Cours de Statistiques Inférentielles

CQLS : cqls@upmf-grenoble.fr

5 juillet 2014

Plan

$\boldsymbol{A}$pproche $\boldsymbol{E}$xpérimentale des $\boldsymbol{P}$robabilités : l'Expérimentateur versus le Matheux

$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$

L'Expérimentateur :

1. Réaliser m expériences

Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

L'Expérimentateur : $1^{ème}$ réalisation $\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$

1. Réaliser m expériences

$$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)}$$

14.6%

Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

L'Expérimentateur :

1. Réaliser m expériences

$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$

$2^{ème}$ réalisation

$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)}$ 14.6%

$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)}$ 16.3%

Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

L'Expérimentateur :

1. Réaliser m expériences

$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$

$m^{\grave{e}me}$ réalisation

$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)}$	$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)}$	$\cdots\cdots$	$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)}$
14.6%	16.3%		14.1%

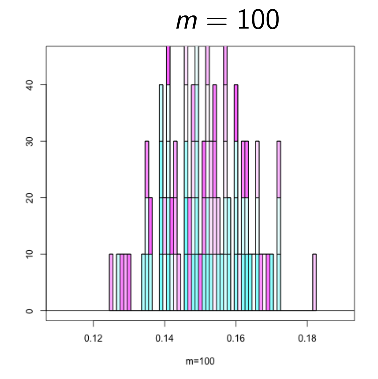
Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

L'Expérimentateur :

1. Réaliser m expériences
2. Répartition des $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$ représentées par m briques de surface $1/m$ et de largeur $1/n$ empilées l'une après l'autre en les centrant en abscisse en leur valeur.)

Future $\widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)$

$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)$	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)$	$\cdots\cdots$	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)$
14.6%	16.3%		14.1%



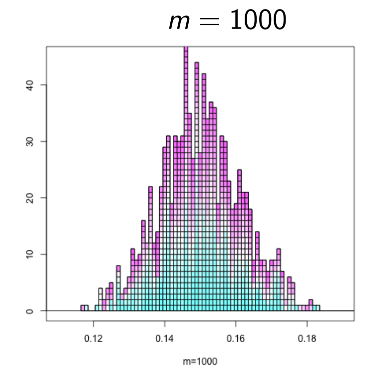
$$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$$

$$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)} \quad \boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)} \quad \cdots\cdots \quad \boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)}$$

14.6% 16.3% 15.1%

L'Expérimentateur :

1. Réaliser m expériences
2. Répartition des $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$ représentées par m briques de surface $1/m$ et de largeur $1/n$ empilées l'une après l'autre en les centrant en abscisse en leur valeur.)



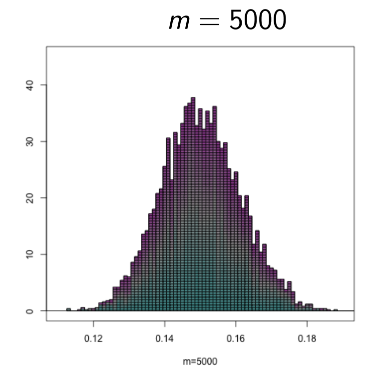
Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

Future $\widehat{\mu^\bullet}(\mathbf{Y}^\bullet)$

$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)$ 14.6% $\quad$ $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)$ 16.3% $\quad \cdots\cdots \quad$ $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)$ 14.7%

L'Expérimentateur :

1. Réaliser m expériences
2. Répartition des $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$ représentées par m briques de surface $1/m$ et de largeur $1/n$ empilées l'une après l'autre en les centrant en abscisse en leur valeur.)



Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

$$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$$

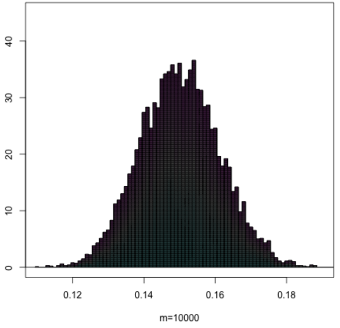
$$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)} \quad \boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)} \quad \cdots\cdots \quad \boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)}$$

14.6% 16.3% 16.3%

L'Expérimentateur :

1. Réaliser m expériences
2. Répartition des $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$ représentées par m briques de surface $1/m$ et de largeur $1/n$ empilées l'une après l'autre en les centrant en abscisse en leur valeur.)

