Cours de Statistiques Inférentielles

CQLS : cqls@upmf-grenoble.fr

19 novembre 2013

Plan

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Avant le jour J, l'industriel a-t-il intérêt
à acheter le jeu de données ?

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Avant le jour J, l'industriel a-t-il intérêt
à acheter le jeu de données ?

Définition du risque d'erreur de seconde espèce

On rappelle que dans la problématique de l'industriel, deux risques étaient en jeu :

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Avant le jour J, l'industriel a-t-il intérêt
à acheter le jeu de données ?

Définition du risque d'erreur de seconde espèce

On rappelle que dans la problématique de l'industriel, deux risques étaient en jeu :

1. Risque d'erreur de première espèce :

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Avant le jour J, l'industriel a-t-il intérêt
à acheter le jeu de données ?

Définition du risque d'erreur de seconde espèce

On rappelle que dans la problématique de l'industriel, deux risques étaient en jeu :

1. Risque d'erreur de première espèce : $\boxed{\text{risque de devenir pauvre}}$

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Avant le jour J, l'industriel a-t-il intérêt
à acheter le jeu de données ?

Définition du risque d'erreur de seconde espèce

On rappelle que dans la problématique de l'industriel, deux risques étaient en jeu :

1. Risque d'erreur de première espèce : risque de devenir pauvre
2. Risque d'erreur de seconde espèce :

Définition du risque d'erreur de seconde espèce et objectif

Objectif

L'analyse du risque d'erreur de seconde espèce permettra de répondre à la question suivante (pour la problématique du produit A) :

Avant le jour J, l'industriel a-t-il intérêt
à acheter le jeu de données ?

Définition du risque d'erreur de seconde espèce

On rappelle que dans la problématique de l'industriel, deux risques étaient en jeu :

1. Risque d'erreur de première espèce : risque de devenir pauvre
2. Risque d'erreur de seconde espèce : risque de ne pas devenir riche

Risque d'erreur de seconde espèce : généralités

pour tous les paramètres d'intérêt ?

L'analyse du risque d'erreur de seconde espèce **ne peut se faire que** pour des paramètres pour lesquels la loi de la **future estimation du paramètre d'intérêt ne dépend que du paramètre d'intérêt** .

Risque d'erreur de seconde espèce : généralités

pour tous les paramètres d'intérêt ?

L'analyse du risque d'erreur de seconde espèce **ne peut se faire que** pour des paramètres pour lesquels la loi de la **future estimation du paramètre d'intérêt ne dépend que du paramètre d'intérêt** .

1. Exemple : une proportion $p^\star$ car (par exemple pour le produit A), on sait que dans un cadre asymptotique

Risque d'erreur de seconde espèce : généralités

pour tous les paramètres d'intérêt ?

L'analyse du risque d'erreur de seconde espèce **ne peut se faire que** pour des paramètres pour lesquels la loi de la **future estimation du paramètre d'intérêt ne dépend que du paramètre d'intérêt** .

1. Exemple : une proportion $p^\star$ car (par exemple pour le produit A), on sait que dans un cadre asymptotique

$$\widehat{p^\star}(\mathbf{Y}) \overset{approx.}{\rightsquigarrow} \mathcal{N}\left(p^\star, \sqrt{\frac{p^\star(1-p^\star)}{n}}\right).$$

2. Contre-exemple : une moyenne $\mu^\star$ car (par exemple pour le produit B), on sait que dans un cadre asymptotique

Risque d'erreur de seconde espèce : généralités

pour tous les paramètres d'intérêt ?

L'analyse du risque d'erreur de seconde espèce **ne peut se faire que** pour des paramètres pour lesquels la loi de la **future estimation du paramètre d'intérêt ne dépend que du paramètre d'intérêt** .

