مقدمة الكتاب

لا شك في أن لـدى كل قـارىء إدراكا معينـا لكلمـة «الإحصاء». وتنشأ هـذه المدارك مما اعتـاد عليه عامة النـاس في أيامنا هـذه من إطلاع شبه يومي على معلـومات إحصائية تقدمها النشرات الإحصائية الرسمية للحكومات والهيئات الدولية ومن خلال الصحافة ووسـائل الإعلام المسموعة والمرئيـة. وهي تشير، في الغالب، إلى أن الإحصاء هو نوع من التجميع لقدر كبير من المعلومات الكمية أو الكيفية واختزالها وتقديمها على شكل جـداول أو رسوم أو أشكـال وخطوط بيـانية معبرة وسهلـة التناول والادراك. كما يمكن أن تتضمن حساب مجاميع أو معدلات أو نسب مئوية أو ماشابهها.

وربما كانت إحـدى الفوائد المتـوخاة لمقـرر ابتدائي في الإحصاء هـي التزود بفهم أكثر شمـولا وعمقا ودقة لكلمة «الإحصـاء» بمعناها العلمي المعاصر، فـالفهم السائد لكلمة الإحصاء يندرج، في الواقع، تحت عنـوان «الإحصاء الوصفي». ولكن الإحصاء يلعب اليـوم دورا مزدوجـا إذ يقـدم إلى جـانب الإحصاء الـوصفي طـرقـا للاستقـراء، فنستخلص من البيان الإحصائي نتائج معينة بطريقة تتسم بالموضوعية. ولا شك في أن جانب الاستقـراء الإحصائي هو الجانب الأكثـر إثارة ومدعاة لـلاهتمام، ويشكل اليوم إحدى أهم الأدوات المعـاصرة لاتخاذ قرار أو القيام بتنبؤ في ظروف تخضـع للمصادفة، أي ظروف لا يمكن معها التنبؤ بالنتـائج أو محاولة التعرف على القرار السليم من خلال قوانين علمية معـروفة. ويهدف كتاب ابتدائي كهـذا، فيما يهدف، إلى نقل القارىء إلى مشارف الاستقراء الإحصائي. وإذا كان الكتاب بأكمله لا يطمح في هذا الخصوص إلى أكثر من ذلك، فمن المستحيل على مقدمة مختصرة وسريعة أن تدعي المقدرة على تقديم

فكرة واضحة ودقيقة عن ما هية الاستقراء الإحصائي. ومع ذلك لا بد لنا من إلقاء بعض الضوء على مصطلحين أساسيين في علم الإحصاء، هما المجتمع والعينة. وسنحاول تلمس العناصر الأساسية للمسألة الإحصائية مهتدين في ذلك بنقاط رئيسة تضمنتها نشرة علمية بعنوان: "Careers in Statistics" أصدرتها عام ١٩٦٢م أكبر هيئتين علميتين إحصائيتين في الولايات المتحدة هما:

"The American Statistical Association"

"The Institute of Mathematical Statistics"

يهدف الإحصاء باعتباره فرعا من فروع الطريقة العلمية إلى دراسة خصائص عددية للمجتمعات. ولكن ماذا نقصد بمصطلح «مجتمع»؟

في معظم الأبحاث العلمية لا ينصب الاهتمام على البيان الإحصائي المدروس وإنما يكتسب البيان أهميته من كونه ممثلا لمجموعة أكبر من المعلومات الإحصائية يشكل البيان المدروس جزءا منها.

وعلى سبيل المثال إذا سألنا مائة طالب من طلاب كلية العلوم عن رأيهم في الدورة المكثفة في اللغة الإنجليزية، فإن آراء الطلاب المائة لذاتها ليس لها أهمية كبيرة، وإنما تأتي أهميتها من كونها مؤشرا للرأي السائد بين مجموعة أكبر بكثير من الطلبة هم جميع طلبة كلية العلوم. ونصطلح في الإحصاء على تسمية الطلاب المائة «عينة» ومجموعة طلبة كلية العلوم «المجتمع». والدراسة تهدف أول ما تهدف إلى التعرف على الرأي السائد بين طلبة كلية العلوم إزاء الدورة المكثفة. أي أن هدف الدراسة هو المجتمع.

