

كيفية الكذب بالإحصاء (*)

How to Lie with Statistics

دليو فضيل*

Delliou Foudil

* كلية علوم الاعلام والاتصال، جامعة قسنطينة 3- صالح بوبنيدر، الجزائر.

fdeliou@yahoo.fr

تاريخ القبول: 2020/09/09

تاريخ الإيداع: 2020/09/06

<p>Abstract :</p> <p>The book of "How to lie with statistics" is considered one of the most famous books on how to lie and claim objectivity by using statistics. Hence the need to summarize its content into Arabic in order to draw attention to the importance of reading it, or at least - in this case - looking at the summary of the statistical tricks in it used to inform the public opinion. Pointing out that the need for his arguments continues today more than ever, as it seems that the increase in the power of computers and "ICT" obsessed with numbers and exciting titles will sink us all in a sea of "statistics", which may mislead us.</p> <p>Keywords: statistical tricks; samples; averages; graphic representations.</p>	<p>ملخص:</p> <p>يعتبر كتاب "كيفية الكذب بالإحصاء" من أشهر ما كُتب حول كيفية الكذب وادعاء الموضوعية باستعمال الإحصاء. ومنه الحاجة إلى أهمية الاطلاع على خلاصة ما ورد فيه من خدع إحصائية تستعمل إعلاميا لتوجيه الرأي العام. مع الإشارة إلى أن استمرار الحاجة إلى حججه الآن أكثر من أي وقت مضى، حيث يبدو أن زيادة قوة الحاسوب وتكنولوجيا الإعلام والاتصال المهووسة بالأرقام وبالعناوين المثيرة ستغرقنا جميعًا في بحر من "الإحصائيات"، التي قد تضللنا. وسنقوم بتلخيصه من خلال عرض أهم محتوياته مدعمة بمقتطفات نصية من كل عنصر من عناصره، متبعين في ذلك خطة الكتاب.</p> <p>الكلمات المفتاحية: الخدع الإحصائية؛ العينات؛ المعدلات، التمثيلات البيانية.</p>
---	--

* المؤلف المرسل.

1. مقدمة

إن المتوسطات والعلاقات والاتجاهات والرسوم البيانية ليست دائماً كما تبدو. فقد توهي بأكثر مما تراه العين أو بأقل بكثير من ذلك. يتم استخدام اللغة السرية للإحصاءات، وتوظيف جاذبيتها بحكم انتمائها لثقافة تدعمها الحقائق، لإحداث الإثارة والتشويه والمغالطة والتبسيط المبالغ فيه. إن الأساليب والمصطلحات الإحصائية ضرورية للإبلاغ عن البيانات الضخمة حول الاتجاهات الاجتماعية والاقتصادية، ووضع الأعمال التجارية، واستطلاعات الرأي، والتعدادات؛ ولكن من دون كتاب يستخدمون الكلمات بأمانة ودقة ومن دون قراء يعرفون مقاصدها، فإن النتيجة ليست سوى مجرد معان غير دالة.

بقليل من "البودرة" وقليل من الطلاء، تجعل الإحصائيات الكثير من الأشياء تبدو حقائق مهمة، بينما هي ليست كذلك.

يعرض هذا الكتاب بعض الأبجديات لكيفية الغش باستخدام الإحصاءات. وقد يبدو، على العموم، كدليل لعديمي الضمير. إن هؤلاء يعرفون بالفعل هذه الحيل، لكن يجب على الشرفاء تعلمها بغية الدفاع عن أنفسهم.

2. العينة التي تحمل عامل التحيز بداخلها¹

تعتبر تقنية أخذ العينات العمود الفقري لمعظم الإحصائيات المنشورة. إن أساس هذه التقنية بسيط، على الرغم من أن خفاياها أدت عملياً إلى جميع أنواع المسارات الغامضة، وبعضها بعيد كل البعد عن الاحترام. فإذا كانت العينة كبيرة بما يكفي وتم اختيارها جيداً، فستمثل المجموع عن كثب في معظم الحالات. أما إذا لم يكن الأمر كذلك، فقد تكون أقل دقة بل ومتحيزة تماماً. إنها حقيقة محزنة بالفعل الاعتراف بأن ما نقرأه أو نعتقد أننا نعرفه هي استنتاجات مستمدة من عينات ذاتية أو صغيرة جداً، أو كليهما في الوقت نفسه.

فعندما تقرأ عدة مرات أن الأمريكي العادي ينظف أسنانه 1.02 مرة في اليوم (وهو رقم افتراضي)، اسأل نفسك، كيف يمكن التأكد منه؟، أو عندما تقرأ في بعض الصحف الأمريكية

أن حوالي 4 ملايين كاثوليكيا أمريكيا اعتنق البروتستانتية في السنوات العشر الأخيرة، حسب نتائج استطلاع رأي عينة من البروتستانت (25000 مفردة)، ردّ منهم على الاستمارة 2219 رجل دين أقرّوا باستقبال 51361 كاثوليكيا سابقا في كنيستهم خلال السنوات العشر الماضية. وبناء عليه تم تّوقّع تحوّل أكثر من 4 ملايين كاثوليكي إلى البروتستانتية على المستوى الوطني. وهو أمر مستحيل لكون عددهم الإجمالي أصلا -حسب بعض الأساقفة- يتراوح بين 2 و3 ملايين كاثوليكي يتردد على الكنائس وقتئذ وقد يقارب الخمس ملايين عند بعض المتقائلين منهم... كما أنه أمر غير صحيح لأن أكثر من 91% من المستطلعين لم يردوا. ولتدمير هذا الاستطلاع كليا، يكفي فقط لفت الانتباه إلى الاحتمال المنطقي بأن أغلب من لم يرد (من 91%)، فعل ذلك بسبب عدم وجود متحول إلى البروتستانتية في كنيستته (يحوّل هذا الاحتمال الـ4 ملايين إلى حوالي 370 ألف)، وأن القائم بالاستطلاع بروتستانتني.

