الفصـــل الرابـــع*

مقاييس بيانات الفترات — INTERVAL SCALES
التوزيعات التكرارية — FREQUENCY DISTRIBUTIONS
والتمثيل البياني — AND GRAPHIC PRESENTATIONS

سوف ينصب اهتمامنا في هذا الفصل على طرق تلخيص البيانات الشبيهة بتلك التى تمّت مناقشتها في الفصل السابق . كما سنقوم بتصنيف البيانات الفاصلة إلى مجموعات Groups حيث نرتّب هذه المجموعات ونستخدمها لتبيان صورة عامة عن توزيع الحالات المدروسة Cases . وبهذه الكيفية يمكننا دمج معلومات أو بيانات هائلة عن عدد كبير من الحالات المدروسة لنحصل على صورة مبسّطة لها فتُمكّن القارئ من تصوّر الكيفية التى بها وُزّعت هذه الحالات .

وسنعرف لاحقاً أنّه بتجميع هذه البيانات بالشكل المشار إليه سنتمكن أيضاً – وإلى حد معقول – من تبسيط إجراء عمليات حسابية بعينها . وفي الفصلين اللاحقين سنهتمّ بطرق تلخيص البيانات بأوجه أكثر تحديداً تمكننا من وصفها بعدة أرقام تمثّل مقاييس للتطابق ودرجة التجانس measures of typicality and degree of homogeneity .

---

* في بعض الترجمات يطلق عليها البيانات « ذات البعد الثابت » أو بيانات «الفترة» .

٤ – ١ التوزيعات التكرارية : تجميع البيانات ٠

قلما نكون قد تعرضنا لاتخاذ قرارات حتمية فيما يتعلّق بتلخيص البيانات في الفصل السابق – هذا إن كان قد حدث شئ من هذا القبيل البتة ٠ ولعلّ ذلك يعزي لحقيقة أنّ الفئات قد تحددت سلفاً وكلّ ما بقي هو أن نرصد عدد الحالات المقابلة لكلّ فئة Class ثم ننسب كلاً منها على حدة إلى مجموع الحالات في العينة عن طريق حساب التناسب ، والنسب المئوية، أو النسبة ٠ ولو قُدِّر لبيانات المتغيّرات الفاصلة أن تلخّص بطريقة متشابهة فإنّه يتعيّن على أية حال أن يُتَّخَذ في البدْء قرار ما يتعلّق بالمجموعات أو الفئات التى ستستخدم ٠ ولأنّ البيانات تكون عادة موزّعة على الطريقة المستمرة The Continuous fashion بدون فوارق تقريبا بين الدرجات المتقاربة ، فلربما يكون مشروع التقسيم إلى فئات Classification Scheme جزافياً إلى حد ما ٠ إذن يكون من الضرورة بمكان ، والحال هذه ، أن نقرر كم عدد الفئات التى نستخدمها ؟ وكيف نحدد طول كلّ فئة من تلك الفئات ؟، وأين نؤسس الحدود الفاصلة بين هذه الفئات ؟ ولسوء الحظ فإنّه لا توجد أسس مبسطة لإجراء تلك العملية (عملية التصنيف إلى فئات) ، لأن القرار في النهاية يعتمد على الغرض الذى سوف تخدمه هذه التصنيفات أو تلك ٠ دعنا نأخذ مثالاً محدداً للتأكيد على طبيعة هذه المشكلة : افترض أن الأرقام الموضّحة أدناه تمثّل نسب الأشخاص المؤهلين للتصويت وهم يدلون بأصواتهم في انتخابات مجلس مدرسة يعطى ٩٣ دائرة تعدادية في مدينة معينة ، كما هو موضح في البيانات أدناه ٠

٣٩,٢	١١,٦	٣٦,٣	٢٦,٣	٣٧,١	١٥,٣	٢٧,٣	٢٣,٥	١٣,٣
٢٨,١	٢٦,٣	٢٧,١	٣٥,١	٢٣,٠	٢٦,١	٣١,٠	٣٦,٣	٢٧,٣
٢٢,٨	٣٣,٤	٢٥,٦	٢١,٦	٤٦,٨	٧,١	١٦,٨	٢٦,٩	٤٦,٦
٤٤,٣	٥٨,١	٣٣,١	١٣,٤	٢٧,٨	٣٣,٤	٢٢,١	٤٢,٧	٣٣,٠
٣٦,٣	٢٠,٧	٩,٣	٢٦,٣	٢٩,٩	٣٩,٤	٥,٣	٢٤,٣	١٧,٨
١٨,٢	٣٧,١	٢١,٦	١٧,٥	١٢,٣	٢٣,٦	٣٧,٢	٣٧,١	٢٥,١
٢٧,١	٢٨,٨	٢٧,٨	٣٣,٦	٢٦,٥	٢٨,٣	٢٦,٩	٢٤,٨	٤١,٠
٣٣,٦	١٩,٣	٤٣,٧	٢٨,٢	١٩,٩	٨٣,٦	٤٧,١	٤,٨	٩,٧
٣٩,٥	٣٢,٣	٢٢,٤	١٥,١	٢٦,٣	٢٦,١	٢٩,٢	١٤,٣	١٤,٦
٢١,٦	٣٧,٩	٣٧,١	٢٤,٩	١٠,٠	٢٠,٧	١١,٨	٢٢,٩	٣٦,٠
٤٦,١	٢١,٥	١٣,٣						

البيانات الخام الموضحة أعلاه تكاد تكون عديمة النفع في إعطاء القارئ صورة واضحة عمّا يحصل ، سيما ، إن كان عدد الحالات التى يشملها البحث كبيرًا جدًا . افترض الآن أنّنا نود مقارنة هذا المجتمع مع مجتمع آخر فيما يتعلّق بنتيجة التصويت . فبنظرة عابرة على البيانات يتّضح أنّ معظم الدوائر التعدادية تحصلّت على نتائج للتصويت تتراوح بين ٢٠ -٤٠٪ ، إلا أنّ هناك دائرة تعدادية واحدة حظيت بنسبة تصويت عالية جدًا (٨٣٫٦٪) . ولكن في الحقيقة نجد أنّه من الصعب على المرء أن ينعم بصورةذهنية واضحة للتوزيع الكلى Total distribution .

