مجلة "آفاق فكرية"

ص. ص. 230 - 245

رت م د: 2353 - 0367

رت م د إ: 2602-5507

الإحصاء ومعالجة البيانات

Statistics and data processing

فضيلة شعوبي

جامعة الوادي (الجزائر)، chaoubi-fadila@univ-eloued.dz

تاريخ الاستلام: 2023-01-27 تاريخ القبول: 2023-12-02 تاريخ النشر: 2023-12-30

ملخص:

يهدف البحث الحالي إلى نقل النُكهة الإحصائية الحديثة إلى أولئك الطلبة الدارسين للإحصاء وكذا للباحثين فيه. ولهذا فقد تناول البحث معنى الإحصاء وطبيعة البيانات، فضلاً عن تناولنا لمجال التفكير الإحصائي وإشكالية استخراج المعنى من تلك البيانات، دون أن ننسى الإشارة إلى الطرق الحديثة المستخدمة في معالجة البيانات الضخمة. هذا وقد جاءت أهدافنا تلك كلها بأسلوب بسيط وجذّاب، يبعث الثقة في نفوس العديد ممن يعتقدون أنّ علم الإحصاء مجال جاف وممل. وإتمامًا للفائدة تناول البحث جمالية الإحصاء ومواده الخام من دون الإشارة أو التلميح لأيّ من صيغه الرياضية أو حتى لمصطلحاته المتخصصة التي يتعذر فهمها، أو لرسومه البيانية المعقدة، وهذا بغرض تبديد الخوف ودحض الخشية البسيطة أو التبلد العقلي التّام لدى الفارين من الأرقام والمعادلات الرياضية. هذا وتضيف الدراسة الحالية إلى الأدبيات الموجودة كيف أنّ معظم الأفكار الإحصائية المهمة ما هي في حقيقتها سوى أفكار بديهية، تستمد قوتها من منطقيتها وقدرتها على استيعاب ظروف الحياة اليومية.

كلمات مفتاحية: إحصاء؛ تفكير إحصائي؛ بيانات؛ بيانات ضخمة.

Abstract:

The aim of the research is to convey the flavor of statistics to students studying statistics and researchers. The research dealt with the meaning of statistics and the nature of data. It also dealt with the field of statistical thinking and the problem of extracting meaning from that data, with broaching to the modern processing of big data. All our goals came in a simple and attractive way. This inspires confidence in the hearts of many who believe that statistics is a dry and boring field. To complete the interest, the research presents the aesthetics and raw materials of statistics, without referring or hinting at any of its mathematical formulas or even its incomprehensible specialized terms or complex graphs, in order to dispel fear and anxiety. Refute simple fear, or complete mental dullness for those fleeing numbers and mathematical equations. This research adds to the existing literature how most important statistical insights are in fact intuitive conditions of everyday life.

Keywords: Counting; statistical thinking; data; Big data.

مقدمة:

يعتبر الإحصاء أحد فروع الرياضيات الأساسية، وهو علم يهتم بالطرق والأساليب العلمية التي تعنى بجمع وعرض ووصف وتحليل البيانات، والاستقراء لصنع القرار. وعلى الرغم من أهمية هذا العلم سواء بالنسبة لأيّ شخص عادي في حياته اليومية، أو للمهتمين بدراسته من الطلاب، إلاَّ أنّنا نجده يمثل عائقا أمام الكثير من هؤلاء، كونه يبدو مجالاً معرفيًا صلبًا وجامدا لا يمكن استيعابه بسهولة، مثلاً إحدى المشكلات المواجهة لدى الإحصائيون في تدريسهم للإحصاء في الجامعات هي كيفية إثارة الدافعية لدى طلبتهم من أجل تعليمهم وتمكينهم منه، ذلك أنّ معظم طلبة الجامعات لاسيما أولئك المنتمين لتخصصات العلوم الاجتماعية أو الإنسانية، يعانون أثناء دراستهم لهاته المادة وليست لديهم أيّة رغبة في تعلمها، وإن حدث ونجحوا فيها فسيصبحون وبلا شك سعداء جدًّا حتى يودعونها للأبد. وفي حقيقة الأمر لسنا مبالغين إذا اعتقدنا أنّ هذا الفرع من فروع المعرفة يمكن أن تدب فيه الحياة إذا ما وقع في أيدي أشخاص مناسبين.

غالبًا ما يغيب عن مدرسي علم الإحصاء الاهتمام بإبراز الجانب الخفي منه، ونقصد بذلك جانب المتعة والإثارة من خلال ربطه بالعالم الحقيقي. فهلاَّ صادفت مثلاً أسئلة امتحان إحصائي صيغت على وجه طريف من قبل: هل يمكنك تعلم الإحصاء بنفسك؟ (يمكن إتقان بعضًا من المفاهيم الأساسية والاحتمالات والتعلم الآلي الإحصائي باستخدام الإنترنت)، كيف يمكنك الكذب في إحصائيات مذكرة تخرجك؟ (تتضمن الإجابة عن هذا السؤال الإشارة لطرق تصحيح البيانات) هل نجح الإحصاء في التعامل مع جائحة كورونا؟ إنَّك لم تصادف ذلك بالطبع. من جهة أخرى نادرًا ما يتضمن الامتحان أسئلة مثل: قارن بين مقاييس النزعة المركزية من حيث أولوية استخدامها؟ ما الإضافة التي يمكن أن يقدمها الإحصاء في تشخيص مشكلة بحثك؟ ما معنى القيم الشاذة وفي أي مرحلة وبأيّة طريقة يمكنك التعامل معها أثناء انجازك لبحثك؟ وغيرها من الأسئلة التي تثير انتباه الطلبة.

قليل من التفكير والتأمل كفيل بأن يقنعنا أنّ علم الإحصاء يمكن أن تكون له طريقته السّهلة وحتى الجذّابة في شرح الظواهر الاجتماعية من حولنا، وهو ما دفعنا لكتابة هذا البحث والذي ابتعدنا فيه عن أيّة معادلات رياضية أو رسوم بيانية معقدة على الرغم من أهميتها، ونود أن نلفت انتباه القارئ هنا أنّه من غير الممكن القيام بشرح واف ومفصل لكل جزئية من جزئيات علم الإحصاء

الطريف في هذه العجالة. وكلنا أمل بأن نسهم ولو بالقدر اليسير في وضع حد لمشاعر الخوف والقلق التي تنتاب بعض الأشخاص أو الطلبة، مقابل نقل متعة هذا العلم ونكهته للفارين منه وحتى المتهكمين عنه.