فعندما يتم تصفية البيانات من خلال مراحل مختلفة من المعالجة الإحصائية وتقليصها إلى متوسط معبر عنه بالأرقام العشرية، تبدأ النتيجة في حيازة هالة من القناعة التي سيتم تشويهها فقط من خلال مراجعة دقيقة للعينة. فلكي يكون لأي تقرير مستند إلى عينة قيمة واقعية يجب أن يستخدم عينة تمثيلية، حيث يتم إزالة جميع العوامل المؤثرة المحتملة. يجب القيام بمراجعة ما نقرأه مرتين على الأقل، وسوف نتجنب تصديق عدد كبير من الأشياء غير الصحيحة. من الجدير بالذكر أن نضع في اعتبارنا أنه يمكن الطعن في مدى تمثيلية العينة بسهولة أكبر، من خلال الأخذ بعين الاعتبار مدى تأثير العوامل المرئية وغير المرئية. أي أنه حتى في حالة عدم وجود عامل مؤثر ملموس، يمكن الاحتفاظ بدرجة ما من الشك في النتائج.

لفحص عينة عشوائية، نتبع المعيار الآتي: هل لدى كل اسم في مجتمع الدراسة نفس الإمكانية ليكون جزءًا من العينة؟ حينها تتحول استطلاعات الرأي إلى معركة ضد التأثيرات المختلفة، وتُشنّ هذه المعركة باستمرار من قبل جميع منظمات الاستطلاع ذات السمعة الطيبة. يجب أن يتذكر قارئ التقرير أن هذه المعركة لم يتم كسبها أبدًا. لا ينبغي قراءة أي

استنتاج مفاده أن "65% من السكان يعارضون شيئاً ما أو شيئاً آخر، دون طرح السؤال، 65% من أي مجموعة من السكان؟". ربما يكون من الأهم أن نتذكر أن أي استبيان ليس سوى عينة أسئلة محتملة، وأن إجابة سيدة ما ليست سوى عينة من مواقفها وخبرتها في كل سؤال. افترض أنك متخصص في إجراء المقابلات وتم تعيينك في زاوية شارع ما لهذا الغرض، فواجهك رجلان يبدو أنهما يتناسبان مع فئة مجموعة تم تحديد خصائصها مسبقاً كما يلي: أكثر من أربعين سنة، من السود، ومن سكان المدينة. يرتدي أحدهما بذلة نظيفة وأنيقة. والآخر قذر ويبدو غير ودي. تتوي إنهاء وظيفتك قريباً، فتقترب من الشخص الذي يبدو لطيفاً، وهو قرار سيتخذه جميع زملائك في جميع أنحاء البلاد.

3. معدل اختيار جيد²

تعتبر كلمة "المعدل" (Average) الميزة الأساسية والصيغة الجمالية للكذب بالإحصاءات. إنها خدعة تُستخدم في كثير من الأحيان: في بعض الأحيان ببراءة، وأحياناً أخرى بقصد مذنب، من قبل الأفراد الذين يريدون التأثير على الرأي العام أو بيع المساحات الإعلانية. عندما يتم إخبارك بأن شيئاً ما عبارة عن معدل، فأنت لا تعرف الكثير عنه حتى الآن، ما لم تتمكن من معرفة ما هو المعدل: المتوسط (مجموع القيم على عددها) أم الوسيط (القيمة التي تتوسط سلسلة القيم مرتبة) أم المنوال (القيمة الأكثر تكراراً) [وثلاثتها مقاييس للنزعة المركزية تدرس للطلبة]. وبالطبع عادة ما تكون أرقامها مختلفة جداً، وسيختار ناشرها الرقم الذي يدعم حجته أكثر ويعتبره معدلاً.

لهذا السبب، عندما تقرأ بياناً من مدير شركة، قائلاً أن معدل راتب الأشخاص الذين يعملون في شركته كبير جداً، فقد يعني الرقم شيئاً وقد لا يعني أي شيء. دعنا نوضح ذلك: أنت أحد الشركاء الثلاثة الذين يمتلكون مصنعاً صغيراً. نحن في نهاية عام جيد جداً. لقد دفعت 198000 دولار إلى التسعين موظفاً الذين يهتمون بصنع وتوزيع الكراسي أو أي شيء آخر. لقد دفعت أنت وشركاؤك لأنفسكم راتباً قدره 11000 دولار لكل منكم. كما اتضح أن

هناك فوائد بقيمة 45 ألف دولار، يتم توزيعها بينكم بمقادير متساوية. كيف ستمثلون ذلك؟ لتسهيل الفهم، ستعبرون عنه في شكل معدلات. نظرًا لأن جميع الموظفين يقومون بنفس النوع من العمل ويكسبون أجورًا متساوية، فلن يكون هناك فرق كبير في استخدام المتوسط أو الوسيط. والنتيجة هي كما يلي: متوسط راتب الموظفين، 2200؛ متوسط راتب وريح أصحاب المصنع 26.000. يبدو الأمر رهيبًا أليس كذلك؟ لنجرب طريقة أخرى. خذ 30.000 دولار من رقم الربح وقم بتوزيعه بين الشركاء الثلاثة، وهذه المرة عند استعمال متوسط الأجر اجعله يشملك أنت وشركائك. وتأكد من استخدام المتوسط. متوسط الراتب أو الأجر: 2,806.45. متوسط ربح أصحاب المصنع: 5000. آه! يبدو هذا أفضل بالفعل. يمكن تحقيق المزيد، ولكن هذا ليس سيئًا.

إن الذين يعملون في مؤسسات سير الآراء لديهم التدريب الكافي والأموال اللازمة لتحقيق درجة معتبرة من الدقة في استطلاعاتهم. لكن ليست كل الأرقام التي تراها قد ولدت في ظل هذه الظروف السليمة، ولا يتم تقديمها جميعًا بمعلومات كافية لفحص دقتها أو عدم دقتها.

