

Test Dotaki

Partie 1 – Statistiques descriptives

En marketing digital, deux mesures importantes de performance d'un site internet sont :

- le « taux de conversion » (*conversion rate*), c'est à dire le pourcentage du trafic allant jusqu'à l'achat sur le site en question. Cette approche peut s'appliquer à l'échelle de pages, par exemple si le site propose le parcours suivant « page produit » —> « panier » —> « paiement » —> « paiement confirmé » alors on peut définir le taux de conversion de la page « panier » comme le ratio entre le nombre de visiteurs étant allés jusqu'à la page « paiement confirmé » (objectif de conversion) et le nombre de visiteurs étant passés par la page « panier ».
- le « panier moyen » qui est le montant moyen commandé par les visiteurs de ce site. Celui-ci est mesuré non pas sur l'ensemble des visiteurs mais seulement sur ceux ayant fait une commande.

Afin d'optimiser les ventes sur les sites, une pratique courante est de réaliser un « AB test » qui consiste à diriger une partie du trafic sur une version (A) d'une page à optimiser, et une autre partie du trafic sur une autre version (B) de cette page. Ensuite, le taux de conversion et le panier moyen de ces deux versions sont comparées et la meilleure version de page (A ou B) peut être choisie. Pour faire un choix robuste, il faut s'assurer que la différence entre les taux de conversion des deux versions et entre les paniers moyens de ces deux versions sont significatifs à plus de 95%.

Question 1 :

Veuillez programmer en Python un programme qui prend 6 valeurs en entrée :

- nombre de visiteurs passés par la variante A d'une page
- nombre de visiteurs passés par la variante A d'une page et étant passé par la page commande validée (utilisateurs dits « convertis »)
- panier moyen des acheteurs étant passés par la variante A de la page, exprimé en euros
- nombre de visiteurs passés par la variante B d'une page
- nombre de visiteurs passés par la variante B d'une page et étant passé par la page commande validée (utilisateurs dits « convertis »)
- panier moyen des acheteurs étant passés par la variante B de la page, exprimé en euros

et qui donne 4 valeurs en sortie :

- le gain (ou la perte) de taux de conversion de la variante B par rapport à la variante A, exprimé en pourcents (ex : si A fait 40% de taux de conversion et B fait 50%, alors le gain relatif de B par rapport à A est de +25%)
- significativité du gain (ou de la perte) entre les taux de conversion, exprimée en pourcents
- le gain (ou la perte) de panier moyen de la variante B par rapport à la variante A, exprimé en pourcents (ex : si A fait 40% de taux de conversion et B fait 50%, alors le gain relatif de B par rapport à A est de +25%)
- significativité du gain (ou de la perte) entre les paniers moyens, exprimée en pourcents

Veuillez détailler votre démarche d'analyse statistique et les choix de code que vous avez réalisés.

À titre d'exemple et pour vérifier le bon fonctionnement de votre programme sur l'analyse de la différence entre les taux de conversion, un outil en ligne existe sur le même principe : <https://neilpatel.com/ab-testing-calculator/> . Note : cet outil n'explique pas exactement comment est calculé le pourcentage de significativité, plusieurs formules existent et donnent des résultats pouvant être différents les uns des autres.

Question 2 :

Comment adapter ce programme pour un nombre de variantes N (supérieur à 2) ?

Question 3 :

Quelle variante retenir dans le cas où la variante B présente un taux de conversion supérieure mais un panier moyen inférieur à la variante A ?

Partie 2 – Machine learning

Le trafic d'un site est analysé par Dotaki et chaque utilisateur est qualifié en temps réel par 5 variables continues (score de 0 à 100 pour chaque variable) et 1 variable discrète (un segment est affecté à chaque utilisateur, parmi 10 segments possibles). Sur ce site, un AB test est réalisé avec 3 variantes, un tiers des visiteurs, tirés aléatoirement, voit la version « Original » du site, un autre tiers la version « Variation 1 », et le dernier tiers voit la « Variation 2 ». Pour chaque utilisateur les données suivantes sont enregistrées dans une base de données :

- caractéristiques : 5 variables continues (nommées A, O, N, E, C) et 1 segment
- variante affectée : Original, Variation 1 ou Variation 2
- conversion (achat réalisé) : oui ou non

Ce test est réalisé sur environ 10 000 visiteurs durant deux semaines. À l'issue de ce test, l'objectif est de prédire quelle variation (Original, Variation 1 ou Variation 2) doit être affichée à chaque nouveau visiteur pour maximiser le taux de conversion global du site (c'est-à-dire le pourcentage d'utilisateurs qui réaliseront un achat sur le site).

Question 1 :

Décrivez la démarche permettant d'atteindre l'objectif fixé et implémentez-la sur le dataset fourni.

Question 2 :

Décrivez les limitations de votre implémentation et les moyens potentiels d'amélioration des résultats.