الفصل

16

الاستغناء عن المتغيرات

تعرض فرانك ويلكوكسن Frank Wilcoxon، الكيميائي في مقر السيناميد Cyanamid، لمعضلة إحصائية في الأربعينات. كان يجري بعض اختبارات الفرضيات بمقارنة آثار العلاجات المختلفة، مستخدماً اختبارات «الطالب»[1] ت وتحليلات فيشر في التباين. كانت هذه الطريقة التقليدية المتبعة آنذاك في تحليل بيانات التجارب، إذ داهمت الثورة الإحصائية المخابر العلمية، وألفت كتب تفسير جداول هذه الفرضيات على رفوف مكتبات معظم العلماء. اهتم ويلكوكسن بما بدا فاشلاً من هذه الأساليب.

كان باستطاعته إجراء سلسلة من التجارب التي يظهر له فيها اختلاف أثر المعالجات. تؤكد

(1) فلنتذكر أن «طالب» كان يحمل الاسم المستعار لوليام س. غوسيت، الذي طور العينة الصغيرة في الاختبارات الإحصائية.

اختبارات ـ ت الدلالة في بعضها وتنفيها في بعضها الآخر. وغالباً ما يحدث، أثناء إجراء التجارب في الهندسة الكيميائية، أن المفاعل الكيميائي أثناء أداء التجربة، لم يتم تسخينه بشكل كافٍ في بداية مراحل المحاولة التجريبية. وقد يحدث اختلاف لإنزيم معين في قدرته على التفاعل. والنتيجة هي قيمة قد تبدو غير صحيحة في التجربة. وغالباً ما تكون رقماً كبيراً أو صغيراً جداً. يمكن في بعض الأحيان معرفة سبب هذه النتيجة البعيدة عن الواقع، وقد تكون النتيجة أحياناً أخرى، ظهور قيمة منفصلة، تختلف بشكل متطرف عن النتائج الأخرى، ولا يظهر سبب واضح لذلك.

لقد نظر ويلكوكسن إلى التركيبات المعدة لحساب اختبارات–ت وتحليلات التباين، ولاحظ التأثير الكبير لهذه القيم المنفصلة، والقيم غير العادية والمُبالغ فيها، على النتائج، مسببة في تصغير قيمة اختبارات الطالب–ت أكثر مما هي عليه. (وبشكل عام، تؤدي القيم الكبيرة لاختبار–ت إلى قيم احتمالية صغيرة.) قد يكون من المفيد إهمال القيمة المنفصلة من مجموعة المشاهدات، وحساب قيمة اختبار–ت من القيم الأخرى. وهذا يقود المسائل إلى الاشتقاق الرياضي لاختبارات الفرضية. كيف يمكن للكيميائي معرفة ما إذا كانت القيمة منفصلة؟ وكم من القيم المنفصلة علينا أن نهمل؟ وهل يستطيع الكيميائي الاستمرار في استخدام جداول الاحتمالية لاختبار الإحصاء القياسي إذا أهملنا القيم المنفصلة؟

مضى فرانك ويلكوكسن في بحثه في كل ما كتب. من المؤكد أن أساتذة الرياضيات الذين قدموا الأساليب الإحصائية قد تعرضوا لهذه المشكلة من قبل! ولكنه لم يجد مرجعاً يذكر ذلك. اعتقد ويلكوكسن أنه قد وجد مخرجاً لهذه المشكلة. تضمنت حسابات سخيفة مبنية على تركيبات وتعديلات للأرقام المراقبة (لقد تم ذكر القيم التوافقية ل ف. ن. ديشيد في الفصل السابق). بدأ العمل لاستنتاج أسلوب يحسب القيم التوافقية.

كان في ذلك حماقة! كيف لعالم كيمياء مثل ويلكوكسن التحقق مع أرقام بسيطة ومضجرة؟ لابد من وجود إحصائي قد فعل ذلك قبله! فبحث في كتب الإحصاء السابقة. لم يجد شيئاً مثل ذلك . قدم بحثاً لمجلة البيومتريكس من أجل التأكد من أسلوبه الرياضي في الدرجة الأولى (من غير الاشتباه بمجلة البيومتريكا لبيرسون). مازال فرانك يعتقد أن هذا ليس بالعمل المبتكر واعتمد في بحثه على الحكام لمعرفة أين تم نشر هذا الموضوع من قبل ليرفض عمله. وبرفضهم ما كتبه، فإنهم سيعلمونه عن المراجع الأخرى. ومن ناحية أخرى، فقد قرر الحكام ورؤساء التحرير، أن العمل مبتكر. ولم يفكر أحد من قبل بهذا الأمر، وتم نشر بحثه سنة 1945.

