

# التنوير ٠٣

## تحليل البيانات: مقدمة قصيرة للعمل مع جداول البيانات



في هذه المذكرة الثالثة من وحدة إضاءة لفك شفرة انعدام العدالة، نستكشف كيفية البدء في إعطاء معنى للبيانات التي جمعناها. هنا، سيجد النشطاء وصانعو التغيير أدوات لبدء تحليل الأرقام التي ستشكل أدلة من أجل إحداث تغيير.

### الأسئلة الرئيسية

كيف يمكن أن تساعدنا جداول البيانات في تحليل البيانات التي جمعناها؟

ما هي الميزات الأساسية لجداول البيانات وكيف يتم استخدامها؟

ما هي المفاهيم الرياضية الرئيسية التي يجب مراعاتها عند تحليل البيانات؟

## تحليل البيانات: مقدمة قصيرة للعمل مع جداول البيانات

### مقدمة

هناك فرق بين البيانات والأدلة. لكي تصبح البيانات دليلاً، يجب علينا تحليلها وتفسيرها. في إطار ما نريده ضمن هذا السياق، يعني هذا طرح السؤال: ماذا تخبرنا البيانات عن المؤشرات التي حددناها لكل بُعد من أبعاد إطار عمل OPERA؟ في بعض الأحيان يمكن قراءة البيانات وتفسيرها بسهولة تامة. في أحيان أخرى، تكون مجموعة البيانات (الأولية أو الثانوية التي تم جمعها للتحليل) كبيرة إلى درجة لا يمكن معها قراءتها بسهولة في صفحة واحدة أو صفحتين. في هذه الحالة، يمكن أن يكون جدول البيانات أداة مفيدة. هذا هو الحال بشكل خاص عند التعامل مع بيانات الميزانية.

يمكن أن يكون العمل باستخدام جداول البيانات مهارة جديدة للعديد من النشطاء. تقدم هذه المذكرة المهارات الأساسية لاستخدام جداول البيانات. كما أنها توضح بعض المفاهيم الرياضية الأساسية التي يمكن استخدامها لتحليل البيانات في الجداول. والغرض من هذه المذكرة أن تكون تمهيدية؛ لذا لا تتردد في تخطي الأجزاء التي تعرفها بالفعل!

الشعور بالثقة في القدرة على التعامل مع الأرقام أمر مهم. يمكن أن يجهزنا بشكل أفضل كناشطين؛ إما لتحليل البيانات بأنفسنا، أو بالتعاون مع الاقتصاديين والإحصائيين وغيرهم لإجراء مثل هذا التحليل. نحن هنا نشجعك على تخصيص الوقت الكافي لتجديد معلوماتك عند الحاجة.

### أساسيات جدول البيانات

يتيح إدخال البيانات في جدول بيانات، تنظيمها وإعادة ترتيبها وتحليلها بمزيد من التفصيل، لاختبار الفرضيات حول كيفية ارتباط أجزاء البيانات المختلفة ببعضها البعض. جدول البيانات هو في الأساس جدول خلايا مرتبة في صفوف وأعمدة. يتم تمثيل الأعمدة عادة بالحروف، بينما يتم تمثيل الصفوف بالأرقام. يمكن الإشارة إلى خلية واحدة من خلال ما تحتله من صف وعمود؛ "C10"، على سبيل المثال.

هناك مجموعة متنوعة من برامج جداول البيانات والتطبيقات المتاحة، مثل Microsoft Excel و Calc from OpenOffice و Google Sheets. في هذه المذكرة، سوف نستخدم Microsoft Excel.

### جلب البيانات الثانوية

ما لم تكن تدخل البيانات الأساسية في جدول البيانات نفسه، فإن الخطوة الثانية بعد إنشاء جدول البيانات هي جلب بياناتك. قد تكون هذه بيانات أولية قمت بتسجيلها في مستند آخر، أو بيانات ثانوية من مصدر خارجي مثل البنك الدولي أو وزارة حكومية. إذا كنت تقوم بتنزيل بيانات ثانوية من مصدر عبر الإنترنت، فتأكد أنها بالتنسيق الصحيح. سيكون لديك خيار تنسيق البيانات التي ترغب في تنزيلها من مصدر البيانات. أحد أكثر التنسيقات شيوعاً لتنزيل البيانات هو CSV، الذي يرمز إلى «comma separated values» قيم مفصولة بفواصل. وهو تنسيق يمكن قراءته بسهولة بواسطة Excel.

لجلب البيانات، اختر علامة التبويب «البيانات» في الجزء العلوي من جدول البيانات. سيظهر لك هذا بعد ذلك خيارات مختلفة لجلب البيانات (على سبيل المثال، «من برنامج Access»، «من موقع إلكتروني»، «من نص معين»). إذا اخترت «من نص»، سيؤدي ذلك إلى فتح مربع حوار، تختار منه الملف الذي قمت بتنزيله مسبقاً.

إذا أظهر مربع الحوار «معالج جلب النص»، يمكنك ترك التحديدات مضبوطة على الإعدادات الافتراضية:

- الخطوة 1 (نوع البيانات): محدد
  - الخطوة 2 (المحددات): علامة التبويب
  - الخطوة 3 (تنسيق بيانات العمود): عام
- تتضمن بعض اختصارات لوحة المفاتيح المفيدة ما يلي:
- نسخ [Ctrl + c]: نسخ الخلايا المحددة في الحافظة.
  - لصق [Ctrl + v]: لصق محتوى الحافظة.



