# أساسيات اختبار A/B

اختبار A/B تجربة مضبوطة تقارن نسختين من شيء واحد، مثل صفحة موقع، بعرض كل نسخة على مجموعة عشوائية مختلفة من الناس. النسخة A هي الحالية وتسمى الضابطة، والنسخة B هي التغيير وتسمى البديلة. وتتيح هذه الطريقة للفريق أن يتعلم من سلوك الناس الفعلي، لا من الآراء. ونتابع رامي، مدير المنتجات في مكتبة إلكترونية.

## 1. البدء بفرضية

كل اختبار جيد يبدأ بسؤال واضح يكتب فرضية، مثل أن تغيير نص الزر سيزيد عمليات الشراء. ونختار مقياس نجاح رئيسيا واحدا مسبقا، مثل نسبة الزوار الذين يشترون. ونختار أيضا مقاييس حماية، مثل سرعة الصفحة أو نسبة المرتجعات، يجب ألا تسوء. واختبار تغيير واحد في كل مرة يجعل النتيجة سهلة الفهم.

**مثال:** لدى خباز سؤال واضح: هل سيختار عملاء أكثر كعكا مصنوعا بوصفة جديدة؟ ويقرر الخباز قياس شيء واحد هو عدد الكعك المبيع.

**في العمل:** فرضية رامي أن نص الزر «الشراء الآن» سيجلب مشتريات أكثر من «إضافة إلى السلة». ومقياسه الرئيسي نسبة الزوار الذين يشترون، ومقياس الحماية عنده نسبة المرتجعات.

## 2. التقسيم العادل

بعد ذلك نوزع الناس على المجموعتين عشوائيا، لتتشابه المجموعتان في كل شيء ما عدا التغيير. وينبغي أن يرى كل شخص النسخة نفسها في كل مرة، وأن تعمل النسختان في الوقت نفسه، حتى لا تحابي العطل أو الأخبار إحدى المجموعتين. والتوزيع العشوائي هو ما يسمح لنا بالقول إن الفرق نتج غالبا عن التغيير.

**مثال:** يقدم مقهى عينة من مشروب جديد لطاولات مختارة عشوائيا، بينما لا تحصل الطاولات الأخرى على شيء، في اليوم نفسه. ولأن الطاولات اختيرت عشوائيا، فالفرق في الطلبات أقرب إلى أن ينتج عن العينة.

**في العمل:** يوزع نظام رامي كل زائر على A أو B بالصدفة، ويرى الزائر النسخة نفسها في كل زيارة. وتعمل النسختان في الوقت نفسه، فلا يحابي تخفيض نهاية الأسبوع إحداهما.

## 3. تخطيط حجم العينة

قبل البدء نحدد كم شخصا نحتاج وكم مدة نشغل الاختبار. فالعينة الصغيرة لا تكشف إلا الفروق الكبيرة، والعينة الأكبر تكشف فروقا أصغر، لذلك تستخدم حاسبة حجم العينة التغير المتوقع لتعطي الجواب. ومن الحكمة تشغيل الاختبار أسابيع كاملة، لتشمل كل أيام الأسبوع. ولا ينبغي أن نتوقف مبكرا لمجرد أن النتيجة تبدو جيدة، لأن النتائج المبكرة تتقلب كثيرا.


حقوق الطبع والنشر محمية © - أكاديمية أساسيات ( شركة أساسيات للخدمات الاستشارية)

**مثال:** قطعة نقدية تسقط على الصورة ثلاث مرات متتالية ليست دليلا على أنها مغشوشة، لأن بضع رميات تخبرنا بالقليل. نحتاج إلى رميات كثيرة قبل أن نثق بنمط، كما نحتاج إلى زوار كثيرين في الاختبار.

**في العمل:** تخبر حاسبة حجم العينة رامي أنه يحتاج إلى نحو ثمانية آلاف زائر في كل مجموعة ليكشف ارتفاعا في الشراء من خمسة في المئة إلى ستة في المئة. فيختار عشرين ألفا في كل مجموعة احتياطا ويشغل الاختبار أسبوعين كاملين. وفي اليوم الثالث تبدو النتيجة رائعة، لكنه ينتظر حتى النهاية كما خطط.

## 4. قراءة النتائج

عندما ينتهي الاختبار نقارن المجموعتين في المقياس الرئيسي، ونتحقق هل الفرق أكبر مما تنتجه الصدفة عادة. ومن الأدوات الشائعة القيمة الاحتمالية: فقيمة أقل من عتبة تحدد مسبقا، غالبا خمسة في المئة، توحي بأن الصدفة وحدها تفسير بعيد الاحتمال. وننظر أيضا إلى حجم الفرق، وهل هو كبير بما يكفي ليهم العمل. وأخيرا نتحقق من مقاييس الحماية ونسجل ما تعلمناه، حتى عندما تخسر النسخة البديلة.

**مثال:** تجرب مدرسة طريقتين للمراجعة. وفرق درجة واحدة بين ستة طلاب قد يكون صدفة بسهولة، أما الفرق نفسه بين ست مئة طالب فيكون أكثر إقناعا.

**في العمل:** في اختبار رامي، اشترى ما نسبته خمسة في المئة من المجموعة A وما نسبته ستة في المئة من المجموعة B. والقيمة الاحتمالية أقل بكثير من خمسة في المئة، والارتفاع كبير بما يكفي ليهم، ولم تزد المرتجعات. فيطلق رامي «الشراء الآن» ويسجل ما تعلمه.

## الخاتمة

وخلاصة القول، يقارن اختبار A/B نسخة ضابطة ونسخة بديلة على مجموعات عشوائية ليتعلم الفريق ما ينجح حقا. ويطبق في أربع خطوات: البدء بفرضية، والتقسيم العادل، وتخطيط حجم العينة، وقراءة النتائج. وعندما يستخدم بانضباط، يحل الدليل محل الرأي.

**المصادر**


• رون كوهافي ودايان تانغ وييه شو (2020): Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing، مطبعة جامعة كامبريدج – https://experimentguide.com/
• دليل NIST/SEMATECH الإلكتروني للأساليب الإحصائية (اختبارات الفرضيات والقيم الاحتمالية) – https://www.itl.nist.gov/div898/handbook/




حقوق الطبع والنشر محمية © - أكاديمية أساسيات ( شركة أساسيات للخدمات الاستشارية)