1. علم الإحصاء الحديث وأهميته:

عرف الإحصاء في الماضي على أنّه علم جمع المعلومات وترتيبها في جداول أو إبرازها في رسوم بيانية أو أشكال تصويرية، غير أنّ هذا التعريف لا يشير إلى مختلف الأدوار التي يلعبها هذا العلم، كونه يتفق وفقط مع العرض البياني والجدولي للبيانات، إذ ما فائدة تلخيص المعلومات في جداول أو رسوم بيانية، من دون استخراج المعنى من هاته البيانات المجمعة، ومن هنا ربما علينا أن نضع تعريفا أكثر دقة يرتكز إلى بعض أساسيات هذا العلم، فيمكن القول إذن أنّ الإحصاء "هو مجموعة من الإجراءات أو العمليات الخاصة بجمع وتنظيم وتحليل وعرض البيانات الكمية" (ريفا بيرمان براون ومارك ساوندرز، Reva Berman Brown ، 2010، ص. 15). ثمّة جانب أساسي يفتقده هذا التعريف كونه لا يشير إلى المصادفة والاحتمال اللّذان يعدان دعامتين أساسيتين للعديد من تطبيقات الإحصاء، وعليه أكدت بعض التعاريف لهذا العلم على أنّ الإحصاء هو "تكنولوجيا التعامل مع عدم اليقين" أو "العلم الرئيس للتنبؤ بالمستقبل" أو "لإنتاج ملخصات مناسبة من البيانات" (ديفيد جيه، DAVID J، 2016، ص.10). في الحقيقة وبسب تفرع علم الإحصاء وكثرة تطبيقاته واستخدامه للرياضيات ونظرية الاحتمالات والرسوم البيانية والحاسب الآلي، فإنّ الحاجة إلى تعريف جديد له أصبحت ضرورة ملحة لكي يكون شاملاً لكل الجوانب المستحدثة. وهكذا يمكننا القول بأنّ الإحصاء هو علم اتخاذ القرار في جميع جوانب الحياة، -في ظل ظروف عدم اليقين- لأنّه يجمع البيانات ويدرسها ويحللها ليستخلص النتائج عن ظاهرة ما، واتخاذ القرار بناءً على هذه النتائج. (هيوبرت، 2000، ص.21)

يستند مفهوم الإحصاء إلى الحصول على كمية كبيرة من البيانات، تدفع بالباحث للسؤال عن طريقة العمل معها بغرض استخراج المعنى منها، وجواب ذلك يكمن في وجوب تلخيصها وتقديمها في أبسط وأنسب صورة ممكنة، حتى يسهل التعامل معها واستخلاص الحقائق العلمية الواضحة والمحددة منها، وهنا يأتي الإحصاء الوصفي كأحد الوظائف الإحصائية المألوفة ليقدم خدماته، ذلك كون أنّ إحصاءات الوصف هي المستخدمة في حياتنا اليومية، ومن خلال هذا القسم من الإحصاء "يستطيع الباحث أن يغير من شكل البيانات بعد تصنيفها وتنظيمها وتلخيصها مستخدما في ذلك

الجانب الوصفي من الإحصاء، إذ يمكنه أن يطبق هنا مجموعة من المقاييس الإحصائية التي لا تتعدى حد الوصف مثل؛ مقاييس النزعة المركزية ومقاييس التشتت ومقاييس الارتباط والانحدار... وبالتالي الخروج بمجموعة من المؤشرات والمعدلات الإحصائية". (القوصي، 2015، ص.27) مع مراعاة أنّ استخدام أيًّا من هذه المقاييس يتوقف على نوعية البيانات ومستوى القياس سواء كان اسميا أو وصفيا أو ترتيبيا أو فئويا أو نسبة.

إنّ اقتصار الوظيفة الإحصائية على تحويل الإحصاءات الرقمية المجمعة من بيانات خام إلى معلومات في جداول إحصائية أو توزيعات تكرارية أو حتى أشكال ورسوم بيانية، إلى جانب الاستعانة ببعض من لمسات الحسابات التقنية غير المعقدة كالمتوسطات الحسابية ودرجات التشتت وغيرها، لا تشير سوى إلى الكيفية التي تتوزع بها البيانات وحسب، فحتى وإن كانت هذه الأساليب مهمّة في مجملها ولا يمكن الاستغناء عنها، فهي لا تحقق شيئا من الاستنتاج أو التنبؤ بالمستقبل واتخاذ القرار، لأنّها لا تمثل سوى جانب واحد من جوانب الإحصاء وهو الجانب الوصفي.

استنادا لما ذكر أعلاه، وجب الاعتماد إذن على جملة من النظريات والأساليب الإحصائية المساعدة على استكمال تقدير معالم وخصائص المجموعات الكبيرة، التي غالبا ما تحتوي على عدد لا نهاية له من المفردات، ومن ثمّ التنبّؤ بمستقبل الظواهر في هذه المجموعات. ومن هنا تقفز نظريتي الاحتمالات وكذا العينات كدعامتين أساسيتين لتحقيق الاستقراء والاستدلال، فضلاً عن اختبارات الفروض وما تتضمنه من خيارات مختلفة من أجل الوصول إلى تعميمات عن المجتمع. وفي هذه الحالة ما دمنا سنجري البحث على العينة وليس على المجموع الأصلي فلا شك أنّنا سوف نتخبط في الظلام، وأنّ نتائجنا ستكون طبعا مثارا للشك، لأنّها لا تخلو من عنصر "التخمين". غير أنّ تطبيق نظرية الاحتمالات في دراسة العينات سيضيء لنا السبيل لمعرفة الحدود التي لا يمكن أن يتجاوزها الخطأ المحتمل تسربه إلى نتائجنا، وبذلك ستزيد ثقتنا في هذه النتائج زيادة عظيمة". (مظلوم، ص. 15)

إنّ ما تمت الإشارة إليه سابقا من أساليب الوصف سوف يقتصر في حالتنا هذه على العينات وحسب، ومن هذا المنطلق سوف تتربع وظيفة الاستدلال للتعميم على المجتمع، وبما أنّ الاستدلال لا يمكن أن يكون مؤكد دوما فإنّ لغة الاحتمال تُستخدم عند عرض النتائج، وعليه لا نظن أنّنا سنكون مغالين حيال قولنا أنّ نظرية الاحتمالات هي جوهر علم الإحصاء. وبالتالي فالعينات والاحتمالات ما هما

سوى همزة وصل بين الوظيفة الأولى للإحصاء أي الإحصاء الوصفي والوظيفة الثانية والمتمثلة في الإحصاء الاستدلالي.