### الإلهام

التحرك من أجل بناء القوة الجماعية ومحاسبة صناع القرار



### التنوير

إلقاء الضوء على المشكلات الأساسية من خلال جمع البيانات وتحليلها وتصورها



### التحقيق

وضع خريطة لفهم المشكلة بعمق باستخدام لتحديد المؤشرات "إطار" أوبرا والمعايير.

تنظيم هذه الوثيقة وفقاً لطريقة مبتكرة لجمع وتحليل وتقديم الدلائل حول ثلاث خطوات:

## معلومة مفيدة

عندما تقوم بتنزيل مجموعة بيانات، فإن أول شيء يجب عليك فعله هو عمل نسخة منها. يجب إجراء أي تغييرات تجريبيًا في هذه النسخة. هذا يعني أنه يمكنك الرجوع والتحقق من البيانات الأصلية في أي وقت. من الممارسات الجيدة أيضًا تدوين من أين حصلت على بياناتك ومتى وكيف يتم استرجاعها.

مجموعة من قواعد الترميز كي تتيح لكل من يعمل على البيانات معرفة ما يلي:

- صغيرة = بين 0 و 99 طالبًا.
  - متوسطة = بين 100 و 499 طالبًا.
  - كبيرة = بين 500 و 999 طالبًا.
- من المهم للغاية التأكد أن كل شخص يُدخل البيانات يتبع القواعد نفسها في كل مرة.

### إقفال الصفوف والأعمدة

قبل أن تبدأ في فرز بياناتك، قد يكون من المفيد «قفل» الصف العلوي و/أو العمود الأول من جدول البيانات، خاصةً إذا كان جدول البيانات الذي تستخدمه كبيرًا. هذا يعني أنه يمكنك دائمًا معرفة ما يشير إليه كل صف و/أو عمود. على سبيل المثال، ربما تكون قد قمت بتنزيل بيانات حول الإنفاق الحكومي على التعليم كنسبة مئوية من إجمالي الناتج المحلي لجميع البلدان للأعوام من عام 1990 إلى عام 2014. سيُشمل هذا العديد من الصفوف والأعمدة. للإبقاء على السنوات وأسماء البلدان مرئية، يمكن قفل هذه الصفوف والأعمدة حتى لا تختفي أثناء التنقل في جدول البيانات.

### فرز البيانات

أول شيء يجب فعله عند النظر إلى مجموعة بيانات جديدة هو التعرف عليها. يتضمن هذا فرز البيانات بطريقة تجعلها أكثر منطقية. لنفترض، على سبيل المثال، أن لديك جدول بيانات يحتوي على بيانات حول إجمالي الناتج المحلي ونفقات الرعاية الصحية ومتوسط العمر المتوقع لجميع البلدان، على مدى عدد من السنوات. للتعرف على البيانات، قد ترغب في إلقاء نظرة على البيانات أولاً بالترتيب حسب البلد، ثم حسب التاريخ. لفرز جدول البيانات بالكامل، استخدم الخطوات التالية:

1. حدد علامة التبويب «البيانات data» ثم الزر «فرز sort». سيؤدي هذا إلى فتح مربع حوار إضافي.
2. حدد العمود الذي ترغب في الفرز به في القائمة المنسدلة drop-down menu؛ يمكنك الفرز تصاعديًا (A-to-Z) أو تنازليًا (Z-to-A)، إذا كان أجنبيًا.
3. يمكنك أيضًا اختيار معايير متعددة للترتيب (على سبيل المثال، حسب البلد أولاً، ثم حسب التاريخ)، وذلك بالنقر فوق الزر «إضافة مستوى Add level»:

### تصفية البيانات

من الشائع تصفية filter out القيم التي لا تريد رؤيتها في مجموعة البيانات. في مجموعة البيانات المذكورة أعلاه، على سبيل المثال، قد تكون مهتمًا فقط بمراجعة أحدث القيم.

لتصفية جدول البيانات بالكامل، استخدم الخطوات التالية:

1. ضمن علامة التبويب «البيانات»، حدد زر «تصفية Filter». يجب أن تشاهد الآن مثلثات بجوار أسماء الأعمدة في الصف الأول.
2. انقر فوق المثلث الموجود بجوار العمود الذي يعينك. يجب أن يظهر مربع حوار، يمكنك من خلاله تحديد وإلغاء تحديد الخيارات التي تريد تضمينها أو استبعادها.

### تحليل البيانات باستخدام جداول البيانات

بمجرد قيامك بفرز جدول البيانات وتصفيته، يحين وقت البدء في التحليل. يساعدك تحليل البيانات في استخراج إجابات لأسئلتك. لنفترض أن لديك بيانات عن الناتج المحلي الإجمالي والسكان ومتوسط العمر المتوقع ونفقات الرعاية الصحية كنسبة مئوية من الناتج المحلي الإجمالي، لجميع البلدان

- قص [Ctrl + x]: نسخ الخلايا المحددة في الحافظة وإزالتها من موضعها الأصلي.
- تراجع عن [Ctrl + z]: التراجع عن آخر تغيير قمت به.
- إعادة [Ctrl + y]: التراجع عن التراجع.