1. Exemple : une proportion $p^\star$ car (par exemple pour le produit A), on sait que dans un cadre asymptotique

$$\widehat{p^\star}(\mathbf{Y}) \overset{approx.}{\rightsquigarrow} \mathcal{N}\left(p^\star, \sqrt{\frac{p^\star(1-p^\star)}{n}}\right).$$

2. Contre-exemple : une moyenne $\mu^\star$ car (par exemple pour le produit B), on sait que dans un cadre asymptotique

$$\widehat{\mu^\star}(\mathbf{Y}) \overset{approx.}{\rightsquigarrow} \mathcal{N}\left(\mu^\star, \sqrt{\frac{\sigma_\star^2}{n}}\right).$$

Retour au produit A

Objectif plus précis de l'industriel

Avant le jour J, l'industriel a-t-il intérêt à acheter l'échantillon (de taille n)

Retour au produit A

Objectif plus précis de l'industriel

Avant le jour J, l'industriel a-t-il intérêt à acheter l'échantillon (de taille n) **sachant qu'** $\boxed{\textbf{a priori}}$**, il pense qu'il y a** $\boxed{p^A = 19\%}$ **d'acheteurs potentiels ?**

Retour au produit A

Objectif plus précis de l'industriel

Avant le jour J, l'industriel a-t-il intérêt à acheter l'échantillon (de taille n) **sachant qu'** $\boxed{\textbf{a priori}}$**, il pense qu'il y a** $\boxed{p^A = 19\%}$ **d'acheteurs potentiels ? et si l'a priori est** $\boxed{p^A = 17\%}$ **?**

Retour au produit A

Objectif plus précis de l'industriel

Avant le jour J, l'industriel a-t-il intérêt à acheter l'échantillon (de taille n) **sachant qu'** $\boxed{\textbf{a priori}}$**, il pense qu'il y a** $\boxed{p^A = 19\%}$ **d'acheteurs potentiels ? et si l'a priori est** $\boxed{p^A = 17\%}$ **?**

et encore plus précis

Retour au produit A

Objectif plus précis de l'industriel

Avant le jour J, l'industriel a-t-il intérêt à acheter l'échantillon (de taille n) **sachant qu'** $\boxed{\textbf{a priori}}$**, il pense qu'il y a** $\boxed{p^A = 19\%}$ **d'acheteurs potentiels ? et si l'a priori est** $\boxed{p^A = 17\%}$ **?**

et encore plus précis

On va s'intéresser à la question suivante :

Retour au produit A

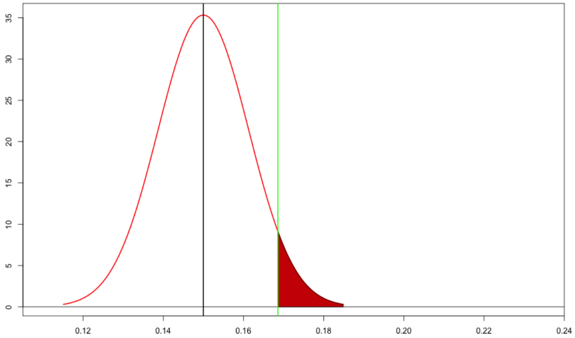
Objectif plus précis de l'industriel

Avant le jour J, l'industriel a-t-il intérêt à acheter l'échantillon (de taille n) **sachant qu'** $\boxed{\textbf{a priori}}$**, il pense qu'il y a** $\boxed{p^A = 19\%}$ **d'acheteurs potentiels ? et si l'a priori est** $\boxed{p^A = 17\%}$ **?**

et encore plus précis

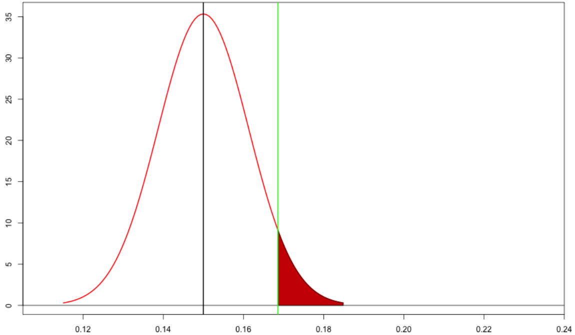
On va s'intéresser à la question suivante :
Etant donnée la règle de décision fixée à $\alpha = 5\%$, si $p^A = 19\%$ **combien** parmi l'infinité des estimations de ce paramètre obtenues avec un échantillon de taille $n = 1000$ conduisent **à ne pas lancer le produit A** ?