مرة أخرى يسمح الإحصاء بتقديم الوصف الدقيق للظواهر المدروسة، وتلخيص النتائج في شكل مفيد ومناسب، واستخلاص الاستنتاجات وتعميمها على مجتمع الدراسة، كما يفيد في التنبؤ ويسمح بتحليل بعض العوامل السببية الكامنة. (عايش، 2016، ص.260) وتكمن أهمية هذا العلم في كونه علمًا محايدًا، فبواسطته يمكن إثبات الكثير من الحقائق، كما يمكن بواسطته أيضا القيام بعمليات تضليل واسعة باسم العلم، لاسيما إذا سُلّط الضوء على موضوع معين بشكل خادع. (احمد حسن عمر، على الموقع: https://www.ahewar.org/debat/show.art.asp?aid=660830). وعمومًا لا يمكن إدارة المجتمع بفاعلية اعتمادا على التخمينات أو مبدأ المحاولة والخطأ بعيدا عن أي منهج ثابت، وهنا يبرز الجهد الإحصائي في شتى مجالات الحياة كأحد المناهج الهامة للتأكد من ضمان استجابة أي مشروع للواقع، مع إمكانية تحديد مدى نجاحه أو إخفاقه. وليس هذا وحسب فأهمية الإحصاء تظهر حتى في حياتنا اليومية؛ مثلاً من التصرفات التي يقوم بها كل إنسان يوميا هو قيامه بالتسوق، فهنا عليه القيام بدراسة السوق بجمع البيانات عن السلعة واستعراض هذه البيانات وإجراء مقارنات عدة للوصول للقرار بالشراء، هذا هو الإحصاء. (عواد، كمرجي، 2020، ص.01)

2. الإحصاء وطبيعة البيانات الإحصائية:

يطلق على علم الإحصاء علم البيانات، ذلك أنّه يتيح لنا استخلاص استنتاجات من مجموعة بيانات تساعد في عملية التنبؤ بالنتائج، "وتعبر البيانات عن الشكل الخام لأي محتوى ننتجه، بل هي المادة الخام للوصول إلى المعلومات، فمثلاً لو كان لديك مائة شخص وقمت بقياس أطوالهم وسجلتها على ورقة، هذه الورقة تحوي بيانات" (Mohamed Hassanenn Badria, 2020, p.29)، وينطبق الأمر تماما كما لو كنت قد سجلت أعمارهم أو علامات التقييم لديهم، أو عن الدخل الفردي لمجموعة معينة.... وعلى نحو أعمق إنّها البيانات الكمية التي تجمل أدبيات الإحصاء. من جهة أخرى يمكنك أن تسجل على ورقتك تلك ما يدلي به المبحوث لسؤال يتضمن احتمالات خمس، مع إعطاء رقم لكل احتمال على أن يكون ذلك بشكل ترتيبي، كأن تسجل مثلاً (1: أمي 2: ابتدائي 3: متوسط 4: ثانوي 5: جامعي) للسؤال التالي: ما هو مستواك التعليمي؟ وفي حالتك هذه أنت تدلي بنوع آخر من البيانات وهو النوع الترتيبي.

ليس شرطا أن تكون بياناتك كمية أو ترتيبية، فأنت حينما تعطي الرقم 1 للجنس الذكر والرقم 2 للجنس أنثى، فإنّك لا تقصد بهذين الرقمين الكم كما أنّك لا تقصد بهما الترتيب. وينطبق الأمر ذاته كما لو أنّك طرحت سؤالا يتضمن اختيارات عدة من قبيل: أي وسائل التواصل الاجتماعي الأكثر استعمالا لديك، 1.الفايسبوك، 2.التويتر، 3.الفايبر، 4.الواتساب؟ فهذه الأرقام الموجودة في الاختيارات هي مجرد رموز فقط وليس لها دلالة عن الكم أو الترتيب، وعليه فإنّ هذا النوع من البيانات يشير إلى النّوع الاسمي.

يمكن القول إذن أنّ البيانات، ما هي في حقيقتها سوى مجموعة من الأرقام أو الحقائق الرقمية التي تزين الإحصاء وتضفي عليه حلّة بشكل فاتن وجذاب، مع أخذنا في الاعتبار أن لا فائدة من هذه البيانات الرقمية في تكوين معرفة ما وهي على صورتها الحالية المجردة، لذا وجب تحويلها إلى معلومات قابلة للاستعمال اعتمادا على مجموعة من الوسائل والأساليب الإحصائية الكمية. ولا يفوتنا في هذا المقام أن نشير إلى أنّ لغة الأرقام التي يحضى بها علم الإحصاء تتمتع بمكانة خاصة لديه، هذه المكانة تؤهلها لأن تتربع على عرش هذا العلم.

تبدو لغة الأرقام قاسية على الطلبة وحتى على بعض الباحثين، لذا ينبغي التعامل معها برفق، إنّنا على حق حين نعتقد أنّ "التعامل مع الأرقام ليس أمرا سهلا، على الرغم من أنّه ليس معضلة لا يمكن حلها، فالحاسبات الصغيرة والكبيرة قد جعلت معالجة الأرقام أمرا هينا، وهذا هو السبب الذي من أجله تعاظمت الأخطاء والأخطار، وكثر من يريد استخدام الطرق الكمية لإضافة ديكورات تزين بحثه وتضفي عليه لباس العلم. ومع كل ذلك ونتيجة لاستخدام الأسلوب الكمي في تحليل المعلومات أصبحت النتائج على درجة عالية من الدقة تصلح أساسا سليما مطمئنا لاتخاذ القرارات. (العزاوي، ص. 16)