حدود الفئات وأطوالها : Number and Size of Intervals

لكى يتضح لنا هذا التوزيع الكلى بجلاء نجد أنّه من المفيد أن نقوم

بتصنيف الدرجات [أو النسب] المتقاربة أو المتشابهة في مجموعة تصنيفية أو فئة واحدة تضمها . وهنا تجابهنا مشكلة على الفور وهى : كم فئة يجب أن نستخدم لتجميع هذه البيانات ؟ وكم يجب أن تكون أطوال هذه الفئات ؟ بالتأكيد ليس هناك ما يدعو لاتخاذ فئات بأطوال غريبة أو شاذة . وهكذا فإنّه يمكننا مثلاً أن نختار فئات بالأطوال ٥ أو ١٠ أو ٢٠ باحتمال أكبر من احتمال اختيارنا لفئات بطول ١٦ر٤ مثلاً . كذلك فإنّه من الطبيعى أن تحتوى نهايات الفئات Class Limits / end points على أعداد صحيحة غير كسرية مثل ٥ ، ١٠ ... إلخ . وحالما نشعر بأننا لسنا متيقنين أيّ نوع من أنواع الفئات نستخدم – الطويلة نسبياً أم القصيرة نسبياً ، فإنها لفكرة جيدة أن نُصَنِّف الدرجات باستخدام فئات عديدة بأطوال قصيرة نسبياً . والسبب واضح : فباستخدام الفئات ذات الأطوال القصيرة فإنّه يمكننا عندما نحتاج مستقبلاً ، أن ندمج التكرارات [الحالات المدروسة] في فئات أطول . ولكن عندما نبدأ بفئات قليلة ومتضخمة [طويلة المدى] فإنّه لا يمكننا أن نعيد تقسيمها مالم نقم بالعملية الحسابية برمتها من جديد . ولهذا السبب ربما نرى تصنيف البيانات إلى فئات طول الواحدة منها ٥ ٪ مثلاً كمااتُّبع في الجدول (٤ –١) أدناه .

جدول رقم (٤ -١)

التوزيع التكراري للبيانات مصنفة إلى فئات بطول ٥ ٪

الفئـــة	التكرار	الفئـــة	التكرار
٠ر٠ - ٩ر٤	١	٠ر٤٥ - ٩ر٤٩	٤
٠ر٥ - ٩ر٩	٤	٠ر٥٠ - ٩ر٥٤	٠
٠ر١٠ - ٩ر١٤	٩	٠ر٥٥ - ٩ر٥٩	١
٠ر١٥ - ٩ر١٩	٨	٠ر٦٠ - ٩ر٦٤	٠
٠ر٢٠ - ٩ر٢٤	١٦	٠ر٦٥ - ٩ر٦٩	٠
٠ر٢٥ - ٩ر٢٩	٢٣	٠ر٧٠ - ٩ر٧٤	٠
٠ر٣٠ - ٩ر٣٤	٨	٠ر٧٥ - ٩ر٧٩	٠
٠ر٣٥ - ٩ر٣٩	١٤	٠ر٨٠ - ٩ر٨٤	١
٠ر٤٠ - ٩ر٤٤	٤		٩٣

بالنظر إلى التكرارات المقابلة لكل فئة يتضّح لنا أنّها - أى التكرارات - لا تبدو متّسقة أو متناغمة . وربما يعزى الإختلاف أو التباين فى التكرارات المقابلة للفئات المتجاورة ، للتذبذبات نتيجة للصدفة . فإن كان عدد الحالات (حجم العينة) أكبر مما هو عليه فإنّنا قطعاً سنتوقع توزيعاً أكثر تناغماً وانسجامًا . وسوف يزداد هذا المعنى وضوحاً فى الفصول اللاحقة ، بيد أنّه يكفى هنا أن نورد بأنّه وجد عمليًا ، أن هذا هو الذى يحدث دائمًا وبهذه

الكيفية . وباعتبار الدوائر التعدادية الثلاث وتسعون [٩٣] فإنّ أحسن ما يمكن عمله ، على أية حال ، للحصول على توزيع أكثر إتساقًا هو أن نختار عددًا أقل من الفئات بأطوال أكبر . وباستخدام الفئات بطول ١٠ نتحصلُ على التوزيع المبين في الجدول رقم (٤ - ٢) أدناه .

جدول رقم (٤ -٢)

التوزيع التكراري للبيانات مصنفة بطول ١٠٪

الفئـــــة	التكرار
٠ر٠ - ٩ر٩	٥
٠ر١٠ - ٩ر١٩	١٧
٠ر٢٠ - ٩ر٢٩	٣٩
٠ر٣٠ - ٩ر٣٩	٢٢
٠ر٤٠ - ٩ر٤٩	٨
٠ر٥٠ - ٩ر٥٩	١
٠ر٦٠ - ٩ر٦٩	٠
٠ر٧٠ - ٩ر٧٩	٠
٠ر٨٠ - ٩ر٨٩	١
	٩٣

ولو استخدمنا فئات بأطوال أكبر ، مثلاً بطول ٢٠٪ ستكون الصورة للتوزيع كما هى موضحة بالجدول رقم (٤ -٣) أدناه .

جدول رقم (٤ - ٣)

التوزيع التكرارى للبيانات مصنفة بطول ٢٠٪

الفئـــــة	التكرار
٠ر٠ - ٩ر١٩	٢٢
٠ر٢٠ - ٩ر٣٩	٦١
٠ر٤٠ - ٩ر٥٩	٩
٠ر٦٠ - ٩ر٧٩	٠
٠ر٨٠ - ٩ر٩٩	١
	٩٣

وواضح هنا أنّنا بدأنا نفقد معظم بياناتنا الأصلية . وهكذا فإنّنا نعرف فقط أن حوالى ثلثى الحالات [٦١ حالة] تقع بين ٢٠ و ٩ر٣٩ . ولكن عندما نتأمّل البيانات في هذا الوضع فإنّه لا يمكننا أن نقول الكثير عن أين تقع أغلبية الحالات التى تضمّها هذه الفئة المتضخّمة والطويلة جداً ؛ من هذا النقاش يتضّح أنّنا يجب أن نصل إلى حلٍّ وسط يوفقُ بين استخدام فئات عديدة جداً بحيث تبقى الصورة العامة للبيانات تفصيلية للغاية أو غير

متسقة ، وبين استخدام فئات قليلة جداً بحيث يصبح فقدان معلومات كثيرة هو المآل المحتوم .