Evaluation du risque d'erreur de seconde espèce



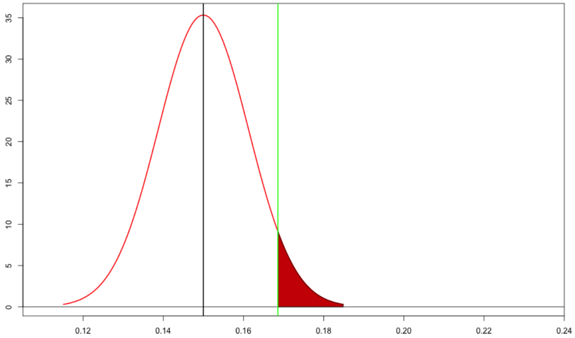
Règle de décision : accepter $\mathbf{H_1}$ si $\widehat{p^A}\left(\mathbf{y^A}\right) > p^+_{lim,5\%}$ avec $p^+_{lim,5\%} \stackrel{R}{=}$ `plim =qnorm(.95,.15,sqrt(.15*.85/1000))`

Evaluation du risque d'erreur de seconde espèce



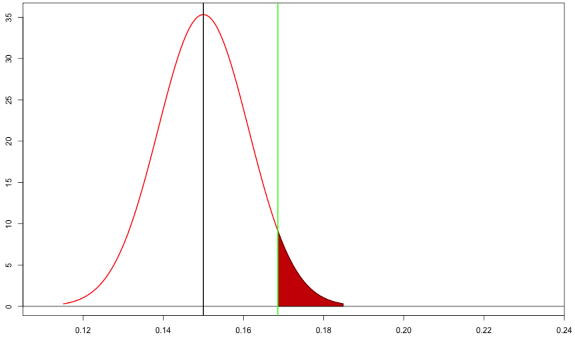
Question : comment interpréter la surface rouge par l'A.E.P. ?

Evaluation du risque d'erreur de seconde espèce



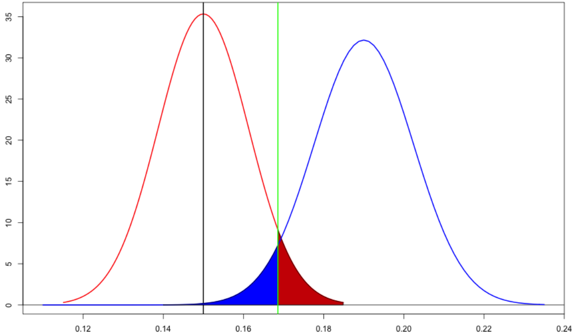
Question : comment interpréter la surface rouge par l'A.E.P. ?
Réponse : la proportion parmi l'infinité des estimations de $p^A = 15\%$ qui sont supérieures à $p_{lim,5\%}$, i.e. la proportion de ces estimations qui conduisent à lancer le produit A à tort (dans la pire des situations)

Evaluation du risque d'erreur de seconde espèce



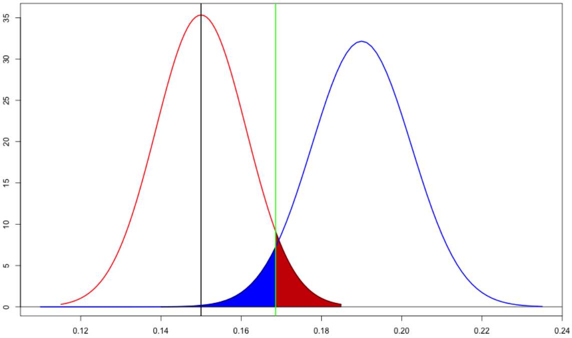
Question : comment représenter le tas de toutes les estimations possibles de $p^A = 19\%$?