4. الأرقام الصغيرة التي لا تظهر³

كما لوحظ سابقًا، يمكن استخدام عينات متحيزة بشكل جيد لإنتاج أي نتيجة قد يرغب فيها أي شخص. تقول بعض عناوين وسائل الإعلام أن تواتر تسوس الأسنان ينخفض بنسبة 23% عند مستخدمي إحدى ماركات معجون الأسنان، والذين يمكنهم العيش بنسبة 23% أقل من وجع الأسنان طيلة حياتهم. هذه النتائج، يؤكدها مختبر "مستقل" ومرموق، وترافقها شهادة صادرة عن خبير معتمد. ماذا تريد أكثر؟ على الرغم من ذلك، إذا لم تكن ساذجًا أو متفائلًا بشكل مفرط، فستتذكر من التجربة أن أي معجون أسنان نادرًا ما يكون أفضل بكثير من أي معجون أسنان آخر [في الدول المتقدمة]. فكيف تسمح لنفسها هذه الماركة بالإبلاغ عن مثل هذه النتائج؟ كيف يمكنها أن تقول أكاذيب بهذا الحجم؟، فهي ليست بحاجة لذلك. هناك طرق أسهل وأكثر فعالية لتحقيق هدفها. دعونا نجعل أي مجموعة صغيرة من الناس تراقب

تسوس أسنانها لمدة ستة أشهر، ثم نخضعها لمعجون أسنان الماركة المعنية. يمكن أن تحدث ثلاثة أشياء: أن لديها عددًا أكبر من التسوس، أو أن لديها عددًا أقل، أو أن لديها نفس العدد تقريبًا. إذا حدثت الحالة الأولى أو الأخيرة، فإن شركة معجون الأسنان المعنية تورشف الأرقام (بعيدا عن المنال) وتعيد الاختبار. عاجلاً أم آجلاً، وعن طريق الصدفة، ستقدم مجموعة تجريبية تحسناً كبيراً يستحق عنواناً وربما حملة إعلانية كاملة. سيحدث هذا سواء كانوا يستخدمون معجون أسنان الشركة المعنية أو إذا كانوا يستخدمون معجون صوديوم عاد أو نفس معجون الأسنان الذي كانوا يستخدمونه من قبل.

وكما هو معروف، فالعينات الأصغر تنتج تبايناً أكبر. فمع 10 رميات قرعة بالعملة (تُرمى في الهوى)، قد تحصل على الصورة أو الكتابة 8 مرات، ولكن من غير المحتمل أن تحصل على أحدهما 80 مرة في 100 رمية. لذا يجب ذكر العدد المرجعي (مجتمع الدراسة): 10 أم 100؟ لحسن تقدير دلالة القيمة العددية.

كيف يمكن تجنب الانخداع بنتائج غير حاسمة؟ هل ينبغي على الجميع أن يصبحوا متخصصين في الإحصائيات وأن يدرسوا جميع البيانات الخام بأنفسهم؟ لا ليس إلى هذا الحد كما يبدو، هناك اختبار دلالي يسهل فهمه. إنها ببساطة طريقة للإبلاغ عن احتمالية أن يمثل رقم الاختبار نتيجة حقيقية وليس نتيجة تم إنتاجها عن طريق الصدفة. إنه الرقم الصغير الذي لا يظهر، على افتراض أي قارئ عادي، لن يفهمه، أو سيفهمه، في الحالات التي يكون فيها من الضروري إخفاء شيء ما.

كما أنه عادة ما يتم حذف رقم صغير آخر، يمكن أن يكون غيابه ضاراً. إنه الذي يشير إلى الفئة التي تنتمي إليها الحالات أو الانحراف الذي تمثله فيما يتعلق بالمعدل. حيث غالباً ما يكون المعدل (سواء كان متوسطاً أو وسيطاً، مع أو من دون مواصفات) تبسيطاً فادحاً لدرجة أنه يكون عديم الفائدة تماماً. فغالباً ما يكون عدم معرفة أي شيء عن شيء ما أكثر صحة من المعرفة غير الدقيقة، لأن معرفة القليل عن شيء ما يمكن أن يكون خطيراً للغاية.

الخطأ هو في عملية تصفية المعلومات، التي تبدأ بالباحث، مروراً بالكاتب المثير أو المضلل، حتى تصل إلى القارئ، الذي لا يستطيع العثور على أرقام مفقودة أثناء العملية. في حين يمكن تجنب الكثير من التفسير الخاطئ بإضافة بيانات الانحراف إلى "قاعدة البيانات" أو المعدل.

ما هو مضلل حول العدد الصغير الذي لا يظهر هو أن غيابه غالباً ما يمر دون أن يلاحظه أحد. هذا، بالطبع، هو سر النجاح. ضع القليل من الثقة في المتوسط أو الرسم البياني وإلا فستكون بنفس درجة عمى الشخص الذي يختار منتجاً صيفياً بناءً على متوسط درجة الحرارة.

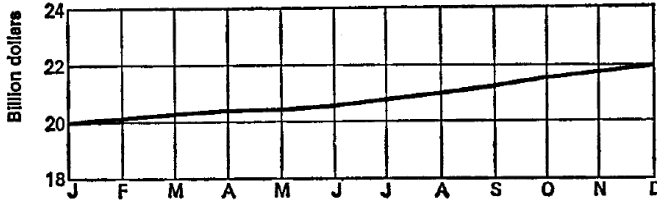
5. تمخض الجبل فولد قأراً⁴

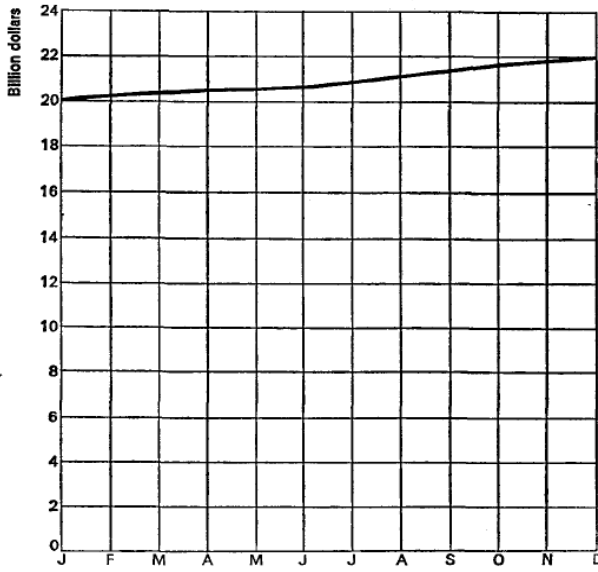
عادة ما يُعتقد في الكميات الملموسة اعتقاداً أعمى إلى درجة عدم الانتباه إلى كون بعضها تقريبياً، وبالتالي، يجب أن يكون الفرق بين كميتين كبيراً بما يكفي لاستنتاج أن هناك اختلافاً حقيقياً.