وهكذا نجد أنّه بتلخيص البيانات الفاصلة لا مفرّ عملياً من فقدان بعض البيانات الهامة . وبالمقابل فإنّ تسجيل كل البيانات يعني عرضها بتفصيل مُخلٍ أكثر من كون عرضها بتفصيل تنويرى .

وبالرغم من أنّ بعض المعادلات الحسابية التى تقود إلى تقدير أمثل لعدد الفئات التى يمكن أن تستخدم متوفرة إلا أنّها دائما تعطى الإنطباع بالدقة في الإختيار رغم أنّ القرار الأمثل فيما يختص باختيار الفئات هو ذلك المبنى على إحساس عام وإلمام تام بالغرض الذى سيخدمه الجدول التكراري. وبغض النظر عن عدد الحالات المدروسة أو إتساق التوزيع التكرارى يكون من الحكمة أن نتبع الحكم البدهى بأنّ طول الفئة يجب ألا يتجاوز مقدار الفرق بين القيم التى يمكن تجاهلها بسهولة . فمثلاً فرق ٥ دولارات في إيجار منزل هو فرق لا يُعْتَدُّ به فى حين أنّه يُعد فرقاً كبيراً إنْ كان في أسعار القمصان . ولذلك فإن الفئة لابد لها أن تشمل حالات تعتبر قيمها الفعلية متماثلة لأسباب عملية .

وهناك مشكلة أخرى تفرض نفسها فى البيانات السابقة وهى : ماذا عن الدائرة التعدادية الوحيدة التى تشمل ٦ر٨٣٪ من الذين أدلوا بأصواتهم ؟ إنّ استخدام فئات بطول ١٠٪ تترك عدة فئات خالية تماماً من التكرارات (أنظر الجدول رقم ٤ -٢) وذلك فقط لأجل أن تشمل الفئات النسبة الوحيدة الشاذة [٦ر٨٣٪] من تلك البيانات . وبالطبع هذا هو الذى يجب أن يفعل إن قدّر لهذه البيانات أن تلخّص بدقة ؛ فهذه الدائرة التعدادية لا مثيل لها ، ومن جانب آخر فربما يكون من المرغوب فيه فى بعض الأحوال أن يُضغط

الجدول . فإن كانت النسب أعلاه تفوق المائة مثلاً وتتخللها عدة نسب متطرفة (عالية أو منخفضة) تتوزع على عشر [١٠] فئات أو أكثر فسنواجه باتخاذ قرار أكثر صعوبة . فهناك عدة بدائل تطرح نفسها .

أولاً : يمكن أن نستخدم فئات بأطوال مختلفة بحيث نسمح لبعض الفئات أن تكون أطول بكثير من بعضها الآخر كي تضم النسب المتطرفة . على ذلك ، فقد نستخدم مثلاً ، الفئة ٥٠ - ٩ر٨٩ والتى ستضم أكبر نسبتين [١ر٥٨ ، ٦ر٨٣] . وبالطبع فإننا وبهذه الطريقة نكون قد فقدنا بعض المعلومات مادمنا الآن نحصل على مؤشر أقل دقة لتكرارى التسجيلين الإستثنائيين .

ثانياً : يمكننا أن نتخذ فئة مفتوحة لتضم كل الحالات الطرفية extreme cases . وهكذا فإنّ الفئة الأخيرة ربما تقرأ ٥٠٪ فأكثر . وهنا نكون قد احتفظنا بمعلومات عن بيانات أقلّ بالرغم من أنّنا وفى هذا المثال بالذات نعلم أن النسب لا تتعدى المائة . وإن كانت البيانات عن الدخل ، مثلاً ، وكانت الفئة الأخيرة هى ٥٠ ألف دولار فأكثر ، فإنّ القارئ سوف لا يتمكن إطلاقاً من التخمين - باستخدام الجدول لوحده - ماذا تكون أعلى الدخول فى الواقع ؟ وتجدر الإشارة إلى أنّه قد لا يكون مهماً على الإطلاق فى بعض الأحوال معرفة هذه الدخول الأعلى . فى هذه الحالة فإن التبسيط المحبّب باستخدام الفئات المفتوحة ربّما يطغى على وجه القصور فى استخذام نفس هذه الفئات خاصة إن كانت التوزيعات تحتوى على عدد قليل من القيم المتطرفة جداً .

ولو أراد الشخص أن يحدّد دخول أثرى أثرياء المواطنين دون أن يُعَتِّم الجدول ، ربما يجد أنّ ذلك سيكون أسهل فقط فى العرض المنهجى . وكما

سنرى فى الفصول اللاحقة فإنّ الفئات المفتوحة ينبغى ألا تستخدم إذا كان الغرض الأساسى من تصنيف البيانات هو تبسيط المعاملات الحسابية أكثر منه عرضاً ذا معنى للبيانات ٠

النهايات الحقيقية : — True Limits

ربما تكون قد لاحظت أنّه وفى تحديد الفئات أنّ نهاياتها الحدِّية Class Limits قد عينّت بشكل لا يسمح بالتداخل بين هذه الفئات ٠ وفى الواقع فإنّ هناك فرقاً بسيطاً بين هذه الفئات ، فالنهايات الحدِّية تُبيَّن بهذه الطريقة لتفادى اللبس لدى القارئ ٠ فإن عُيِّنَت ، مثلاً ، كما يلى :