### إدخال البيانات الأولية

إذا كنت تقوم بإدخال بياناتك الخاصة في جدول بيانات، فمن الضروري عمل ذلك بطريقة متسقة وموحدة. البيانات التي يتم إدخالها بشكل غير متسق تصعب من عملية البحث، والعد، والفرز، والتصفية. على سبيل المثال، إليك المعلومات الديموغرافية التي تم إدخالها في جدول بيانات:

من السهل تحديد المشكلة هنا: لقد تم تسجيل جنس الأفراد بطرق مختلفة. يؤدي إدخال البيانات بطريقة موحدة إلى تجنب هذه الأخطاء. هذا يعني اتخاذ خيارات حول كيفية تمثيل البيانات بشكل متسق. تتضمن البيانات التي يحتمل تسجيلها بشكل غير متسق ما يلي:

- **التواريخ والأوقات:** 1 يوليو 2005، الأول من يوليو 005، 1/7/2005 أو 20050701 كلها طرق تمثل التاريخ نفسه.
- **الأسماء:** على سبيل المثال، هل ستستخدم «الأمر المتحدة» أم اختصار لها «UN»؟ هل ستضع الاسم الأول والثاني لشخص في أعمدة مختلفة أم في العمود نفسه؟

الجنس	الاسم
م	هارون
مذكر	هايون
م	إليانور

- **الأماكن:** ما مدى الدقة التي تحتاجها عند وصف البيانات الجغرافية؟ هل ينبغي أن تستخدم الجغرافيا الإدارية لبلد ما (على سبيل المثال، بلدة، مدينة، مقاطعة)، أم جغرافيا انتخابية (جناح، دائرة انتخابية) أم جغرافيا تشغيلية؟
- النقطة الأساسية هي النظر في الطرق المختلفة لتسجيل البيانات، ثم الاختيار، والمواظبة على تطبيقه باتساق. هذا سيوفر قدرًا هامًا من الوقت ويجنبك الإحباط.

المسألة الثانية هي كيفية «ترميز» (أو تصنيف) البيانات. تزداد أهمية هذه النقطة إذا كنت تسجل بيانات نوعية ستخضع لتحليل كمي. على سبيل المثال، لنفترض أنك تريد تسجيل معلومات حول المدارس التي قمت بزيارتها، وتصنيفها على أنها إما صغيرة أو متوسطة أو كبيرة. عندما تقوم أنت وزملاؤك بمراجعة جدول البيانات، يجب أن تكون واثقًا من أنه في كل مرة ترى فيها مدرسة توصف بأنها «صغيرة»، فهذا يعني الشيء نفسه. لضمان ترميز البيانات بشكل متسق، يمكنك تصميم

الوصف	الرمز
هذا هو أول شيء يجب إدخاله في خلية الصيغة الخاصة بك، لأنه يخبر جدول البيانات أنك تكتب صيغة.	=
جمع	±
طرح	-
ضرب	*
قسمة	/

على مدى عدد من السنوات. هناك العديد من الأسئلة المتعلقة بالحق في الصحة التي يمكنك الإجابة عليها باستخدام هذه المعلومات. فمثلاً:

- ما المبلغ الإجمالي الذي تنفقه كل دولة على الرعاية الصحية (بالدولار الأمريكي)؟
- كيف تقارن دولة معينة بغيرها من الدول؟
- ما نصيب الفرد من الإنفاق في كل بلد (بالدولار الأمريكي)؟
- كيف تغير الإنفاق في كل بلد بمرور الوقت؟

من الأهمية تذكّر أن البيانات الموجودة في جدول البيانات الخاص بك لن توفر دائماً جميع المعلومات التي تحتاجها للإجابة على أسئلة البحث الذي تعمل عليه. على سبيل المثال، بناءً على البيانات المتوفرة لدينا، لا نعرف مقدار الأموال التي تُنفق على الرعاية الصحية في كل بلد. ومع ذلك يمكننا حساب هذا الرقم باستخدام البيانات المتوفرة لدينا، أي الناتج المحلي الإجمالي لكل بلد والنفقات الصحية كنسبة مئوية من الناتج المحلي الإجمالي. يمكنك حساب قيم جديدة مثل هذه باستخدام صيغ جداول البيانات spreadsheet formulae.

## استخدام الصيغ

الصيغ هي وظائف رياضية أساسية؛ نوع الأشياء التي قد تجدها في آلة حاسبة بسيطة. عن طريق كتابة هذه الصيغ في خلية جدول بيانات، يمكنك إجراء عمليات حسابية في جدول البيانات الخاص بك. تشمل الرموز الأكثر شيوعاً التي ستستخدمها في هذه الصيغ ما يلي:

تنطبق القواعد الرياضية الأساسية المتعلقة بترتيب الوظائف أيضاً عند العمل باستخدام جداول البيانات. على سبيل المثال، الصيغة

تنطبق القواعد الرياضية الأساسية المتعلقة بترتيب الوظائف أيضاً عند العمل باستخدام جداول البيانات. على سبيل المثال، الصيغة

=3+5 ستساوي 13 وليس 16. إذا كنت تريد تغيير ترتيب الدالة، فيجب عليك تضمين أقواس. سيتم حساب الصيغ الموجودة داخل الأقواس قبل أي معادلة أخرى. على سبيل المثال، إذا كنت تريد أن ينتج عن الصيغة أعلاه 16، فيجب إدخالها على النحو التالي: =2\*(5+3).

تعني الإضافة/الجمع إضافة القيم في نطاق من الخلايا في جدول بيانات. هناك طريقة سريعة للقيام بذلك وهي استخدام الرمز =SUM (بدلاً من ربط كل خلية على حدة برمز «+»). على سبيل المثال، إذا كنت تريد حساب إجمالي الناتج المحلي لمجموعة من 10 دول، يمكنك استخدام الخطوات التالية:

توجد بيانات إجمالي الناتج المحلي في العمود B. في الصف الأول الفارغ في العمود B، في نهاية قائمة إجمالي الناتج المحلي للبلدان في عام 2010، اكتب =SUM ثم قوس مفتوح؛ =SUM (.