هناك محررون في المجالات يقدسون استطلاعات الرأي، وهم لا يفهمونها. فمثلاً، عندما تشير نتائجها مثلاً إلى أن 40% من القراء مهتمون بأحد المقالات و35% منهم مهتمون بمقال آخر، فإنهم يطلبون المزيد من المقالات المشابهة للمقال الأول. بينما قد يكون الفرق بين 35 و45% من القراء كبيراً بالنسبة للمجلة، ولكن قد لا يكون هذا الفرق في الاستطلاع حقيقياً أو دالاً. فغالباً ما تفرض تكاليف الاستطلاعات جعل عينات القراء تصل إلى بضع مئات من الأشخاص فقط (يستثنى منهم عدد الذين لا يقرؤون)، كما أن قراءة المجالات تستهوي عادة النساء أكثر من الرجال، فإذا تم توزيع ما تبقى على الخيارات الجوابية: يقرأ كل المقالات، معظمها، بعضها، ولا يقرأ المقال المعني بالاستطلاع، فإن نتيجة الـ35% قد يكون أساسها حفنة من القراء فقط. تقول حكمة قديمة: يعتبر الفرق فرقاً فقط إذا أحدث فرقاً.

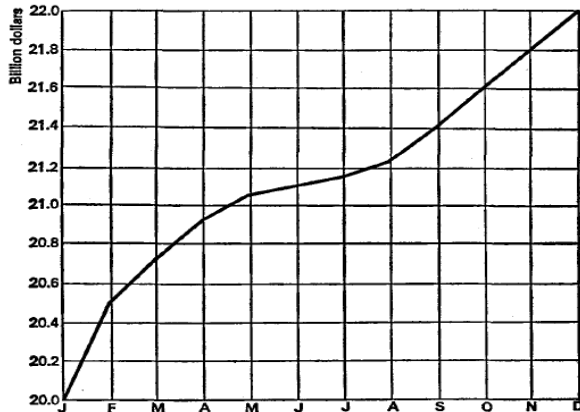
6. الرسم البياني المخادع⁵

هناك خوف رهيب من الأرقام. ربما نعاني من صدمة مبكرة ناجمة عن دراسة مادة الرياضيات في المدرسة الثانوية. ومهما كان السبب في ذلك، فإن الأرقام تستعمل لاستمالة القراء وفي الإشهار عن المنتجات، وعندما تكون الأرقام في شكل جدول غير مقبولة والكلمات لا تفي بالغرض، كما يحدث في كثير من الأحيان، هناك حل واحد فقط: رسم بياني إحصائي. ومن أبسط أنواعه المنحنى البياني والذي قد يرسم ليُظهر اتجاهًا طفيفًا على أنه مهم جدًا، وذلك من خلال مثال: قص الجزء السفلي من الرسم البياني وتمديد المحور الرأسي للمبالغة في الفرق. فالحالة الأولى يشخصها الرسم البياني الموالي الخاص بتمثيل زيادة الدخل الوطني المقدرة بـ 10% في السنة (من 20 إلى 22 مليار دولار ومن جانفي إلى ديسمبر)، والذي أبان الزيادة بنسبتها الطبيعية (انطلاقًا من خط الصفر في الأدنى بغية المقارنة). هذا هو التمثيل البصري الحقيقي لهذه الزيادة، لكن الذي يريد تضخيمها يقوم بقص الجزء السفلي من الرسم (أنظر الرسم البياني الثاني). لم يتم تغيير الأرقام ولا الأشهر والمنحنى هو نفسه، فلا شيء تم تزويره - باستثناء الانطباع الذي يتركه الرسم (الثاني) لدى مشاهده: الدخل الوطني زاد مقدار نصف ورقة الرسم الأخير خلال 12 شهرًا... بسبب قص ثلثي الورقة الأصلية.





والآن بعد أن مارست الخداع، لماذا تقتصر على هذا التشويه؟ لا تزال هناك خدع تستحق العشرات مثل هذا... مع ذلك، يمكنك أن تجعل الزيادة المتواضعة بنسبة 10% تبدو أكثر إبهارًا بكثير من الزيادة الحقيقية بنسبة 100%. وببساطة يمكن تعديل النسبة بين إحداثيات السينات الرقمية والصادات الحرفية. فليس هناك قاعدة تمنعه وهو يجعل التمثيل البياني يبدو أفضل وأفيد لاتجاه صاحبه.



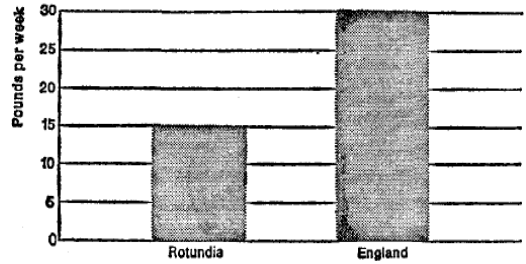
إن الرسوم البيانية تيسر الموضوعية الوهمية والتي يمكن استغلالها بسهولة من خلال الخداع البصري.

7. الصورة أحادية البعد⁶

يعتبر المخطط البياني بالأعمدة (أحادي البعد) أبو جميع الرسوم البيانية الوصفية، وهو طريقة بسيطة ومعروفة لتمثيل البيانات الكمية عندما يتعين مقارنة اثنين أو أكثر منها. يمكن أن يكون الرسم البياني بالأعمدة خدعة أيضاً. انظر بشكل مريب إلى أي تمثيل بياني تتغير أعمدته في العرض أو الطول، رغم تمثيلها نفس المتغير، وكذلك أي رسم بياني يمثل كائنات ثلاثية الأبعاد لا يمكن مقارنة أحجامها بسهولة.

إنه يتم التلاعب بالتمثيلات البيانية بشكل متكرر. لقول الحقيقة، أريدك أن تستنتج شيئاً ويتكون لديك انطباع مبالغ فيه، بغية لفت الانتباه...

بدلاً من الأعمدة، ضع تمثيلات ثلاثية الأبعاد (مثل أكياس النقود)، ولكن ضع في اعتبارك الارتفاع والعرض لتمثيل المتغير. مما يترك انطباعاً بصرياً مبالغاً فيه، وذلك بخلاف لم أخذنا بعين الاعتبار الارتفاع فقط. والصورتان الموائيتان لأجرة عاملين تعكسان الفرق البصري بين التمثيلين المقارنين (يبدو الفرق أقل في التمثيل بالأعمدة):



قد تكون بعض هذه التمثيلات المغرضة ناتجة عن عيوب في الرسم، ولكن هناك احتمال وجود سوء نية مبيتة.