ما لم يعرفه ويلكوكسن ولا رؤساء تحرير البيومتريكس، أن عالم الاقتصاد واسمه هنري ب. مان Henry B. Mann، وطالب جامعي في كلية الإحصاء، في جامعة ولاية أوهايو، واسمه د. رانسوم ويتني D. Ransom Whitney، كانا يعملان

على مسألة مشابهة. كانا يحاولان تنظيم التوزيعات الإحصائية لدرجة معقولة، قد يقول أحدهم فيها إن توزيعات الرواتب في سنة 1940 أقل منها في سنة 1944. توصلا إلى أسلوب في التنظيم يشمل سلسلة من أساليب الحساب السهلة والممتعة.

قاد هذا مان وويتني إلى إحصاء اختباري يتم فيه حساب التوزيعات من الحسابات التوافقية، مثل نوع الحساب الذي استخدمه ويلكوكسن. وقاما بنشر بحث يصفون فيه تكنولوجيتهم الجديدة سنة 1947، بعد ظهور بحث ويلكوكسن بسنتين. اتضحت لاحقاً الصلة القوية بين اختبار ويلكوكسن واختبار مان-ويتني بإعطاء القيم الاحتمالية نفسها. و كلا الاختبارين تضمن شيئاً جديداً. وحتى ظهور بحث ويلكوكسن، كان يعتقد أنه يجب على الإحصاءات الاختبارية أن تعتمد على تقديرات للمتغيرات والتوزيعات. لقد كان هذا اختباراً، بالرغم من عدم تقديره للمتغير. لقد قارن بين البيانات المبعثرة تحت المراقبة، وبين ما يمكن توقعه من البعثرة النقية والعشوائية. إنه اختبار لا علاقة له بالمتغيرات[2].

وبهذه الطريقة، تخطت الثورة الإحصائية أفكار بيرسون

(2) في الحقيقة ولمزيد من الإنصاف لقانون ستيغلر في الأسماء المستعارة، لم يكن ويلكوكسن أول من اقترح أساليب من دون المتغيرات. اتضح أن بعض أعمال كارل بيرسون سنة 1914 اقترحت بعض هذه الأفكار. ومن ناحية أخرى فإن اتجاه من دون المتغيرات، لم يكن مفهوماً ليشكل ثورة عنيفة إلى أن ظهر عمل ويلكوكسن في هذا المجال.

الأساسية. أصبحت تتعامل الآن مع توزيعات القياسات من غير اللجوء إلى المتغيرات. ومن غير معرفة الكثير في الغرب، قام أندريه كولموغوروف في الإتحاد السوفييتي، وطالب ن.ف. سميرنوف N.V.Smirnov بالبحث في عدة اتجاهات للمقارنة بين التوزيعات التي لم تستخدم المتغيرات. لقد فتحت أعمال ويلكوكسن ومان وويتني، نافذة جديدة في البحث الرياضي عن طريق توجيه الاهتمام للطبيعة الضمنية للسلاسل المنظمة، وسرعان ما انضمت إليه أعمال سميرنوف ـ كولموغوروف.

تطورات إضافية

بافتتاح نافذة جديدة على الأبحاث الرياضية، بدأ الباحثون بالنظر إليها بطرق مختلفة. لقد تبع عمل ويلكوكسن الأصلي اتجاهات بديلة. اكتشف هيرمان تشيرنوف وأ. ريتشارد سافاج Herman Chernoff and I. Richard Savage إمكانية النظر في اختبار ويلكوكسن من خلال القيم المتوقع نسبتها للإحصاء المنظم؛ وبإمكانها توسعة الاختبار الذي لا يعتمد على المتغيرات إلى مجموعة من الاختبارات، تشمل توزيعات ضمنية مختلفة، لا يحتاج أي منها إلى حساب للمتغير. وأصبح مع بداية سنة 1960 هذا النوع من الاختبارات (والتي يعبر عنها الآن «بالاختبارات الخالية من التوزيعات»)، الموضوع الساخن في مادة البحث. ملأ طلاب الدكتوراه الجوانب الصغيرة في النظرية لدعم رسائلهم، وتم تسخير الاجتماعات خصيصاً لهذه النظرية الجديدة. استمر ويلكوكسن بالعمل في هذا المجال متوسعاً في