حدد الخلية التي تحتوي على الدولة الأولى في المجموعة. أثناء الضغط باستمرار على المفتاح SHIFT، حدد الدولة الأخيرة في المجموعة. يجب أن تضيف خلية الصيغة أرقام الخلايا تلقائياً.

أكمل الصيغة بإضافة قوس قريب في النهاية؛ على سبيل المثال، =SUM (B2: B11) سيضيف هذا قيم الناتج المحلي الإجمالي في النطاق المحدد.

الضرب والقسمة مفيدان إذا كنت تريد تحويل قيمة عددية إلى نسبة مئوية، أو العكس. على سبيل المثال، لنفترض أنك تريد معرفة المبلغ (بالدولار الأمريكي) الذي يتم إنفاقه على الرعاية الصحية، بشكل إجمالي، في بلدك. أنت تعلم أن ناتجها المحلي الإجمالي يبلغ 100357.000.000 دولار أمريكي وأن الإنفاق الحكومي على الصحة يبلغ 3.48% من الناتج المحلي الإجمالي. يمكنك الإجابة على سؤالك باستخدام المعادلة التالية:

الناتج المحلي الإجمالي \* (الإنفاق على الصحة/100)

أو في جدول بيانات = 100357000000\*(3.48/100)

لا تحتاج إلى كتابة الأرقام يدوياً؛ سيؤدي النقر فوق الخلية إلى إتمام ذلك نيابةً عنك. تبدو الصيغة كما يلي: =D2\*(D2/100)، حيث تمثل الخلية C2 الناتج المحلي الإجمالي لبلدك والخلية D2 النفقات الصحية كنسبة مئوية من الناتج المحلي الإجمالي.

يكون نسخ الصيغ مفيداً عندما تريد تطبيق نفس الحساب عبر الصفوف أو عبر الأعمدة. لتنفيذ ذلك، ما عليك سوى نسخ الصيغة التي كتبها للتو (باستخدام Ctrl + c) ولصقها في الخلية أدناه (باستخدام Ctrl + v). بدلاً من ذلك، انقر فوق الزاوية اليمنى السفلية للخلية (المربع الأزرق) واسحب الصيغة لأسفل إلى نهاية العمود.

يمكن أن تكون الخلايا الفارغة مشكلة عندما تبدأ في تحليل بياناتك؛ لأنها قد تنتهي بإنتاج قيم صفرية. تعتبر المعالجة الصحيحة للقيم المفقودة خطوة مهمة في تنظيف البيانات وتحليلها. نادراً ما تكتمل مجموعات البيانات الكبيرة؛ لذا يجب أن يكون لديك استراتيجية للتعامل مع الأجزاء المفقودة. تتمثل إحدى طرق البحث عن الأخطاء في تصفية البيانات وتحديد جميع القيم «صفر» وحذفها.

## مفاهيم رياضية مفيدة لتحليل البيانات

قد يكون طلب البيانات للمساعدة في الإجابة عن أسئلة البحث أمراً شاقاً إذا كنت تحاول مقارنة متغيرات متعددة. على سبيل المثال، إذا كنت تبحث في البيانات المتعلقة بالحق في الغذاء، فقد ترغب في إلقاء نظرة على: (أ) نصيب الفرد من التوافر؛ (ب) من المواد الغذائية الرئيسية؛ (ج) عبر البلدات. قد تكون مهتماً بمعرفة متوسط التوفر لهذه العناصر، والحد الأدنى والحد الأقصى المتاح، والفرق، أو الاختلاف، بين الحد الأدنى والحد الأقصى. عادة ما يشار إلى معرفة الإجابات على هذه الأسئلة بالإحصاء الوصفي descriptive statistics.

## النطاق

أول جزء من المعلومات التي قد ترغب في العثور عليها من بياناتك هو النطاق؛ بمعنى آخر، من أين إلى أين تمتد بياناتك؟ هل تبدأ بأعداد صغيرة؟ أعداد كبيرة؟ هل تتحول من السلبي إلى الإيجابي؟ هذه كلها معلومات أساسية ستساعدك في التعامل مع بياناتك.

سيساعدك النظر إلى النطاق أيضاً في العثور على أخطاء في بياناتك. على سبيل المثال، لنفترض أنك تبحث في البيانات المتعلقة بمؤشر «متوسط سنوات الدراسة وسط السكان البالغين». قد تجد أن بياناتك تتراوح من 4 إلى 58. هنا سيكون من الواضح وجود خطأ ما؛ لأن احتمالية أن يكون لدى أي شخص بالغ في أي بلد 58 عاماً من التعليم هو احتمال ضئيل للغاية. يجب عليك العودة إلى بياناتك والتحقق منها.

كيفية تجد النطاق الخاص بك؟ ما عليك سوى الاطلاع على بياناتك وإيجاد القيم الدنيا والقصى: الأدنى والأعلى، على التوالي. على سبيل المثال، لنفترض أن لديك البيانات التالية حول متوسط سنوات الدراسة بين السكان البالغين في بلدك، وكذلك في العديد من البلدان المجاورة:

8,5, 5,8, 6,5, 7,6, 10,2, 8,4, 7,3, 7,2, 9,2, 9,3

• السؤال: ما هو نطاق مجموعة البيانات الخاصة بك؟

• الإجابة: أقل رقم (أدنى) هو 5.8 وأعلى رقم (أقصى) هو 10.2. وبالتالي، فإن النطاق يتراوح من 5.8 إلى 10.2.