Evaluation du risque d'erreur de seconde espèce



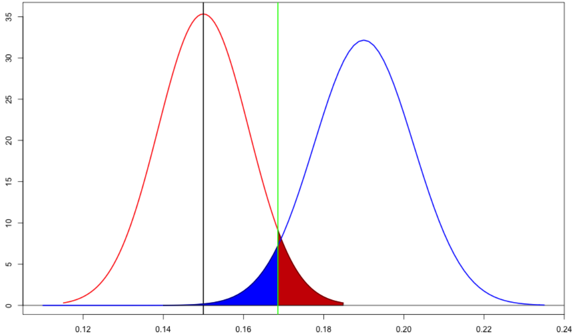
Question : à quoi correspond alors la surface bleue via l'A.E.P

Evaluation du risque d'erreur de seconde espèce



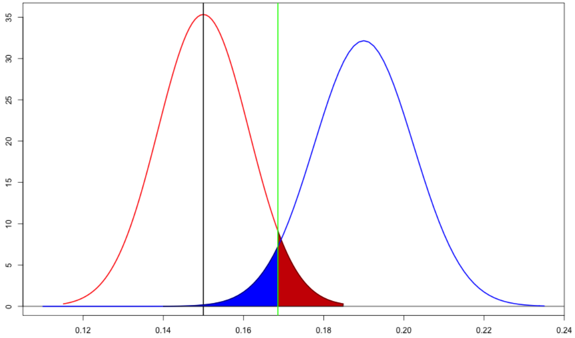
Question : à quoi correspond alors la surface bleue via l'A.E.P
Réponse : à la proportion parmi l'infinité des estimations de $p^A = 19\%$ qui sont inférieures à $p_{lim,5\%^+}$, i.e. à la proportion de ces estimations qui conduisent à ne pas lancer le produit A (à tort !)

Evaluation du risque d'erreur de seconde espèce



Question : comment définir mathématiquement cette surface ?

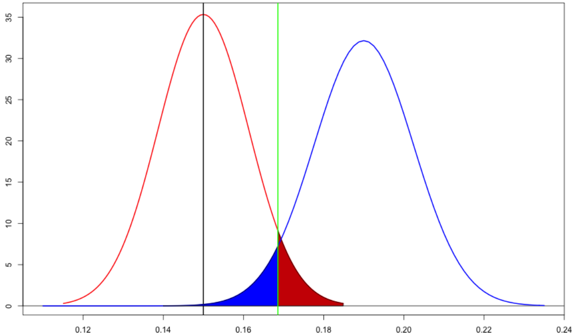
Evaluation du risque d'erreur de seconde espèce



Question : comment définir mathématiquement cette surface ?

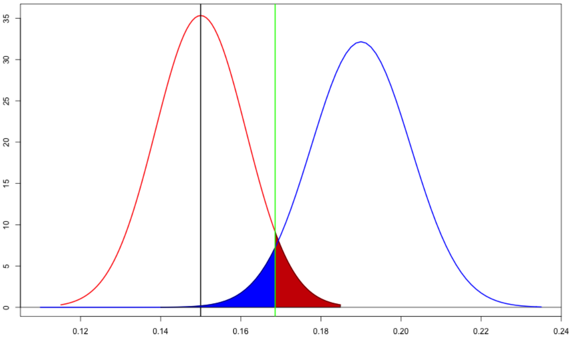
Réponse : $\beta(19\%) = \mathbb{P}_{p^A=19\%}\left(\widehat{p^A}\left(\mathbf{Y^A}\right) < p^+_{lim,5\%}\right)$

Evaluation du risque d'erreur de seconde espèce



Question : conseillez-vous l'industriel d'acheter $\mathbf{y^A}$?

