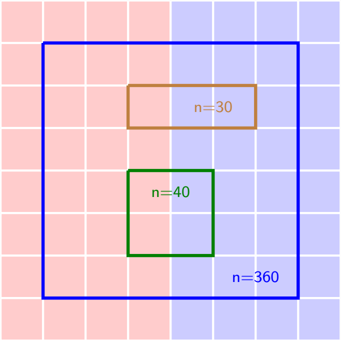
$\hat{p} = \frac{1}{2}$, $\hat{\sigma} \approx 0,026$.

$\hat{p} = \frac{1}{3}$, $\hat{\sigma} \approx 0,086$.

Données binaires : estimer une proportion

Pour un échantillon de n données, de valeurs 0 (échec) ou 1 (succès) :

- un estimateur de la proportion de succes est : $\hat{p} = \frac{\#succes}{n} \in [0, 1]$,
- la variance (de l'estimateur) est : $\hat{\sigma}^2 = \frac{p \times (1-p)}{n}$,
- un IC* à 95% pour $\hat{p}$ est : $[\hat{p} \pm 1.96 \times \hat{\sigma}]$.



$\hat{p} = \frac{1}{2}$, $\hat{\sigma} \approx 0,079$, $IC_{95}(\hat{p}) = [0.35; 0.65]$.

$\hat{p} = \frac{1}{2}$, $\hat{\sigma} \approx 0,026$, $IC_{95}(\hat{p}) = [0.45; 0.55]$.

$\hat{p} = \frac{1}{3}$, $\hat{\sigma} \approx 0,086$, $IC_{95}(\hat{p}) = [0.16; 0.50]$.

*$n \geq 30, n\hat{p} \geq 5, n(1-\hat{p}) \geq 5$.

Test d'hypothèse : comparer une propotion à une référence

Contexte : comparer une proportion estimée $\hat{p}$ sur un n-échantillon, avec une proportion de référence p_0.

Test d'hypothèse : comparer une propotion à une référence

Contexte : comparer une proportion estimée $\hat{p}$ sur un n-échantillon, avec une proportion de référence p_0.

Exemple : Depuis la réforme du lycée, la part de filles, parmis les élèves de maths, est elle significativement différente de la parité, $p_0 = 50\%$?

Test d'hypothèse : comparer une propotion à une référence

Contexte : comparer une proportion estimée $\hat{p}$ sur un n-échantillon, avec une proportion de référence p_0.

Exemple : Depuis la réforme du lycée, la part de filles, parmis les élèves de maths, est elle significativement différente de la parité, $p_0 = 50\%$?

Hypothèses : $\mathcal{H}_0 : p_0 = \hat{p}$ contre $\mathcal{H}_1 : p_0 \neq \hat{p}$.

Test d'hypothèse : comparer une propotion à une référence

Contexte : comparer une proportion estimée $\hat{p}$ sur un n-échantillon, avec une proportion de référence p_0.

Exemple : Depuis la réforme du lycée, la part de filles, parmis les élèves de maths, est elle significativement différente de la parité, $p_0 = 50\%$?

Hypothèses : $\mathcal{H}_0 : p_0 = \hat{p}$ contre $\mathcal{H}_1 : p_0 \neq \hat{p}$.

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim \mathcal{N}(0,1)$ sous l'hypothèse $\mathcal{H}_0$.

Test d'hypothèse : comparer une propotion à une référence

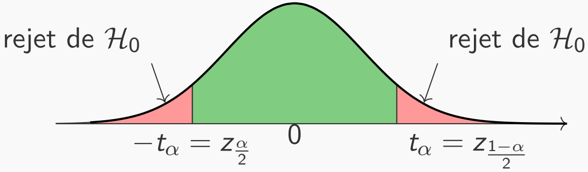
Contexte : comparer une proportion estimée $\hat{p}$ sur un n-échantillon, avec une proportion de référence p_0.

Exemple : Depuis la réforme du lycée, la part de filles, parmis les élèves de maths, est elle significativement différente de la parité, $p_0 = 50\%$?

Hypothèses : $\mathcal{H}_0 : p_0 = \hat{p}$ contre $\mathcal{H}_1 : p_0 \neq \hat{p}$.