وإليك مثال ثان. لنفرض أن عدد المستجدين في الجامعة هو خمسة آلاف طالب، وأن باحثا يرغب في معرفة مجموع أوزان المستجدين. فالمجتمع هنا هو كافة المستجدين في الجامعة ويمكن للباحث أن يقوم بوزنهم واحدا فآخر ويصل إلى ما

يريد، كما يمكنه اتباع طريقة أخرى، فيختار مائتي طالب، مثلا، ويقيس أوزانهم، ومن هذه القياسات يحاول تقدير الوزن الكلي لجميع الطلبة المستجدين. ويشكل الطلاب المائتان الذين اختارهم عينة من مجتمع المستجدين، ويسمى وزن الطالب، قياسا أو ملاحظة أو مشاهدة.

ومثال ثالث. لنفترض أن باحثا في العلوم الطبية يرغب في تثمين دواء جديد لمرض معين. وقد طبق المعالجة الجديدة على عشرين مريضا، فمن وجهة نظر الباحث لا يشكل المرضى العشرون المجتمع الذي يهدف إلى دراسته، وإنما يشكلون عينة منه فقط. وهو لا يهتم بنتائج المعالجة بين هؤلاء المرضى العشرين لذاتهم وإنما يهمه معرفة مدى نجاح المعالجة من أجل أي مصاب بذلك المرض. والمجتمع الذي يهمه هو إذا مجتمع جميع المصابين بهذا المرض ويمكنهم تلقي العلاج، سواء من كان منهم موجودا الآن ومن سيوجد في المستقبل. والمجتمع هنا هو نوع من المجتمع التصوري، إذ لا وجود له في الواقع المحسوس، ومع ذلك فهو المجتمع الذي ينصب عليه الاهتمام، لأن الباحث يريد تثمين معالجته وهي تطبق على المصابين بهذا المرض بصورة عامة، وليس على المرضى العشرين الذي يشكلون العينة.

وعندما يكرر باحث في العلوم الفيزيائية، مثلا، تجربة قياس ثابت فيزيائي معين، عشر مرات، فنصطلح على اعتبار القياسات العشرة، التي يحصل عليها، عيّنة من مجتمع تصوري يتضمن جميع القياسات التي كان سيحصل عليها الباحث لو أنه استمر في تكرار تجربته عددا لا نهائيا من المرات. والمجتمع في هذه الحالة تصوري وغير محدود (لا نهائي).

وبصورة عامة، يمكن القول إن المجتمع هو جملة الأشياء أو العناصر التي تشكل هدف الدراسة، أما العينة فهي الجزء من المجتمع الذي يخضع بالفعل للدراسة.

والسؤال الذي يفرض نفسه الآن هو: لماذا لا تتناول الدراسة المجتمع كله؟ وإذا كنا نريد معلومات تتعلق بالمجتمع كله فلماذا نكتفي بجمع معلومات من عينة منه

فقط؟ وللإجابة نقول إن المجتمعات غالبا ما تكون من الضخامة بحيث يكون إخضاع كل عنصر فيها للدراسة نوعا من المستحيل. وحتى عندما يكون ذلك ممكنا من الناحية النظرية، على الأقل، فإن ما تتطلبه الدراسة من جهود وزمن ونفقات طائلة تجعلها من الناحية الواقعية أمرا غير عملي البتة. لا بل قد تقدم لنا دراسة متأنية ودقيقة للعينة، من المعلومات، أفضل مما تقدمه دراسة تتناول المجتمع كله، ولكنها دراسة تنقصها الدقة وتسودها الفوضى. ففي مثال المستجدين يمكن للباحث أن يقوم بوزن كل طالب من طلاب العينة المائتين بدقة، ولكنه إذا حاول الحصول على أوزان المستجدين بآلافهم الخمسة فقد يضطر إلى الاقتناع بتوجيه سؤال إلى الطالب عن وزنه ويكتفي بتسجيل الإجابة، وقد يكون الجواب بعيدا كل البعد عن الدقة. أما عندما يكون المجتمع تصوريا فنجد أنفسنا ملزمين بالإعتماد على عينة، ولا خيار لنا في ذلك.