في جدول البيانات، يمكنك القيام بذلك عن طريق فرز البيانات من الأصغر إلى الأكبر، أو باستخدام الصيغ = MIN و = MAX، باستخدام الأقواس لتحديد الخلايا التي تريد تضمينها في الحساب؛ على سبيل المثال، = (MIN (D12: D84)).

### عملية العد

المعلومة المهمة التالية التي قد ترغب في تحديدها هي عدد الأشياء التي لديك بيانات عنها. كم عدد البلدان؟ كم عدد الأسر؟ وهلم جرا.

كيف تحصل على هذه المعلومات من بياناتك؟ ببساطة عدّها. في مجموعة البيانات أعلاه، على سبيل المثال، هناك 10 ملاحظات.

إذا كانت مجموعة البيانات كبيرة جدًا بحيث لا يمكن عدّها، يمكنك استخدام الصيغ = COUNT عندما يكون لديك أرقام في الخلايا أو = COUNTA عندما لا يكون لديك أرقام في الخلايا. مرة أخرى، استخدم الأقواس لتحديد الخلايا التي تريد تضمينها في الحساب؛ على سبيل المثال = COUNTA (A5: A2089).

قد يبدو هذا بسيطًا. لكن عندما يتعلق الأمر بتحليل البيانات وتفسيرها، فإنه يصبح مهمًا للغاية. على سبيل المثال، إذا كنت تقارن بياناتك عن المستشفيات في بلدك، فهل بيانات 10 مستشفيات كافية لإجراء هذه المقارنة؟

### المتوسّطات

الجزء التالي من المعلومات التي قد ترغب في إلقاء نظرة عليها هو القيمة المركزية وكيفية توزيع البيانات بالنسبة لتلك القيمة. هل تعطي القيمة المركزية مؤشرًا جيدًا لمجموعة البيانات بأكملها، مع توزيع متساو لنقاط البيانات فوقها وتحتها؟ أم أن التوزيع «منحرف»، بمعنى أن هناك ذروة في أحد طرفي نطاق البيانات مع ذيل طويل باتجاه الطرف الآخر؟ يخبرك التوزيع بنوع الوصفات de-scribers الإضافية العملية للاستخدام. هناك عدد من الطرق المختلفة للإجابة على هذه الأسئلة.

المتوسط mean (أو «average») هو الطريقة الأكثر شيوعًا للنظر إلى «القيمة المركزية». وهي أمر مألوف في التقارير؛ على سبيل المثال، ارتفع متوسط البطالة في البلد «س» أو انخفض متوسط معدلات معرفة القراءة والكتابة في البلد «ص».

كيف يتم حساب المتوسط؟ المتوسط هو مجموع كل القيم الواردة في مجموعة البيانات مقسومًا على عدد القيم الموجودة. على سبيل المثال، لنفترض أنك مهتم بتحديد متوسط دخل الأسرة. لديك البيانات التالية حول متوسط الدخل السنوي للأسرة، محسوبًا بالدولار، لعدد من الأسر: 1120، 241، 876، 201، 112، 345، 567، 156، 154، 1345.

• سؤال: ما هو متوسط مجموعة البيانات الخاصة بك؟

• الإجابة: اجمع الدخل الذي لديك لجميع الأسر (1120 + 241 + 876 + 201 + 112 + 345 + 567 + 156 + 154 = 5117). ثم قسّم هذا الرقم على عدد الأسر التي لديك (5117/10). إجابتك هي 511.7.

في جدول بيانات، يمكنك حساب ذلك باستخدام الصيغة = AVERAGE.

يمكن أن يمنحك المتوسط تقديرًا جيدًا لما هو «طبيعي» عندما يتم توزيع باقي البيانات بالتساوي فوقها وتحتها. ومع ذلك، إذا كانت بقية بياناتك «منحرفة» لأي جانب، فقد يكون مقياسًا مختلفًا للمتوسط أكثر ملاءمة. على سبيل المثال، إذا كان لديك عدد قليل من ذوي الدخل المرتفع للغاية في مجموعة سكانية، فإن استخدام المتوسط سيجعل متوسط دخل الفرد يبدو أعلى مما هو عليه بالفعل.

الوسيط هو القيمة العددية التي تفصل النصف الأعلى من القيم في مجموعة البيانات عن النصف السفلي. يكون مفيدًا عندما لا يتم توزيع باقي بياناتك بالتساوي على جانبي الوسط. في مثال متوسط دخل الأسرة أعلاه، متوسط قيمة الدخل هو 511.7. ومع ذلك، يعد هذا عددًا كبيرًا جدًا بالنسبة إلى معظم القيم الموجودة في مجموعة البيانات. ينتج عن وجود عدد قليل من الأسر ذات الدخل المرتفع للغاية، مما يؤدي إلى تحريف البيانات. في هذه الحالة، قد يوفر الوسيط تقديرًا أفضل لما هو «طبيعي» من المتوسط.

كيف يتم حساب الوسيط؟ أولًا، قم بفرز البيانات (لا يهم تصاعديًا أو تنازليًا) والقيمة الموجودة في منتصف مجموعة البيانات هي الوسيط. إذا كان هناك عدد زوجي من القيم في مجموعة البيانات، فاحسب متوسط القيمتين الوسييتين.