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim \mathcal{N}(0,1)$ sous l'hypothèse $\mathcal{H}_0$.

Conclusion : au risque $\alpha \in [0,1]$, $\exists! t_\alpha > 0$ tq $P(-t_\alpha \leq \hat{z} \leq t_\alpha) = 1 - \alpha$.
On rejette $\mathcal{H}_0$ si $\hat{z} < -t_\alpha$ ou $\hat{z} > t_\alpha$.



Test d'hypothèse : comparer une propotion à une référence

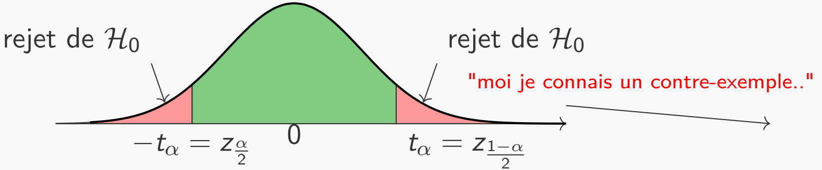
Contexte : comparer une proportion estimée $\hat{p}$ sur un n-échantillon, avec une proportion de référence p_0.

Exemple : Depuis la réforme du lycée, la part de filles, parmis les élèves de maths, est elle significativement différente de la parité, $p_0 = 50\%$?

Hypothèses : $\mathcal{H}_0 : p_0 = \hat{p}$ contre $\mathcal{H}_1 : p_0 \neq \hat{p}$.

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim \mathcal{N}(0,1)$ sous l'hypothèse $\mathcal{H}_0$.

Conclusion : au risque $\alpha \in [0,1]$, $\exists ! t_\alpha > 0$ tq $P(-t_\alpha \leq \hat{z} \leq t_\alpha) = 1 - \alpha$. On rejette $\mathcal{H}_0$ si $\hat{z} < -t_\alpha$ ou $\hat{z} > t_\alpha$.



Exemples (1) : part des filles en maths au lycée

Sur les 149540 élèves qui ont choisi les mathématiques comme spécialité au baccalauréat 2021, 62390 sont des filles. [source]

Exemples (1) : part des filles en maths au lycée

Sur les 149540 élèves qui ont choisi les mathématiques comme spécialité au baccalauréat 2021, 62390 sont des filles. [source]

Proportion : estimation de la part de filles parmis les élèves de maths :

$$\hat{p} = \frac{62390}{149540} \approx 0,417 \text{ avec } IC_{95}(\hat{p}) \approx [0.414; 0.419].$$

Exemples (1) : part des filles en maths au lycée

Sur les 149540 élèves qui ont choisi les mathématiques comme spécialité au baccalauréat 2021, 62390 sont des filles. [source]

Proportion : estimation de la part de filles parmis les élèves de maths :

$$\hat{p} = \frac{62390}{149540} \approx 0,417 \text{ avec } IC_{95}(\hat{p}) \approx [0.414; 0.419].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 65.$

Exemples (1) : part des filles en maths au lycée

Sur les 149540 élèves qui ont choisi les mathématiques comme spécialité au baccalauréat 2021, 62390 sont des filles. [source]

Proportion : estimation de la part de filles parmis les élèves de maths :

$$\hat{p} = \frac{62390}{149540} \approx 0,417 \text{ avec } IC_{95}(\hat{p}) \approx [0.414; 0.419].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 65.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.

Exemples (1) : part des filles en maths au lycée

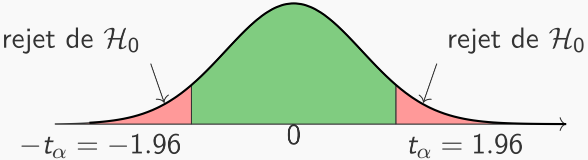
Sur les 149540 élèves qui ont choisi les mathématiques comme spécialité au baccalauréat 2021, 62390 sont des filles. [source]

Proportion : estimation de la part de filles parmis les élèves de maths :

$$\hat{p} = \frac{62390}{149540} \approx 0,417 \text{ avec } IC_{95}(\hat{p}) \approx [0.414; 0.419].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 65.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Exemples (1) : part des filles en maths au lycée