مجال الاختبارات أثناء تطويره الحسابات العشرية بالغة الذكاء، لاستعمالها في الحسابات التوافقية.

وفي سنة 1971، أصدر التشيكي ياروسلاف هايك Jaroslav Hajek كتاباً تعريفياً يعطي نظرة موحدة لهذا المجال بأكمله. اكتشف هايك، الذي توفي سنة 1974 عن عمر يناهز الثامنة والأربعين، تعميماً ضمنياً لكل الاختبارات التي لا تستخدم المتغيرات، وربط هذا الاتجاه العام لحالات ليندبيرغ ليڤي لنظرية النهاية المركزية. عادة هذه هي الطريقة المتبعة في الأبحاث الرياضية. وبمعنى آخر فإن الرياضيات كلها مرتبطة ببعضها، ولكن تستغرق الطبيعة المحددة لهذه الروابط، وأبعاد استثمارها عدة سنوات كي تظهر للعيان.

وبتعقبه للمعاني المتضمنة حول اكتشافه الإحصائي، ترك ويلكوكسن مجال عمله الأصلي في علم الكيمياء، ليرعى مجموعة الخدمات الإحصائية في السيناميد الأمريكي، وقسم المختبرات ليديرل Lederle. التحق سنة 1960 بكلية قسم الإحصاء في جامعة ولاية فلوريدا Florida State University، والذي أثبت فيها جدارته العالية، كأستاذ قدير وباحث ومرشح للدكتوراه عالمياً لمرات متعددة. ترك وراءه بعد وفاته سنة 1965 تلاميذه وابتكاراته الإحصائية التي استمر أثرها بشكل واضح في هذا المجال.

مسائل غير محلولة

أدى التطور في طريقة الحساب بدون المتغيرات إلى صيحة

من النشاطات في هذا المجال الجديد. ومع ذلك، لم يكن هناك رابطة قوية بين أساليب استخدام المتغيرات التي كانت تستخدم سابقاً وبين الأساليب بدون المتغيرات. يظهر تساؤلان لا حل لهما:

1. إذا كان للبيانات المعطاة توزيع متغير معروف، مثل التوزيع الطبيعي، فما مدى سوء وقوع التحليلات في الخطأ إذا استخدمنا أساليب من غير متغيرات؟
2. إذا لم تتفق البيانات تماماً مع نموذج المتغيرات، فإلى أي مدى يجب أن تكون البيانات بعيدة عن النموذج قبل أن تكون الأساليب من دون المتغيرات أفضل للاستخدام؟

تلقى سنة 1948 رؤساء تحرير مجلة السجلات الرياضية الإحصائية Annals of Mathematical Statistics بحثاً من بروفيسور رياضيات غير معروف في جامعة تاسمانيا University of Tasmania، في جزيرة على الشاطئ الجنوبي من أستراليا. قام هذا البحث الرائع بحل كلتا المسألتين. نشر إدوين جيمس جورج بيتمان Edwin James George Pitman ثلاثة أبحاث سابقة في مجلة المجتمع الإحصائي الملكي Journal of the Royal Statistical Society وأخرى في أحداث مجتمع كامبردج الفلسفي Proceedings of the Cambridge Philosophical Society والتي، باستعادة الأحداث الماضية، وضعت أسس أعماله التالية ولكنها أُهملت أو نُسيت. باستثناء هذه الأبحاث الأربعة، وبيتمان الذي كان عمره اثنتين وخمسين سنة عندما قدم مقالته التي لم تُنشر ولم تُعرف، للمجلة السنوية.