١٠ – ٢٠

٢٠ – ٣٠ ٠٠٠٠٠ إلخ

فسوف يبرز السؤال بمَاذا نفعل بدرجة مثل ٢٠ بالضبط ؟ وفى الحقيقية فإنّنا نقع دائماً في بعض اللّبس كيفما حددنا الفئات ٠ ويمكن أن يتضّح ذلك بسؤالنا عمّا يمكن عمله في حالة أنّ درجة إحدى الحالات تقع بين ١٩٫٩ وَ ٢٠ ٠ نلاحظ أنه بالطبع ليس هناك حالات لها مثل هذه الدرجات٠ لكن بتفكير بسيط يمكن أن نقتنع بأنّ هذا نتيجة لحقيقة أن البيانات قد أختصرت أو جُبرت لأقرب كسر عشرى مئوى ٠ لذلك يمكننا توجيه السؤال التالى : أىُّ الحالات تقع فى فئة معينة مع الأخذ فى الإعتبار أن البيانات قد جبرت (rounded) ؟ هكذا يتضّح أن نهايات الحدود الحقيقية للفئات ليست هى نفس النهايات المبينة ٠ ولو اتبعنا الطرق العادية لجبر الكسر إلى أقرب رقم صحيح فإنّ دائرة تعدادية بنسبة أعلى بقليل عن ١٩٫٩٥ يمكن أن تجبر إلى رقم صحيح هو ٢٠ ؛ وهكذا ستوضع في الفئة ٢٠ – ٢٩ ٠ فلو كانت النسبة أقلّ بقليل من ١٩٫٩٥ تَعيَّنَ علينا أن نجبر الدرجة إلى ١٩٫٩ ٠

وهكذا سنضع الدائرة التعدادية في التصنيف الأقل . ولهذا فإنّ النهايات الحقيقية المستعملة هى كالآتى : -

٠٫٠٥ - ٩٫٩٥

٩٫٩٥ - ١٩٫٩٥

١٩٫٩٥ - ٢٩٫٩٥

...... إلخ

وهكذا نرى أنّه وباستخدامنا النهايات الحدِّية الحقيقية فإنّ كل فئة تكون بالضبط ١٠ [وليس ٩٫٩] . وإنّ النهاية الحدية العليا للفئة الواحدة تتطابق مع النهاية الحدية الدنيا للفئة التى تليها . وإذا كانت هناك درجة تساوى بالضبط ٩٫٩٥ لا تّبعنا الطريقة التقليدية وجبرنا الكسر مادام الرقم الذى يلى الـ ٥ مباشرة هو أحادى وليس زوجياً . ولذلك يمكننا أن ندخل أى حالة من الحالات فى الفئة التى تناسبها دون لبس . لاحظ أنّه إذا كان الجبر لأقرب عدد صحيح كما هو الحال عموماً ، فالفئات الحقيقية دائماً تقبل تقسيم الفرق بين الحدود الدنيا والعليا المعنية لكل من الفئتين المتجاورتين . مثلاً : إذا أردنا تقسيم مجموع الدرجتين ١٩٫٩ وَ ٢٠ إلى نصفين متساويين فسنحصل على ١٩٫٩٥ . فالمعتاد أن نُسَجِّل الأرقام بالكيفية التى تسمح بتحديد درجة الدقة فى القياس . فمثلاً الرقم «١٠٫٤٥٠» حُدِّدَت درجة الدقة فيه بثلاث خانات للكسور العشرية والرقم «١٠٫٤» حددت درجة الدقة فيه بخانة واحدة للكسر العشرى .

لذلك فإنّ الدقة فى القياس يجب أن تسجّل في الفئات المعنية ليتأكد القارئ من الفئات الحقيقية إذا كان يود أن يستفيد من هذه النهايات (العليا والدنيا) في إجراء العمليات الحسابية . فمثلاً إذا عُيِّنت النهايتان العليا

والدنيا لفئة ما بـ ٠٠ر١٠ إلى ٩٩ر١٩ نعرف أن القياس دقيق إلى غاية خانتين عشريتين ٠ وإن كان الجبر إلى أقرب $\frac{١}{١٠٠}$ من ١٪ فإنّه بذلك تكون النهايتان العليا والدنيا الحقيقيتان هما ٩٩٥ر٩ إلى ٩٩٥ر١٩ ٠ أما إذا حدّدت النهايات مثلاً ، بـ ١٠ إلى ١٩ ، فالنهايتان الحقيقيتان ستكونان بالطبع ٥ر٩ إلى ٥ر١٩ ٠

وفي حالات قليلة مثل العمر في آخر عيد ميلاد ، ربما لا تجبر البيانات بالطريقة التقليدية المعتادة ٠ وإذا ماسألنا أنفسنا إلى أية فئة – في الحقيقة – تنتمي أيٍّ من الحالات المرصودة ؟ فإنّ الإجابة ، على أية حال ، يجب أن تكون واضحة ٠ فبما أن الشخص الذى يكون عمره ٢٠ سنة غداً سيقول بأن عمره ١٩ سنة اليوم ، فمن الواضح أن الفئة التى تقرأ « ١٥ إلى ١٩» يكون لها النهايات الحقيقية بالقيم ١٥ وَ ٢٠ سنة ٠ وحتى إن بدا لنا أنّنا كمن يسَرِّح في شعره ونحن نحاول التمييز بين النهايات العليا والدنيا الحقيقية والمعينة للفئات ، سيتأكد لنا من الفصول اللاحقة أن النهايات العليا والدنيا الحقيقية للفئات لابد أن تستخدم لأغراض العمليات الحسابية ، حتى وإن كانت لا تحـــدد بوضـــوح عندما تعرض البيانات في شكل توزيع تكرارى ٠

البيانات المنقطعة

والبيانات المستمرة Discrete and Continuous Data

البيانات التى كنا نستخدمها في الأمثلة السابقة هي بيانات مستمرة Continuous بمعنى أنّ أية قيمة يمكن أن يتحصّل عليها كنسبة معينة [على الأقل نظرياً] بافتراض أن الدقة في القياس محددة تحديداً متناهياً وأن الدوائر التعدادية كثيرة جداً ٠ هكذا فإنّ القيمة ٤٥٣١ر١٧ ٪ ممكنة كما أنّ