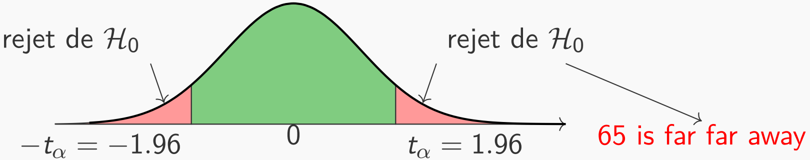
Sur les 149540 élèves qui ont choisi les mathématiques comme spécialité au baccalauréat 2021, 62390 sont des filles. [source]

Proportion : estimation de la part de filles parmis les élèves de maths :

$$\hat{p} = \frac{62390}{149540} \approx 0,417 \text{ avec } IC_{95}(\hat{p}) \approx [0.414; 0.419].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 65.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Exemples (2) : part des femmes (MCF) en section 26

Sur les 1147 MCF en section 26, 387 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{387}{1147} \approx 0,34 \text{ avec } IC_{95}(\hat{p}) \approx [0.31; 0.37].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 11.$

Exemples (2) : part des femmes (MCF) en section 26

Sur les 1147 MCF en section 26, 387 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{387}{1147} \approx 0,34 \text{ avec } IC_{95}(\hat{p}) \approx [0.31; 0.37].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 11.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.

Exemples (2) : part des femmes (MCF) en section 26

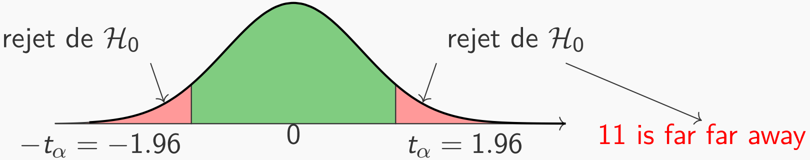
Sur les 1147 MCF en section 26, 387 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{387}{1147} \approx 0,34 \text{ avec } IC_{95}(\hat{p}) \approx [0.31; 0.37].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 11.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Exemples (2) : part des femmes (MCF) en section 26

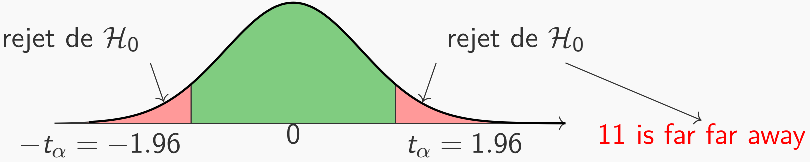
Sur les 1147 MCF en section 26, 387 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{387}{1147} \approx 0,34 \text{ avec } IC_{95}(\hat{p}) \approx [0.31; 0.37].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 11.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Si $n = 1147$, il faudrait 541 femmes, soit $\hat{p} = 0.471$ pour ne pas rejeter $\mathcal{H}_0$.

Exemples (3) : part des femmes (PR) en section 26

Sur les 629 MCF en section 25, 101 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{101}{629} \approx 0,16 \text{ avec } IC_{95}(\hat{p}) \approx [0.13; 0.19].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Exemples (3) : part des femmes (PR) en section 26

Sur les 629 MCF en section 25, 101 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{101}{629} \approx 0,16 \text{ avec } IC_{95}(\hat{p}) \approx [0.13; 0.19].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.

Exemples (3) : part des femmes (PR) en section 26

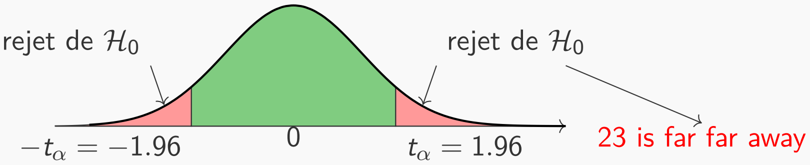
Sur les 629 MCF en section 25, 101 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{101}{629} \approx 0,16 \text{ avec } IC_{95}(\hat{p}) \approx [0.13; 0.19].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Exemples (3) : part des femmes (PR) en section 26