يكمن فن التعامل مع البيانات الرقمية في جودتها أو رداءتها، فإذا كانت هاته البيانات حسنة الجودة جاءت على غرارها النتائج الحسنة، أمّا إن كانت البيانات رديئة كانت النتائج كذلك، وعلى الرغم من قبولنا بهذا الرأي واستسلامنا له إلاّ أنّنا "نكون سذجا حينما نعتقد أنّه بالإمكان الحصول على بيانات مثالية، ذلك أنّ البيانات ما هي إلاّ انعكاس للعالم الحقيقي والعالم الحقيقي معقد من حولنا". (ديفيد جيه،DAVID J ، 2016، ص.57). والأهم من ذلك كله وحفاظا على تحقيق نتائج صحيحة ودقيقة، عندما نقوم بإنتاج ملخصات مناسبة من البيانات أي القيام بتنفيذ تحليل

إحصائي، علينا أن نعرف شيئًا عن كيفية الحصول على هذه البيانات الرقمية. هل كانت مكتملة بحيث شملت جميع جوانب الاستبيان أم أن هناك بيانات مفقودة؟ هل تم التأكد من أنّ جميع المستجوبين قد خضعوا للإجابة عن الاستبيان أم لا؟ لأنّ ذلك سيجعلنا محظوظين للغاية في حالة ما إذا حصلنا على ردود من جميع الأشخاص الذين نتواصل معهم. فهناك دوما مقدار من عدم الاستجابة. وهكذا... نستمر في طرح الأسئلة مثلا؛ هل حدث وأن تكررت إجابة بعض الأشخاص عددا من المرات عن الاستبيان؟ هل هناك تحيز في الإجابة بأي حال من الأحوال؟ هل البيانات المتحصل عليها صحيحة أم خاطئة؟ كالذي يحدث في بعض الحالات من قبيل تقريب آحاد أعمار المستجوبين إلى صفر أو خمسة، أو من قبيل وضع البيانات في العمود الخطأ في النموذج.

ثمة إذن عدد لا حصر له من مثل هذه الأسئلة والتي تدور من حولنا عند جمعنا للبيانات، والجدير بالذكر أنّ البيانات غير الصحيحة أو النّاقصة ستؤثر حتمًا على النتائج، وإن كان في حقيقة الأمر "من النّادر أن نقابل مجموعة من البيانات ليس بها مشاكل في الجودة من أي نوع، ربّما إلى حد أنّنا إذا قابلنا مجموعة من مثل هذه البيانات "المثالية" فلا بد أن نشك فيها. ربما يجب علينا وقتها أن نسأل عن عمليات الإعداد التي خضعت لها مجموعة البيانات والتي تجعلها تبدو مثالية. (ديفيد جيه، J DAVID، 2016، ص. 41) وعليه كان لزامًا علينا دائما إجراء فحص دقيق للبيانات المحصل عليها بغية معالجتها قبل التنفيذ. "فتمحيص البيانات الإحصائية والوصول بها إلى درجة عالية من المصداقية وتحقيق الموثوقية هو أمر أهم بكثير من معالجتها رياضيا ونمذجتها على شكل خطط مستقبلية، ذلك لأنّ البيانات غير الدقيقة تؤدي إلى نتائج غير صحيحة". (حيدر، 2007، ص. 188)

هناك العديد من طرق تصحيح البيانات التي يمكن القيام بها قبل الشروع في عملية التحليل، كأن يتم مثلاً إدخال البيانات للحاسوب عبر شخصين مختلفين ومقارنة النتائج، لأنّه غالبا ما يُتحجج بخطأ الحاسوب كأحد التخمينات الساذجة في رداءة البيانات، هذا مع العلم أنّ براءته تكمن في تنفيذ ما يؤمر به وفقط. يمكن أيضا القيام باستكمال البيانات الناقصة -إن وجدت- بقيم بديلة عن طريق متوسط القيم المسجلة، مع العلم أنّه في هذه الحالة نكون قد اختلقنا بيانات جديدة من الأساس، وعليه يفضل دوما التركيز جيدا منذ البداية أثناء "تصميم الاستمارات على نحو يقلل الأخطاء؛ فلا ينبغي أن تكون معقدة على نحو مفرط ويجب أن تكون جميع الأسئلة واضحة. ومن الواضح أنّه من

الأفكار الجيدة إجراء دراسة مسحية تجريبية صغيرة للتعرف على أية مشكلات في عملية جمع البيانات قبل الانتقال إلى التنفيذ الفعلي". (ديفيد جيه، DAVID J ، 2016، ص. 49)

3. التفكير الإحصائي في العلوم الاجتماعية وإشكالية استخراج المعنى من البيانات:

يمتد التفكير الإحصائي لمختلف التفاعلات الاجتماعية المصادفة لنا في حياتنا اليومية، مثلاً نحن نوظف بعض الحسابات البسيطة ومبادئ الإحصاء الأولية في معرفة متوسط الزمن المستغرق للوصول إلى مكان معين، ونتابع الأخبار حول عدد الإصابات بجائحة كوفيد 19 وعدد الوفيات وحالات الشفاء، فضلا عن مراقبتنا لاحتمالات نجاح أو فشل اللقاحات المستخدمة ومتوسط الفترة الزمنية المقدرة لذلك، إنّنا نهتم بمتوسط عدد التعليقات أو مرات الإعجاب لمنشوراتنا على مواقع التواصل الاجتماعي، هذا ونعتمد الإحصاء في تقييم طلبتنا حين نبدي مزيدا من الثقة في التكهن بالمتوسط النهائي لمجموع درجاتهم والتنبؤ بالعوامل المساهمة في نجاحهم أو فشلهم، وهكذا... غالبا ما تراودنا أفكار إحصائية من دون أن نشعر بها على أنّها من أساسيات علم الإحصاء، وتأتي هاته الأفكار تزامنا مع ظروف الحياة اليومية التي نعيشها، وإن دل ذلك على شيء فإنّما يدل على بداهة المبدأ الإحصائي الذي أصبح اليوم يستنجد بأدوات وأساليب إحصائية كانت متسترة في الخلفية في زمن مضى، وبناءً على ذلك نتخذ قراراتنا في الحياة ونحن فخورين بها لأنّنا استخدمنا الأسلوب الكمي الإحصائي، وبمعنى بسيط إنّه التفكير الإحصائي الذي يدور من حولنا.