القيمة ١٧٫٠٠٠٪ أيضًا ممكنة . هنالك بيانات أخرى نطلق عليها «بيانات منقطعة Discrete » لأنّه ليست كلّ القيم ممكنة . فالمرأة مثلاً يمكن أن يكون لديها صفر ، ١ ، ٢ أو حتى ١٧ طفلاً ولكن لا يمكن أن يكون لديها ٢٫٣١ طفلاً . والدخل والحجم السكانى لمدينة ما لا يمكن إلا أن يكونا متغيرين ببيانات منقطعة - بمعنى أنه لا يمكن أن يكسب الشخص مثلاً ٣٢١٩٫٥٦١٨ دولاراً أو أن يكون حجم المدينة ٤٣٦٣٥٫٧ شخصاً . ونسبة لمحدودية أية أداة للقياس والضرورة التى تتبع ذلك وتملى جبر الكسور فى نقطة ما ، تأتى البيانات دائماً في شكل بيانات منقطعة لكن في معظم الأحوال نكون على أسوأ الفروض قادرين على تخيل توزيع مستمر يمكن أن يتحصّل عليه بأداة قياس مُثلى . وكما سنرى فى الفصل الخاص بالمنحنى المتماثل (أو المنحنى الطبيعى) (المركزي) فإنّ علماء الرياضيات يُجْبَرون دائماً على تطوير توزيعات نظرية تفترض متغيرات مستمرة . في بعض الحالات ، مثل الدخل وحجم المدينة السكاني ، ليس صعباً جدًا أن تتخيل بيانات مستمرة حتى وإن يكن في الحقيقة هناك وحدات صغيرة جداً (مثل أقل وحدة من أى عملة وأيضاً الأشخاص) لا يمكن تجزئتها .

ولكن ماذا عن عدد الأطفال في أسرة ما ؟ هنا سيخيل إلينا أنّنا نكون غير عادلين على الإطلاق إن أقدمنا على اعتبار مثل تلك البيانات على أنّها بيانات مستمرة ، وتعرض البيانات في هيئة توزيع تكرارى ، وسوف لا نكون بالطبع غير عقلانيين بدرجة تحثنا على استخدام فئات تقرأ ٠٫٥ إلى ٢٫٤ أو ٢٫٥ إلى ٤٫٤ من الأطفال مثلاً ! بل وببساطة شديدة نُعَيِّن فئات مثل صفر إلى ٢ ؛ و ٣ إلى ٤ .. إلخ . وسوف لا يكون هناك أىّ لُبْس فيما يختص بالمسافات بين الفئات . لإجراء بعض العمليات الحسابية ، يكون من الضروري ، ولأسباب عملية ، معاملة البيانات كأنّها مستمرة وتنشر

الدرجات المنقطعة على فئات صغيرة – أى ، قصيرة الأطوال ، أو ضيِّقة . ربما يبدو غريباً في الظاهر حينما نود أن نتأمَّل كيف أن عدد الأمهات (اللائى لم يكن للواحدة منهُنَّ إلا طفل واحد) يتراوح طبقًا لعدد الأطفال من ٠ر٥ إلى ١ر٥ . ولمعظم الأغراض سوف نتحصَّل بالضرورة على نفس النتائج التى يمكن الحصول عليها بالإبقاء على البيانات في شكلها المنقطع. ربما كان ضرورياً أن نتجه نحو حلٍّ توفيقى هو أقرب إلى الواقع في هذه الحالة ، وفي حالات أخرى مماثلة لتوفيق النماذج المعدة بواسطة الرياضيين، وبافتراض أننا نعى تماماً ما نحن فاعلون يمكن ألا يحدث سوى قليل جداً من الخلط أو لا يحدث البتة .

٤ – ٢ توزيعات التكرار المتجمع Cumulative Frequency Distributions

لخدمة بعض الأهداف أو الأغراض قد يُرْغَبُ فى عرض البيانات بطريقة مختلفة إلى حد ما . فبدلاً من إعطاء عدد الحالات فى كلّ فئة على حدة ، ربما نرغب فى تحديد عدد الحالات أو الدرجات التى تقل عن قيمة معينة أو تفوقها . فى حالة البيانات التى استخدمناها نجد من الواضح أنَّه لا توجد أى دوائر تعدادية بدون أية نسبة معينة ؛ وتوجد خمس دوائر تعدادية بأقل من ٩٫٩٥٪ ، ٢٢ بأقل من ١٩٫٩٥٪ ؛ وكل الدوائر الـ ٩٣ تكون لديها تسجيلات أقل من ٨٩٫٩٥٪ . وبهذه الطريقة يمكننا أن نعرض البيانات في شكل تجمعى كما هو موضحّ فى الجدول رقم (٤ –٤) أدناه .

جدول رقم (٤ -٤)

التوزيع التكراري المتجمع الصاعد والنازل

المتجمع الصاعد			المتجمع النازل		
عدد التكرارات التى لها نسب أقل من :	تكرار المتجمع الصاعد F	النسبة المئوية ٪	التكرارات التى لها نسب أعلى من :	تكرار المتجمع النازل F	النسبة المئوية ٪
٠,٠	صفر	٠,٠	٠,٠	٩٣	١٠٠,٠
٩,٩٥	٥	٥,٤	٩,٩٥	٨٨	٩٤,٠
١٩,٩٥	٢٢	٢٣,٧	١٩,٩٥	٧١	٧٦,٣
٢٩,٩٥	٦١	٦٥,٦	٢٩,٩٥	٣٢	٣٤,٤
٣٩,٩٥	٨٣	٨٩,٢	٣٩,٩٥	١٠	١٠,٨
٤٩,٩٥	٩١	٩٧,٨	٤٩,٩٥	٢	٢,٢
٥٩,٩٥	٩٢	٩٨,٩	٥٩,٩٥	١	١,١
٦٩,٩٥	٩٢	٩٨,٩	٦٩,٩٥	١	١,١
٧٩,٩٥	٩٢	٩٨,٩	٧٩,٩٥	١	١,١
٨٩,٩٥	٩٣	١٠٠,٠	٨٩,٩٥	صفر	٠,٠

لاحظ أنّ باستطاعتنا الجمع تصاعديًا أو تنازليًا بالسؤال عن كم حالة من الحالات يكون لِديها قيم أعلى في مجموعها من قيمة معينة . والتكرارات

