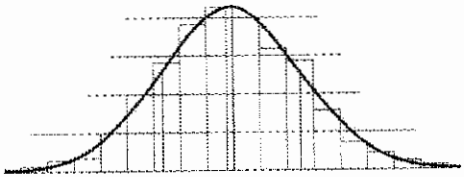

الفصل

11

اختبار الفرضية

سأل إيغون بيرسون جيرزي نيمان في بداية تعاونهما، كيف له أن يتأكد من صحة توزيع مجموعة من البيانات، إذا لم ينجح في استخراج دلالة قيمة-ب أثناء اختباره مدى صحة الحالة، ومع أن تعاونهما ابتدأ من هذا التساؤل إلا أنه فتح أبواباً واسعة. ما معنى أن تكون لدينا نتائج لا دلالة لها في اختبار الدلالة؟ هل نستنتج صحة الفرضية أنها ثابتة إذا لم نتمكن من تفنيدها؟

عرض ر.آ. فيشر التساؤل بطريقة غير مباشرة، فأخذ قيماً احتمالية كبيرة لـ ب (فاشلا في استخراج الدالّة) لتشير إلى أن البيانات كانت غير ملائمة للجزم بها. لم يقر فيشر أن الفشل في استخراج الدلالة يعني صحة الفرضية المراد اختبارها، فيقول:

obeikandl.com

إن الإيمان بثبوت الفرضية وصحتها لمجرد أنها لم تتعارض مع الحقائق المتوفرة، إيمان وهمي ولا حق له بإقحام نفسه في الاستنتاجات الإحصائية كما في الاستنتاجات العلمية الأخرى . . . ويزيد من وضوح الرؤيا كذلك، إذا أخذنا بعين الاعتبار أن اختبارات الدلالة، إذا استخدمت بدقة، لديها القدرة على رفض أو تفنيد النظريات، ما دامت تتعارض مع البيانات المعطاة: لكنها غير قادرة مطلقاً على إثبات صحتها. .

لطالما استخدم كارل بيرسون اختبار تربيع تشاي لجودة التوافق كي «يثبت» أن البيانات تتبع توزيعات معينة، فهو يطرح أساليب إحصائية رياضية أكثر شدة وصلابة، ولم تعد فرضيات كارل بيرسون مقبولة بعدها. لكن ما يزال التساؤل ما إذا كان ضرورياً افتراض توافق البيانات مع توزيع معين من أجل معرفة ما هي المتغيرات المطلوب تقديرها، وتقرير مدى صلتها بالمسألة العلمية المطروحة. يتمنى الإحصائيون اللجوء إلى اختبار الدلالة لإثبات ذلك.

اكتشف إيغون بيرسون وجيرزي نيمان من خلال مراسلاتهما، الكثير من المفارقات التي نشأت من اختبار الدلالة، كحالات من الاستعمال غير المنهجي لاختبار الدلالة أدّت إلى رفض نظريات صحيحة لا غبار عليها. لم يسمح فيشر لنفسه بدخول هذه المفارقات لأنه سيكون واضحاً لديه عدم تطبيق اختبار الدلالة بشكل صحيح. تساءل نيمان عن ماهية

المقاييس المستعملة لتقرير صحة استخدام اختبار الدلالة، فانبثقت الأفكار الرئيسية لاختبار الفرضية تدريجياً من خلال رسائلهما، والزيارات التي أجراها نيمان إلى إنجلترا في الصيف، وزيارات بيرسون إلى بولندا[1].

من الممكن الحصول على نسخة مبسطة لاختبار الفرضية الذي شكله نيمان-بيرسون في كتب الإحصاء الابتدائية. فهي سهلة التركيب ويسهل فهمها لطلاب السنة الأولى، وأصبحت تلك النسخة دقيقة ومعدة للتعليم منذ تصنيفها. وهذه هي الطريقة التي يجب أن تتبع، إذ يلمّح المنهج تلميحاً، فهذه هي الطريقة الوحيدة. تقبّلت هذه المنهجية لاختبار الفرضية، وكالات رسمية مثل الوكالة الأمريكية للغذاء والأدوية U.S. Food and Drug Administration ووكالة حماية البيئة Environmental Protection Agency، كما يدرّسها الباحثون في المدارس الطبية، ومهدت طريقها في الإجراءات القانونية لبعض الأنواع الخاصة من حالات التمييز.