يمكن تقديم شواهد أخرى تدل على أنّ معظم الأفكار الإحصائية المهمة ما هي في حقيقتها سوى أفكار بديهية، تستمد قوتها من منطقيتها وقدرتها على استيعاب ظروف الحياة اليومية.

أولا/ إنّ المتسول الذي يأتيك صباح كل يوم وأنت تنتظر في محطة القطار، ليطلب منك مالاً مرددا العبارات ذاتها، يراهن في الحقيقة معك على قوانين الاحتمالات، لأنّه يعتقد أنّه كلما زادت عدد فرص المرور من حولك، كلما زاد احتمال تحقق حصوله على بعض النقود، وهو ما يعني أنّ المتسول قد اهتدى لهذا المفهوم الإحصائي بالفطرة.

ثانيا/ إنّك كلما حاولت اختيار طلبة أو لاعبي كرة بناءً على تفوقهم في اختبار واحد فقط من أجل تقديمهم لاجتياز مسابقة وطنية، فإن حدث وأخفق هؤلاء في المسابقة فإنّ ذلك يشير إلى قلة إلمامك بالمعلومات الإحصائية، ذلك أنّ تفوقهم في الاختبار كان بشكل استثنائي وهو نوع من التوفيق

الذي لا يعول عليه، وتسمى هذه الظاهرة في علم الإحصاء "الانحدار إلى المتوسط"، بمعنى أنّ أي وضع استثنائي يعود إلى طبيعته.

ثالثا/ إذا كنت أحد المرشحين للفوز في الانتخابات المحلية مثلا فإنك ستبحث وبلا شك أثناء عملية الفرز عن عدد الأصوات الأكثر تكرار لتحديد إمكانية فوزك، وبعبارة لافتة جدًّا أنت توظف أحد مقاييس النزعة المركزية الهامة، إنه المنوال الذي يقوم على البحث عن القيمة الأكثر شيوعا.

وأخيرا وليس آخرا/ إذا كنت طالبا في الجامعة فإنّك ستهتم بداهة بمقدار انخفاض درجاتك في الاختبار في حالة حصولك على قسط غير كاف من النَّوم، ومن هذا المنطلق إنّك لن تتحسر على شيء، حين حصولك على 3 ساعات نوم فقط مع علامة حسنة، مقارنة بحصولك على علامة قريبة من الحسن مع 9 ساعات نوم كاملة، غير أنّ الحسرة والقلق سيعتريان طريقك حين حيازتك على علامة ضعيفة وقد خسرت معها 6 ساعات من النوم، ففي هذه الحالة من المحتمل أن توافق على أنّ قلة النوم لها عظيم الأثر على درجاتك. وبلغة الإحصاء فأنت تحاول قياس حجم الأثر، هذا الأخير الذي يوضح مدى الاختلاف الواقع بين القيم المتوسطة للمتغيرات.

في الواقع يتمثل التفكير الإحصائي في الطريقة التي يفكر بها الناس في المفاهيم الإحصائية التي تتضمن تفسير مجموعة من البيانات أو تمثيل البيانات الذي يؤدي إلى إصدار حكم وتفسير نتائج (السلمي، 2019، ص.174) ويعرف على أنّه القدرة على رؤية العمليات الإحصائية ككل، بما في ذلك تفسير البيانات الإحصائية، واكتشاف البيانات بطرائق غير متوقعة، وتوليد أسئلة جديدة. وهناك من يعرفه على أنّه "نشاط ذهني مرن، ومنظم يقوم به المعلم أو الطالب بهدف حل التمارين والمشكلات الإحصائية من خلال استخدام بعض مهارات الاستقراء والاستنباط والتفسير وإدراك العلاقات" (الرواحي، 2017، ص. 621)

وفي سياق الاهتمام بالتفكير الإحصائي، نادى بعض الباحثين بوجوب ممارسة الطلبة لأنشطة تحث على التفكير الإحصائي مثل: استخدام العينات لعمل استدلالات حول المجتمعات والتوصل لاستنتاجات حقيقية، واستخدام مقاييس النزعة المركزية والمقارنة بينها والمفاضلة لاستخدام بعضها لتناسب طبيعة البيانات، والتمييز بين المتغيرات المستقلة والتابعة. (الرفاعي، 2015)

قليل من التأمل كفيل بأن يخبرنا بأن تدريس الإحصاء في تخصصات العلوم الاجتماعية لا يزال يقدم بالطرق الكلاسيكية المعهودة، والتي لا تسعى لربط تفكير الطالب بما يجب أن يطبقه في أبحاثه من الناحية الإحصائية، وذلك لاقتصار التعليم الرسمي للإحصاء عادة على المعادلات والقوانين المستخدمة والإجراءات الحسابية من دون النظر إلى ممارسة وتطبيق التفكير الإحصائي في المواقف الحياتية. وبالفعل "ففي مجال الثقافة الإحصائية والتواصل الإحصائي وتنمية مهارات التفكير الإحصائي، فثمَّة مشكلات ظهرت منها: نقص المعرفة الإحصائية وعدم قدرة الطلبة على تطبيق الإحصاء في الحياة اليومية، والقصور في تطبيق مهارات الرياضيات المطلوبة مسبقا والتفكير المجرد" (الرواحي، 2017، ص. 618)، وهذه حقيقة نلمسها في العديد من مذكرات التخرج التي يشقى فيها الأستاذ المشرف لشرح أو ربما للقيام بالعمليات الإحصائية لهذه الأبحاث بنفسه لاستخراج المعنى من البيانات المجمعة، إضافة إلى ذلك فإنّ هذا العلم يُدرَّس بعيدا عن تطبيق برمجيات الحاسوب الإحصائية، بل من النّادر التطرق للاستخدامات العملية. وفي الحقيقة فإنّ معظم كليات هذا التخصص تفتقد لقاعات مجهزة، وحتى إن وجدت فإنّ أغلب الحواسب معطلة -هذا إن كانت تغطي عدد المستخدمين أصلا-، وكل هذه التحديات ساعدت على بقاء تطبيق الأسلوب العلمي الإحصائي واستخراج المعنى من البيانات هاجسا أمام الطلاب. على الرغم من سعي العديد من المدرسين نحو تمكين طلابهم من التفكير الإحصائي والاستقصاء الإحصائي، باعتباره أهم الأهداف في تدريس الإحصاء.

