

Vous venez d'obtenir un résultat vous indiquant que :

Résultat :

Il vous faut plus d'observations pour pouvoir affirmer que votre taux de défaillance prévisionnel est bien compatible avec votre retour d'expérience

Précision : (>100%)

[Plus de détails](#)

Nous allons, dans ce complément d'information vous indiquer la raison d'un tel résultat.

[Pour un accès direct à la table de la qualité de la compatibilité ... voir p.16](#)

Dans ce test, nous cherchons à montrer si le taux de défaillance prévisionnel FIDES est compatible ou non avec notre retour d'expérience (REX), que nous avons donc suffisamment de certitudes pour le considérer comme vraisemblable pour l'ensemble de nos produits ou non.

En statistiques, lorsque l'on a besoin de définir un critère d'adéquation entre un modèle de fiabilité prévisionnel par rapport aux informations recueillies en exploitation, on peut passer par l'utilisation d'un test d'adéquation, le test d'hypothèses.

« *Qu'est-ce qu'un test d'hypothèses ?* »

Très souvent, un problème de décision consiste à trancher, au vu d'observations, entre une hypothèse ou une autre.

Par exemple, lorsque l'on tire à « Pile ou Face », qu'on choisit toujours pile, et que l'on a l'impression de perdre un peu trop souvent, ou de gagner un peu trop souvent, on peut se demander si la pièce est truquée ou si elle ne l'est pas.

L'une de ces hypothèse sera appelée hypothèse nulle (notée H_0), et l'autre hypothèse qu'on lui oppose sera dite « hypothèse alternative » (notée H_1).

En général, une et une seule de ces deux hypothèses peut être vraie. Un test d'hypothèses est donc une procédure qui permet de choisir entre ces deux hypothèses.

Lors de cette procédure, on peut noter deux types d'erreurs de décision possible :

- l'**erreur de première espèce**, qui est de rejeter l'hypothèse H0 à tort
- l'**erreur de deuxième espèce**, qui est de rejeter l'hypothèse H1 à tort.

L'idéal serait évidemment de trouver une procédure qui minimise les deux risques d'erreurs en même temps. Malheureusement, on peut démontrer qu'ils varient en sens inverse, c'est-à-dire que toute procédure diminuant l'un va en général, augmenter l'autre et réciproquement.

Dans notre exemple, on se demande si la pièce de monnaie est truquée lors de notre tirage. On peut donc poser les hypothèses suivantes :

- H0 : La pièce de monnaie n'est pas truquée.
- H1 : La pièce de monnaie est truquée.

On pose nos hypothèses ainsi car on veut établir une certaine « présomption d'innocence ». De même que dans un procès d'assises, il y a présomption d'innocence, en théorie des tests, il y a **présomption de H0**, c'est-à-dire qu'il faut montrer que H0 est peu probable pour décider H1.

Par contre accepter H0 peut venir du fait que soit H0 est effectivement vraie, soit on n'a pas les moyens nécessaires de prouver H1. (Ex: pas assez d'observations).

On note α , la probabilité de l'erreur de première espèce, et β la probabilité de l'erreur de deuxième espèce.

Dans la pratique des tests statistiques, il est de règle de **fixer** α (ex : 5%, 1%,10%) car c'est le risque que l'on veut contrôler, celui que l'on est prêt à prendre en rejetant à tort H0, c'est-à-dire le **niveau de risque** que l'on s'accorde quant à l'affirmation des décisions que l'on va prendre par la suite. Ce choix est basé sur la perception que l'on a de la gravité des conséquences d'un rejet injustifié de H0.

Pour mettre en relation la décision prise avec la réalité de la situation (dont on ne peut jamais être sûr à l'avance, voir ne jamais être sûr du tout selon les cas), on pose le tableau suivant :

Décision \ Réalité	H0 : La pièce de monnaie n'est pas truquée.	H1 : La pièce de monnaie est truquée.
H0 : La pièce de monnaie n'est pas truquée.	Probabilité = $1-\alpha$	Probabilité = β
H1 : La pièce de monnaie est truquée.	Probabilité = α	Probabilité = $1-\beta$

Effectivement, on peut maintenant mieux distinguer tous les cas qui peuvent arriver lors d'une prise de décision pareille.