8. الرقم المرتبط بشكل غير مباشر⁷

إذا لم نستطع إثبات ما نريد، نثبت شيئاً آخر ونجعله يبدو كما الذي نريده. في الوهج الذي يلي صراع الإحصائيات مع الدماغ البشري، لن يلاحظ أحد الفرق تقريباً. الرقم المرتبط بشكل غير مباشر هو خدعة مضمونة ستكون مفيدة. لقد كانت دائماً كذلك.

فمثلاً، قد لا يمكنك أن تثبت أن تحضيراتك الصيدلانية تشفي من نزلات البرد، ولكن يمكنك نشر تقرير (بأحرف كبيرة)، مع ضمان مختبر، يؤكد أن 30 غرام من هذا المنتج تقتل أزيد من 31000 جرثومة في أنبوب اختبار في أحد عشر ثانية. إذا قمت بذلك، فتأكد من أن يكون المختبر ذا شهرة معروفة أو اسم مثير للإعجاب. قم بتصوير طبيب في معطف أبيض وضع صورته على هامش التقرير. ولكن لا تذكر الحيل المختلفة التي استخدمتها. ليس من شأنك أن تذكر أن المطهر الذي يعمل بشكل جيد في أنبوب الاختبار قد لا يكون له أي تأثير على الإنسان بعد أن يتم تخفيفه بشكل مناسب لمنع حرق أنسجة الحلق. لا تعرّض النجاح للفشل بالقول ما هي أنواع الجراثيم التي دمرتها، من يدري ما هي الجراثيم التي تسبب نزلات البرد، خاصة عندما يكون من المحتمل أنها ليست أي جراثيم؟ باختصار، لا توجد علاقة معروفة بين الجراثيم في أنبوب الاختبار وسبب نزلات البرد، لكن الناس لن يدققوا بشكل جيد عندما يحتاجون إلى معالجة عطسهم.

أو خذ هذا المثال: 27% من عينة كبيرة من الأطباء البارزين يدخنون إحدى العلامات التجارية أكثر من أي علامة تجارية أخرى. قد تكون دلالة الرقم مزيفة لأسباب عديدة، ولكن هذا ليس له أهمية. الاحتجاج الوحيد على الرقم الذي قد لا يبدو مهماً للغاية هو: وماذا بعد؟ مع كل الاحترام الواجب للفئة الطبية، هل يعرف الأطباء عن ماركات السجائر أكثر مما تعرفه؟ هل لديهم أي معلومات خاصة تسمح لهم باختيار أقل سيجارة ضارة بين الجميع؟ بالطبع ليس لديهم ذلك، وسيكون طبيبك أول من يخبرك بذلك. على الرغم من هذا، فإن نسبة 27% هذه تُعرّض بطريقة أو بأخرى لتبدو وكأنها تعني شيئاً (كلنا لا نعرف أي شيء عن نسبة الـ 73% المتبقية. تخيل للحظة أن 73% المتبقية من الأطباء الذين يدخنون

يلجأون إلى علامة تجارية أخرى غير ماركة سجاير الـ27% المختارة، لأنهم يعتبرون هذه الأخيرة ضارة بالصحة بشكل خاص).

بنفس نوع الهراء الذي ورد في المثال السابق، يمكنك إظهار أن الطقس الصافي أكثر خطورة من الضبابي. فعدد الحوادث يكون أكبر عندما يكون الطقس صافياً، لأن بكل بساطة- الأيام الصافية في السنة أكثر عدداً من الأيام الضبابية. ومع ذلك، وبشكل عام، يمكن أن تكون القيادة في الضباب أكثر خطورة بالفعل.

يمكنك أيضاً تمثيل البيانات والأرقام نفسها بطرق مختلفة: فعلى سبيل المثال، يمكن التعبير عن نفس الحقيقة تماماً من خلال تسميتها عائداً (ربحاً) بنسبة 1% من المبيعات، وعائداً بنسبة 15% من الاستثمار، وربحاً بقيمة عشرة ملايين دولار، وزيادة في الأرباح بنسبة 40% (مقارنةً بمتوسط فترة زمنية مثلاً)، أو بانخفاض قدره 60% عن العام الماضي.

ومثال آخر: إذا اشترت سلعة كل صباح مقابل 99 سنتاً وقمت ببيعها كل مساءً مقابل دولار واحد، فسوف أحصل على 1% فقط من إجمالي المبيعات، ولكنني أحصل على 365% من الأموال المستثمرة بعد عام.

كما قد يستعمل ذلك في بعض المقارنات غير المناسبة: كانت نسبة الوفيات في البحرية خلال الحرب بين الولايات المتحدة وإسبانيا 9%. وكانت نسبة وفيات عدد السكان المدنيين في نيويورك خلال نفس الفترة 16%. استخدمت السلطات المجندة للمتطوعين في البحرية هذين الرقمين للإبلاغ عن أن التواجد في القوات البحرية كان أكثر أماناً من التواجد على اليابسة. دعونا نعترف بأن الأرقام صحيحة. وبالتأكيد هي كذلك. فكّر للحظة وانظر ما إذا كان بإمكانك معرفة سبب عدم منطقيتها، على الأقل فيما يتعلق بالاستنتاج الذي توصلت إليه السلطات المجندة. المجموعتان غير قابلتين للمقارنة.

9. العلائقية مقابل السببية⁸

عندما يكون هناك العديد من التفسيرات المعقولة للحقيقة نفسها، لا شيء يسمح لك باختيار التفسير الذي يناسب ذوقك، والإصرار عليه، ولكن الكثير من الناس يفعلون ذلك. وبالتالي، قد يرتبط متغيران، "أ" و"ب"، لأن "أ" يسبب "ب"، لأن "ب" يسبب "أ"، لأن هناك سلسلة من الأسباب والآثار بين الاثنين أو لأن كليهما ناتج عن متغير آخر... وبالطبع ليس لأن "أ" ذكر قبل "ب"... ومنه الافتراض غير المبرر الذي مفاده: بما أن التدخين ودرجات الحرارة المنخفضة يلتقيان معًا، فإن التدخين يسبب درجات منخفضة. ألا يمكن أن يكون العكس هو الصحيح؟.