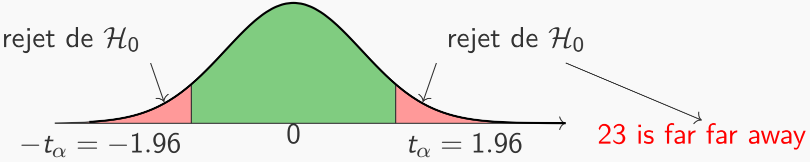
Sur les 629 MCF en section 25, 101 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{101}{629} \approx 0,16 \text{ avec } IC_{95}(\hat{p}) \approx [0.13; 0.19].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Si $n = 629$, il faudrait 290 femmes, soit $\hat{p} = 0.461$ pour ne pas rejeter $\mathcal{H}_0$.

Exemples (4) : part des femmes (MCF) en section 25

Sur les 823 MCF en section 25, 155 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{155}{823} \approx 0,188 \text{ avec } IC_{95}(\hat{p}) \approx [0.162; 0.215].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Exemples (4) : part des femmes (MCF) en section 25

Sur les 823 MCF en section 25, 155 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{155}{823} \approx 0,188 \text{ avec } IC_{95}(\hat{p}) \approx [0.162; 0.215].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.

Exemples (4) : part des femmes (MCF) en section 25

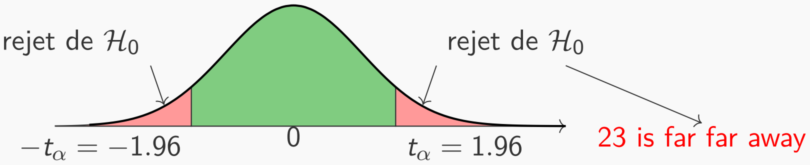
Sur les 823 MCF en section 25, 155 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{155}{823} \approx 0,188 \text{ avec } IC_{95}(\hat{p}) \approx [0.162; 0.215].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Exemples (4) : part des femmes (MCF) en section 25

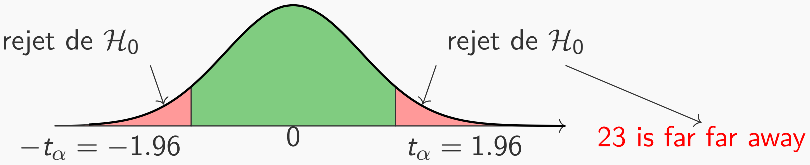
Sur les 823 MCF en section 25, 155 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{155}{823} \approx 0,188 \text{ avec } IC_{95}(\hat{p}) \approx [0.162; 0.215].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 23.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Si n=823, il faudrait 383 femmes, soit $\hat{p} = 0.466$ pour ne pas rejeter $\mathcal{H}_0$.

Exemples (5) : part des femmes (PR) en section 25

Sur les 498 PR en section 25, 31 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{31}{498} \approx 0,06 \text{ avec } IC_{95}(\hat{p}) \approx [0.04; 0.08].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 40.$

Exemples (5) : part des femmes (PR) en section 25

Sur les 498 PR en section 25, 31 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{31}{498} \approx 0,06 \text{ avec } IC_{95}(\hat{p}) \approx [0.04; 0.08].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 40.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.

Exemples (5) : part des femmes (PR) en section 25

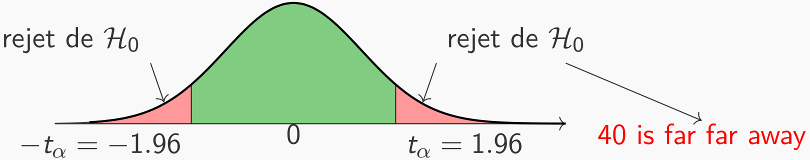
Sur les 498 PR en section 25, 31 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{31}{498} \approx 0,06 \text{ avec } IC_{95}(\hat{p}) \approx [0.04; 0.08].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 40.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Exemples (5) : part des femmes (PR) en section 25