$m = 10000$



Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$

$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)}$ 14.6% $\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)}$ 16.3% $\cdots\cdots$ $\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)}$ $\cdots$

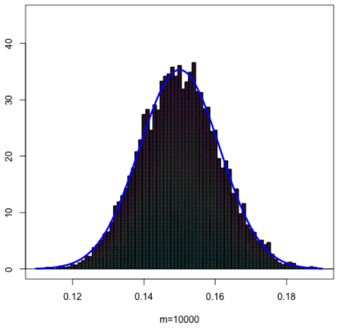
L'Expérimentateur :

1. Réaliser m expériences
2. Répartition des $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$ représentées par m briques de surface $1/m$ et de largeur $1/n$ empilées l'une après l'autre en les centrant en abscisse en leur valeur.)

Le Matheux :

3. Je le savais **à l'avance** pour $m \to +\infty$

$m = +\infty$



Approche Expérimentale des Probabilités : l'Expérimentateur versus le Matheux

L'Expérimentateur :

1. Réaliser m expériences
2. Répartition des $\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$ représentées par m briques de surface $1/m$ et de largeur $1/n$ empilées l'une après l'autre en les centrant en abscisse en leur valeur.)

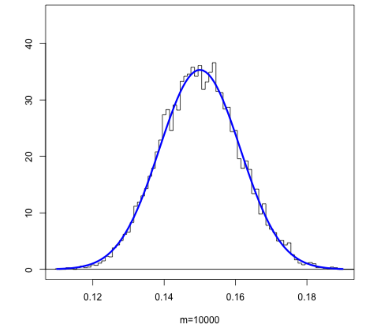
Le Matheux :

3. Je le savais **à l'avance** pour $m \to +\infty$
4. $\widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right) \overset{approx.}{\rightsquigarrow} \mathcal{N}\left(\mu^\bullet, \frac{\sigma_\bullet}{\sqrt{n}}\right)$

$$\boxed{\text{Future } \widehat{\mu^\bullet}\left(\mathbf{Y}^\bullet\right)}$$

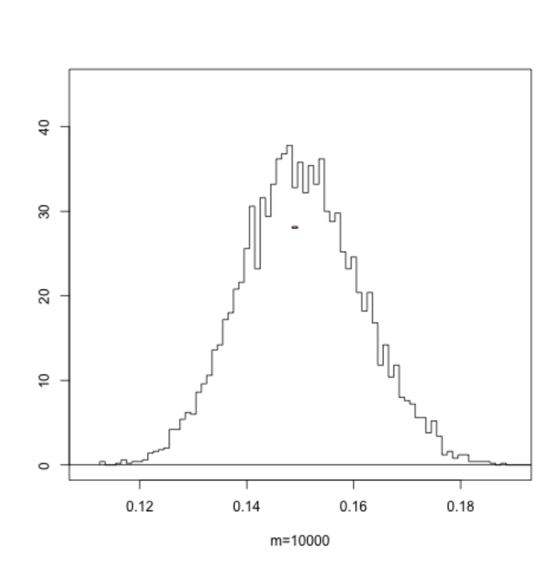
$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[1]}\right)}$	$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[2]}\right)}$	$\cdots\cdots$	$\boxed{\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[m]}\right)}$
14.6%	16.3%		$\cdots$