(1) لقد عزوت الأفكار الرياضية الأساسية لنيمان خلال هذا الفصل لأنه كان وراء تركيبتها النهائية المصقولة، ومن ثم التطور الرياضي الدقيق بعدها. بيد أن المراسلات بين إيغون بيرسون ووليام سيلي غوسيت William Sealy Gosset والتي بدأت قبل ستة شهور من لقاء نيمان ببيرسون، تشير إلى أن بيرسون كان يفكر مسبقاً بالفرضيات البديلة والأخطاء المتنوعة، لذا قد يكون غوسيت هو أول من اقترح الفكرة. بالرغم من أن بيرسون كان أول من وضع المعلومات الأولية، فقد اعترف أن نيمان زوده بالأصول الرياضية بدلاً من أفكاره «غير المحكمة».

حرّفت طريقة تدريس نموذج نيمان-بيرسون بنسختها المبسطة والصلبة كما طورها نيمان، اكتشافاته وذلك بالتركيز على الجوانب الخاطئة من هذا النموذج. كان اكتشاف نيمان الأساسي هو أنه لا معنى لاختبار الدلالة ما لم يكن هناك فرضيتان ممكنتان على الأقل. لذلك فإنك لا تستطيع فحص في ما إذا كانت البيانات توافق التوزيع الطبيعي إلا إذا كان هناك توزيع آخر، أو مجموعة أخرى من التوزيعات تعتقد أنها تتلاءم، ويملي علينا الاختيار بين الفرضيتين البديلتين، الطريقة التي يجري فيها اختبار الدلالة. أُطلق مسمّى «قوة» الاختبار على احتمالية تنبؤ هذا البديل. تتضح الأفكار في علم الرياضيات بتخصيص مسميات واضحة وبيّنة لمفاهيم معيّنة. أطلق نيمان وبيرسون مصطلح «فرضية اللاشيء» على تلك الجاري فحصها، وأطلقا مصطلح «البديل» على غيرها، وذلك كي يفرّقا بين الفرضيات المستعملة لحساب قيمة ـ ب الفيشرية، وبين الفرضيات الأخرى الممكنة. يتم حساب قيمة ـ ب حسب صياغتهما لاختبار فرضية اللاشيء، بينما تعود القوة إلى الكيفية التي تتشكل فيها قيمة ـ ب فيما إذا كانت البدائل صحيحة.

قاد هذا نيمان إلى نتيجتين، كانت أولاهما أن قوة الاختبار تحدد مقياس جودته، فتكون الأفضلية للاختبار ذي القوة الأكبر. أما النتيجة الثانية فهي أنه لا يمكن لمجموعة من البدائل أن تكون كبيرة. ولا يستطيع المحلل الجزم أن البيانات أتت من توزيع طبيعي (فرضية اللاشيء)، أو من أي توزيع ممكن آخر،

إذ يمثل هذا مجموعة بدائل واسعة جداً، ولايمكن لأي اختبار مهما كانت قوته، الوقوف أمام كل البدائل الممكنة.

أوضح في سنة 1956 كل من ل.ج. سافاج L. J. Savage وراج راغو باهادور Raj Raghu Bahadur من جامعة شيكاغو University of Chicago أن اتساع رقعة البدائل لا تشكل سبباً في فشل اختبار الفرضية، فأنشأ مجموعة صغيرة نسبياً من الفرضيات البديلة، بحيث لا يكون هناك قوة لأي اختبار منها. طور نيمان في الخمسينيات فكرة الاختبارات الفرضية المحصورة محدداً فيها الفرضيات البديلة تحديداً ضيقاً، وأوضح أن مثل هذه الاختبارات أقوى من تلك التي تحوي مجموعة من الفرضيات الشاملة.