Il devient évident que l'on peut définir les probabilités de ces différents cas de figure comme dans le tableau précédent.

En effet, si on part du principe qu'il y a forcément une réalité absolue, c'est que l'autre cas de figure n'est donc pas possible.

Donc si la réalité est « H_0 : La pièce de monnaie n'est pas truquée. », c'est qu'il ne peut en être autrement, et qu'il ne tient plus qu'à l'homme concerné de prendre la bonne décision ou non.

Il en ressort donc que sa probabilité de prendre la mauvaise décision est égale à « α » (l'erreur de première espèce), et que du coup sa probabilité de prendre la bonne décision est égale à « $1-\alpha$ ».

De même, si la réalité est « H_1 : La pièce de monnaie est truquée. », il ne peut là non plus en être autrement, et là encore il ne tient plus qu'à l'homme concerné de prendre la bonne décision ou non.

Il en ressort donc que sa probabilité de prendre la mauvaise décision est dans ce cas égale à « β » (l'erreur de deuxième espèce), et que du coup, sa probabilité de prendre la bonne décision est égale à « $1-\beta$ ».

Cette dernière probabilité « $1-\beta$ » est appelée **la puissance du test**. En effet, vu que nous avons préalablement posé notre test de façon à ce que la probabilité de l'erreur de première espèce « α » soit la probabilité de déclarer à tort que la pièce est truquée, il va en découler que **la puissance du test** « $1-\beta$ » soit la probabilité de « **détecter** » à **raison** que la pièce est bien truquée.

« *Comment peut-on maintenant exprimer ce test d'hypothèses de manière statistique ?* »

Si la pièce n'est effectivement pas truquée et que l'on réalise N tirages (N assez grand), on devrait normalement obtenir un nombre de réalisations de « Pile » et de « Face » plus ou moins égales, car la probabilité d'obtenir l'une ou l'autre des situations est de 50%.



Par exemple, $N = 100$, on lance 100 fois la pièce et on note X la variable qui compte le nombre de fois que l'on obtient « pile ».

Si on réalise un test **bilatéral**, on peut donc adopter la règle de décision suivante :

H_0 : la pièce n'est pas truquée est,

- acceptée si $X \in [40,60]$,
- rejetée si $X \notin [40,60]$.

Dans le cadre de ce test, on a :

$$P(\text{rejeter } H_0 / H_0 \text{ vrai}) = P(X \notin [40,60]) = \alpha \approx 0,05$$

et donc

$$P(\text{accepter } H_0 / H_0 \text{ vrai}) = P(X \in [40,60]) = 1 - \alpha \approx 0,95.$$

Il s'agit donc d'un test où l'on s'est fixé $\alpha = 0,05$.

→ Si la valeur de X que l'on obtient se trouve dans la zone de rejet (région critique, c'est-à-dire en dehors de $[40,60]$), on dira que la statistique X observée est statistiquement significative au seuil α . Cette valeur est anormalement élevée ou basse et ne permet pas d'accepter H_0 . On rejette H_0 . On ne peut pas accepter que la pièce de monnaie ne soit pas truquée.