$m = 10000$ vs $m = +\infty$



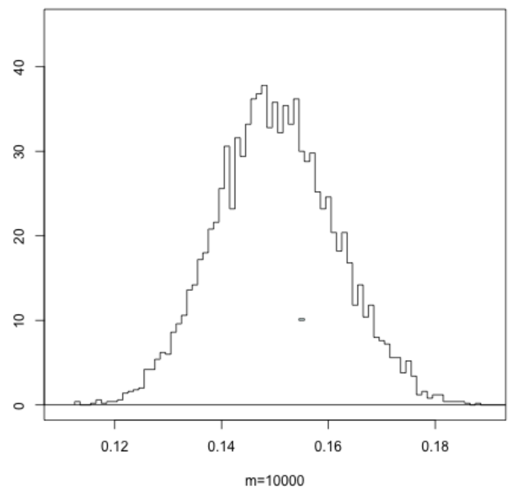
*Réalisation d'une future estimation par l'**Expérimentateur***

j	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$
$\vdots$	$\vdots$
4150	14.4%
4151	17.2%
4152	15%
4153	14.9%
4154	13.7%
4155	15.8%
4156	14.6%
$\vdots$	$\vdots$



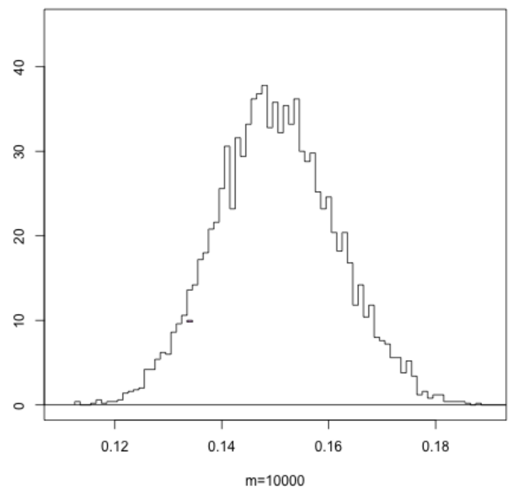
*Réalisation d'une future estimation par l'**Expérimentateur***

j	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$
⋮	⋮
2105	15.3%
2106	14.1%
2107	13.2%
2108	15.5%
2109	16.7%
2110	15.5%
2111	14.5%
⋮	⋮



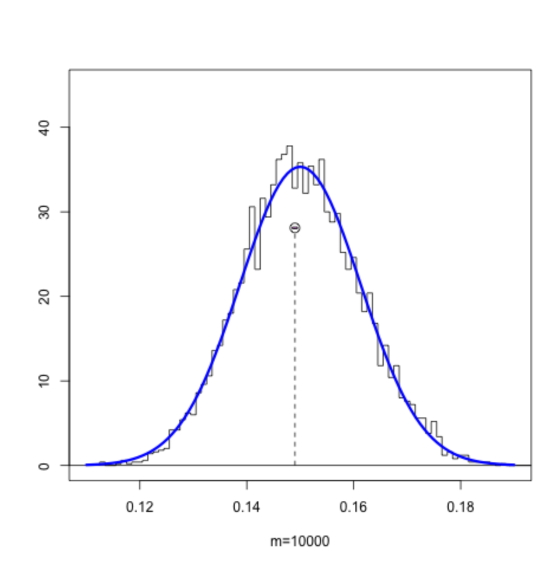
*Réalisation d'une future estimation par l'**Expérimentateur***

j	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$
⋮	⋮
3728	14.9%
3729	14.4%
3730	14.8%
3731	13.4%
3732	14.9%
3733	14.4%
3734	16.4%
⋮	⋮



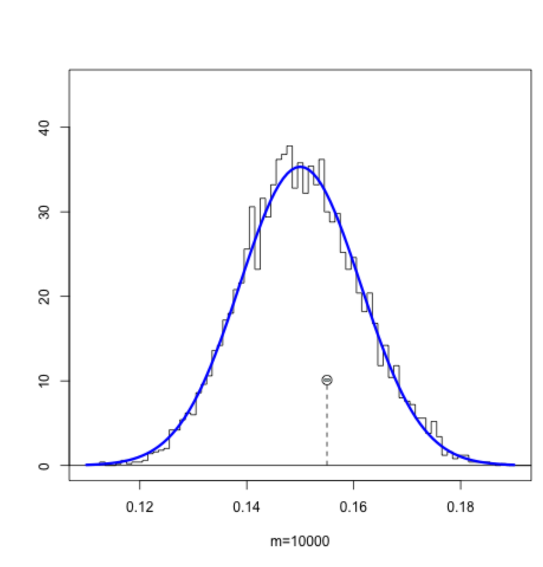
*Réalisation d'une future estimation par le **Matheux***