Evaluation du risque d'erreur de seconde espèce

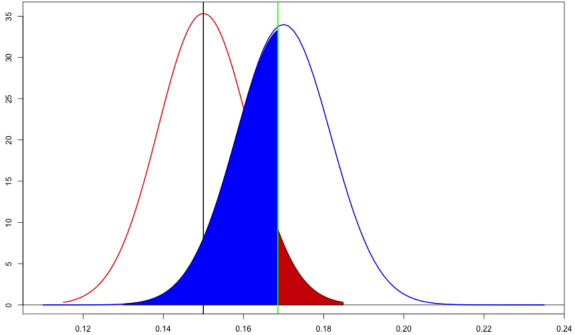


Question : conseillez-vous l'industriel d'acheter $\mathbf{y}^{\mathbf{A}}$?
Réponse : puisque

$$\beta(19\%) \stackrel{R}{=} \texttt{pnorm(plim,0.19,sqrt(0.19*0.81/1000))} \simeq 4.2\%$$

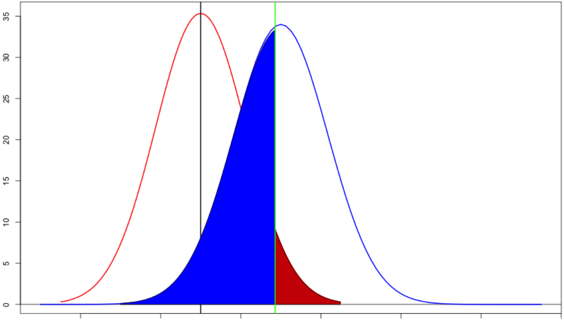
est relativement faible, on peut lui conseiller d'acheter $\mathbf{y}^{\mathbf{A}}$

Evaluation du risque d'erreur de seconde espèce



Question : et si l'a priori de l'industriel était de 17% ?

Evaluation du risque d'erreur de seconde espèce



Question : et si l'a priori de l'industriel était de 17% ?
Réponse : puisque

$$\begin{aligned}\beta(17\%) &= \mathbb{P}_{p^A=17\%}\left(\widehat{p^A}\left(\mathbf{Y^A}\right) < p^+_{lim,5\%}\right)\\ &\overset{R}{=} \texttt{pnorm(plim,0.17,sqrt(0.17*0.83/1000))} \simeq 45.2\%\end{aligned}$$

est très élevé, on ne lui conseillerait pas d'acheter $\mathbf{y^A}$.

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) = \mathbb{P}_{p^\star = p}\left(\widehat{p^\star}\left(\mathbf{Y}\right) < p^+_{lim,5\%}\right)}$$

définition

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) \overset{R}{=} \texttt{pnorm(plim,p,sqrt(p*(1-p)/1000))}}$$

définition

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) \stackrel{R}{=} \text{pnorm(plim,p,sqrt(p*(1-p)/1000))}}$$

définition

Pour tout $0 < p < 1$, on définit la **fonction puissance** notée $\gamma(p)$. Cette quantité représente les chances que l'on a si $\boxed{p^\star = p}$ de lancer le produit A (si la règle de décision est fixée à $\alpha = 5\%$), i.e.

$$\boxed{\gamma(p) = \mathbb{P}_{p^\star = p}\left(\widehat{p^\star}\left(\mathbf{Y}\right) > p^+_{lim,5\%}\right)}$$

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) \stackrel{R}{=} \text{pnorm(plim,p,sqrt(p*(1-p)/1000))}}$$

définition

Pour tout $0 < p < 1$, on définit la **fonction puissance** notée $\gamma(p)$. Cette quantité représente les chances que l'on a si $\boxed{p^\star = p}$ de lancer le produit A (si la règle de décision est fixée à $\alpha = 5\%$), i.e.