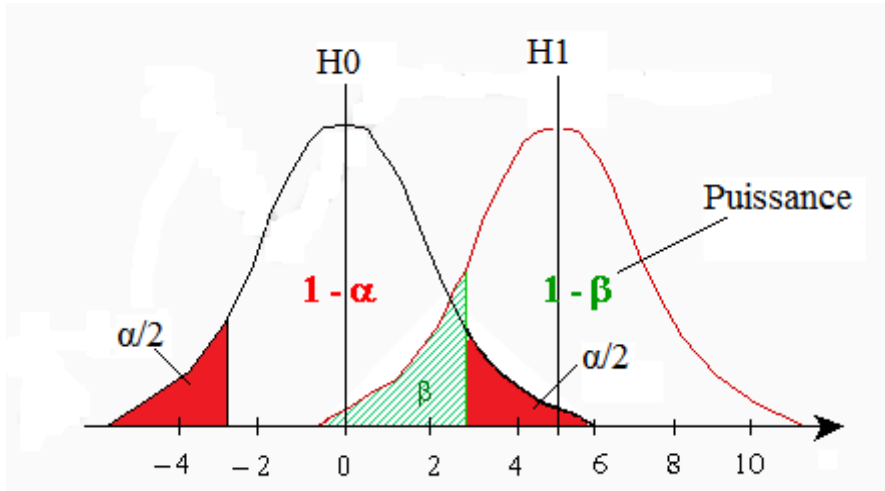
→ Si la valeur de X que l'on obtient se trouve dans la zone d'acceptation (c'est-à-dire en à l'intérieur de $[40,60]$), on dira que la statistique X observée n'est pas significative au seuil α . Cette valeur est imputable aux fluctuations d'échantillonnage. On accepte H_0 . On ne peut pas rejeter le fait que la pièce de monnaie ne soit pas truquée.

Remarque :

On revient sur une notion mentionnée plus tôt, α étant fixé, β sera déterminé comme le résultat d'un calcul.

β varie en sens contraire de α : si l'on veut diminuer α , on est conduit à ne rejeter H_0 que dans des cas rares donc à conserver H_0 bien souvent à tort, donc on augmente β (proba sous H_1 d'accepter H_0).

Cela s'exprime sous la forme d'un graphique (cas du test bilatéral) :



On peut donc grâce à ce graphique, observer les différentes possibilités représentées par des aires sous les courbes de distribution.

La courbe de gauche représente la distribution dans le cas où H0 est vraie, celle de droite dans le cas où H1 est vraie.

On a bien évidemment fixé α la région critique représentée en rouge. On sait que cette aire représente la probabilité de rejeter H0 à tort, que l'on cherche bien évidemment à minimiser.

On peut voir que, si on diminue α , alors β augmente, car la zone où on va accepter l'hypothèse H0 alors que H1 est vraie augmente et donc par conséquent, la puissance diminue.

Une autre remarque est que, plus la distribution sous H0 est proche de la distribution sous H1, plus β augmente et donc plus la puissance diminue. Cela se traduit par le fait que plus les deux hypothèses sont représentées par des distributions proches, plus il sera difficile de « détecter » les cas où l'on peut rejeter H0 à raison car les différences paraissent moins significatives.

Évidemment, dans le cas inverse, c'est-à-dire les deux distributions sont très éloignées, il sera plus facile de « détecter » les cas où l'on peut rejeter H0 à raison, car la puissance augmente et β diminue.

Imaginons le cas extrême où H0 et H1 sont représentées par des distributions avec des courbes quasiment juxtaposées, alors on aurait environ « $\beta=1-\alpha$ » et « Puissance= α », ce qui veut dire que les différences entre les deux hypothèses seraient quasiment indétectables, car la probabilité de détecter les cas où l'on peut rejeter H0 avec raison seraient faibles, environ égale à α (valeur préalablement minimisée).

Cas du REX et des modèles de fiabilité prévisionnelle :

Maintenant que la définition d'un test d'hypothèse est plus clairement définie, nous pouvons revenir à son utilisation dans le cadre de l'adéquation entre notre REX et un modèle de fiabilité prévisionnelle.

On pose les hypothèses suivantes :

- H0 : Le modèle de fiabilité prévisionnelle est compatible du REX.
- H1 : Le modèle de fiabilité prévisionnelle n'est pas compatible du REX.

Ce qui donne le tableau suivant :

Décision \ Réalité	H0 : Le modèle de fiabilité prévisionnelle est compatible du REX.	H1 : Le modèle de fiabilité prévisionnelle n'est pas compatible du REX.
H0 : Le modèle de fiabilité prévisionnelle est compatible du REX.	Probabilité = 1- α	Probabilité = β (erreur deuxième espèce)
H1 : Le modèle de fiabilité prévisionnelle n'est pas compatible du REX.	Probabilité = α (erreur première espèce)	Probabilité = 1- β (puissance test)

La puissance de ce test nous permet donc de détecter les cas où l'on rejette la compatibilité entre le modèle de fiabilité prévisionnelle et le REX à raison. C'est-à-dire qu'elle nous indique **la probabilité de détecter** les cas où il y a « réellement » incompatibilité (la différence est significative et détectable).