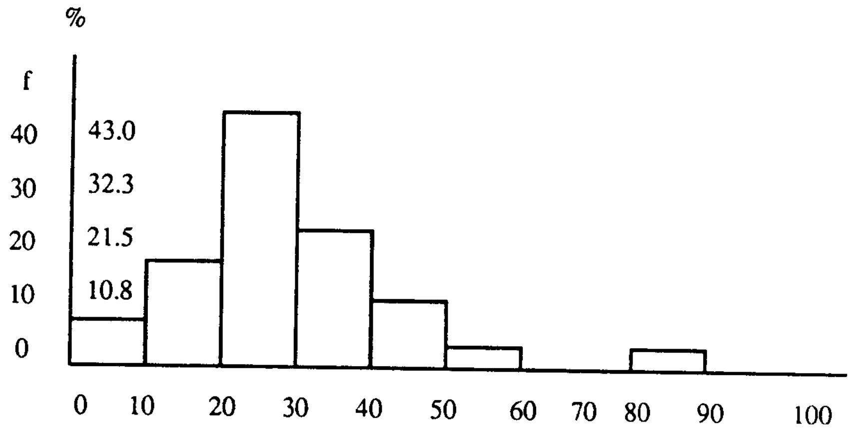
المتجمعة يرمز لها بالحرف الإنجليزي "F" الضخم بدلاً من صورة الحرف الرفيع "f" . ويمكننا إذا أردنا أن نُحيل التكرارات الملاحظة observed frequencies إلى نسب مئوية . وسوف نتعرض إلى بعض الأمثلة التى توضّح كيف يمكننا الإستفادة من التوزيعات المتجمعة وذلك في الفصل الخامس عندما نودّ حساب الوسيط وأيضًا في الفصل الرابع عشر بالجزء الثانى من الكتاب .

٤ – ٣ العرض البيانى للبيانات : المُدَرَّجات والمُضلّعات والمنْحنَيات التكرارية : –

نجد على الدوام أن بعضنا يتردّد عند قراءة الجدول فيما يكون الآخر أكثر استعدادًا لتفهّم مضامين هذه الجداول لو أنّ بياناتها عرضت في شكل رسم بيانى أو أية صورة إيضاحية أخرى مناسبة . إنّ واحدة من أفْيَد وأسهل الطرق لعرض البيانات بحيث تبرز الفروقات فى التكرارات بصورة واضحة هى إستخدام أشكال ذات مساحات أو ارتفاعات منسوبة للتكرارات في كلّ فئة من الفئات المرصودة . فمثلاً يمكننا أن نستخدم «أعمدة» يُمَثّل كلّ عمود منها تكرارات فئة معينة ، وإرتفاع العمود يشير إلى الحجم النسبى لتكرارات الفئة . فإن كان مقياس البيانات مقياسًا إسميًا nominal scale فإنّ الترتيب الفعلى للأعمدة يكون غير وارد . أما بالنسبة للمقياسين، الرُّتَبِي ordinal scale ومقياس البيانات الفاصلة interval scale فيمكن أن تنظم الأعمدة برُتَبِها الحقيقية لتعطى مؤشرًا مرئياً جيدًا للتوزيع التكرارى ، والشكل الناتج عن رسم تلك الأعمدة يسمى « المدرج التكراري» histogram. وفى المقياس العمودى لهذا الشكل توضع إما التكرارات المطلقة وإما التكرارات النسبية في كلّ فئة كما هو موضّح في الشكل رقم (٤–١) أدناه :

الشكل رقم (٤ -١)

مدرج تكراري بفئات متساوية

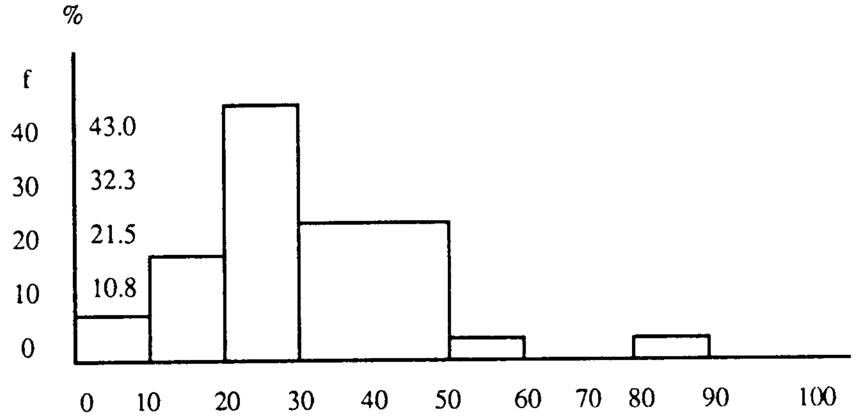


ولابد من الإشارة إلى أنّه في حالة كون إرتفاعات الأعمدة في المدرج متناسبة مع عدد التكرارات في كلّ فئة فإنّ الصورة المرئية يمكن أن تكون مضللة ما لم تكن كل الفئات مغلقة وبأطوال متساوية . لنفترض ، مثلاً ، أن إحدى الفئات الوسطية بطول ٢٠ بدلاً من ١٠ . هنا سنجابه بحالات أو درجات أكثر في كلّ فئة عمّا كان عليه عددها سابقًا ، وسيكون المدرج الجديد كما هو موضّح في الشكل (٤ -٢) أدناه واضحًا تمامًا .

الشكل رقم (٤ -٢)

« مدرج تكرارى بفئات غير متساوية

وارتفاعات تتناسب مع التكرارات»

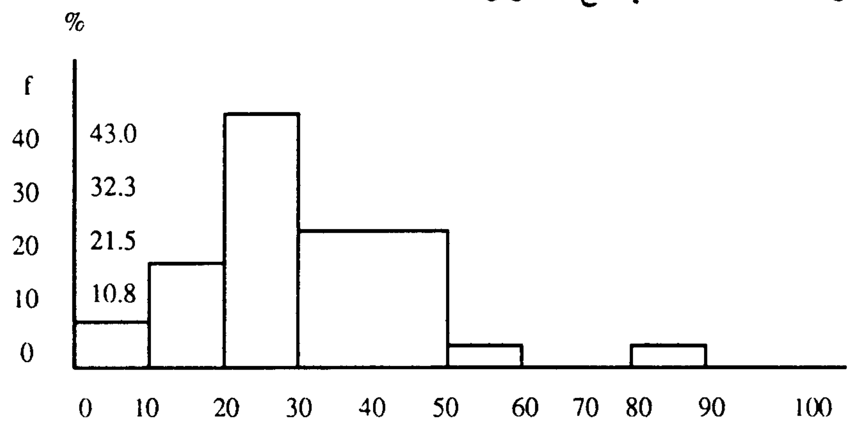


إننا لو أردنا أن نرسم مدرجًا يمثّل البيانات بصورة أكثر صدقًا ، يجب علينا أن نجعل العمود الأخير بطول يساوى - في الحقيقة - نصف طوله المبيين ، حيث أننا ضاعفنا طول الفئة وبذلك نكون - فى المتوسط - قد أدخلنا في هذه الفئة ضعف الدرجات التى يمكن أن تضمّها كلا الفئتين المكونتين لهذه الفئة الطويلة في حالة عدم دمجها في فئة واحدة بالطول المشار إليه . وهذا المعنى سيعطينا مدرجًا كالذى يعكسه الشكل (٤ -٣) أدناه ، ويكون هذا المدرج أكثر شبهًا بالمدرج الأصلى الأول (شكل رقم٤-١) .