تُستخدَم اختبارات الفرضية في كثير من الحالات مقابل فرضية اللاّشيء الوهمية. فإذا قمنا مثلا بتجربة إكلينية للمقارنة بين عقارين، تكون فرضية اللاّشيء المراد اختبارها أن لديهما التأثير ذاته، بيد أنه إذا كان هذا صحيحاً فلا حاجة إذن لإجراء الدراسة. وفرضية اللاّشيء أن للعلاجين نفس الأثر، هي حجة وهمية يتم دحضها من نتائج الدراسة. لذلك يؤيد نيمان فكرة تصميم دراسة تهدف إلى زيادة قوة البيانات المستخرجة لأقصى درجة داحضة هذا الوهم وموضحة مدى اختلاف الأثر بين العقارين.

ما هو الاحتمال؟

كان على نيمان لسوء الحظ، التعامل مع مسألة أهملها

فيشر من أجل تطوير منهج رياضي متماسك لاختبار الفرضية، وهذه مسألة أثرت سلباً على اختبار الفرضية رغم دقة ونقاء حلول نيمان الرياضية. فهي مشكلة في تطبيق الأساليب الإحصائية على العلوم بشكل عام، ويمكن عموماً اختصارها بهذا السؤال: ما يعني الاحتمال في حياتنا العادية؟

يمكننا استخدام التركيبات الرياضية الإحصائية لحساب الاحتمالات، وتمكننا هذه الاحتمالات من تطبيق الأساليب الإحصائية على المسائل العلمية. إن الاحتمالات معرّفة جيداً في الاستعمال الرياضي، فكيف نربط هذا المفهوم النظري بالواقع؟ وكيف يمكن للعالم تفسير احتمالات التحاليل الإحصائية عند تقرير ما هو صحيح أم لا؟ سأقوم في الفصل الأخير من هذا الكتاب بمناقشة المسألة العامة والمحاولات التي أجريت للإجابة عن هذه الأسئلة. لكننا سنقوم الآن بفحص الظروف التي أجبرت نيمان على العثور على إجابته الخاصة به.

فلنتذكر اختبار فيشر للدلالة الذي أنتج رقماً أسماه فيشر قيمة-ب، وهذا احتمال محسوب، وهو احتمال مرتبط بالبيانات المشاهدة على افتراض صحة فرضية اللاّشيء. فلنفترض مثلاً أننا نريد اختبار عقار جديد وقدرته على منع الإصابة بسرطان الثدي ثانية عند من استُأصل لهن ثدي من قبل، ومقارنته مع الأدوية المهدئة. حسب فرضية اللاّشيء الوهمية لا يمتاز العقار عن المهدئ. ولنفترض أنه بعد مضي خمس سنوات، أصيبت نسبة 50٪ ممن تناولن المهدئات بسرطان الثدي ثانية، ولم

تُصب اللاتي تناولن العقار الآخر بشيء. هل يثبت ذلك «فعالية» العقار؟ تعتمد الإجابة بالطبع على عدد المريضات الممثلات في نسبة الـ 50٪.

فإذا تضمنت الدراسة أربع نساء فقط في كل مجموعة، يعني هذا أنه لدينا ثماني مريضات، اثنتان منهما عاودهما السرطان. فلنفترض أننا أخذنا مجموعة من ثماني نساء وميّزنا اثنتين منهما، وقمنا بتقسيم الثماني عشوائياً إلى مجموعتين كل مجموعة من أربع، يكون احتمال وقوع المميّزتين في مجموعة واحدة قرابة 0،30 وإذا كان في كل مجموعة أربع نساء فقط يصبح وقوع حالات تكرار الإصابة في مجموعة اللاتي تعاطين المهدئات غير ذي دلالة. أما إذا تضمنت الدراسة 500 امرأة في كل مجموعة يكون من غير الممكن وقوع جميع الـ 250 امرأة اللاتي عاودهن السرطان في مجموعة اللاتي تناولن المهدئات، اللهم إلا في حالة فعالية العقار. واحتمال وجود كل الـ250 امرأة في مجموعة واحدة في حالة كان العقار مشابهاً لأثر المهدئ هي القيمة-ب والتي هي أقل من 0,0001.

