

LE TEST A/B, ET QUAND LE VOLUME NE SUFFIT PAS

Deux versions d'un même geste, envoyées au hasard, une seule différence, et tout écrit avant : la mesure, la taille, la durée. Quand le volume manque, on suit une carte de contrôle à la place.

AUSSI APPELÉ test A/B, expérimentation contrôlée, test randomisé PRINCIPE QU'IL APPLIQUE le cycle PDCA, fiche P-06

SE PRÉPARE EN une demi-journée, pour écrire la mesure, les garde-fous et la taille SE FAIT EN 2 à 8 semaines, selon le volume et le délai de la mesure

LEVIER PRINCIPAL ROI D'APRÈS R. Kohavi, D. Tang et Y. Xu, Trustworthy Online Controlled Experiments, 2020

CE QUE ÇA RAPPORTE

Impact sur le TTM		Faible
Impact sur la qualité (FTR)		Moyen
Impact sur l'Opex / Capex		Moyen

CE QUE ÇA COÛTE

Coût d'implémentation		Faible
Difficulté d'implémentation		Moyen
Difficulté de maintien		Moyen

Faible Moyen Élevé N/A (sans objet)

1 Le problème auquel ça s'applique

Et ce qu'il vous coûte chaque année

CE QUE VOUS VOYEZ DÉJÀ

- Les paiements montent depuis la nouvelle relance. Rien ne dit que c'est elle.
- Trois choses ont changé dans le même courrier. Personne ne sait laquelle a joué.
- Le test a été arrêté le jour où la courbe était belle.

OÙ PART L'ARGENT

Une version généralisée sur une fausse preuve coûte deux fois : le gain annoncé n'arrive pas, et la bonne version n'est jamais essayée.

L'ORDRE DE GRANDEUR, À LA LOUCHE

Comptez-le sur vos chiffres. Un ordre de grandeur suffit pour décider d'essayer.

COMBIEN DE FOIS / AN
envois du geste
testé

CE QUE ÇA COÛTE
À CHAQUE FOIS
coût d'un client qui
ne répond pas

PART QU'ON ÉVITE
points gagnés par
la meilleure
version, une fois
prouvés

GAIN / AN
à écrire au
crayon, à la date
prévue

QUAND ÇA NE SERT À RIEN

Sous quelques milliers de clients par version, l'écart ne se voit pas. On change pour tous et on suit l'effet sur une carte de contrôle (fiche O-18).

2 Comment ça marche

La mécanique, en clair

L'IDÉE EN UNE PHRASE

Le hasard fait deux groupes semblables en tout, sauf la version reçue. L'écart mesuré à la date prévue vient donc de la version, à un risque connu près.

LES GESTES, DANS L'ORDRE

- Changer une seule chose entre A et B, et écrire pourquoi elle devrait jouer.
- Écrire avant la mesure principale, deux garde-fous et le plus petit écart à pouvoir détecter.
- Calculer la taille par version : $n \approx 2 \times (1,96 + 0,84)^2 \times p(1 - p) / d^2$.
- Tirer au hasard par le système, client par client, jamais par agence ou par gestionnaire.
- Fixer la durée d'avance, et ne pas s'arrêter le jour où le résultat plaît.
- À la date prévue, décider sur la mesure principale, garde-fous vérifiés.

CE QU'IL FAUT AVOIR SOUS LA MAIN

Un outil d'envoi qui sait tirer au hasard, la mesure disponible client par client, et quelqu'un qui sait refaire le calcul de taille.

LE SIGNE QUE ÇA A MARCHÉ

Le comité attend la date prévue pour conclure, alors que la courbe de la deuxième semaine lui plaisait.

3 Les pièges de l'implémentation

Ce qui fait rater, et ce qu'on fait à la place

AU LANCEMENT Changer trois choses à la fois

B gagne, et personne ne sait si c'est l'objet, le ton ou la date d'envoi.

→ Une seule différence par test. Les autres attendent le suivant.

AU LANCEMENT Lancer sans calculer la taille

Le test finit sur un écart « prometteur » qui tient dans le bruit.

→ Écrire le plus petit écart à détecter, et en déduire le nombre de clients par version.