Dans le cadre de notre test, **l'estimateur du taux de défaillance** pour les informations recueillies (REX) est donné par :

$$\hat{\lambda} = \frac{\text{Nombre de pannes}}{\sum \text{Hrs}} = \frac{r}{T}$$

Donc « r » est le nombre de pannes (défaillances) que l'on a observées sur nos données durant notre expérience et « T » est la somme des heures cumulées.

Il s'agit des deux paramètres d'entrées que l'on renseigne dans la zone suivante :

Retour d'expérience (REX)

Nombre de pannes observées

Somme des heures

Nous sommes dans un cadre où nous définirons nos données comme censurées à droite (Type I), avec la possibilité de remplacement c'est-à-dire que dès qu'un produit observé est en panne, il est soit immédiatement remplacé soit immédiatement réparé.

Il s'agit donc d'essais stoppés après un temps « t_0 » fixé à l'avance, quel que soit le **nombre « r » de pannes observées**.

Si on a « N » **entités en observation** et que l'on ne considèrerait pas la possibilité de remplacement en cas de panne, la somme des heures cumulées « T », s'exprimerait sous la forme :

$$T = t_1 + t_2 + \dots + t_r + (N - r)t_0$$

Cependant, nous sommes dans le cadre de remplacement ou réparation immédiate en cas de panne (le temps de remplacement ou réparation est négligeable dans le calcul), donc **la somme des heures cumulées « T »**, s'exprime simplement sous la forme :

$$T = N * t_0$$

On peut donc calculer (grâce à un certain nombre de règles statistiques importantes), un intervalle de confiance en fonction de α pour notre taux de défaillance observé (avec un niveau de risque α qu'on s'est fixé et donc un $NC=1-\alpha$ niveau de confiance que l'on s'est fixé) :

$$[\hat{\lambda}_{min}; \hat{\lambda}_{max}] \text{ tel que } P(\text{accepter } H_0 | H_0 \text{ vrai}) = P(\lambda_{FIDES} \in [\hat{\lambda}_{min}; \hat{\lambda}_{max}]) = NC = 1 - \alpha$$

Il s'agit du niveau de risque α que vous avez renseigné comme paramètre d'entrée :

Précision du test

Niveau de risque :

10%	▼
10%	
5%	
1%	
0.5%	
0.1%	

On dit qu'un retour d'expérience (REX) est **significatif** si le taux de défaillance prévisionnel donné par la méthodologie FIDES **n'est pas compris** dans l'intervalle de confiance du taux de défaillance observé.

J'affirme donc que, si le taux de défaillance prévisionnel « λ_{FIDES} » est en dehors de l'intervalle de confiance du REX, le REX est significatif en ce sens que l'on peut rejeter l'hypothèse comme quoi le taux de défaillance prévisionnel est compatible du REX.