الفرق الترتيبي	الفرق عن المتوسط	القيمة
2.25	-1.5	1
0.25	-0.5	2
0.25	0.5	3
2.25	1.5	4

• سؤال: ما هو متوسط دخل الأسرة؟

• الإجابة: أولًا، قم بفرز البيانات: 112، 154، 156، 201، 241، 345، 567، 876، 1120، 1345. توجد 10 قيم والقيمتان الوسييتان هما 241 و 345. أوجد المتوسط بين هذين العددين (293 = 345/2 + 241). الوسيط هو 293.

في أي جدول بيانات، يمكنك حساب ذلك باستخدام الصيغة = MEDIAN الوسيط.

المنوال هو أكثر قيمة تظهر في مجموعة من البيانات. في بعض الأحيان لا يخبرنا المتوسط ولا الوسيط حقًا بما نريد معرفته. على سبيل المثال، إذا كنت ترغب في معرفة متوسط عدد الأطفال لكل أسرة مسجلة في المدرسة، فقد يكون لديك مجموعة البيانات التالية:

0، 1، 1، 1، 1، 2، 2، 2، 3، 5

متوسط عدد الأطفال المسجلين في المدرسة لكل أسرة هو 1.8، والوسيط 1.5. لكن ما تريد معرفته حقًا هو عدد الأطفال المسجلين في المدرسة، في غالبية الأسر. يمكنك أن ترى أن طفل «1» هو الإجابة الأكثر شيوعًا. هذا هو المنوال.

في الحالة التي تتكرر فيها أكثر من قيمة واحدة، يمكن أن تكون مجموعة البيانات ثنائية المنوال bimodal (قيمتان متوسطتان) أو متعددة المناول multimodal (أكثر من قيمتين متوسطتين).

### التباين Variation

الجزء المهم التالي من المعلومات التي قد ترغب في تحديدها هو حجم التباين في مجموعة البيانات. هذا أمر بالغ الأهمية عند النظر إلى البيانات المجمعة والمصنفة. على سبيل المثال، لنفترض أنك تحقق في أعمال الحق في العمل. تريد معرفة متوسط معدل البطالة. ومع ذلك، قد ترغب أيضًا في اختبار مدى تمثيل المتوسط للبيانات المختلفة داخل بلدك. هناك نوعان من

الإجراءات الشائعة لعمل ذلك.

الانحراف المعياري هو مقياس لمدى انحراف قيم البيانات عن المتوسط. توضح الخطوات الثلاث التالية كيفية حسابه:

2. احسب متوسط الفروق.  
لنأخذ مثال البطالة نفسه على النحو الوارد أعلاه، لكن مع بيانات عن بلدية إضافية واحدة، لتصبح بياناتك 1، 2، 3، 4، 5.

1. اجمع مربع الفروق بين القيم والمتوسط.
2. قسّم هذا المجموع على عدد القيم مطروحاً منه واحد.
3. احسب الجذر التربيعي.

### تطبيع Normalization

بمجرد أن تكون لديك فكرة عن البيانات التي تتعامل معها، قد تبدأ في إجراء مقارنات. على سبيل المثال، في استقصاء عن مدى إعمال الحق في الصحة، قد ترغب في مقارنة الإنفاق الحكومي على الصحة في بلدك بإنفاق بلد آخر يختلف في عدد من الجوانب. ستخبرك مقارنة القيمة الإجمالية للإنفاق الحكومي على الرعاية الصحية في هذه الحالة بالكثير، على سبيل المثال إذا كان بلدك كبيراً جداً والبلد الذي تقارنه به صغيراً جداً، من المرجح أن تنفق الدولة الأكبر على الصحة أكثر بكثير من الدولة الأصغر. هل هذا يعني أن الدولة الأكبر تفي بالتزامها بالحق في الصحة بشكل أفضل من الدولة الأصغر؟ ليس بالضرورة. للإجابة على هذا السؤال، يجب مقارنة البلدين على قدم المساواة. يتم ذلك عادةً باستخدام مؤشر يخبرنا عن حجم الدولة؛ في كثير من الأحيان حجم سكانها. لمقارنة الإنفاق الحكومي على الصحة، يمكنك تقسيم إجمالي الإنفاق على السكان. هذا يسمى التطبيع أو التسوية.

### تطبيق هذه المفاهيم عند تحليل بيانات الميزانية

يتضمن تحليل ميزانية الحكومة من منظور حقوق الإنسان بعض الحسابات المحددة. هذا لأن الأرقام في الميزانية دائماً ما تكون نسبية. على سبيل المثال، تحليل ما إذا كانت أرقام الميزانية مرتفعة أم منخفضة، يتضمن أن نسأل مرتفع أم منخفض نسبة إلى ماذا؟ لنأخذ مثالاً افتراضياً، لنفترض أن ميزانية الإسكان الاجتماعي لبلد ما انتقلت من 100 مليون دولار أمريكي في عام 2015 إلى 200 مليون دولار أمريكي في عام 2020. قد تبدو مضاعفة ميزانية الإسكان الاجتماعي أمراً بالغ الأهمية. ومع ذلك، ماذا لو تضاعفت الميزانية الإجمالية للحكومة ثلاث مرات في نفس الفترة؟ كنسبة مئوية، تكون ميزانية الإسكان الاجتماعي قد تقلصت بالفعل. ماذا لو زادت تكلفة بناء المساكن الاجتماعية بسبب التضخم بنسبة 150% خلال نفس الفترة؟ في هذا السيناريو، تنخفض القوة الشرائية للحكومة، لذا لا تمتد الميزانية إلى هذا الحد. بسبب هذه النسبية، غالباً ما يكون من الضروري القيام بعملية تحويل أرقام الميزانية.