PENDANT Regarder chaque jour, et s'arrêter quand ça plaît

À force de regarder, un écart dû au seul hasard finit par passer le seuil.

→ La durée fixée d'avance, ou un test séquentiel, conçu pour être lu en cours de route.

PENDANT Répartir par agence ou par gestionnaire

Les groupes diffèrent avant même le test. L'écart mesure l'agence, pas la version.

→ Le tirage au hasard par le système, client par client.

DANS LA DURÉE Oublier les garde-fous

La relance plus ferme fait payer plus vite, et les réclamations doublent trois mois plus tard.

→ Deux garde-fous écrits avant, et suivis après la généralisation (fiche O-18).

4 Un exemple

Cas type du tertiaire

L'OBJET

Le courrier de première relance des factures impayées d'un fournisseur d'énergie, clients particuliers.

CE QUI CLOCHE

Trois refontes en deux ans, chacune déclarée gagnante. Le taux de paiement n'a pas bougé.

CE QU'ILS REGARDAIENT AVANT

Le taux de paiement du mois, comparé au mois précédent. Il variait de quatre points sans que rien change.

LES CHIFFRES DE DÉPART

À relever avant d'écrire le protocole.

180 000 relances d'impayés par an
38% payées sous 15 jours avec le courrier actuel
240 € d'impayé moyen par relance
9 € par relance de niveau 2, recommandé et appel

CE QU'ILS ONT FAIT

Un seul changement, 4 500 clients par version tirés au hasard, une date de lecture fixée d'avance.

LE DÉCLENCHEUR

Une quatrième refonte du courrier est proposée, avec un gain annoncé de cinq points.

LE GESTE

Protocole écrit avant. Seul changement : un lien de paiement en tête du courrier.

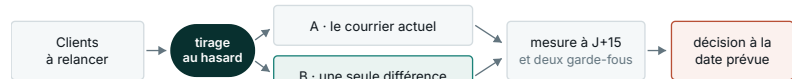
CE QU'ILS ONT VU

Au quinzième jour, 41,6 % contre 38,0 %. Réclamations au même niveau.

LA DÉCISION

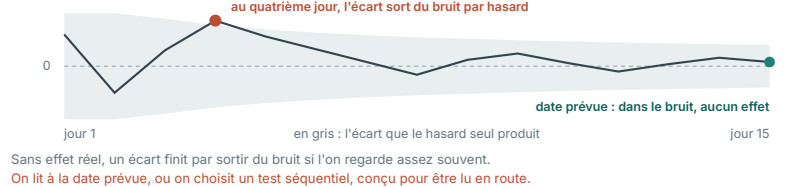
B généralisée. Le test suivant portera sur la date d'envoi, et sur elle seule.

1 · UNE DIFFÉRENCE, UN TIRAGE, UNE DATE



Écrit avant le premier envoi : la mesure, deux garde-fous, l'écart à détecter, la taille, la durée.

SANS EFFET RÉEL, L'ÉCART OBSERVÉ JOUR APRÈS JOUR



Sans effet réel, un écart finit par sortir du bruit si l'on regarde assez souvent. On lit à la date prévue, ou on choisit un test séquentiel, conçu pour être lu en route.

2 · LA RELANCE D'IMPAYÉS

LA TAILLE, CALCULÉE AVANT
 $n \approx 2 \times (1,96 + 0,84)^2 \times 0,38 \times 0,62 / 0,03^2$
 ≈ 4100 clients par version. 4 500 retenus.

PAYÉES SOUS 15 JOURS

A 38,0 %
B 41,6 %

+3,6 points

Réclamations et désinscriptions : au même niveau dans A et B.

LE CAS CONTRAIRE : 300 CLIENTS ENTRE ENTREPRISES

À 150 clients par version, seul un écart de 16 points se verrait.

On change pour tous, et on suit le délai de paiement en carte de contrôle.

nouveau courrier

haute
moyenne
basse

Neuf semaines de suite sous la moyenne : le signal (fiche O-18).

À gauche le protocole, et pourquoi on ne lit pas le résultat en cours de route. À droite la relance, puis le cas où le volume ne suffit pas.