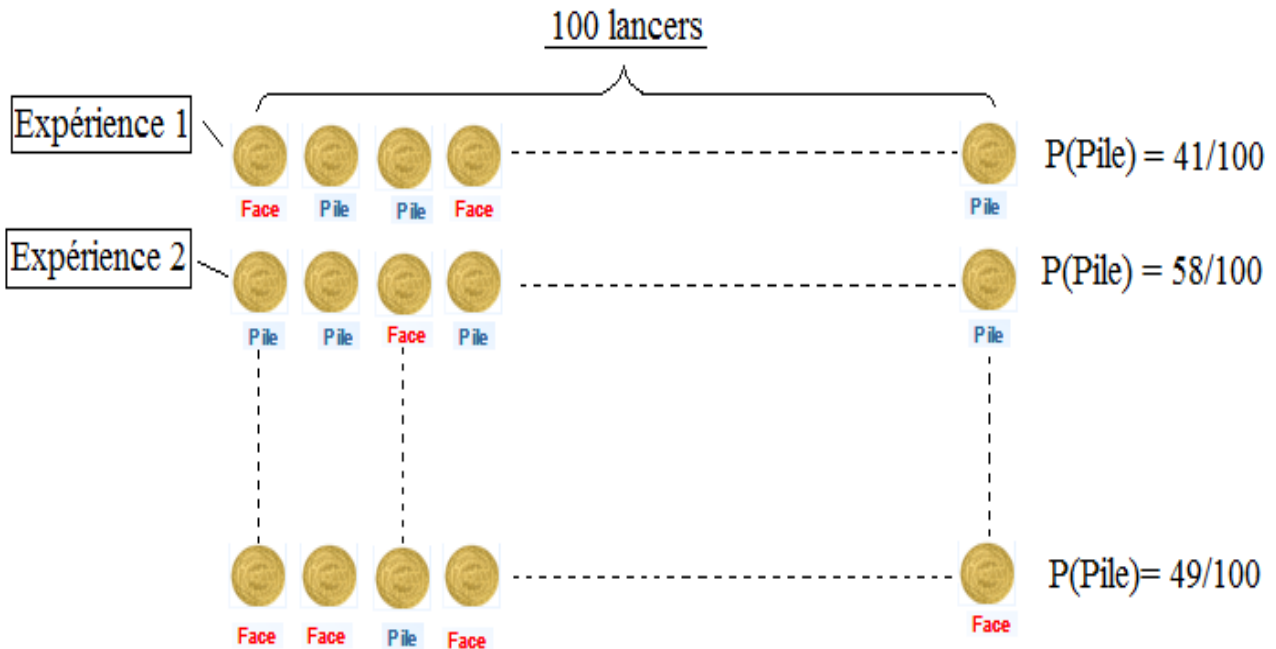
Si le taux de défaillance prévisionnel « λ_{FIDES} » est dans l'intervalle de confiance du REX, on ne peut pas rejeter l'hypothèse comme quoi le taux de défaillance prévisionnel est compatible avec REX, mais cependant, pour un certain nombre de raisons qui vont suivre, on ne peut pas être sûr qu'il soit **compatible** au REX non plus.

Dans le cadre du rejet de l'hypothèse H_0 , α est suffisamment petit (car on l'a préalablement minimisé) pour que l'estimateur du taux de défaillance pour les informations recueillies par notre REX et l'intervalle de confiance qui en découle soient suffisant pour statuer sur le rejet de l'hypothèse nulle. Le fait que « λ_{FIDES} » se retrouve dans la zone critique alors qu'il est pourtant bien compatible du REX est suffisamment rare pour qu'on puisse décider le rejet de l'hypothèse nulle dans ce cas.

Cependant, lorsque « λ_{FIDES} » se trouve dans l'intervalle de confiance du REX, β (l'acceptation de la compatibilité à tort) n'ayant pas été minimisé, le REX seul n'est pas une source suffisante pour statuer sur l'acceptation de l'hypothèse nulle (compatibilité). On s'est rendu compte en analysant notre intervalle de confiance, qu'il n'y a donc pas de prise en compte de l'aspect aléatoire des données d'exploitation car il n'y a pas de prise en compte de l'occurrence des défaillances.

Pour illustrer cette problématique, reprenons le cas simple où l'on réalise 100 lancers de pièces. A chaque répétition de cette expérience, l'estimateur de la probabilité d'avoir « Pile » est différent. On cherche à estimer la probabilité d'avoir « Pile ».

Schéma représentatif :



On peut observer que pour chaque estimateur de la probabilité d'avoir « Pile » peut donner une valeur différente. Parfois, les variations entre une valeur et une autre peuvent être suffisamment grandes, pour que si on se contente d'une seule expérience, on puisse obtenir une interprétation et une prise de décision totalement différente que celle qu'on aurait obtenue avec une autre expérience.

De la même façon, dans le cadre de notre recherche de compatibilité, l'estimateur du taux de défaillance **ne correspond qu'à une seule expérience** et on ne peut pas se fier aveuglément à cette valeur.