عموما يظل التفكير الإحصائي طريقة مهمة لفهم ما يدور في العالم وفهم العلوم الحديثة، وتقييم المعلومات بصورة ناقدة، وهو أسلوب مهم للتعلم يوضح للطالب أسباب نجاحه وفشله، ويساعد في اتخاذ القرار بناءً على النتائج، بالإضافة إلى ارتباطه بمواقف حياتية مختلفة (هجاد سعد، ص. 1070). هذا وتتضمن عمليات التفكير الإحصائي أربع مهارات يمكن إيجازها في الآتي:

1.3. وصف البيانات:

وتتضمن القدرة على قراءة البيانات المعروضة في قوائم أو جداول أو بيانات ممثلة في رسوم بيانية، ومقارنة تمثيلات مختلفة لنفس البيانات وتقويم درجة توضيح كل تمثيل للمظاهر المهمة للبيانات. (الرواحي، 2017، ص. 621)، والحقيقة التي لا مناص منها أن عملية قراءة التمثيلات البيانية (الجداول، القوائم، الرسوم) تعد أحد الأشكال الأساسية للتفكير الإحصائي؛ حيث تتمثل تلك

العمليات في كل من القراءة المباشرة للبيانات والقراءة بين البيانات والقراءة لما وراء البيانات، وتعد قراءة وبناء الرسوم الإحصائية جزء من الثقافة الإحصائية التي يحتاجها الأفراد خاصة الدارسين لمقررات الإحصاء. (سعود الإسلامية، على الموقع https://mirror.academia-arabia.com/ar/reader/2/81397)

2.3. تنظيم البيانات وتلخيصها:

"تتضمن هذه المهارة ترتيب وتجميع البيانات وفق معايير محددة" (النمراوي، 2020، ص.1404)، "وتنطوي في جملتها على بعض العمليات الذهنية مثل وصف شكل وأهمية خصائص مجموعة من البيانات، ومقارنة مجموعات البيانات المترابطة مع التركيز على كيفية توزيعها باستخدام مقاييس النزعة المركزية أو مقاييس التشتت وتصنيفها واستخدامها". (الرواحي، 2017، ص.622)

3.3..تمثيل البيانات:

وتتضمن بعض المهارات مثل ترتيب البيانات في رتب حتى يمكن تمثيلها بيانيا، واختيار الشكل البياني المناسب، وتمثيل البيانات بالجداول والرسوم مثل خط الانتشار والأعمدة البيانية والخطوط البيانية وكذا المدرجات التكرارية، والمضلعات التكرارية، والقطاعات الدائرية.....

4.3. تحليل البيانات وتفسيرها:

وتتضمن هذه العملية مجموعة من المهارات منها: التوصل لاستنتاجات من خلال الجداول أو الرسوم البيانية، والمقارنة بين مجموعة من البيانات ومقارنة البيانات المعروضة في رسوم بيانية وتقديم الاستدلالات والتنبؤات المبنية على البيانات. (الرواحي، 2017، ص.622)، وهذا هو المغزى الأسمى من علم الإحصاء الأعظم الساعي وراء التنبؤ بالمستقبل وصنع استنتاجات حول المجهول من أجل تبديد الضباب ودحض الارتباك الموجود في العالَم من حولنا، هذا "وتستند صحة نتائج التحليل العلمي على المدخلات (البيانات) وعلى المعالجة (التحليل). فعندما تكون المدخلات قد أختيرت بعناية، وتأكد الباحث من صوابها ودقتها وتناسبها مع طريقة المعالجة وهدف الدراسة، عندها يكون قد ضمن سلامة ثلث الإجابة وترتبط سلامة الثلث الثاني باختياره وسائل المعالجة المناسبة لطبيعة البيانات وهدف الدراسة. (العزاوي، ص. 36)

ولا بد من الإشارة هنا إلى نقطة هامة ساعدت كثيرا على استخراج المعنى من البيانات على اعتبارها الحجر الأساس للحضارة الحديثة، هاته النقطة تتمثل في الانتشار الهائل للبرمجيات الأكثر تقدما والمتخصصة في الأساليب الإحصائية منها على سبيل المثال لا الحصر؛ برنامج excel وبرنامج Statistica وبرنامج SAS وبرنامج Minitab وبرنامج SPSS الأوسع انتشارا واستخداما في العلوم الاجتماعية. ولا تعبر هذه البرمجيات في مجملها سوى عن البديل للمعالجة اليدوية للبيانات، وعليه يمكن القول أن عملية التدريس في المجال الإحصائي باستخدام الحاسوب أو المساعدات التكنولوجية المساعدة على تنمية وتطوير مهارات التفكير الإحصائي إلى أنشطة إحصائية تبقى أحد الأهداف المنشودة والغايات السامية لطلاب هذا العلم.

إننا نبوح برأينا هذا، كوننا على ثقة من أن البرمجيات المذكورة آنفا لا تعالج كافة البيانات المحيطة بعالمنا، والتي هي في تزايد مستمر وبشكل متسارع ومطرد، كتلك البيانات الناتجة عن شبكة الانترنت وعبر شبكات التواصل الاجتماعي، وكذا البيانات المنتجة من مختلف الأجهزة الالكترونية التي من حولنا، ونقصد بكلامنا هذا البيانات الضخمة التي يتمحور حولها مجال علم الإحصاء الحديث.

4. المعالجات الحديثة للبيانات الإحصائية الضخمة:

تعرف البيانات الضخمة بأنّها "مجموعات ضخمة من البيانات ذات قيمة من الصعب تخزينها وتحليلها وإدارتها باستخدام أدوات إدارة البيانات أو تطبيقات معالجة البيانات التقليدية، ويرجع ذلك بسبب مشكلات الحجم والتعقيد بجانب قيود التكلفة وغياب الآليات المناسبة". (مرسي وآخرون، 2023. ص 6). وهي بيانات حلّت إيجابيا على علم الإحصاء ودفعته نحو التطور والنمو، ومنحته فرصة هائلة ليواكب تطور العلوم المختلفة، فمكنته بذلك من أن يشهد عصرا ذهبيا لم يعرفه من قبل.