j	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$
$\vdots$	$\vdots$
4150	14.4%
4151	17.2%
4152	15%
4153	14.9%
4154	13.7%
4155	15.8%
4156	14.6%
$\vdots$	$\vdots$



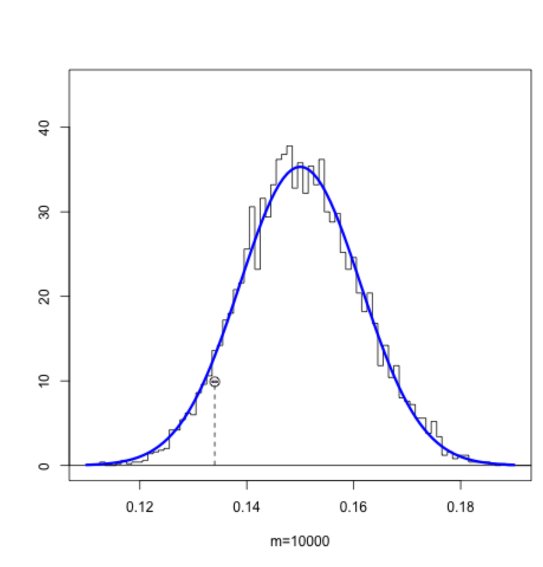
*Réalisation d'une future estimation par le **Matheux***

j	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$
⋮	⋮
2105	15.3%
2106	14.1%
2107	13.2%
2108	15.5%
2109	16.7%
2110	15.5%
2111	14.5%
⋮	⋮



Réalisation d'une future estimation par le ***Matheux***

j	$\widehat{\mu^\bullet}\left(\mathbf{y}^\bullet_{[j]}\right)$
$\vdots$	$\vdots$
3728	14.9%
3729	14.4%
3730	14.8%
3731	13.4%
3732	14.9%
3733	14.4%
3734	16.4%
$\vdots$	$\vdots$



Comment l'industriel doit-il utiliser ces informations ?

Réalisation d'une future estimation

→ L'industriel s'imagine être le **jour J** dans la situation où $\mu^{\bullet} = 0.15$ (juste pas le marché)

Comment l'industriel doit-il utiliser ces informations ?

Réalisation d'une future estimation

→ L'industriel s'imagine être le **jour J** dans la situation où $\mu^\bullet = 0.15$ (juste pas le marché)

→ Il prend alors conscience que ce qui peut lui arriver **le jour J**, c'est équivalent (ou presque) à :

Comment l'industriel doit-il utiliser ces informations ?

Réalisation d'une future estimation

→ L'industriel s'imagine être le **jour J** dans la situation où $\mu^\bullet = 0.15$ (juste pas le marché)

→ Il prend alors conscience que ce qui peut lui arriver **le jour J**, c'est équivalent (ou presque) à :

1. Choisir au hasard une brique (i.e un $\widehat{\mu^\bullet}\left(\mathbf{y_{[j]}}\right)$ parmi les m)

Comment l'industriel doit-il utiliser ces informations ?

Réalisation d'une future estimation

→ L'industriel s'imagine être le **jour J** dans la situation où $\mu^\bullet = 0.15$ (juste pas le marché)

→ Il prend alors conscience que ce qui peut lui arriver **le jour J**, c'est équivalent (ou presque) à :

1. Choisir au hasard une brique (i.e un $\widehat{\mu^\bullet}\left(\mathbf{y_{[j]}}\right)$ parmi les m)
2. Choisir au hasard un point sous la "courbe $\mathcal{N}(\mu^\bullet, \frac{\sigma_\bullet}{\sqrt{n}})$" associé à son abscisse représentant une réalisation au hasard de $\widehat{\mu^\bullet}\left(\mathbf{Y}\right)$ choisie parmi une infinité.