أو المغالطة اللاحقة والتي مفادها أن بعض الأرقام تظهر أن ذهابك إلى الجامعة، من المحتمل جدا أن يكسبك أموالاً أكثر مما لو قررت قضاء السنوات الثلاث المقبلة بطريقة أخرى. إن هذا الاستنتاج غير المبرر أساسه الافتراض غير المبرر أيضا أنه منذ أن حصل بعض الأشخاص الذين دخلوا الجامعة على المزيد من المال، فإنهم يكسبونه لأنهم دخلوها. في الواقع، نحن لا نعرف ولكن هؤلاء هم الأشخاص الذين كانوا سيحصلون على المزيد من المال حتى لو لم يذهبوا إلى الكلية، بحكم خصائصهم الذاتية مثلا.

وربما تكون أكثر الحالات المضللة هي الحالة الشائعة جدًا التي لا يؤثر فيها أي متغير على الآخر، ولكن هناك مع ذلك ارتباط حقيقي. باستخدام هذه الوسيلة، تم القيام بالكثير من الأعمال القذرة. كمثال على علاقة سخيفة أو مشوهة حول حقيقة إحصائية فعلية، أشار شخص مضحك إلى وجود علاقة وثيقة بين رواتب القائمين على الكنيسة المشيخية في "ماساتشوستس" وسعر نبيذ الروم في "هافانا"...

شيء آخر يجب الانتباه إليه هو الاستنتاج الذي مفاده أن الارتباط يتخطى البيانات المستخدمة لإثباته. من السهل الإثبات بأنه كلما أمطرت الأمطار في منطقة ما، زاد نمو القمح وزاد الحصاد. لكن موسم الأمطار الغزيرة يمكن أن يضر القمح بل ويدمره.

من جهته، ابتكر البروفيسور "هيلين م. ووكر" (Helen M. Walker) قصة ممتعة لاختبار السخافة التي ينطوي عليها الادعاء بأنه يجب أن يكون هناك سبب وتأثير عندما يختلف

شيثان في نفس الوقت. عند دراسة العلاقة بين العمر وبعض خصائص النساء، ابدأ بقياس الزاوية التي شكلتها أقدامهن عند المشي. ستجد أن الزاوية تميل إلى أن تكون أكبر بين النساء الأكبر سنًا. قد تعتقد في البداية أن النساء يتقدمن في السن لأنهن يباعدن بين أقدامهن عند المشي.

10. كيفية التلاعب إحصائياً⁽⁹⁾

يمكن أن يسمى تقديم معلومات خاطئة، باستخدام مواد إحصائية، تلاعباً إحصائياً (وفي كلمة واحدة، حتى لو لم تكن جيدة جداً: "إحصائياً") (Statisticulating). فمهما كان المذنب في قضية معينة، يصبح من الصعب إلقاء اللوم على عجزه البريء. طالما أن الأخطاء كلها في جانب واحد (المستفيد عادة)، فليس من السهل نسبها إلى نقص المعرفة أو الصدفة. لتحقيق جو من الدقة العلمية التي ستعطي الاتساق للإحصائيات الأكثر احتياطاً، استخدم الكسور العشرية. فالنسب المئوية أرض خصبة للتضليل، مثلها مثل الكسور العشرية التي تثير الإعجاب دائماً، حيث يمكن أن توفر هالة من الدقة لأمر غير دقيق. من المرجح أن تكون أي نسبة مئوية، تستند إلى عدد صغير من الحالات، مضللة. من المفيد أكثر إعطاء الرقم نفسه. وعندما يتم التعبير عن النسبة المئوية بالكسور العشرية، توقع التعامل مع مقياس يتراوح بين العبثي والاحتمالي.

عند حساب النسبة المئوية للريح، يمكنك الاختيار من بين عدة طرق: تقول إحدى الإعلانات المناسبة: "اشترى هدية العيد الآن ووفر 100% من مصروفك". يوحي هذا الإعلان عن عرض بالمجان، لكنه عبارة عن مغالطة مرجعية. فنسبة التخفيض 50% فقط. ونسبة التوفير 1% من السعر الجديد أو المخفض. هذه هي الحقيقة وليس ما يقوله الإعلان.

ينتج الكثير من الضجيج والخداع من إضافة أشياء لا يمكن تلخيصها ولكنها تبدو قابلة لذلك. على مدى أجيال، يستخدم الأطفال حيلة لإثبات عدم ذهابهم إلى المدرسة. ربما تتذكرها. انطلاقاً من 365 يوماً في السنة، يمكنك طرح 122 يوماً للثلاث من الوقت الذي

تقضيه في السرير و45 يوماً أخرى للساعات الثلاث التي تقضيها يومياً في تناول الطعام. ومن الـ198 يوماً المتبقية، أطرح 90 يوماً للعطلة الصيفية و21 لعطلة الشتاء والربيع. الأيام المتبقية لا تغطي حتى أيام عطلة نهاية الأسبوع: الجُمعات والأسبات.

كما تظهر الكذبة القديمة والبديهية المتمثلة في الإضافة اللامتناهية للنسب المئوية في عالم الأعمال أيضاً: ففي كل مرة ينظم فيها إضراب، تعلن غرفة التجارة أن الإضراب يكلف ملايين الدولارات في اليوم. وهي تحصل على الرقم بإضافة جميع السيارات التي كان من الممكن أن تصنع لو عمل المضربون طوال الوقت. ثم تضيف كل خسائر الموردين حينها. مع إضافة جميع الخسائر المحتملة، بما في ذلك رسوم وقوف السيارات وخسائر البائعين... والأرضية الخصبة الأخرى للخداع هي الخلط بين النسب المئوية والنقاط المئوية. إذا بلغت أرباحك 3% من الاستثمار في عام واحد و6% في العام التالي، فستجعلها تبدو متواضعة جداً من خلال وصفها بزيادة قدرها ثلاث نقاط مئوية. بنفس الصلاحية، كان يمكن أن تصفها بأنها زيادة مائة بالمائة. تلعب استطلاعات الرأي العام بحرية مع هذين النظامين.

وهو خلط شبيه بنتيجة تفسير بائع سندويشات دجاج بسعر بخس، والذي فسر ذلك بإضافة لحم الحمير: 50% - 50%: حمار واحد مقابل دجاجة واحدة... (وليس 50% من نوع اللحمين في كل سندويش).