ولد إ. ج. بيتمان E.J.G. Pitman سنة 1897 في ملبورن Melbourne باستراليا Australia. كان طالب بكالوريوس في جامعة ملبورن University of Melbourne، فجاءت الحرب العالمية الأولى وأعاقت تعليمه، فالتحق بالجيش لمدة سنتين. ثم عاد لإتمام تعليمه. «في تلك الأيام»، كما كتب لاحقاً، «لم يكن هناك كليات رياضيات يتخرج فيها الطلبة من جامعات أستراليا». بعض الجامعات كانت تعطي أموال البعثات للطلبة المتفوقين للتخرج من جامعات إنجلترا، ولكن لم تفعل جامعة ملبورن ذلك. وحيث إنني لم أتلق أي تدريب في مجال الأبحاث، غادرت جامعة ملبورن بعد أربع سنوات من الدراسة هناك؛ ولكنني كنت اعتقد أنني تعلمت كيف أدرس واستخدم الرياضيات، وكنت مستعداً لمواجهة أي مشكلة تعترضني... وكانت أول مشكلة كيف أعيل نفسي.

كانت جامعة تاسمانيا تبحث عن أستاذ رياضيات. قُدم بيتمان إليها وأُطلق عليه بروفيسور الرياضيات. كان القسم بأكمله يتألف من البروفيسور الجديد ومحاضر يعمل وقتاً إضافياً. والمطلوب من هذا القسم إعطاء دورات في رياضيات ما قبل التخرج لطلاب الأقسام الأخرى، وكان البروفيسور الجديد مشغولاً بمجموعة من الدورات أخذت معظم وقته. وعندما قررت لجنة الأمناء استئجار بروفيسور رياضيات يعمل وقتاً كاملاً، سمع أحد أعضاء اللجنة بنوع جديد من الرياضيات اسمه الإحصاء؛ لذلك فإن المتقدم الجديد سوف يتم سؤاله عن

استعداده لتدريس مادة الإحصاء (مهما كانت).

أجاب بيتمان، «لم أستطع الادعاء أن لدي معرفة خاصة بالنظرية الإحصائية؛ ولكنني سأكون مستعداً إذا تم تعييني لإعطاء محاضرة في هذه المادة سنة 1927». لم يكن لديه معرفة خاصة، أو أي نوع من المعرفة عن النظرية الإحصائية. لقد تلقى في ملبورن دورة في علم المنطق المتقدم، وخصص البروفيسور محاضرتين في مادة الإحصاء. وكما عبر عنها بيتمان، «لقد قررت في وقتها أن مادة الإحصاء لا تجذبني بحال من الأحوال، ولم اكن لأهتم لها أبداً».

وصل الشاب إ. ج. بيتمان في خريف 1926 إلى هوبارت Hobart بتسمانيا، ولم يكمل تعليمه الجامعي ليقوم بمقام البروفيسور، في كلية إقليمية بعيدة كل البعد عن التقدم الفكري الموجود في لندن وكامبريدج. «لم أتمكن من نشر أي شيء»، كتب قائلاً، «حتى سنة 1936. كان هناك سببان لتأخري في النشر؛ ثقل العمل الذي على عاتقي، والطريقة التي نشأت فيها»، وما يعنيه قلة تدربه في أساليب البحث الرياضي.

توسع قسم الرياضيات في جامعة تسمانيا، عندما أرسل بحثه الرائع إلى مجلة السجلات الإحصائية الرياضية سنة 1948. ولم يكن لديهم سوى بروفيسور واحد (بيتمان)، وبروفيسور آخر مشارك، ومحاضرين ومعلمين. كانوا يدرسون مجالات واسعة في الرياضيات التطبيقية والنظرية. كان بيتمان يعطي اثنتي عشرة محاضرة أسبوعياً وبعض الدروس في يوم السبت. أصبح لديه

الآن بعض الدعم لبحثه. كما ابتدأت حكومة الكومنويلث منذ سنة 1936 بإعطاء نحو ثلاثين ألف جنيه كل عام من أجل البحث العلمي في جامعات أستراليا. ويتم توزيع ذلك حسب الولايات ونسب السكان فيها؛ ولأن تسمانيا كانت من الولايات الصغيرة، فقد بلغت القيمة نحو ألفين وأربعمائة جنيه سنوياً لكل أقسام الجامعة. ولم يصرح بيتمان كم كان نصيبه منها.