En effet, en ce qui concerne le cas où l'on refuse la compatibilité entre le taux de défaillance prévisionnel et le REX, on peut se permettre de le faire en ne considérant que l'unique valeur que nous avons à notre disposition car vu que l'on a nous-même minimisé α (le rejet de H_0 à tort), une expérience suffit pour prouver la non compatibilité.

En outre, comme α est minimisé, β (l'acceptation à tort de H_0) ne l'est pas, on ne veut donc pas faire de conclusions hâtives quant à l'acceptation de H_0 , uniquement parce qu'on n'a pas réussi à rejeter l'hypothèse H_0 dans le cadre d'une unique expérience.

On va donc procéder à une deuxième étape de validation pour palier à ce problème.

L'idéal serait de pouvoir réaliser un grand nombre d'expériences comme celle de notre retour d'expérience (REX) pour pouvoir conclure avec plus de certitudes quant à la possible compatibilité.

Mais cela est couteux et on ne peut pas se permettre de réaliser un grand nombre de fois les mêmes expériences. On va donc mettre en place une autre alternative, la simulation de plusieurs expériences ayant la même propriété « T » que nôtre REX.

Par un certain nombre de procédés mathématiques, on va réaliser une simulation à partir d'un paramètre λ , pendant un temps cumulé total T.

On obtiendra bien une simulation semblable à la situation obtenue avec notre retour d'expérience. On obtiendra alors la simulation d'un nombre « r » de défaillances pour **une somme des heures cumulées** « T » (le même T que celui de notre REX).

On peut alors recommencer cette simulation autant de fois que l'on veut, on obtiendra un nombre « r » de pannes aléatoirement différents pour chaque simulation.

« *Quel complément d'information ces simulations vont-elles nous apporter ?* »

On considère que le vrai λ correspondant à nos pièces n'est pas connu mais se trouve autour de « λ_{FIDES} » car ce dernier a passé la première étape de validation avec succès.

On va donc, pour chaque valeur d'un ensemble de λ autour de « λ_{FIDES} » simuler un grand nombre de fois la variables aléatoires « r » en fonction de « T » et calculer à chaque reprise l'estimateur du taux de défaillance pour chacune de ces simulation de REX. :

$$\hat{\lambda} = \frac{\text{Nombre de pannes}}{\sum \text{Hrs}} = \frac{r}{T}$$

On pourra aussi calculer dans chacun de ces cas, l'intervalle de confiance de ce taux de défaillance (en utilisant la formule de l'intervalle utilisée jusqu'à présent).

Quand on fait des simulation de variables aléatoires un grand nombre de fois, on sait grâce à la loi des grands nombres, que les proportions que l'on obtient convergent presque sûrement vers la probabilité.

Si pour un certain taux de défaillance λ autour de « λ_{FIDES} », on calcule le pourcentage de cas où le taux de défaillance prévisionnel « λ_{FIDES} » est en dehors de l'intervalle de confiance calculé à chaque tirage, on obtient alors grâce à ce pourcentage **une estimation de la puissance de test**. En d'autres termes, il s'agit du pourcentage de cas (de la probabilité par loi des grands nombres) de détecter que λ et « λ_{FIDES} » ne sont pas compatibles à raison.

On va donc, pour une puissance qu'on s'est fixée (ex : 80%),

Puissance :

80%	▼
50%	
60%	
70%	
80%	
90%	

obtenir un intervalle autour de « λ_{FIDES} » qui aura pour bornes la première **valeur à gauche** et la première **valeur à droite** de « λ_{FIDES} » pour lesquels on a au moins 80% de chances de détecter à raison que les λ correspondants à ces bornes ne sont pas compatibles de « λ_{FIDES} ».

On appellera cet intervalle, **l'intervalle d'incertitude** ou **intervalle de précision** (noté IP), car toutes les valeurs de λ à l'intérieur de cet intervalle ont moins de 80% de chances d'être détectées à raison qu'elles ne sont pas compatibles de « λ_{FIDES} ». Évidemment, par logique toutes les valeurs à l'extérieur de cet intervalle ont plus de 80% de chance d'être détectées à raison qu'elles ne sont pas compatibles de « λ_{FIDES} ».