إن قيمة ـ ب هي قيمة احتمالية وهكذا يتم حسابها. فما الذي تعنيه إذا كانت تستخدم لتوضح أن الفرضية التي قامت بحسبها فرضية خاطئة؟ إنها احتمال نظري مرتبط بالمشاهدات تحت ظروف خاطئة في الغالب. فلا شأن لها بالواقع. إنها قياسات غير مباشرة للتظاهر. فهي لا تمثل احتمال خطئنا بالقول بفعالية العقار، ولا تمثل احتمالات للخطأ أياً كانت، كما أنها

لا تمثل احتمال تحسن المريضة بالعقار أو المهدئ، لكنها تُستعمل لتقرير أي الاختبارات أفضل من غيرها. كان على نيمان أن يجد طريقة لوضع اختبار الفرضية بإطار يمكن بواسطته حساب الاحتمالات المرتبطة بالقرارات المأخوذة من الاختبار. كان عليه الربط بين قيمة-ب في اختبار الفرضية مع واقع الحياة.

التعريف الترددي للاحتمالية

قدم الفيلسوف البريطاني جون فن John Venn سنة 1872 منظومة احتمالية رياضية قريبة من أرض الواقع، وأدار نظرية أساسية في الاحتمالات رأساً على عقب. هذا هو قانون الأرقام الكبيرة الذي يقول إذا كان لحدث ما احتمال معين (مثل رمي النرد واستقراره على وجه يشير إلى رقم ستة) وإذا قمنا بمحاولات مشابهة عدة مرات، فإن نسبة وقوع الحدث ستقترب من الاحتمالية.

يقول فن إن احتمالية الحادثة هي نسبة حدوثه لمرات كثيرة. وحسب منظومته فإن النظرية الرياضية للاحتمال لا تستوجب قانون الأرقام الكبيرة بل إن قانون الأرقام الكبيرة يستوجب الاحتمال. وهذا هو التعريف الترددي للاحتمال. دحض جون مينارد كينز[1] John Maynard Keynes سنة 1921

(1) هناك تزييف في التسمية في ما يتعلق بكينز، فقد عُرف كعالم اقتصاد ومؤسس المدرسة الكينزية في الاقتصاد، يعالج أمورا ككيفية استخدام الحكومات للسياسات النقدية وأثرها على الاقتصاد. لكنه حاز درجة =

ذلك لعدم فائدته أو إمكانية تفسيره منطقياً، موضحاً وجود تضارب أساسي يمنع تطبيق التعريف الترددي في جميع الحالات التي تطبق فيها الاحتمالية.

يلجأ نيمان لتعريف فن الترددي عند بناء الاختبارات الفرضية بالطريقة الرياضية التقليدية، ويستخدم نيمان ذلك ليبرر تفسيره لقيمة-ب في اختبار الفرضية. يقوم العالم حسب نموذج نيمان-بيرسون، بتحديد رقم ثابت مثل 0،05، ويرفض فرضية اللاشيء كلما كانت قيمة-ب في اختبار الدلالة أقل من أو تساوي 0،05، وبهذه الطريقة وعلى المدى الطويل، يرفض العالم فرضية اللاشيء بنسبة 5٪. يتم التركيز الآن حسب الطريقة التي يدرس فيها اختبار الفرضية، على تطبيق نيمان للاتجاه الترددي. من السهل اعتبار نموذج نيمان-بيرسون لاختبار الفرضية كجزء من الاتجاه الترددي للاحتمال وإهمال الرؤيا المهمة التي قدمها نيمان عن الحاجة إلى مجموعة من الفرضيات البديلة لاختبار الوهم بفرضية اللاشيء.

أخطأ فيشر في فهم رؤيا نيمان. لقد ركز على تعريف مستوى الدلالة، ولم يهتم بالأفكار الهامة للقوة والحاجة لتعريف مستوى البدائل. كتب ينقد نيمان:

---

* الدكتوراه في علم الفلسفة وكانت رسالته المنشورة سنة 1921 بحثاً في الاحتمالية A Treatise on Probability، علماً أساسياً في تطوير الأسس المنطقية في استخدام الإحصائيات الرياضية. سوف نتعرض لأقوال كينز في فصول لاحقة، ولكنها ستكون من كينز عالم الاحتمالية، وليس كينز عالم الاقتصاد.