Comment l'industriel doit-il utiliser ces informations ?

Réalisation d'une future estimation

→ L'industriel s'imagine être le **jour J** dans la situation où $\mu^\bullet = 0.15$ (juste pas le marché)

→ Il prend alors conscience que ce qui peut lui arriver **le jour J**, c'est équivalent (ou presque) à :

1. Choisir au hasard une brique (i.e un $\widehat{\mu^\bullet}\left(\mathbf{y_{[j]}}\right)$ parmi les m)
2. Choisir au hasard un point sous la "courbe $\mathcal{N}(\mu^\bullet, \frac{\sigma_\bullet}{\sqrt{n}})$" associé à son abscisse représentant une réalisation au hasard de $\widehat{\mu^\bullet}\left(\mathbf{Y}\right)$ choisie parmi une infinité.

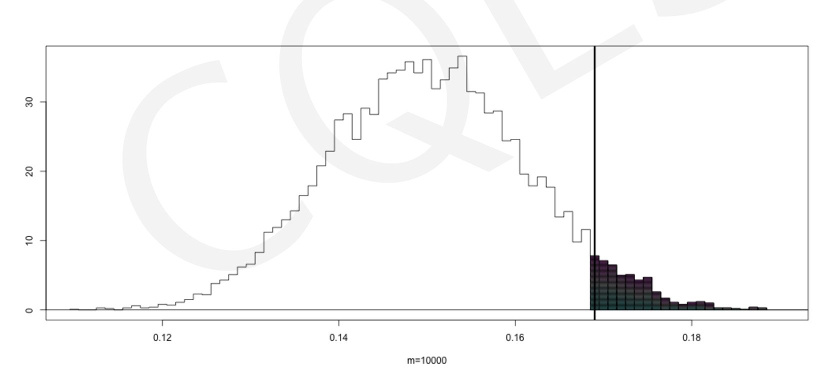
⇒ Il voit clairement la "courbe $\mathcal{N}(\mu^\bullet, \frac{\sigma_\bullet}{\sqrt{n}})$" comme un empilement d'une infinité de briques ("devenues des points") associées à une infinité de réalisations possibles de $\widehat{\mu^\bullet}\left(\mathbf{Y}\right)$.

Produit A : Risque 1ère espèce

$\overline{\left(\widehat{p^A}\left(\mathbf{y}^{\mathbf{A}}_{[\cdot]}\right) > 16.9\%\right)_m} = \text{Prop. des } \left(\widehat{p^A}\left(\mathbf{y}^{\mathbf{A}}_{[\cdot]}\right)\right)_{10000} \text{ supérieurs à } 16.9\%$

$= \frac{1}{m} \times \left(\text{Nbre des } \left(\widehat{p^A}\left(\mathbf{y}^{\mathbf{A}}_{[\cdot]}\right)\right)_m \text{ supérieurs à } 16.9\%\right)$

$= \text{Surface des } \textbf{briques} \text{ associées aux } \left(\widehat{p^A}\left(\mathbf{y}^{\mathbf{A}}_{[\cdot]}\right)\right)_m \text{ supérieurs à } 16.9\%$