$$\boxed{\gamma(p) = \mathbb{P}_{p^\star = p}\left(\widehat{p^\star}(\mathbf{Y}) > p^+_{lim,5\%}\right)}$$

1. lorsque $p < 0.15$,

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) \stackrel{R}{=} \text{pnorm(plim,p,sqrt(p*(1-p)/1000))}}$$

définition

Pour tout $0 < p < 1$, on définit la **fonction puissance** notée $\gamma(p)$. Cette quantité représente les chances que l'on a si $\boxed{p^\star = p}$ de lancer le produit A (si la règle de décision est fixée à $\alpha = 5\%$), i.e.

$$\boxed{\gamma(p) = \mathbb{P}_{p^\star=p}\left(\widehat{p^\star}(\mathbf{Y}) > p^+_{lim,5\%}\right)}$$

1. lorsque $p < 0.15$, $\boxed{\gamma(p) = \text{risque de devenir pauvre si } p^\star = p.}$

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) \stackrel{R}{=} \texttt{pnorm(plim,p,sqrt(p*(1-p)/1000))}}$$

définition

Pour tout $0 < p < 1$, on définit la **fonction puissance** notée $\gamma(p)$. Cette quantité représente les chances que l'on a si $\boxed{p^\star = p}$ de lancer le produit A (si la règle de décision est fixée à $\alpha = 5\%$), i.e.

$$\boxed{\gamma(p) = \mathbb{P}_{p^\star = p}\left(\widehat{p^\star}(\mathbf{Y}) > p^+_{lim,5\%}\right)}$$

1. lorsque $p < 0.15$, $\boxed{\gamma(p) = \text{risque de devenir pauvre si } p^\star = p.}$
2. lorsque $p \geq 0.15$?

Fonction puissance

plus généralement

Avant le jour J, on peut calculer le risque d'erreur de seconde espèce pour tout $p \geq 0.15$ par

$$\boxed{\beta(p) \overset{R}{=} \text{pnorm(plim,p,sqrt(p*(1-p)/1000))}}$$

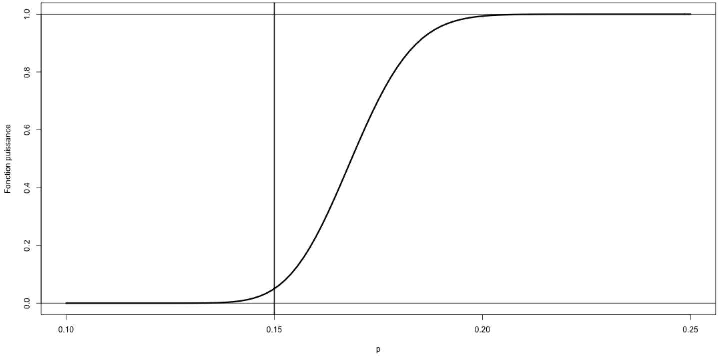
définition

Pour tout $0 < p < 1$, on définit la **fonction puissance** notée $\gamma(p)$. Cette quantité représente les chances que l'on a si $\boxed{p^\star = p}$ de lancer le produit A (si la règle de décision est fixée à $\alpha = 5\%$), i.e.

$$\boxed{\gamma(p) = \mathbb{P}_{p^\star=p}\left(\widehat{p^\star}(\mathbf{Y}) > p^+_{lim,5\%}\right)}$$

1. lorsque $p < 0.15$, $\boxed{\gamma(p) = \text{risque de devenir pauvre si } p^\star = p.}$
2. lorsque $p \geq 0.15$? $\boxed{\gamma(p) = 1 - \beta(p)}$