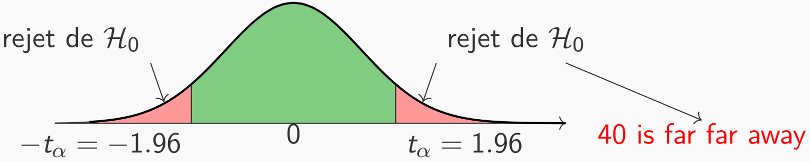
Sur les 498 PR en section 25, 31 sont des femmes. [source]

Proportion : estimation de la part des femmes :

$$\hat{p} = \frac{31}{498} \approx 0,06 \text{ avec } IC_{95}(\hat{p}) \approx [0.04; 0.08].$$

Statistique de test : $\hat{z} = \frac{p_0 - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx 40.$

Conclusion : au risque $\alpha = 5\%$, on a $t_\alpha = 1.96$, et ici **on rejette** $\mathcal{H}_0$.



Si $n = 498$, il faudrait 228 femmes, soit $\hat{p} = 0.457$ pour ne pas rejeter $\mathcal{H}_0$.

Pré-requis : Vocabulaire en statistiques

Test statistique : comparaison de deux proportions

Test d'indépendance : de deux caractéristiques

Conclusions

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

O_{ij}	F	G	total
maths	62 390	87 150	149 540
~~maths~~	147 663	74 502	222 165
total	210 053	161 652	371 705

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

O_{ij}	F	G	total
maths	62 390	87 150	149 540
~~maths~~	147 663	74 502	222 165
total	210 053	161 652	371 705

$$T_{ij} = \frac{(O_{i+} \times O_{+j})}{n}$$
$$O_{i+} = \sum_j O_{ij}$$
$$O_{+j} = \sum_i O_{ij}$$

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

O_{ij}	F	G	total
maths	62 390	87 150	149 540
~~maths~~	147 663	74 502	222 165
total	210 053	161 652	371 705

$$T_{ij} = \frac{(O_{i+} \times O_{+j})}{n}$$
$$O_{i+} = \sum_j O_{ij}$$
$$O_{+j} = \sum_i O_{ij}$$

T_{ij}	F	G	total
maths	84 506	65 034	149 540
~~maths~~	125 547	96 618	222 165
total	210 053	161 652	371 705

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

O_{ij}	F	G	total
maths	62 390	87 150	149 540
$\overline{\text{maths}}$	147 663	74 502	222 165
total	210 053	161 652	371 705

$$T_{ij} = \frac{(O_{i+} \times O_{+j})}{n}$$
$$O_{i+} = \sum_j O_{ij}$$
$$O_{+j} = \sum_i O_{ij}$$

T_{ij}	F	G	total
maths	84 506	65 034	149 540
$\overline{\text{maths}}$	125 547	96 618	222 165
total	210 053	161 652	371 705

Hypothèses : $\mathcal{H}_0 : X \perp\!\!\!\perp Y$ vs $\mathcal{H}_1 : X \not\!\perp\!\!\!\perp Y$, $\hat{z} = \sum_{i,j} \frac{(O_{i,j} - T_{i,j})^2}{T_{i,j}} \sim \mathcal{X}_1^2$ sous $\mathcal{H}_0$.

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

O_{ij}	F	G	total
maths	62 390	87 150	149 540
~~maths~~	147 663	74 502	222 165
total	210 053	161 652	371 705

$$T_{ij} = \frac{(O_{i+} \times O_{+j})}{n}$$
$$O_{i+} = \sum_j O_{ij}$$
$$O_{+j} = \sum_i O_{ij}$$

T_{ij}	F	G	total
maths	84 506	65 034	149 540
~~maths~~	125 547	96 618	222 165
total	210 053	161 652	371 705

Hypothèses : $\mathcal{H}_0 : X \perp\!\!\!\perp Y$ vs $\mathcal{H}_1 : X \not\perp\!\!\!\perp Y$, $\hat{z} = \sum_{i,j} \frac{(O_{i,j} - T_{i,j})^2}{T_{i,j}} \sim \mathcal{X}_1^2$ sous $\mathcal{H}_0$.

Conclusion : au risque α=5%, on rejette $\mathcal{H}_0$ si $\hat{z} > t_\alpha = 3.84$.

Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

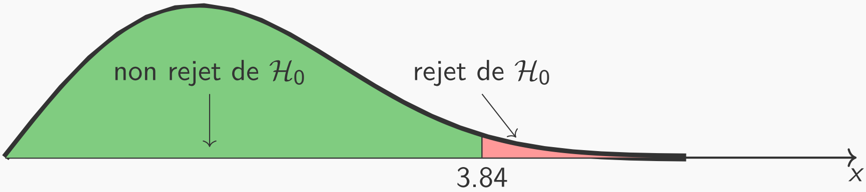
O_{ij}	F	G	total
maths	62 390	87 150	149 540
~~maths~~	147 663	74 502	222 165
total	210 053	161 652	371 705

$$T_{ij} = \frac{(O_{i+} \times O_{+j})}{n}$$
$$O_{i+} = \sum_j O_{ij}$$
$$O_{+j} = \sum_i O_{ij}$$

T_{ij}	F	G	total
maths	84 506	65 034	149 540
~~maths~~	125 547	96 618	222 165
total	210 053	161 652	371 705

Hypothèses : $\mathcal{H}_0 : X \perp\!\!\!\perp Y$ vs $\mathcal{H}_1 : X \not\perp\!\!\!\perp Y$, $\hat{z} = \sum_{i,j} \frac{(O_{i,j} - T_{i,j})^2}{T_{i,j}} \sim \mathcal{X}^2_1$ sous $\mathcal{H}_0$.

Conclusion : au risque α=5%, on rejette $\mathcal{H}_0$ si $\hat{z} > t_\alpha = 3.84$. Ici, $\hat{z} = 22266$.



Test d'indépendance de deux variables

Contexte : comparer des effectifs observés de deux caractéristiques X et Y, avec des effectifs théoriques issus de variables indépendantes.

Exemple : On note X la v.a *genre* et Y la v.a *faire des maths au lycée*.

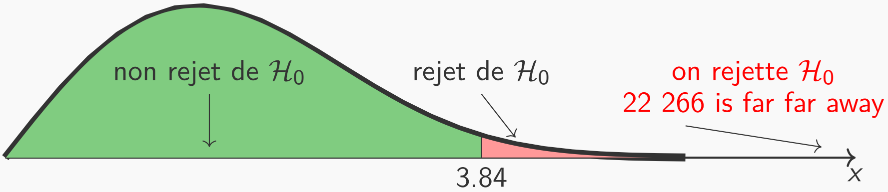
O_{ij}	F	G	total
maths	62 390	87 150	149 540
~~maths~~	147 663	74 502	222 165
total	210 053	161 652	371 705

$$T_{ij} = \frac{(O_{i+} \times O_{+j})}{n}$$
$$O_{i+} = \sum_j O_{ij}$$
$$O_{+j} = \sum_i O_{ij}$$

T_{ij}	F	G	total
maths	84 506	65 034	149 540
~~maths~~	125 547	96 618	222 165
total	210 053	161 652	371 705

Hypothèses : $\mathcal{H}_0 : X \perp\!\!\!\perp Y$ vs $\mathcal{H}_1 : X \not\perp\!\!\!\perp Y$, $\hat{z} = \sum_{i,j} \frac{(O_{i,j} - T_{i,j})^2}{T_{i,j}} \sim \mathcal{X}_1^2$ sous $\mathcal{H}_0$.

Conclusion : au risque α=5%, on rejette $\mathcal{H}_0$ si $\hat{z} > t_\alpha = 3.84$. Ici, $\hat{z} = 22266$.



Pré-requis : Vocabulaire en statistiques

Test statistique : comparaison de deux proportions

Test d'indépendance : de deux caractéristiques

Conclusions

Quelques ressources

- chiffres filles et maths au lycée : [test]https://femmes-et-maths.fr/2022/03/17/30-des-filles-et-54-des-garcons-ont-presente-la-specialite-maths-au-baccalaureat-2021/
- chiffres femmes et maths (MCF & PR 25, 26) : `https://femmes-et-maths.fr/wp-content/uploads/2020/02/journéeParité4BROZE.pdf`
- Pour un regard objectif/quantifié sur les stéréotypes : `https://femmes-et-maths.fr/wp-content/uploads/2021/03/Brochure_Grand_Public_interactif.pdf`