الشكل رقم (٤ -٣)

مدرج تكرارى بفئات غير متساوية

ومساحات تتناسب مع التكرارات



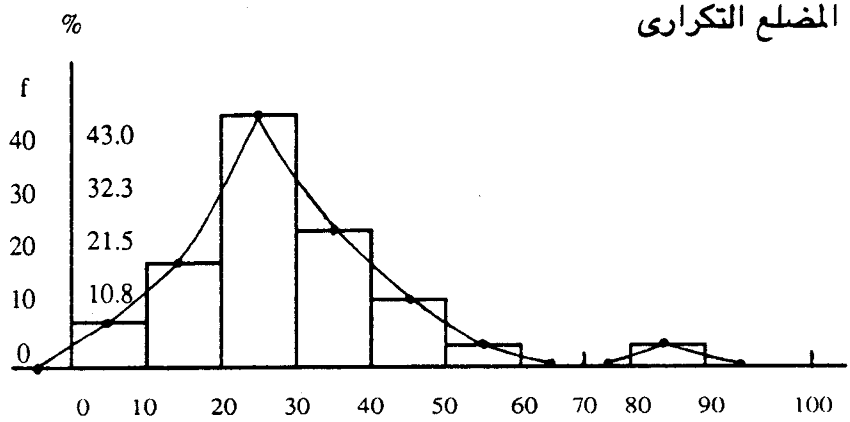
وبقليل من التفكير يمكنك أن تقتنع بأنّه إن كنا نتأمل المساحات أكثر من تأملنا للإرتفاعات فإننا سوف نكون أكثر إستعدادًا لمعالجة البيانات التى تحويها فئات ليست ذات أطوال متساوية ٠ وبمعنى آخر ، نجعل مساحات المستطيلات تكون متناسبة مع عدد الحالات ٠ وبالطبع فإنّه فى الحالة الخاصة والمهمّة عندما تكون الفئات بأطوال متساوية فإنّ إرتفاعات الأعمدة أيضًا تكون بنسبة التكرارات فى الفئة المعيّنة من مجموع التكرارات فى كلّ الفئات ٠ ولو أخذنا عرض كلّ واحدٍ من المستطيلات كوحدة واحدة ومثّلت إرتفاعاتها كنسب ، فإن مجموع المساحة تحت المدرج تكون وُحْدَة واحدة unity ٠ وهكذا فإن :

$$١ \times (\frac{٥}{٩٣}) + ١ \times (\frac{١٧}{٩٣}) + ١ (\frac{٣٩}{٩٣}) + ... ١٠٠ \times (\frac{١}{٩٣}) = ١$$

وعندما نتطرق للمنحنى الطبيعى Normal Curve في الفصل السابع سنجد أنَّنا بالضرورة نتحدّث عن المساحات بدلاً من الإرتفاعات ، ويكون من الأسهل أن نأخذ المساحة تحت المدرج كوحدة واحدة unity . وهناك طريقة أخرى ممثالة لعرض التوزيع التكرارى بيانياً وذلك باستخدام المضلع التكراري Frequency polygon . وللحصول على المضلع التكرارى فإنَّنا وبكل بساطة نعين نقاط الوسط فى قمة إرتفاع كل مستطيل من المستطيلات ثم نصل هذه النقاط كما هو موضح فى الشكل رقم (٤ -٤) أدناه .

الشكل رقم (٤ -٤)

المضلع التكرارى



لاحظ أن نقطتى نهايتى قاعدة المضلع التكرارى (اليسرى واليمنى) قد وضعتا على المحور الأفقى عند منتصف كلا الفئتين الأولى والأخيرة . ونحن عادة لا نستفيد من كلا الرسمين الإيضاحيين (المدرج والمضلع) متفرقين كُلاً منهما على حدة ، ولكن بإدخال المضلع التكرارى فى المدرج التكرارى يتبيّن لنا أن المساحة تحت كلا الرسمين لابد أن تكون مثلاً بِمثيل . ذلك أن لكل مثلّث صغير (الشكل ٤ -٤) يقع داخل المضلع التكرارى وخارج المدرج

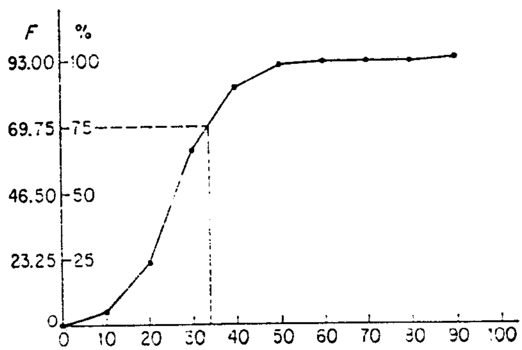
التكرارى في نفس الوقت ، هناك مثلث مثله تماماً يقع تحت المدرج التكرارى وأعلى المضلع التكرارى في نفس الوقت . وهكذا أيضاً يمكننا أن نعتبر المساحة تحت المضلع التكرارى تساوي وحدة . ولاحظ أنّنا لم نفعل أكثر من أنّنا قد أوصلنا عدداً من النقاط بخطوط مستقيمة . وكما أنّ النقاط نفسها يمكن أن تمثّل عدد الحالات (الدرجات) فى كلّ فئة ولكن يجب أن نكون حريصين على ألا نستخلص من ذلك أن هناك عدداً معيناً من الحالات عند كلّ نقطة من نقاط تقاطع خطوط المضلّع التكرارى مع أعمدة المدرج التكرارى . فمثلاً لا يجب أن نستخلص من تقاطع خط المضلع « الثالث من الجانب الأيسر في الشكل (٤ - ٤) » مع خط المدرج « الجانب الأيسر من العمود الثالث» أن التقاطع يشير إلى أنّ هناك حوالى ٢٨ حالة لها بالضبط ٤٠٪ من الدرجات .