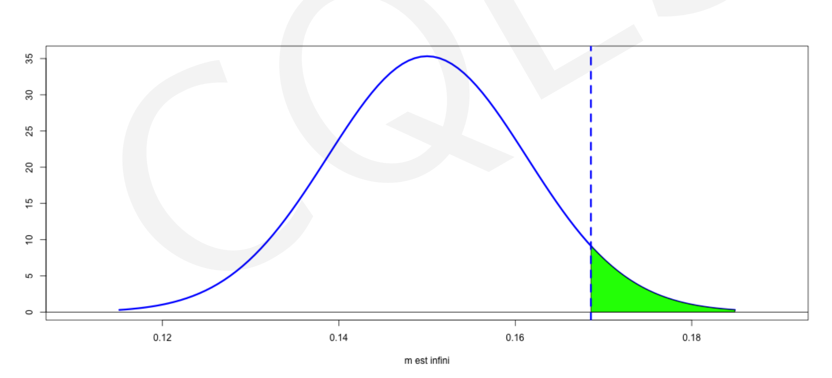


Produit A : Risque 1ère espèce

$P(\widehat{p^A}(\mathbf{Y^A}) > 16.9\%) = \overline{\left(\widehat{p^A}\left(\mathbf{y^A_{[\cdot]}}\right) > 16.9\%\right)}_\infty$

$= \lim_{m\to\infty} \frac{1}{m} \times \left(\text{Nbre des } \left(\widehat{p^A}\left(\mathbf{y^A_{[\cdot]}}\right)\right)_\infty \text{ supérieurs à } 16.9\%\right)$

$\simeq$ Surface des **points** associés aux $\left(\widehat{p^A}\left(\mathbf{y^A_{[\cdot]}}\right)\right)_\infty$ supérieurs à 16.9%

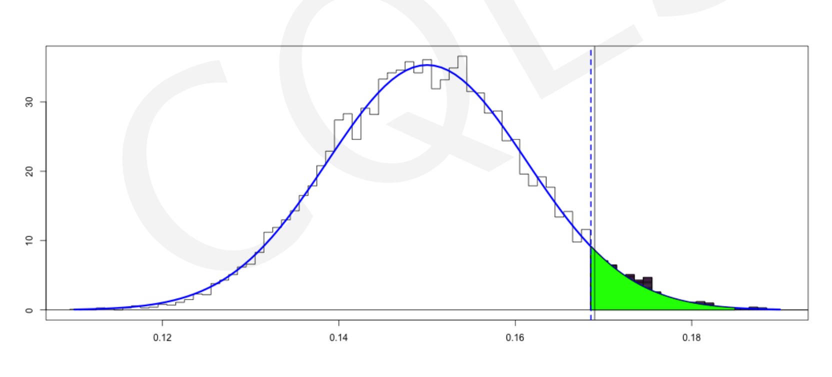


Produit A : Risque 1ère espèce

$$P(\widehat{p^A}(\mathbf{Y^A}) > 16.9\%) \simeq \overline{\left(\widehat{p^A}\left(\mathbf{y^A_{[\cdot]}}\right) > 16.9\%\right)_m}$$

$$= \frac{1}{m} \times \left(\text{Nbre des } \left(\widehat{p^A}\left(\mathbf{y^A_{[\cdot]}}\right)\right)_m \text{ supérieurs à } 16.9\%\right)$$

$$= \text{Surface des } \textbf{briques} \text{ associées aux } \left(\widehat{p^A}\left(\mathbf{y^A_{[\cdot]}}\right)\right)_m \text{ supérieurs à } 16.9\%$$

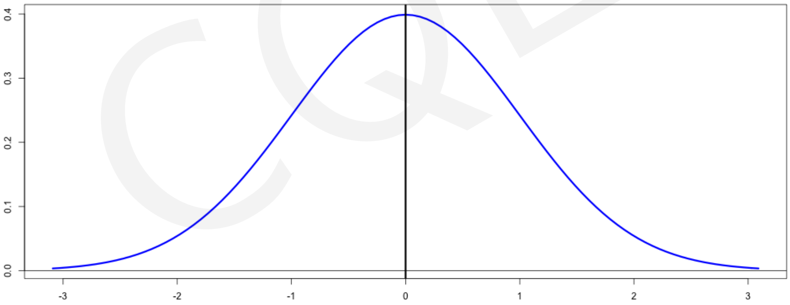


Plan

Exemple diététicien

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ sous $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$