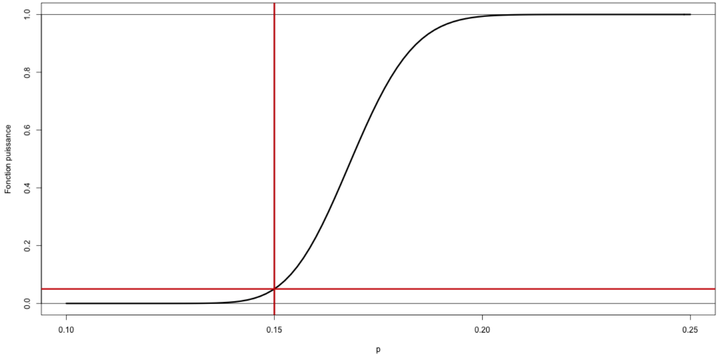
La fonction puissance en image



Questions

Que vaut $\gamma(15\%)$?

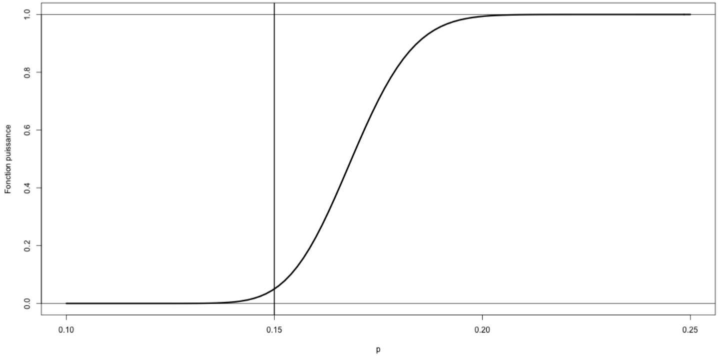
La fonction puissance en image



Questions

Que vaut $\gamma(15\%)$? Réponse : $\boxed{5\%}$

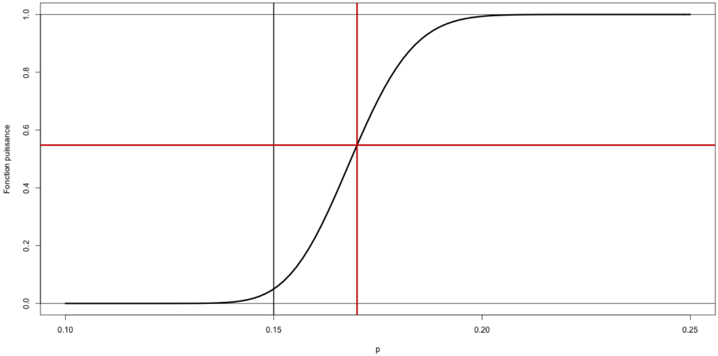
La fonction puissance en image



Questions

Que vaut $\gamma(17\%)$?