### تحويل الأرقام إلى نسب

يعد استخدام النسب المئوية إحدى طرق جعل البيانات قابلة للمقارنة. على سبيل المثال، قد ترغب في مقارنة عائدات الضرائب لبلدين. تتمثل إحدى طرق القيام بذلك في تطبيع الإيرادات الضريبية كنسبة مئوية من الناتج المحلي الإجمالي للبلدين، مما قد يُظهر أن الدولة الأولى تجمع قدرًا مرتفعاً نسبياً من الإيرادات، في حين أن الدولة الثانية لا تفعل ذلك.

### معلومة مفيدة

قامت منظمة شراكة الموازنة الدولية Int. Budget Partnership بتطوير حاسبة للتضخم. لاستخدامها، ما عليك سوى إدخال المبالغ الاسمية لكل عام ومؤشر أسعار المستهلك (CPI) لكل عام وحسابها.

انظر: <http://openbudgetsgame.org/inflationguide>

في أي جدول بيانات، يمكنك القيام بذلك باستخدام الصيغة =STDEV.  
لنفترض أن لديك البيانات التالية حول معدلات البطالة كنسبة مئوية من إجمالي السكان الفاعلين في أربع بلديات مختلفة:

- 1، 2، 3، 4
- ما هو الانحراف المعياري لمجموعة البيانات لديك؟
- المتوسط 2.5، لذا اتبع الخطوات من 1 إلى 3:
1. اجمع مربع الفروق بين القيم والمتوسط = 5.
2. قسّم هذا المجموع على عدد القيم ناقص واحد =  $5 / (4-1) = 5/3$ .
3. احسب الجذر التربيعي = 1.291.

إذا تم توزيع البيانات بشكل طبيعي (يتم توزيع جميع نقاط البيانات بالتساوي على جانبي المتوسط)، فإن 68.27% من نقاط البيانات ستقع ضمن انحراف معياري واحد عن المتوسط و95.45% من نقاط البيانات ستقع ضمن انحرافين معياريين عن الوسط. هذا يعني أنه كلما زادت قيمة الانحراف المعياري، زاد التباين في مجموعة البيانات.

باستخدام مثال البطالة، يعني الانحراف المعياري البالغ 1.291 أن 68.27% من نقاط البيانات ستقع ضمن هذه المسافة من المتوسط (بافتراض التوزيع الطبيعي). يشير هذا إلى احتمال وجود تباين جغرافي معقول في البطالة؛ لذا فإن الإبلاغ عن المتوسط وحده قد يكون مضللاً.

يشبه متوسط الانحراف المطلق median absolute deviation الانحراف المعياري، لكنه يستخدم مع الوسيط بدلاً من المتوسط. توضح الخطوات التالية كيفية حسابه:

1. احسب الوسيط والاختلافات المطلقة بين كل قيمة والوسيط.

### معلومة مفيدة

إذا انتهى بك الأمر إلى العمل باستخدام مجموعة بيانات كبيرة جداً، قد تكون هناك نقطة «تجاوز فيها» مساحة جدول بياناتك. إذا حدث ذلك، لا بد أن نفكر في استخدام برامج قواعد بيانات أخرى مثل SPSS أو Stata أو R. بخلاف جدول البيانات المصمم ليكون قادراً على «القراءة» على الشاشة، غالباً ما تكون الطريقة التي يتم بها تخزين البيانات في قاعدة البيانات مخفية تماماً عن المستعمل. يتيح ذلك طرقاً مجردة ومعقدة لتخزين كميات أكبر من البيانات ويمنح المستخدم مزيداً من المرونة في كيفية استخدامها. ومع ذلك، تعد قواعد البيانات طريقة أكثر تقنية لتخزين البيانات وتحليلها، وإذا كنت تستخدمها، فقد تحتاج إلى العمل مع الإحصائيين والمبرمجين والمصممين.

## صيغة لحساب التكيف مع التضخم

القيمة الحقيقية =  $\frac{\text{القيمة الاسمية للسنة المستهدفة} \times \text{مؤشر أسعار المستهلك لسنة الأساس (CPI)}}{\text{التكلفة لكل تثبيت للعام المستهدف}}$

على سبيل المثال، تحسب أموال 2010 مضروبة في قيم 2000 على النحو التالي

القيمة الحقيقية =  $\frac{\text{قيمة 2010} \times \text{مؤشر أسعار المستهلك}}{\text{مؤشر أسعار المستهلك 2010}}$

### حساب مخصصات نصيب الفرد

للكشف عن الأنماط التمييزية للمخصصات، يجب أيضًا تقييم الميزانية من حيث كيفية توزيعها للمنافع بين الأسر والأفراد. لتحقيق ذلك، قد تحتاج، على سبيل المثال، إلى حساب إنفاق الفرد حسب المنطقة أو البلدية. يمكنك عمل ذلك ببساطة عن طريق قسمة إجمالي المخصصات للمنطقة أو البلدية على إجمالي عدد السكان في تلك المنطقة (أو حسب إجمالي مجموعة السكان التي تهتم بها، مثل الأطفال في سن المدرسة أو النساء في سن الإنجاب).