ويمكن أيضًا استخدام المضلعات التكرارية في تمثيل التوزيعات التكرارية المتجمّعة ، والشكل الناتج عن ذلك يسمّى بالمنحنى التكرارى المتجمع ogive . على محور الصادات (المحور الرأسى) يمكننا أن نعيّن التكرارات أو النسب ، ونضع درجات متغيّر البيانات الفاصلة على طول محور السِّينات (المحور الأفقى) كما كان سابقاً ، مع الأخذ فى الإعتبار أنّ التكرارات الممثّلة تشير إلى عدد الحالات التى هى أقل من القيمة المعيّنة على محور السينات . ومثالاً لذلك نرى أنّه فى الشكل (٤ -١) توجد تقريبًا ٧٥٪ من الدرجات أقلّ من ٣٤ . لذلك يمكن استخدام المنحنيات التكرارية المتجمّعة كطريقة بيانية لتحديد عدد الحالات التى لها قيم أعلى من قيمة معينة أو أقل منها . ومن الواضح أن شكل المنحنى التكرارى المتجمع (منحنى التكرارات المتجمعة) عادة ، إما أن يكون متزايدًا (مرتفعًا) باستمرار أو متناقصاً (نازلاً) باستمرار ، معتمدًا على كيفية تجميع

التكرارات من أعلى إلى أسفل أو من أسفل إلى أعلى . ويكون منحنى التكرار المتجمع أفقياً عند الفئات الخالية من التكرارات .وإن كان التوزيع التكرارى من النوع الذى تعكسه بياناتنا هذه - حيث أن أكبر عدد من الحالات يتمركز فى الفئات التى تُجاور مركز التوزيع - فإنّ منحنى التكرارات المتجمّعة ogive سيكون في شكل الحرف الإنجليزي "S" بحيث أنّ أشد انحدار فيه يكون بالقرب من الفئات التى تحتوى على أكبر عدد من الحالات .

الشكل رقم (٤ -٥)

«منحنى يمثل توزيعاً تكرارياً متجمعاً »



التماريــــن :

١ - إفترض أنّ البيانات أدناه تمثّل الدخول السنوية (بالدولار الأمريكى) لعينة من سكان مجتمع ما :

٩٨٦٠	١٨٧٤٠	١٠٣٤٠	١٠٨٥٠	٩٧٦٠
١١٣٥٠	١٨٧٤٠	١١٣٥٠	١١٣٦٠	١١٤٣٠
١٤٣١٠	١٠٥٦٠	٩٦١٠	٩١٤٠	١٢٢١٠
١٠٥٥٠	٩٧٤٠	١٥١٩٠	١٠٣٣٠	١٠٤١٠
١١٢١٠	١٤١١٠	١١٢٥٠	١٤٨١٠	١١٥٧٠
١٢٤٩٠	١٧٣٠٠	١٠٤٦٠	١٢٣٤٠	١٦٣٠٠
٩١١٠	١١٤٤٠	٢٦٣٠	٩٩٧٠	١٠٣٢٠
٣٠٤٠٠	١٠٣٧٠	٩٦٧٠	١١١٤٠	٨٧٩٠
١٠٧٦٠	١٢١٧٠	١٠١٠٠	١٠٠٠٠	١١٥٦٠
١١١٧٠	١٠١٦٠	١٢١٣٠	٨٦١٠	١٠٨٠٠
١٣١٧٠	٩٨٠٠	٨٧١٠	١١٥٧٠	٢٠٤٦٠
٩٣٥٠	١٠١٨٠	١١٣٢٠	٨٩٤٠	١٢٢١٠
١٥٣٤٠	١١٢٤٠	١٦٨٣٠	٩٧٨٠	٩٦٩٠

أ - أنشئ التوزيع التكرارى والتوزيع التكرارى المتجمع .

ب - ماهى النهايات الحقيقية للفئات .

جـ - ارسم المدرج والمضلّع والمنحنى التكرارى .

٢ - في مسح لأنماط الزيارات وسط الأصدقاء الحميمين والأقارب ، طلب من ٨١ مستجيبا أن يسجلوا عدد هؤلاء الأصدقاء والأقارب والذين يتبادلون معهم الزيارات على الأقل مرة واحدة خلال الشهر الواحد . وكانت النتائج كما يلى (الأرقام تشير إلى العدد الفعلى للأشخاص الذين تمّت زيارتهم) :

٣	٥	٢	٣	٣	٤	١	٨	٤
٢	٤	٢	٥	٣	٣	٣	صفر	٣
٥	٦	٤	٣	٢	٢	٦	٣	٥
٤	١٤	٣	٥	٦	٣	٤	٢	٤
٩	٤	١	٤	٢	٤	٣	٥	صفر
٤	٣	٥	٧	٣	٥	٦	٢	٢
٥	٤	٢	٣	٦	١	٣	١٦	٥
٣	١١	٤	٥	١٩	٤	٥	٢	٢
٤	٣	١٤	٥	٢	١	٤	٣	٤

أ - احسب التوزيع التكرارى والتكرارى المتجمع .

ب - بقدر ما تستطيع دعِّم بالحجة اختيارك للفئات بالكيفية التى اخترتها .

ج - ارسم المدرج ، المضلّع ، والمنحنى التكرارى .

٣ - عيّن النهايات الحقيقية (أو الفعلية) للفئات التالية : -

أ - ١٠٠٠ - ١٩٠٠

٢٠٠٠ - ٢٩٠٠

ب - ١٠٠٠ - ١٩٩٩

٢٠٠٠ - ٢٩٩٩

جـ - ١ر٠٠٠ - ١ر٩٩٩ (الإجابة ٠ر٩٩٩٥ - ١ر٩٩٩٥)

٢٫٠٠٠ - ١٫٩٩٩

د - ١٠٫٠٠ - ٠٫١٩ (الإجابة ٠٫٠٩٥ - ٠٫١٩٥)

٠٫٢٩ - ٠٫٢٠

ما الذى افترضته في طريقة الجبر فى كلّ حالة من الحالات أعلاه ؟