انشغل بيتمان تدريجياً بأنواع مختلفة من البحث. لقد عالج بحثه الأول الذي نشر مسألة القوة المائية (الهيدروديناميكية)، وتناولت أبحاثه الثلاثة الأخرى جوانب معينة في نظرية اختبار الفرضيات. لم تكن هذه الأبحاث رائعة بحد ذاتها، ولكنها كانت تشمل رسائل بيتمان التعليمية. كان يبحث عن كيفية تطوير الأفكار وربط البنية الرياضية مع بعضها بعضاً.

طور بيتمان أسلوباً واضحاً في الاستنتاج حول طبيعة اختبارات الفرضية الإحصائية، والعلاقات المتبادلة بين الاختبار القديم (باستخدام المتغيرات)، والاختبار الجديد (من دون متغيرات)، وذلك بينما كان يعمل فيه ببحثه سنة 1948. فقام بمواجهة المسألتين الواضحتين بأساليبه الجديدة.

اندهش الجميع مما توصل إليه، إذ كانت جودة الاختبارات من دون المتغيرات، بمثل جودة اختبارات المتغيرات، حتى عند صحة الافتراضات الأصلية. تمكن بيتمان من الإجابة عن السؤال الأول: مامدى سوء الأمر عند استخدامنا اختبارات من دون المتغيرات، في حالة معرفتنا لنموذج

المتغيرات وعند ضرورة استخدامنا اختبار المتغيرات المحدد؟ ليس بهذا السوء، أجاب بيتمان.

أما إجابته عن السؤال الثاني فكانت أكثر دهشة. كم يتوجب علينا أن نبتعد عن نموذج المتغيرات من أجل صحة اختبارات من دون المتغيرات في حال لم تتلاءم البيانات مع نموذج المتغيرات؟ أوضحت حسابات بيتمان أنه بقليل من الانحراف عن نموذج المتغيرات تصبح نتائج اختبارات من دون المتغيرات أفضل بكثير من اختبارات المتغيرات.

يبدو الأمر وكأن فرانك ويلكوكسن، عالم الكيمياء الذي كان متأكداً من أن أحداً لم يسبقه في هذا الاكتشاف البسيط، قد تعثر من على صخرة فيلسوف حقيقية. أشارت نتائج بيتمان أنه يجب على كل اختبارات الفرضية أن تكون من دون متغيرات. كان اكتشاف بيرسون للتوزيع الإحصائي المبني على المتغيرات بمثابة الخطوة الأولى. واستطاع علماء الإحصاء الآن من التعامل مع التوزيع الإحصائي من غير أن يقلقوا بشأن متغيرات محددة.

هناك أمور دقيقة تتضمنها أمور أدق في علم الرياضيات. وفي أعماقها تظهر أمور بسيطة. توصل ويلكوكسن، ومان وويتني، وبيتمان إلى فرضيات حول توزيع البيانات. قد يتطلب فهم هذه الفرضيات خمساً وعشرين سنة أخرى. اكتشف ر.ر. بهادور R.R.Bahadur و ل.ج. («جيمي») سافاج («jimmie» L.J. Savage في جامعة شيكاغو سنة 1956 أول مسألة في التوزيع. أشار إلي صديق هندي كنت قد عرضت بحث بهادور وسافاج

عليه قبل عدة سنوات، إلى تطابق أسمائهم، فبهادور تعني «محارب» في اللغة الهندية، فتطلب الأمر محارباً وهمجياً لإثارة أول انفجار لنظرية الاختبارات الإحصائية من دون المتغيرات.

أتت المسائل التي أوضحها سافاج وبهادور من صميم المسألة التي أشارت إلى ويلكوكسن باستخدام اختبارات من دون المتغيرات: مسألة القيم المنفصلة. وإذا كانت القيم المنفصلة نادرة و«خاطئة» في المراقبة، فإن أساليب من دون المتغيرات تقلل من تأثيرها في التحليلات. وأما إذا كانت القيم المنفصلة جزءاً من النظام المتشابك للبيانات، فإن الانتقال إلى أساليب من دون المتغيرات يزيد الأمر سوءاً. سوف نبحث في مسألة التوزيعات المتشابكة في الفصل الثالث والعشرين.