لقد اعتقد نيمان أنه قام بتصحيح وتحسين أعمالي السابقة في اختبار الدلالة، كوسيلة «لتحسين المعرفة الطبيعية»، ولكنه في الواقع قام بإعادة صياغتها بما يتماشى مع الأجهزة التقنية والتجارية بما يُعرف بإجراءات القبول، والتي لها أهمية بالغة في العصر الحديث. فعندما تتلقى مؤسسة كبيرة مثل البحرية الملكية Royal Navy مواد من شركة هندسية فهي، كما اعتقد معرضة للتفتيش الوافي وللاختبار لتقليل تكرار قبول البضائع الخاطئة ذات العيوب... لكن تبدو الاختلافات المنطقية بين هذه العمليات وبين أعمال الاكتشافات العلمية عن طريق التجارب الفيزيائية والبيولوجية واسعة جداً لدرجة تكون المقارنة بينها غير مجدية، وتصبح المطابقة بين النوعين مضللة حتماً.

أصبح اختبار الفرضية أكبر وسيلة إحصائية مستخدمة في الأبحاث العلمية، بالرغم من تحريفات أفكار نيمان الأساسية، وأصبح أسلوب جيرزي نيمان الدقيق، أسلوباً ثابتاً في كثير من أقسام العلوم. تطلب معظم المجلات العلمية من كتّاب المقالات أن يضموا اختبار الفرضية في تحليل بياناتهم. بل وتعدّى الأمر إلى أبعد من المجلات العلمية، إذ تطلب السلطات التنظيمية للأدوية في الولايات المتحدة وكندا وأوروبة، استخدام اختبارات الفرضية كوسيلة مُسلّم بها. وقبلت محاكم القضاء اختبار الفرضية كأسلوب مناسب للإثبات، وسمحت للادعاء استخدامه لإثبات التمييز الوظيفي، وتفشّى هذا الأسلوب في كل فروع الإحصاء العلمي.

لم يمض صعود نموذج نيمان-بيرسون إلى قمة الإحصاء هكذا من غير تحدٍ، فهاجمها فيشر منذ البداية والى آخر رمق لديه. نشر بحثاً سنة 1955 بعنوان «الأساليب الإحصائية والاستقراء العلمي Statistical Methods and Scientific Induction» في مجلة المجمع الإحصائي الملكي Journal of the Royal Statistical Society ثم توسع في كتابه الأخير، الأساليب الإحصائية والاستدلال العلمي Statistical Methods and Scientific Inference. نشر ديفيد كوكس David Cox في نهاية الستينات لما كان محرراً لمجلة البيومتريكا، تحليلاً واضحاً عن كيفية استعمال اختبارات الفرضية في العلوم، موضحاً عدم ملاءمة صياغة نيمان الترددية للواقع، وقام أدوارد ديمنغ Edwards Deming في الثمانينيات بمهاجمة الفكرة الأساسية لاختبار الفرضية على أنها عمل أحمق. (سوف نعود إلى تأثير ديمينغ على الإحصاء في الفصل الرابع والعشرين). ومع مضي السنوات، استمر ظهور الأبحاث في الأدب الإحصائي والتي اكتشفت أخطاءً جديدة في نموذج نيمان-بيرسون مجمدة في المناهج.

لم يكن نيمان نفسه جزءاً من التمجيد الحاصل لنموذج نيمان- وبيرسون لاختبار الفرضية، فقد نشر مقالة بالفرنسية في بداية سنة 1935 في نشرة المجتمع الرياضي الفرنسي Bulletin de la Société Mathematique de France ، أثار فيها الشكوك حول إمكانية الحصول على اختبارات الفرضية المثلى. وقلما استخدم نيمان في أبحاثه الأخيرة اختبارات الفرضية بشكل

مباشر، بل تضمنت طريقته الإحصائية اشتقاق التوزيعات الاحتمالية من المبادئ النظرية، ومن ثم تقدير المتغيرات من البيانات المعطاة.