A partir de cet intervalle, on calculera une certaine valeur de précision :

$$Précision = \frac{(\lambda_{FIDES} - \lambda_{Gauche})}{\lambda_{FIDES}} * 100 + \frac{(\lambda_{Droite} - \lambda_{FIDES})}{\lambda_{FIDES}} * 100$$

Précision à gauche

Précision à droite

En fonction de cette valeur, nous pourrons statuer sur **la qualité de la compatibilité** entre le REX et le standard FIDES.

Exemple 1 :

Si on a un taux de défaillance prévisionnel donné par FIDES tel que :

$\lambda_{\text{FIDES}}=0,001$ (soit en moyenne, une panne toutes les milles heures)

On entre,

Modèle de fiabilité prévisionnelle FIDES

Lambda FIDES

On réalise un essai et on obtient comme retour d'expérience,

Retour d'expérience (REX)

Nombre de pannes observées :

Somme des heures cumulées:

On désire réaliser notre calcul avec les critères suivants,

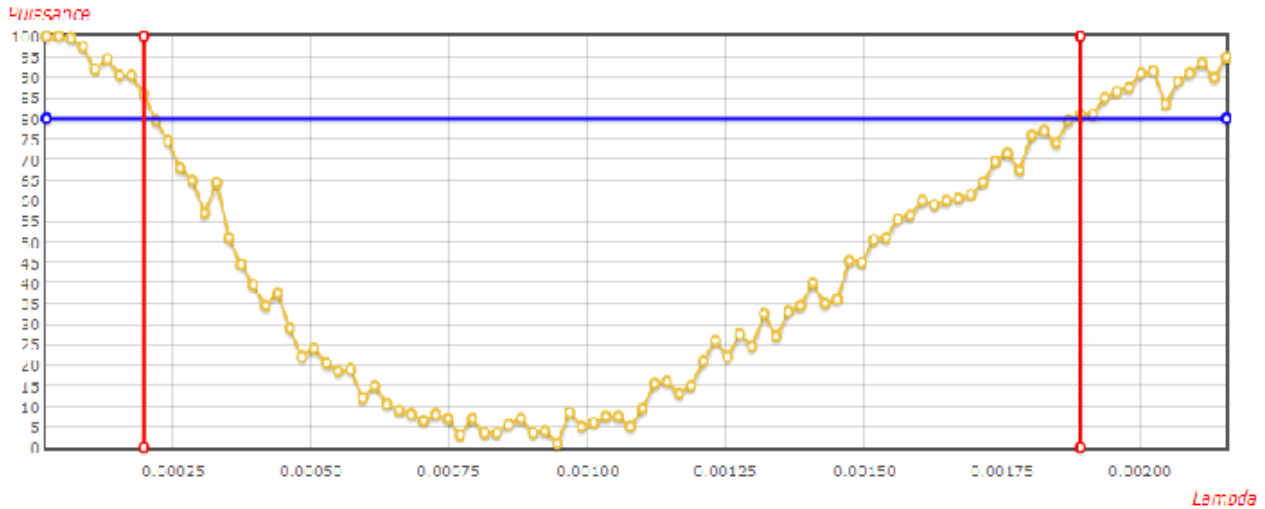
Précision du test

Niveau de risque :

Puissance :

On lance notre test,

et on obtient,



avec comme résultat,

Résultat :

Il vous faut plus d'observations pour pouvoir affirmer que votre taux de défaillance prévisionnel est bien compatible avec votre retour d'expérience

Précision : **171** (>100%)

En effet à partir de ces résultats, on peut voir qu'il faut une Précision de 171% autour de « λ_{FIDES} » pour pouvoir détecter dans plus de 80% des cas qu'un λ serait incompatible du « λ_{FIDES} ». Nous considérons que ce n'est pas assez précis car nous voulons une précision au moins inférieure à 100% pour commencer à statuer sur la qualité de la compatibilité.