كما تعتبر الأرقام القياسية (الأسعار المرجعية) مهمة بالنسبة لملايين الأشخاص لمعرفة مستوى دخلهم ومعيشتهم. وهو أمر عرضة للتلاعب أيضاً. فمثلاً، إذا انخفض سعر الحليب 10 سنوات وارتفع سعر الخبز 10 سنوات. فما الذي نريد أن نتأكد منه؟ أن تكلفة المعيشة زادت؟ أن تكلفة المعيشة قد انخفضت؟ أم أنه لم يكن هناك تغيير؟ فإذا اعتبرنا العام الماضي الفترة المرجعية للمقارنة، مما يجعل الأسعار حينها تمثل 100%. وبما أن سعر الحليب انخفض إلى النصف (50%) وتضاعف سعر الخبز (200%) ومتوسط 50 و200 هو 125، فقد ارتفعت الأسعار بنسبة 25%. أما إذا اعتبرنا العام الحالي فترة مرجعية، فإن الأمر سيختلف. كان الحليب يكلف 200% أكثر مما يكلفه الآن وكان الخبز يباع مقابل

50% أقل. المتوسط: 125%. أي أن الأسعار كانت أيضا أعلى بنسبة 25% مما هي عليه الآن. ولإثبات أن مستوى المعيشة لم يتغير، نستعمل بدلا من المعدل الحسابي (أو المتوسط) المعدل الهندسي مع اعتبار الفترتين معا كقاعدة مرجعية.

الحقيقة هي أنه على الرغم من أساسها الرياضي، فإن الإحصائيات هي فن وعلم. العديد من التلاعبات وحتى التحريفات ممكنة في حدود اختصاصك وتقديرك. فغالبا ما يجب على الخبير الإحصائي الاختيار من بين الأساليب المختلفة، التي لا تزال عملية ذاتية، والعثور على الطريقة التي يجب أن يستخدمها لتمثيل الحقائق.

11. كيفية التعامل مع الإحصائيات¹⁰

حتى الآن، خاطبتك وكأني قرصان وأرغب في إرشادك حول كيفية التعامل مع الأخطاء الفادحة. في الفصل الذي يغلق هذا الكتاب، سأستغني عن هذه الحيلة الأدبية. سوف أشير إلى الغرض الجاد الذي أعتقد أنه يمكن أن تستشفه في ثنايا هذا الكتاب؛ وذلك من خلال شرح كيفية النظر إلى إحصائية مشوهة وكشفها؛ والأهم من ذلك، كيفية التعرف على البيانات المفيدة والمحددة بين مجموعة عمليات الاحتيال التي خصصت لها الفصول السابقة. والأسئلة الخمس الآتية تساعدك على تجنب الوقوع في مصيدة الإحصائيات:

1.11. من قال هذا؟

أول شيء يجب النظر إليه هو كيف يمكن أن تكون المعلومات متحيزة. اجتث عن التأثير المقصود. وابتث بعناية أكثر عن التأثير غير المقصود. فهو غالبا ما يكون أكثر خطورة. وقد يستغرق الأمر على الأقل فحصا ثانياً لمعرفة ذلك. فعندما يذكر اسم مؤسسة مرموقة، على سبيل المثال، تأكد من حقيقة ذلك ومن عدم ورود المعلومات كشيء معروض بجانبه(أ). وهل هناك داع ما للتحيز: "مختبر يريد إثبات شيء من أجل نظرية أو سمعة أو ضرائب؛ صحيفة هدفها قصة جيدة؛ عمال أو إدارة مع مستوى أجر على المحك".

2.11. كيف تم معرفة ذلك؟

لا يمكنك أنت كقارئ تطبيق اختبارات الدلالة أو الوصول إلى استنتاجات دقيقة حول مدى ملاءمة العينة (تم تحديدها بشكل صحيح يسمح باستنتاج موثوق به). ومع ذلك، وحول العديد من الأشياء التي يتم تقديمها في التقارير، ستتمكن من القول بلمحة (أو بنظرة متفحصة) أنه لم تكن هناك حالات كافية لإقناع الفرد العقلاني بأي شيء على وجه الخصوص. لن يتم إخبارك دائماً بعدد الحالات. إن غياب هذا الرقم، خاصة عندما يكون مصدر المعلومات طرفاً معنياً، يكفي للشك في كل التقرير. كما تُفقد العديد من الأرقام دلالتها من خلال فقدان مصطلح المقارنة. ففي بعض الأحيان، يتم ذكر النسب من دون المادة الرقمية المرجعية، والتي يمكن أن تكون مضللة أيضاً. إذا أعطوك مؤشراً ما، فاسأل عما هو مفقود. ربما هي قاعدة البيانات المختارة من أجل تحريف النتيجة.

3.11. ما المفقود؟

ابحث عن معدل ما، عن تنوع غير مفصل، في أي مسألة حيث من المتوقع أن يختلف المتوسط والوسيط بشكل كبير. في بعض الأحيان تكون النسب المئوية معطاة والأرقام الأولية مفقودة، وهذا يمكن أن يكون خادعاً أيضاً. منذ فترة طويلة، عندما بدأت جامعة "جونز هوبكنز" في قبول الطالبات، أبلغ شخص غير معجب بشكل خاص بالتعليم المختلط عن صدمة حقيقية: تزوجت ثلاث وثلاثون بالمائة وثلث النساء في "هوبكنز" من أعضاء هيئة التدريس! أعطت الأرقام الأولية صورة أوضح. كانت هناك ثلاث نساء مسجلات في ذلك الوقت، وتزوجت إحداهن من رجل هيئة تدريس. كما أن إغفال معدل البيانات قد يكون مضللاً: إن الإعلان عن زيادة كبيرة في عدد الوفيات بسبب السرطان في الربع الأخير من القرن الماضي قد يكون مضللاً إذا لم تأخذ بعين الاعتبار عوامل مرضية أخرى غير متوافرة الآن، وإذا كنت تنظر إلى إجمالي الوفيات بدلاً من معدل الوفيات، فلا تهمل حقيقة أن هناك عدداً من الناس الآن أكثر مما كان عليه في السابق.