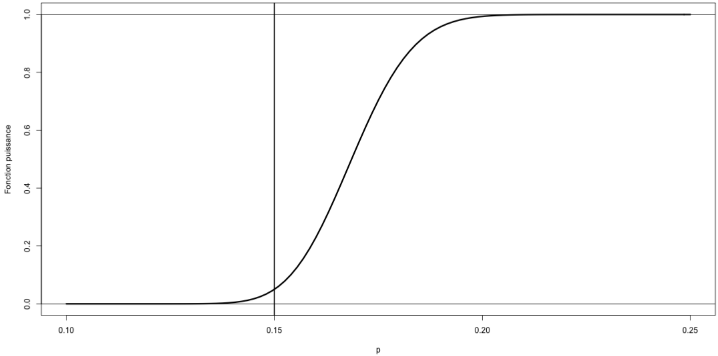
La fonction puissance en image



Questions

Que vaut $\gamma(17\%)$? Réponse : $\gamma(17\%) = 1 - \beta(17\%) \simeq 55\%$

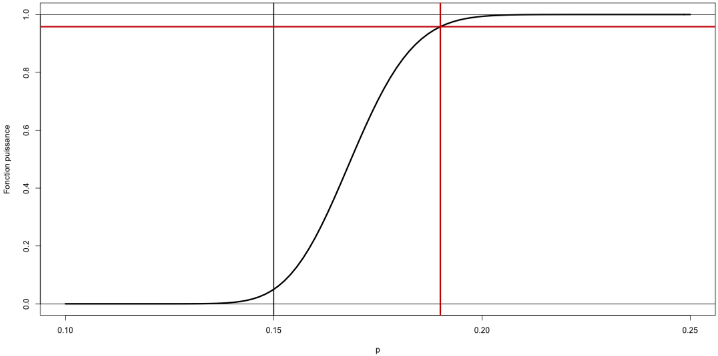
La fonction puissance en image



Questions

Que vaut $\gamma(19\%)$?

La fonction puissance en image



Questions

Que vaut $\gamma(19\%)$? Réponse : $\gamma(19\%) = 1 - \beta(19\%) \simeq 4\%$

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 17% d'acheteur potentiels.

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 17% d'acheteur potentiels.
Comment **ajuster la taille d'échantillon** de telle sorte que $\alpha = 5\%$ et $\beta(17\%) \leq 5\%$?

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 17% d'acheteur potentiels.

Comment **ajuster la taille d'échantillon** de telle sorte que $\alpha = 5\%$ et $\beta(17\%) \leq 5\%$?

Merci qui ? le matheux évidemment

Le mathématicien nous dit qu'il suffit de prendre n tel que

$$\boxed{n \geq \left(\frac{q_{1-\alpha} \times \sqrt{0.15(1-0.15)} + q_{\beta(17\%)} \times \sqrt{0.17(1-0.17)}}{17\% - 15\%} \right)^2}$$

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 17% d'acheteur potentiels.
Comment **ajuster la taille d'échantillon** de telle sorte que $\alpha = 5\%$ et $\beta(17\%) \leq 5\%$?

Merci qui ? le matheux évidemment

Le mathématicien nous dit qu'il suffit de prendre n tel que

$$\boxed{n \geq \left(\frac{q_{1-\alpha} \times \sqrt{0.15(1-0.15)} + q_{\beta(17\%)} \times \sqrt{0.17(1-0.17)}}{17\% - 15\%} \right)^2}$$

Application en R : il faut que $n \geq$ **3632.0**

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 19% d'acheteur potentiels.

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 19% d'acheteur potentiels.
Comment **ajuster la taille d'échantillon** de telle sorte que $\alpha = 5\%$ et $\beta(19\%) \leq 5\%$?

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 19% d'acheteur potentiels.
Comment **ajuster la taille d'échantillon** de telle sorte que $\alpha = 5\%$ et $\beta(19\%) \leq 5\%$?

Merci qui ? le matheux évidemment

Le mathématicien nous dit qu'il suffit de prendre n tel que

$$\boxed{n \geq \left(\frac{q_{1-\alpha} \times \sqrt{0.15(1-0.15)} + q_{\beta(19\%)} \times \sqrt{0.19(1-0.81)}}{19\% - 15\%} \right)^2}$$

Choix de la taille d'échantillon

les contraintes de l'industriel

L'industriel souhaite construire une règle de décision avec un risque de devenir pauvre $\alpha = 5\%$ et pense qu'a priori son produit sera acheté par au moins 19% d'acheteur potentiels.
Comment **ajuster la taille d'échantillon** de telle sorte que $\alpha = 5\%$ et $\beta(19\%) \leq 5\%$?

Merci qui ? le matheux évidemment

Le mathématicien nous dit qu'il suffit de prendre n tel que

$$\boxed{n \geq \left(\frac{q_{1-\alpha} \times \sqrt{0.15(1-0.15)} + q_{\beta(19\%)} \times \sqrt{0.19(1-0.81)}}{19\% - 15\%} \right)^2}$$

Application en R : il faut que $n \geq$ **950.0**