Il faut donc observer plus longtemps pour obtenir plus de pannes et pour qu'ainsi les fluctuations d'échantillonnage n'aient plus suffisamment d'importance sur les pannes aléatoirement simulées. Nous pourrions ainsi obtenir une précision plus acceptable autour de « λ_{FIDES} ».

Exemple 2 :

Si on reprend le même « λ_FIDES » que dans « l'exemple 1 ». On reprend aussi les mêmes paramètres de « Précision du test » que dans « l'exemple 1 ».

On réalise juste plus longtemps notre essai et on finit par obtenir les observations suivantes lors de notre REX :

Retour d'expérience (REX)

Nombre de pannes observées :

109

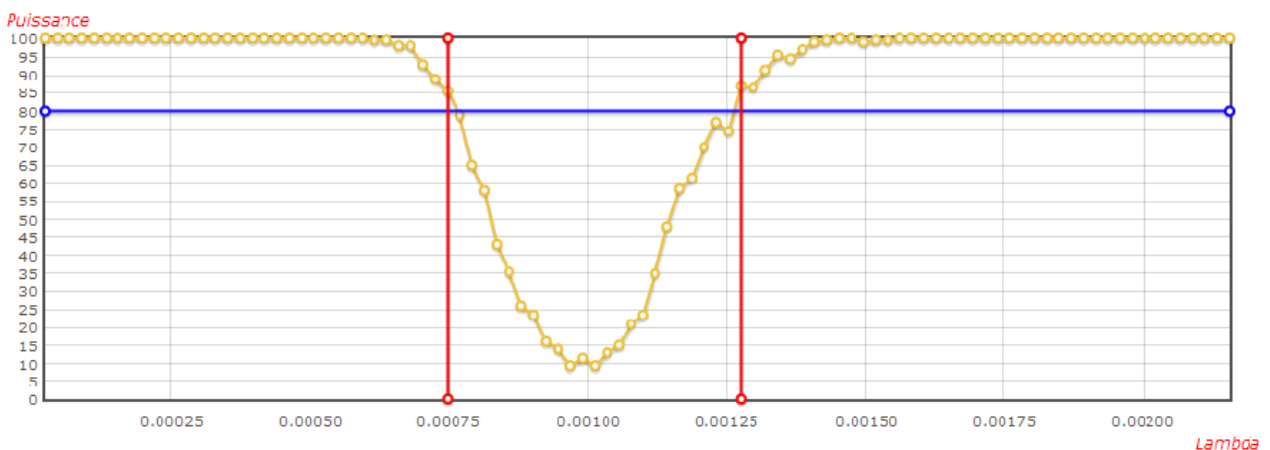
Somme des heures cumulées:

103271

On lance le test,

Lancer test

et on obtient,



avec comme résultat,

Résultat :

Votre taux de défaillance prévisionnel Fides est bien compatible avec votre retour d'expérience

Précision : **54 %** (26 % à gauche du lambda FIDES, 28 % à droite du lambda FIDES)

En effet, à partir de ces résultats, on peut voir qu'il faut une Précision de 52% autour de « λ_{FIDES} » pour pouvoir détecter dans plus de 80% des cas qu'un λ serait incompatible du « λ_{FIDES} ». Nous considérons que c'est assez précis pour pouvoir statuer sur la qualité de la compatibilité entre le taux de défaillance prévisionnel FIDES et le REX. Pour plus de renseignements sur la qualité de la précision, vous pouvez vous référer à la table de qualité à venir, mais vous pouvez aussi statuer en fonction de vos propres critères.

Table de la qualité de la compatibilité entre le taux de défaillance prévisionnel FIDES et le REX :

Précision	Qualité de la compatibilité
0% < Précision < 25%	Très Bonne
25% < Précision < 50%	Bonne
50% < Précision < 75%	Assez Bonne
75% < Précision < 100%	Passable
Précision > 100%	Médiocre

C'est la raison pour laquelle vous obtenez le résultat suivant :

Résultat :

Il vous faut plus d'observations pour pouvoir affirmer que votre taux de défaillance prévisionnel est bien compatible avec votre retour d'expérience

Précision : (>100%)

[Plus de détails](#)