4.11. هل قام أي شخص بتغيير البيانات؟

عند التحقق من إحصائية ما، ابحث عن الاتجاه المحتمل الذي أدخله شخص ما في إجمالي الأرقام أو في الاستنتاجات. فغالبًا ما يتم تقديم شيء بدلاً من شيء آخر. فقد تحدث أشياء غريبة عندما تستند الأرقام إلى ما يقوله الناس... حتى عندما يتعلق الأمر بأشياء تبدو موضوعية. فمثلاً، أظهرت بيانات التعداد السكاني أن عدد الأشخاص الذين يبلغون من العمر خمسة وثلاثين عاماً، على سبيل المثال، يزيد عن عدد ذوي أربعة وثلاثين أو ستة وثلاثين عاماً. وقد تحدث هذه الصورة الزائفة لأن فرد العائلة الذي يبلغ عن أعمار الآخرين يميل إلى تقريبهم إلى مضاعفات خمسة (30-35-40... سنة). طريقة مواجهة ذلك هي أن تطلب تسجيل تواريخ الميلاد.

مثال آخر: نقل البعض أن عدد سكان منطقة في الصين بلغ 28 مليون نسمة. بعد خمس سنوات بلغ عددهم 105 مليون. القليل جداً من تلك الزيادة كان حقيقياً. لكن لا يمكن تفسير الفرق الكبير إلا من خلال مراعاة أغراض التعدادين والطريقة التي يميل الناس إلى الشعور بها في الحساب في كل حالة. التعداد الأول كان للأغراض الضريبية والعسكرية، والثاني لتخفيف المجاعة.

كما تعتبر طريقة التفسير البعدي (post hoc) طريقة أخرى عبثية وسخيفة لتغيير فكرة ما دون أن يبدو عليها ذلك. حيث يتم تقديم تغيير طراً على شيء ما وحدث بالتزامن مع تغيير في شيء آخر، كما لو كان أحدهما هو سبب الثاني. ففي بعض الأحيان يتم التلاعب بالألفاظ (بدلالاتها) لتغيير الفكرة.

12. الخاتمة: هل لذلك معنى؟

سيقل هذا السؤال من أهمية الإحصائيات عندما يعتمد تقريرها على افتراض غير مثبت. فقد تزور الإحصائيات ويتم نشرها فقط لأن سحر الأرقام يتجاوز الحس السليم. ومن ذلك مثلاً:

ما ورد في التعديلات على قانون الضمان الاجتماعي، والتي قد يعتبر بيانها ذا معنى عندما لا يتم النظر فيه عن كثب. تمثلت حجة التعديلات فيما يلي: نظرًا لأن متوسط العمر المتوقع يبلغ حوالي 63 عامًا فقط، فمن الغش والاحتيال وضع خطة للضمان الاجتماعي تعتمد سن تقاعد من خمسة وستين عامًا، لأن كل الأشخاص سيموتون تقريبًا قبل ذلك.

وقد نجد مثالاً آخر في إحصائيات طبيب مسالك بولية مشهور يذكر وجود ثمانية ملايين حالة من سرطان البروستاتا في الولايات المتحدة؛ وهو ما يكفي للتنبؤ بوجود 1.1 غدد سرطانية لكل واحد من جميع الرجال الذين هم في سن عرضة للإصابة بهذا المرض. إن هذا الرقم الدقيق المثير للإعجاب أحيانًا مع الحس السليم.

قد يفيد التنبؤ الاستقرائي المسمى "التنبؤ بالاتجاهات المستقبلية"، ولكن عند فحص الأشكال أو الرسوم البيانية المستمدة منه، يجب مراعاة شيء واحد: قد يكون الاتجاه الحالي حقيقة، ولكن لا يمكن توقع الاتجاه المستقبلي إلا بفرضية معقولة.

مما يجعلنا نختم كلامنا بالقول-تبعاً لـ"لانغ" (Andrew Lang)- أنه لا يجب استخدام الإحصائيات مثل السكارى الذين يستخدمون أعمدة الإنارة للاتكاء عليها بدلاً من الاستفادة من إضاءتها¹¹.

الهوامش المراجع

(*) العمل عبارة عن مقتبسات مترجمة وأفكار ملخصة بالعربية من كتاب بالإنجليزية صدر عدة مرات منذ عام 1954 ولا زال يباع حتى الآن، بل إن بعض الجامعات (جامعة بوغوتا الكولومبية) ضمنت ملخصه برنامج مادة الصحافة الاستقصائية في طور الماستر:

Darrell Huff, (1991) How to Lie with Statistics. Penguin; New Ed edition, 130 pp. ISBN 0-14-013629-0

مع ملاحظة: لقد تم انتقاد الكاتب لتحريره مسودة كتاب آخر في الستينيات تراجع عن طباعته، وكان من تمويل شركات التبغ التي كلفته بإنتاج كتاب بعنوان كيف تكذب مع إحصائيات التدخين، وذلك لمحاولة استخدام الحجج الإحصائية لتقويض القول بأن التدخين ضار بالصحة¹². مما يفيد في معرفة أن التشهير بالإحصائيات سيف ذو حدين: توعوي بإمكانية تضليلها للقارئ أو تبريري لإمكانية نفي فوائدها.

- 1- Darrell Huff, (1991), How to Lie with Statistics. Penguin; New Ed edition, pp. 13-28.
- 2-Maldonado Rayma Ileri (2011). Cómo mentir con estadística. <http://raymastudillo.webcindario.com/unidac/UAP2/ESTADISTICA/ComoMentirConEstadisticas.pdf>. p.14. 03.6/2020
- 3- Darrell Huff, Op. Cit. pp. 37-51.
- 4- Ibid. pp. 52-57.
- 5- Ibid. pp. 58-63.
- 6- Ibid. pp. 64-71
- 7-Sala de prensa (2012). Cómo mentir con estadística. <http://www.saladeprensa.org/art794.ht>. 04/6/2020;Darrell Huff, Op. Cit. pp. 72-83.
- 8- Ibid. Pp. 84-93
- 9- <https://www.nateliason.com/notes/lie-statistics-darrell-huff>. 04/6/2020; Darrell Huff, Op. Cit. pp. 94-109.
- 10- <https://thepowermoves.com/how-to-lie-with-statistics-summary>. 04/6/2020.
- 11- Bolaños, Ricardo (2014).¿Qué dicen y qué no dicen las estadísticas?. <http://app4stats.com/wp-content/uploads/2013/06/descrip1.jpg>. 03/6/2020.
- 12- https://en.wikipedia.org/wiki/How_to_Lie_with_Statistics. 03/6/2020.