### التكيف مع التضخم

من المهم مقارنة مخصصات الميزانية بمرور الوقت وتقييم ما إذا كان هناك زيادة أو نقصان في المبالغ المخصصة لمختلف القطاعات. ترجع أهمية ذلك، لما يفيد عما إذا كانت الحكومة تتخذ إجراءات لإعمال هذه الحقوق «بشكل تدريجي» أم لا.

على الرغم من ذلك، فإن الأمر ليس ببساطة مجرد النظر إلى أرقام الميزانية من سنوات مختلفة. فأرقام الميزانية ترد بالقيمة «الاسمية»، مما يعني أنها لا تأخذ التضخم في الاعتبار. لهذا السبب، من الضروري تحويل المخصصات من المبالغ «الاسمية» إلى المبالغ «الحقيقية». وهذا يجعل أرقام الميزانية من سنوات مختلفة «مكافئة» للقيم السارية في إحدى السنوات، ويتيح إجراء مقارنات صحيحة مع مرور الوقت. بمعنى آخر، القيمة الحقيقية = القيمة الاسمية المعدلة للتضخم.

### التفسير وكيفية تجنب المفاهيم الخاطئة الشائعة

في حين أن التبسيط مطلوب لفهم ما تعنيه البيانات، كما هو الحال عند تقدير الدليل في تنسيق قارئ على الجرافيك أو الرسم، فمن الأهمية بمكان أن تظل قريبًا قدر الإمكان من «القصة الكاملة».

### الارتباط ليس سببية

بشكل عام، من الصعب للغاية تحديد علاقة سببية بين ملاحظتين مترابطتين. هناك أسباب عديدة تجعل الاستنتاجات المنطقية حول السبب والنتيجة خاطئة. على سبيل المثال، قد ترغب في الإفادة عن العلاقة بين التعليم والصحة. باستخدام مخطط التشتت scatter plot، يكون لديك متوسط سنوات الدراسة على أحد المحاور ومتوسط العمر المتوقع على المحور الآخر. يكشف مخطط التشتت أن المتغيرين مرتبطين بشكل إيجابي للغاية: يرتبط متوسط سنوات الدراسة الأعلى بمتوسط العمر المتوقع. لكن هل يمكن القول إن ارتفاع متوسط سنوات الدراسة يؤدي إلى ارتفاع متوسط العمر المتوقع؟ لا، إذ يمكن أن يكون هناك عدد من الأسباب لهذا الارتباط:

- ارتفاع متوسط سنوات الدراسة قد يؤدي إلى ارتفاع متوسط العمر المتوقع.
- ارتفاع متوسط العمر المتوقع قد يؤدي إلى ارتفاع متوسط سنوات الدراسة.
- ارتفاع متوسط سنوات الدراسة وارتفاع متوسط العمر المتوقع هما نتيجتان لسبب مشترك، لكنهما لا يسببان بعضهما البعض. قد يكون السبب الشائع هو ارتفاع متوسط الدخل، على سبيل المثال، أو توزيع أكثر إنصافاً للدخل.
- لا توجد علاقة بين متوسط سنوات الدراسة ومتوسط العمر المتوقع. هذا الارتباط هو من قبيل الصدفة.

### تغيير النسبة المئوية وتغيير النسبة المئوية للنقطة

يمكن أحيانًا الخلط بين التغيير في النسبة المئوية وتغيير النقطة المئوية في تفسير البيانات. على سبيل المثال، إذا تغيرت قيمة من 5% إلى 10%، فما هي النسبة المئوية للتغيير؟ الخطأ الشائع هو الإجابة بنسبة 5%. هذا غير صحيح. النسبة المئوية للتغيير في السؤال هي في الواقع 100%. ومع ذلك، فهو تغيير في خمس نقاط مئوية. يعتمد اختيار الإفادة عن النسبة المئوية للتغيير أو التغيير في النسبة المئوية على السؤال الذي تطرحه. غالبًا ما يكون من المفيد الإبلاغ عن كليهما.

## أفكار استخلاصية

يساعد بناء الثقة اللازمة للعمل مع الأرقام على تجهيزنا كناشطين؛ إما لتحليل البيانات بأنفسنا أو للتعاون مع الاقتصاديين والإحصائيين وغيرهم لإجراء مثل هذا التحليل. يمكن أن يستغرق التحليل بعض الوقت. وتتضمن حالاته أن يكون تجريبيًا واستكشافيًا، مما يعني التلاعب بالبيانات التي لدينا. على سبيل المثال، قد نقارن المتغيرات، ونفصلها، وننظر إليها بمرور الوقت، وما إلى ذلك. وهذا يساعدنا على معرفة ما هو الأكثر صلة ببحثنا.

على الرغم من ذلك، من المهم جدًا الاحتفاظ بسجل لأي تغييرات يتم إجراؤها على البيانات. هذا هو الحال بشكل خاص إذا كان هناك أشخاص مختلفون يعملون على مجموعة البيانات. وهذا يجعل من الممكن تتبع أي أخطاء حدثت، مما يسهل الرجوع إليها وتصحيحها.

من المهم أيضًا تذكر أن تحليل الأرقام ليس هو الخطوة الوحيدة في تفسير البيانات. غالبًا ما يتبقى المزيد للبحث فيه، من أجل الإجابة عن سبب كون الوضع على ما هو عليه. تعد هذه العملية أمرًا بالغ الأهمية في تحديد الاستنتاجات التي يمكن استخلاصها من بحثك. وسوف نتناول هذا الموضوع في المذكرة الخامسة من هذه الوحدة.