وفي الإحصاء نعتمد عادة على عينات نختارها عشوائيا ونسميها عينات عشوائية. ولكن ماذا نقصد بكلمة عشوائية؟ ولماذا نريد للعينة أن تكون عشوائية؟ لنفرض، على سبيل المثال، أننا نريد تقديم جائزة لطالب نختاره عشوائيا من فصل يتضمن ثلاثين طالبا، فكيف يتم مثل هذا الاختيار العشوائي؟ إن أي طريقة اختيار نقتنع جميعا أنها خالية تماما من التحيز لمصلحة طالب دون آخر هي طريقة يمكن أن توصف بالعشوائية. لنقم، مثلا، بتسجيل اسم كل طالب على قطعة واحدة من الورق، ثم نطوي هذه الأوراق ونضعها في قبعة، ثم لنخلطها جيدا قبل أن نختار واحدة منها، دون النظر إلى القبعة، ونقدم الجائزة للطالب الذي كُتب اسمه عليها. وسنوافق على وصف هذه الطريقة بأنها عشوائية إذا لم تتضمن أي عمل أو تصرف يمكن أن يساعد على التحيز في الاختيار لمصلحة طالب أو طلاب معينين. فالقطع من الورق يجب أن تكون من الحجم والملمس ونوع الورق نفسه، وتطوى بالطريقة نفسها بحيث تكون متماثلة في كل شيء باستثناء الإسم الذي كتب عليها، وبحيث يمتنع على من يختار الاستفادة بأي صورة من الصور من حاسة اللمس أو النظر. ولا بد أن تخلط الأوراق خلطا جيدا قبل الشروع في اختيار إحداها. وبالمعنى الاصطلاحي للكلمة تطلق كلمة «عشوائي» على أي طريقة اختيار لا هدف لها ولا غاية. ونتحدث عادة عن اختيار أسماء من قبعة عشوائيا، وعن اختيار سنابل قمح عشوائيا من حقل قمح، واختيار أسرة عشوائيا من مجتمع من الأسر في مدينة، الخ. ونعني بذلك أن يتم

الاختيار بفعل المصادفة البحتة وأن تتاح الفرصة نفسها عند كل سحب لكل عنصر من عناصر المجتمع الذي نسحب منه.

وسبب اعتمادنا على العشوائية في علم الإحصاء هو أنها تسمح بتطبيق الطرق الرياضية بسهولة، مما يؤدي إلى استخلاص نتائج تتعلق بالمجتمع بطريقة تتسم بالموضوعية. والجدير بالذكر أنها تقي من آثار التحيز الشخصي، إذ لا يجوز بالطبع أن نترك للباحث الحرية في اختيار عينته، فقد يختارها عندئذ بصورة متحيزة تدعم نظريته.

وفي المجتمعات المحدودة التي يمكن ترقيم عناصرها من 1 إلى عدد محدود N، حيث N عدد الوحدات أو العناصر في المجتمع، توجد جداول للأرقام العشوائية هي جداول كل رقم فيها اختير عشوائيا من بين الأرقام 9 ,..., 2, 1, 0. وقد أعدت بحيث يكون لكل رقم من هذه الأرقام الفرصة نفسها في أن يكون الرقم المسحوب وذلك عند كل سحب. ومن بين الجداول الأكثر انتشارا نجد تلك التي نشرتها مؤسسة راند (RAND) عام ١٩٥٥م، وتتضمن مليون رقم. ويعرض الجدول (ا) التالي ألف رقم عشوائي للتوضيح.

وعند استخدام هذه الجداول لاختيار عينة عشوائية بسيطة تكون الخطوة الأولى هي ترقيم الوحدات في المجتمع من 1 إلى N، حيث N عدد وحدات المعاينة في المجتمع، وإذا كان الرقم الأول (من اليسار) للعدد N بين 5 و 9 تكون الطريقة التالية مناسبة. فلنفرض للتوضيح أن المجتمع يتضمن 528 وحدة، أي N = 528، ونريد عينة عشوائية بسيطة حجمها 10 وحدات، فنختار، لا على التعيين، أحد أعمدة الجدول الخمسين، وليكن مثلا العمود 25 ونأخذ العمودين التاليين له وهما العمود 26 والعمود 27، فتعطينا الأرقام المتجاورة (الواقعة على السطر نفسه) من الأعمدة الثلاثة عددا من ثلاثة أرقام. نستعرض هذه الأعداد من الأعلى إلى الأسفل، ونختار الأعداد المتميزة العشرة الأولى الواقعة بين 001 و 528 فنجد 36، 509، 364، 417، 348، 127، 149، 186، 290، و162. وتكون العينة العشوائية المطلوبة هي الوحدات التي تحمل هذه الأرقام. (من أجل العددين الأخيرين قفزنا إلى الأعمدة 30، 31، 32). وعند اختيار عينات مختلفة يستحسن تغيير النقطة التي نبدأ عندها في الجدول من عينة إلى أخرى.[1]

(١) انظر كتاب "Sampling Techniques" لمؤلفه W. Cochran، الطبعة الثالثة، صفحة 19.