تبنى آخرون أفكارا من نموذج نيمان-بيرسون وطوروها، وسّع آبراهام والد Abraham Wald، إبّان الحرب العالمية الثانية استخدام نيمان للتعريف الترددي لثمن في تطوير مجال نظرية القرار الإحصائي. وأصدر إيريك ليمان Eric Lehmann مقياساً بديلاً للاختبارات الجيدة سنة 1959 عندما كتب كتاباً تعريفياً عن اختبار الفرضية، والذي بقي يمثل الوصف الكامل لفرضية نيمان - بيرسون.

قدم نيمان إلى الولايات المتحدة قبيل احتلال هتلر لبولندا وإسداله ستارة الشيطان على قارة أوروبة، وبدأ برنامجاً إحصائياً في جامعة كاليفورنيا في بيركلي Berkeley، ولبث هناك لحين موته سنة 1981، بعد أن أنشأ أشهر قسم أكاديمي إحصائي في العالم، إذ جلب إليه كبار شخصيات هذا المجال، كما أظهر نبوغ آخرين غير معروفين آنذاك ليمضوا قدماً بإنجازاتهم. فكان ديفيد بلاكويل David Blackwell مثلاً يعمل منفرداً في جامعة هاورد Howard University، بعيداً عن غيره من علماء الإحصاء الرياضيين. وبسبب التفرقة العنصرية، لم يتمكن من أخذ مركز في مدارس «البيض» بالرغم من قدراته؛ فدعاه نيمان إلى بيركلي. كما أنه جلب خريجاً فرنسياً جاء من عائلة قروية غير متعلمة؛ وأصبح بعدها لوسيان لو كام Lucien le Cam من أحد رواد علم الاحتمالات.

أعار نيمان اهتمامه بطلبته وزملائه من الهيئة التدريسية. لقد وصفوا أمسيات الشاي الممتعة في القسم، والتي ترأسها نيمان بكل تهذيب ولطف. كان يحث أحد الطلبة أو أعضاء التدريس لشرح الأبحاث الجديدة، ومن ثم وبكل لباقة يشق طريقه في الغرفة ليتلقى الانتقادات مساعداً في إدارة النقاش، وغالباً ما كان ينهي شرب الشاي برفع الأكواب في صحة «السيدات»، فكان واضحاً اهتمامه «بالسيدات»، وحثهن لمواصلة أعمالهن. يبرز من بين اللاتي تلقين دعم نيمان عالمات مثل الدكتورة إليزابيت سكوت Dr. Elizabeth Scott، التي عملت معه وألفا سوية، منشورات عدة في مجالات علم الأرصاد وعلم السرطان وعلم الحيوان، وكذلك الدكتورة إيفيلين فيكس Dr. Evelyn Fix التي كان لها دور كبير في علم الأمراض.

ظل نيمان تحت نيران ر.آ. فيشر إلى أن تُوفي هذا العبقري المزاجي سنة 1962، إذ كانت كل أعمال نيمان موضع نقد لفيشر. حتى ولو نجح في إثبات فكرة ما لفيشر، لم يتوان الأخير عن مهاجمته مدعياً عدم فهم نيمان لما كتب. ولو أن نيمان تبحر في فكرة لفيشر، فسيقوم أيضا بمهاجمته لاستخدامه النظرية في سبيل ميؤوس منه. لم يحبه نيمان أبداً سواءً في نشر مقالات أو بشكل شخصي، حسب ادعاء من كان يعمل معه.

وصف نيمان حقبة من الزمن في الخمسينيات خلال مقابلة أجريت معه في نهاية عمره عندما كان يستعد لتقديم دراسة

بالفرنسية في مؤتمر عالمي. لاحظ أثناء صعوده المنصة أن فيشر كان من بين الحضور، لذلك أعدّ نفسه لتلقي هجمات فيشر المتوقعة. إذ كان على ثقة أن فيشر سيستغل بعض نقاط بحثه غير ذات الأهمية ليمزقه وبحثه إرباً. انتهى نيمان من إلقائه منتظراً أسئلة الحضور، سأله البعض لكن فيشر لم ينبس بكلمة. اكتشف بعدها نيمان أن فيشر لا يتكلم الفرنسية.

obeikandl.com