يشهد سوق البيانات الضخمة في الوقت الحالي نموّا كبيرا بسبب التسجيل التلقائي لهاته البيانات، ففي الحياة اليومية في كل مرة تقوم فيها بالنّقر على زر الإعجاب بمنتج أو منشور معين أو مشاركة الحالة أو فيديو مع الأصدقاء أو الدردشة على الفايسبوك، وكتابة الرسائل النصية أو كتابة التغريدات على تويتر، ... ستخزن تفاصيل عمليتك تلك تلقائيا، ممّا يصعب معالجتها وإدارتها باستخدام أداة واحدة من أدوات إدارة قواعد البيانات أو بالطرق التقليدية لمعالجة البيانات التي

ذكرناها سابقا، لذا يكافح الإحصائيون دوماً ومنتجو البيانات عموما لاكتشاف طرق بديلة عن الطرق التقليدية تمكن من تخزين البيانات وإدارتها واستخدامها لتحقيق أقصى استفادة منها.

ليس من الغريب أن تطور أدوات وحزم برامج إحصائية سهلة الاستعمال تبدد مشاكل التعامل مع الكميات الهائلة من البيانات، ذلك أنّ البداية الكلاسيكية لمعالجة البيانات في الأصل كانت عبر الحاسوب، فنجد من الأدوات والتقنيات الحديثة مثلا " MapReduce, Hadoop, Cassandra, HPCC, GridGain إلاّ أنّ الهادوب Hadoop يعتبر من أشهر هذه الأدوات، وهو باختصار إطار عمل برمجي مفتوح المصدر يدعم التطبيقات الموزعة للبيانات الكثيفة، حيث دعم عملية توسيع نطاق البيانات من خادم واحد ليشمل خمسة آلاف جهاز". (ملياني، سفاحلو ، 2019. ص 65). "بالإضافة إلى ذلك تمّ تقديم أنواع مختلفة من تقنيات الذكاء الاصطناعي، والأساليب القائمة على البحث لتقديم نتائج أسرع وأكثر دقة لعملية تحليل البيانات الكبيرة مثل: التعلم الآلي، وقد أدّى الجمع بين أدوات البيانات الضخمة وتقنيات الذكاء الاصطناعي إلى خلق فرص جديدة في تحليل البيانات الضخمة" (الشوابكة، 2018).

في الواقع إنّنا لا نخشى شيئا من هذا التطور لتقنيات المعالجة الإحصائية الحديثة، بل إنّنا نثمن هذه المجهودات ونحن فخورين بهذا الإنجاز، وندعو للمزيد لاسيما في ظل توسع الذكاء الاصطناعي، إنّنا لا نعتقد أبدا أنّ هذه الأدوات من شأنها أن تزيل الحاجة للإحصائيين، بل على العكس من ذلك تماما، فقط إنّ ما نخافه اليوم هو غياب الوعي والثقافة الاحصائية لمنتجي البيانات.

يقودنا مستقبل التعامل مع البيانات الضخمة إلى عرض الفرص والتحديات التي تواجه العمل الإحصائي لاسيما تلك المتعلقة بمنتجي البيانات والأرقام الإحصائية. فمن أجل تعزيز وتطوير العمل الإحصائي وضمان استجابته بشكل مُستدام وخدمةً للمسار التنموي، لابدّ من العمل على إشاعة الوعي الإحصائي لدى منتجي البيانات، وذلك من خلال "تكوين المهندسيين والتقنيين السامين في الإحصاء طبقا لمراحل التكوين بالمدارس والمعاهد العليا المختصة. وتكوين إطارات سامين في الإحصاء بالمدارس الوطنية والمعاهد المؤهلة" (يعلى، ص.157). وستساهم هذه العملية عموما في تحقيق مستقبل تقوده البيانات، هذا مع الأخذ في الاعتبار أن الوعي الإحصائي لن يقتصر على منتجي البيانات وحسب، بل يمتد ليشمل المستخدمين وحتى سائر أفراد المجتمع، وخلاصة القول لن نكون مغالين أبدا

حينما نعتقد أنه إذا ما أردنا أن نرتقي بالإحصاء ليكون أسلوب حياة وثقافة يومية فإنَّ الوعي الإحصائي هو الطريق لذلك الهدف.

خاتمة:

أصبح علم الإحصاء في عصرنا الحالي أحد أهم العلوم وأكثرها استعمالا وتطبيقا في هذا العالم، بل إنّنا لن نكون مغالين حين نعتقد بأنّ علم الإحصاء هو علم يساعد الأفراد في مختلف مجالات الحياة من أجل العثور على إجابات لأسئلة بحثية مرتبطة بالعلم أو العمل، وهو علم يساعد على التنبؤ بالنتائج ومن ثمَّ اتخاذ القرار. ومن هذا المنطلق خلصت الدراسة إلى نتائج أهمها: أنّ علم الإحصاء يمكن أن يقدّم بطريقة سهلة وجذابة لشرح الظواهر الاجتماعية، ممّا يساعد على ربط تفكير طلبة الجامعات بما يجب أن يطبقونه في أبحاثهم من الناحية الإحصائية. من جهة أخرى توصلت الدراسة إلى أنّ نظريتي الاحتمالات والعينات يمثلان الدّعامتين الأساسيتين لتحقيق الاستقراء والاستدلال، كما أنهما همزة وصل بين الإحصاء الوصفي والإحصاء الاستدلالي. وأشارت الدراسة أيضا إلى أنّ علم الإحصاء يمكن أن تدب فيه الحياة إذا ما اعتمد على استخدام البرمجيات الحديثة والتي تعبر عن البديل لكل ما هو تقليدي في التعامل مع البيانات الضخمة.

اقتراحات وتوصيات:

سعيا منّا لجعل الإحصاء مادة مثيرة ورائعة أثناء تدريسه في الجامعات، وكذا وثيقة الصلة بشتى جوانب الحياة كونه علما غاية في التشويق، خلصت دراستنا إلى الاقتراحات الآتية:

- العمل على توفير كتابات عن علم الإحصاء تتميز بإبراز جانب المتعة والإثارة فيه من أجل أن يستمتع بقراءتها الطلاب، تماما مثلما يستمتعون بقراءاتهم للقصص أو سير الأعلام والنبلاء.
- زيادة الاهتمام بعلم الإحصاء والسعي نحو تجسيده على أرض الواقع، من خلال فتح أقسام وتخصصات أكاديمية لتدريس هذا العلم في قسم العلوم الاجتماعية (تخصص الإحصاء الاجتماعي).
- العمل على تطوير المناهج الدراسية ذات الصلة بالإحصاء، وتضمينها تدريبات متنوعة في الإحصاء لاسيما الإحصاء الاستدلالي. دون إغفال تدريبات الإحصاء الوصفي.