جدول (ا) يوضح ألف رقم عشوائي

	00-04	05-09	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39965	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30286	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067

وطريقة اختيار عينة عشوائية تصبح مسألة معقدة عندما لا تتوفر قائمة بوحدات المجتمع، أو لا يمكن ترقيم تلك الوحدات. وقليل من التأمل في كيفية اختيار عينة عشوائية من المنازل في مدينة كبيرة، أو من الأشجار في غابة، أو من السمك في بحيرة، أو من المرضى المصابين بعلة معينة، ينبغي أن يقنعنا بأن الصعوبات عديدة ومتنوعة. ويبقى ابتكار طريقة مناسبة تضمن عشوائية العينة أمرا مطلوبا من الباحثين العاملين في حقل المعاينة الإحصائية. وتقنيات اختيار عينة أو ما يسمى بتقنية المعاينة الإحصائية هو عنوان بارز وضخم في أدبيات الإحصاء.

وعندما يكون المجتمع تصوريا وغير محدود، كما في مثال تجربة قياس ثابت فيزيائي، نصطلح على اعتبار التكرارات الـ n الأولى للتجربة عينة عشوائية حجمها n من ذلك المجتمع، شريطة أن تتم التكرارات تحت الشروط والظروف نفسها وأن يكون بعضها مستقلا عن بعض.

العناصر الرئيسة لمسألة إحصائية

سنحاول الآن عرض مـزيد من الأمثلة نستعرض من خـلالها أشكالا من المسائل الإحصائية ونتلمس منها العناصر الرئيسة في مسـألة إحصائية. ونسوقها هنا على سبيل المثال لا الحصر، ويحتاج فهم الطـرق المتبعة في هذه المسائل إلى العـديد من المقررات في نظرية الإحصاء وتطبيقاتها.

١) تقوم إدارة مصنـع بتفتيش شحنات البضـاعة الخام الواردة إلى المصنـع وعلى أساس هذا التفتيش تتخذ قرارا بقبول البضاعة أو رفضها وإعادتها إلى الممول. ويمكن أن يتضمن التفتيش اختيـار عينة عشـوائيـة من عشرين وحـدة، مثلا، من الشحنـة الواردة وفحصهـا بدقـة للوصول إلى عـدد الوحـدات غير المقبولـة من بينها. وعلى أساس هـذا العدد يُتخـذ قرار برفـض الشحنة أو قبولها. إن طـريقة اختيـار العينة وتحديد حجمها وطريقة اتخاذ القرار هي كلها مسائل إحصائية.

٢) يتوقف إنتاج منشأة للصناعات الكيميائيـة على عوامل عدة. ويمكن وضع معادلة تنبؤ تربط بين الإنتـاج وبين مستويات هذه العوامل وذلك بعـد ملاحظة وتسجيل قيمة الإنتـاج وقيم هذه العوامل لفترة زمنيـة معينة. ولكن كيف نضع معادلـة تنبؤ جيدة؟ وعنـد استخدام المعادلـة للتنبؤ بالإنتـاج لن يكون التبنـؤ مساويا لـلإنتاج الفعلي، بل سيكون هنـاك دائما فارق أو حيـدان بين قيمة التنبؤ والقيمـة الفعلية، فكيف يمكن التحكم بهذا الفـرق أو الحيـدان ووضع حـدود دنيـا وعليـا لمقـدار الحيـدان؟ وأخيرا ما هي العـوامل الأكثـر أهمية في عمليـة الإنتـاج؟ وهذه جميعهـا مسائل إحصـائية. ونواجـه مثل هذه المسائل في العـديد من ميـادين المعرفة نـذكر منها، على سبيل المثـال لا الحصر، العلوم السلوكيـة (علم التربية، علم الاجتماع، علم النفس. . .)، العلوم الحيوية، العلوم الهندسية والصناعية، العلوم الزراعية، العلوم الاقتصادية، الخ.