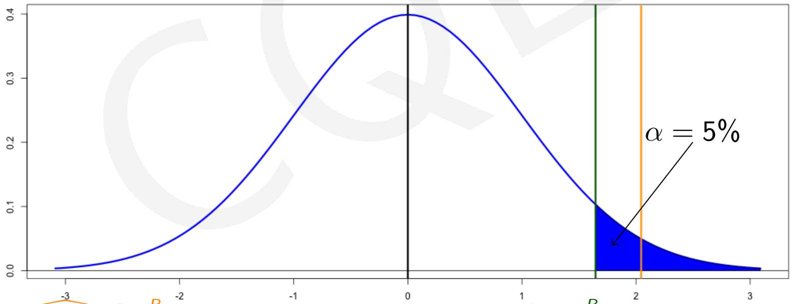
Exemple diététicien

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)

Décision (au vu des $n=50$ données) : Accepter $\mathbf{H_1}$ si $\widehat{\delta_{\mu^D,4}}\left(\mathbf{y^D}\right) > \delta^+_{lim,\alpha}$

Question : Conclure pour $\alpha = 5\%$?

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ sous $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$



$\widehat{\delta_{\mu^D,4}}\left(\mathbf{y^D}\right) \overset{R}{=}$ `(mean(yD)-4)/seMean(yD)=2.045` $> \delta^+_{lim,5\%} \overset{R}{=}$ `qnorm(.95)=1.645`

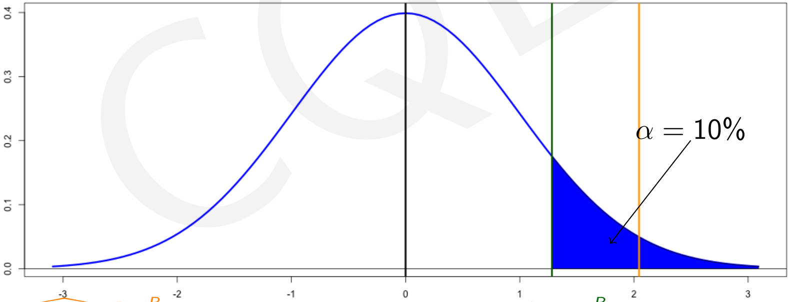
Exemple diététicien

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)

Décision (au vu des $n=50$ données) : Accepter $\mathbf{H_1}$ si $\widehat{\delta_{\mu^D,4}}\left(\mathbf{y^D}\right) > \delta^+_{lim,\alpha}$

Question : Conclure pour $\alpha = 10\%$?

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ sous $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$



$\widehat{\delta_{\mu^D,4}}\left(\mathbf{y^D}\right) \overset{R}{=}$ `(mean(yD)-4)/seMean(yD)=2.045` $> \delta^+_{lim,10\%} \overset{R}{=}$ `qnorm(.9)=1.282`

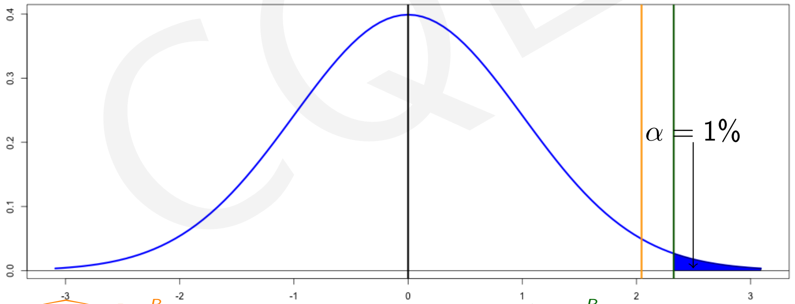
Exemple diététicien

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)

Décision (au vu des $n=50$ données) : Accepter $\mathbf{H_1}$ si $\widehat{\delta_{\mu^D,4}}\left(\mathbf{y^D}\right) > \delta^+_{lim,\alpha}$

Question : Conclure pour $\alpha = 1\%$?