- السعي وراء تنمية مهارات التفكير الإحصائي إلى أنشطة إحصائية، والعمل على خلق طرق بديلة مساعدة على محو الأمية الكمية والتي تزيد من القدرة على القراءة والفهم والمناقشة واتخاذ القرارات من المعلومات، من ذلك مثلا تدريب الطلاب على كيفية استخدام البرمجيات الإحصائية المختلفة.

- التأكيد على الحاجة إلى استخدام علم الإحصاء في مذكرات الطلبة، لأنّ ما ينطبق على هذا العالم الواسع ينطبق على إجراء بحث في هذا العالم الواسع.

- الاهتمام بموضوع البيانات الضخمة، ومحاولة وضع الخطط الرامية للرفع من الوعي والثقافة الإحصائية لدى منتجي البيانات.

- التركيز على أهمية تحليل البيانات الضخمة، باستخدام أدوات التحليل غير التقليدية مع محاولة الاستفادة من تقنية الذكاء الإصطناعي في عملية معالجة هذا النوع من البيانات.

قائمة المصادروالمراجع:

1. أحمد حسن، عمر، خداع الأرقام الإحصائية، الحوار المتمدن-العدد: 6452 - 2020 / 1 / 1 - 12:56) على الموقع: https://www.ahewar.org/debat/show.art.asp?aid=660830؛
2. حيدر، أحمد عباس، (2007). تصحيح التشويه المحتمل في البيانات الإحصائية، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، سوريا، مجلد 23، ع 2، ص. ص. 185-216؛
3. الرفاعي، أحمد محمد رجاء. (2015). مستويات التفكير الإحصائي لدى طلاب كلية العلوم بجامعة الإمام محمد بن سعود الإسلامية، *المجلة الدولية التربوية المتخصصة*، المجلد 04، ع 04، على الموقع https://mirror.academia-arabia.com/ar/reader/2/81397؛
4. الرواحي، منصور بن ياسر، (2017). أثر استخدام الويب كويست في تنمية مهارات التفكير الإحصائي لدى طلبة الصف العشر الأساسي بسلطنة عمان، *مجلة الدراسات النفسية والتربوية*، مجلد 11، ع 3، جامعة السلطان قابوس، ص. ص. 617-644؛
5. السلمي، عزة، (2019). نموذج انحدار خطي متعدد للتنبؤ بالتفكير الإحصائي في ضوء بعض المتغيرات، *مجلة المنارة*، مجلد 26، ع1، ص. ص. 171-195؛
6. الشوابكة، يونس، (مارس، 2018). *الوعي بمفهوم البيانات الضخمة لدى العاملين في المكتبات الأكاديمية: دراسة حالة لمكتبة الجامعة الأردنية*، المؤتمر السنوي الرابع والعشرون لجمعية المكتبات المتخصصة، البيانات الضخمة وآفاق استثمارها: الطريق نحو التكامل المعرفي، فرع الخليج العربي؛
7. عايش، صباح، (2016). أخلاق استعمال الإحصاء في البحوث النفسية والتربوية، *مجلة نقد وتنوير*، ع4، ص. ص. 256-274؛
8. العزاوي، علي، *الأساليب الكمية الإحصائية في الجغرافية*، دار اليازوري العلمية للنشر والتوزيع؛
9. عواد منذر، كمرجي حسام، (2020). *الإحصاء والاحتمالات*، منشورات الجامعة الافتراضية السورية، على الموقع https://pedia.svuonline.org/ (بتاريخ 10/مارس/2022)؛

10. القوصي، محمد مفيد، (2015). *الإحصاء الوصفي والاستدلالي*، ط1، مركز الكتاب الأكاديمي، عمان. حمدي محمد مظلوم، ط2، دار المعارف، مصر؛

11. مرسي، نورهان وآخرون، (2023). النظم والبرمجيات السحابية المستخدمة في معالجة البيانات الضخمة، *المجلة العلمية بكلية الآداب*، ع 50، جامعة طنطا، ص ص 1-28؛

12. ملياني، فتيحة، سفاحلو رشيد، (2019). البيانات الضخمة: الفرص، التحديات، ومجالات التطبيق، *مجلة أبحاث كمية ونوعية في العلوم الاقتصادية والإدارية*، المجلد 01، ع 02، جامعة غرداية، ص ص 61-75. على الموقع https://www.asjp.cerist.dz/en/article/150075 (بتاريخ 28/جويلية/2023)؛

13. النمراوي، زياد محمد، (2020). مستويات التفكير الإحصائي في مقاييس النزعة المركزية، *المجلة التربوية*، ع 8، الأردن. ص. ص. 1400-1428؛

14. هاند، ديفيد جيه، (2016). *علم الإحصاء: مقدمة قصيرة جدا*،ترجمة أحمد شكل، (ط 1)، مؤسسة هنداوي للتعليم والثقافة، القاهرة؛

15. هجاد، سعد مسفر، فاعلية استخدام نموذج أدى وشاير في تدريس الرياضيات على تنمية مهارة تنظيم وتلخيص البيانات لدى طلاب المرحلة الثانوية، ص. ص. 1069-1099؛

16. هيوبرت، م. بلالوك، (2000). *الإحصاء الاجتماعي*، ترجمة عثمان الحسن محمد نور وسليمان محمد رضوان، (ط 2)، مكتبة العبيكان؛

17. يعلى، عادل، اقتراح استراتيجية لتطوير المنظومة الإحصائية في الجزائر بقياسات دولية، *مجلة الحقوق والعلوم الانسانية – دراسات اقتصادية-*، المجلد 20، ع 02، جامعة زيان عاشور بالجلفة، ص ص 147-160. على الموقع: https://www.asjp.cerist.dz/en/downArticle/417/2/3/85640 (بتاريخ 29/جويلية/2023)؛

18. Mohamed Hassanenn, Badria, (2020). Internet of Things and Big Data, *Revolution in Education*, Sys 8, No1, 23-43.