٣) يدعي فريق من الباحثين في العلوم الطبية أنهم توصلوا إلى لقاح جديد فعّال في مجال الوقاية من الـزكام. فهل ترفض دعواهم أم يجاز تصنيع اللقاح وطـرحه للاستهلاك

على نطاق واسع؟ ولنفرض للتبسيط أن اللقاح أعطي لعشرة أشخاص روقبوا طيلة فصل الشتاء وقد جانب الزكام ثمانية منهم فهل يكون اللقاح فعالا؟ إن تصميم التجربة واختيار الأشخاص وتحليل النتائج للوصول إلى قرار حول صلاحية اللقاح هي جميعها مسائل إحصائية. وكم من المواقف المشابهة يتعرض لها الباحثون يوميا ويتعين عليهم الحكم أو اتخاذ قرار بين بديلين مطروحين!

٤) لنفرض أننا قدمنا موضوعا معينا بطريقتين مختلفتين في التدريس إلى مجموعتين من الطلاب لا تتفوق إحداهما على الأخرى في مقدرتها العامة. ثم حصلنا في نهاية الفترة الدراسية على قياس معين لما أنجزته كل من الطريقتين، وعلى أساس من هذه المعلومات نتساءل عما إذا كانت نتائج التجربة تقدم دلالة كافية على تفوق إحدى الطريقتين على الأخرى؟

٥) وعلى مستوى أعم يمكن أن تتطرق الدراسة إلى عدد من المعالجات التي يعتقد أن لها أثرها على ناتج نهائي. فلنفرض ثلاثة أنواع من الأسمدة تختلف في تركيبها من حيث نسبة الأزوت والبوتاس والفوسفات في كل منها. ويمكن تطبيقها في حقول القمح بثلاثة مستويات مختلفة، فنرش مساحة معينة من الأرض، بعدد من الكيلوغرامات أو ضعفي ذلك، أو ثلاثة أضعاف ذلك، كما يمكن توقيت رش السماد في فترتين مختلفتين، فأي الأسمدة، وأي مستويات التطبيق، وأي توقيت للرش أفضل بالنسبة لزيادة إنتاج القمح؟ وإذا كان المستوى الأعلى هو الأجود، مثلا، فهل هناك مجال لمزيد من تحسين الإنتاج من خلال رفع مستوى التطبيق؟ وإلى أي حد يمكن أن نمضي في مثل هذه العملية؟

وتختلف الأمثلة السابقة في طبيعتها ودرجة تعقيدها. إلا أنها تشترك في أن كلا منها ينطوي على تنبؤ أو اتخاذ قرار. بالإضافة إلى أننا في كل من هذه الأمثلة قد أخذنا عينة من كيان أكبر بكثير يدعى المجتمع. والجدير بالذكر أن نتائج العينة لا تُملي علينا القرار أو التنبؤ، فعند مقارنة طريقتين مختلفتين في التدريس، مثلا، لا نلجأ إلى المقارنة الظاهرية المباشرة بين أداء الطريقتين لتفضيل إحداهما على الأخرى، وإنما نلجأ إلى طرق

إحصائية تسمح لنا باتخاذ القرار في سياق العينة التي بين أيدينا وجميع العينات الأخرى الممكنة من الحجم نفسه التي كنا سنحصل عليها لو أننا كررنا تجربة أخذ العينة مرة بعد أخرى. ونعتمد هنا اعتمادا حاسما على نظرية الاحتمالات، ويبدو أساسيا إذاً أن نقوم بتحليل البيان الإحصائي الملحوظ ثم نستقرىء، استنادا إلى التحليل، المجتمع الذي جاءت منه العينة. وثمة عنصر أساسي ثالث لا يبدو بوضوح، فالبيان الإحصائي يحوي قدرا معينا من المعلومات عن الخاصة المدروسة من خصائص المجتمع. وقد تم الحصول على هذه المعلومات نتيجة جهد مبذول كلف مالا ووقتا صرفناهما في تجربة معينة. ولا بد أن قدرا معينا من النفقات والجهود سينتج مقادير مختلفة من المعلومات تبعا لطرق تجريبية مختلفة. ولذلك فمن الواجب تصميم التجربة أو تصميم إجراءات أخذ العينة بحيث نحصل على أكبر قدر من المعلومات المطلوبة لقاء نفقة معينة، ونلخص بقولنا إن المسألة الإحصائية تتضمن:

١ ـ تصميم التجربة أو طريقة أخذ العينة وتجميع البيانات.
٢ ـ تحليل البيان الإحصائي الناتج.
٣ ـ الإستناد إلى هذا التحليل للقيام باستقراء المجتمع الذي جاءت منه العينة.