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ sous $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$

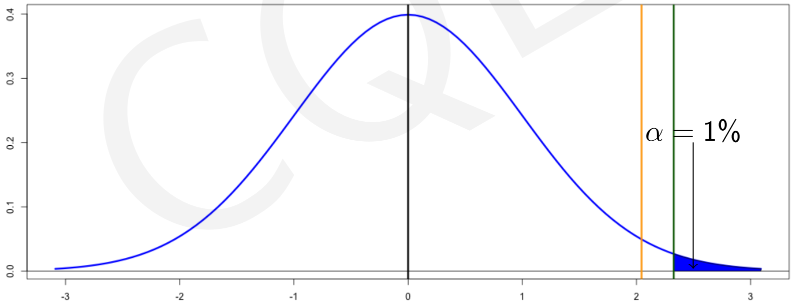


$\widehat{\delta_{\mu^D,4}}\left(\mathbf{y^D}\right) \overset{R}{=}$ (mean(yD)-4)/seMean(yD)=2.045 $\not> \delta^+_{lim,1\%} \overset{R}{=}$ qnorm(.99)=2.326

Exemple diététicien

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)
Question : Quel est le plus petit α (i.e. risque maximal de décider à tort $\mathbf{H_1}$) à encourir pour accepter $\mathbf{H_1}$ (i.e. l'assertion d'intérêt) au vu des $n = 50$ données ?

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ **sous** $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$

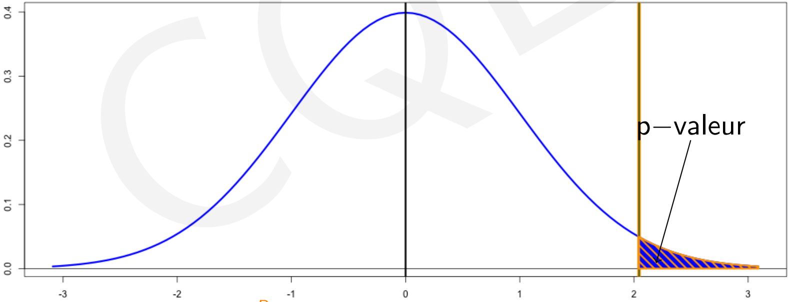


Exemple diététicien

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)

Réponse : **p−valeur= le** plus petit **risque** maximal de décider à tort $\mathbf{H_1}$ à encourir **pour accepter $\mathbf{H_1}$** (i.e. l'assertion d'intérêt).

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ sous $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$

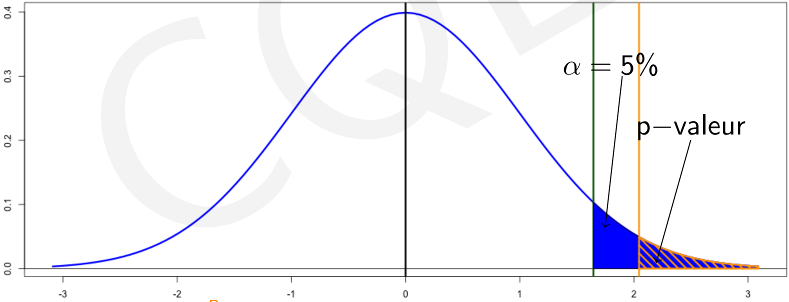


p−valeur(droite)$\overset{R}{=}$ `1-pnorm((mean(yD)-4)/seMean(yD))=2.04%`

Exemple diététicien (fin)

Assertion d'intérêt : Le régime permet une perte de poids de 2 kilos par semaine $\Leftrightarrow$ $\mathbf{H_1}$: $\mu^D > 4$ (en 2 semaines)
Décision (au vu des $n = 50$ données) : Accepter $\mathbf{H_1}$ si **p − valeur** $< \alpha$
i.e. si **le risque pour accepter $\mathbf{H_1}$** est **raisonnablement petit**

Loi de $\widehat{\delta_{\mu^D,4}}\left(\mathbf{Y^D}\right)$ sous $\mathbf{H}_0$: $\mu^D = 4 \Leftrightarrow \delta_{\mu^D,4} = 0$



p−valeur(droite)$\overset{R}{=}$ `1-pnorm((mean(yD)-4)/seMean(yD))=2.04%` $< \alpha = 5\%$