جامعة البعث

المحاضرة الثالثة

الاحصاء الحيوي

كلية التمريض

المعهد التقاني لطب الطوارئ

الدكتور ياسر العمر

المقاييس الرقمية (١)
NUMERICAL MEASURES

تشمل قياسات الإحصاء الوصفي جمع البيانات وتبويبها وتلخيصها ومن ثم عرضها . وهذا يعني أن هذا النوع من الإحصاء يهتم بوصف البيانات الإحصائية بدون الدخول في عملية التحليل الإحصائي للوصول إلى الاستنتاجات العلمية وتعتبر أهمَ وسائل الإحصاء الوصفي هي مقاييسُ الميول أو النزعة المركزية Measures of Central Tendency ومقاييس التشتت Measures of Dispersion .فعندما ننظر في توزع قيم بيانات إحصائية لتعداد الكريات البيضاء عند عينة عشوائية لعدة أشخاص في منطقة من المناطق سوف نجد أن هذه القيم تتوزع حول قيمة متوسطة /مركزية/ إذ يكثر عددها حول هذه القيمة المتوسطة ويقل تدريجياً مع الابتعاد عنها . تسمى المقاييس التي تقيس مثل تلك القيمة المتوسطة مقاييسَ النزعة المركزية . إلا أن دراسة قيم النزعة أو الميول المركزية والتي تلقى الضوء على بعض الجوانب الهامة من الظاهرة تبقي في الخفاء جوانب أخرى على قدرٍ كبيرٍ من الأهمية .

إذ إننا غالباً ما نحتاج إلى دراسة مدى تبعثر أو بعد أو انحراف قيم الظاهرة عن تلك القيمة المتوسطة . تسمى المقاييس التي تقيس انحراف القيم عن القيمة المتوسطة مقاييسَ التشتت . وتعتبر المرحلة الأولى في أي دراسة هي حساب تلك المقاييس والذي يعتبر بداية أولية في العمل الإحصائي من حيث تحليل البيانات الإحصائية بعد عمليات جمع البيانات وتصنيفها وعرضها بالطرق المناسبة .

١-المقاييس الموضعية Measures of Location

أو ما يسمى بمقاييس الميول أو النزعة المركزية Measures of central Tendency أو

ما يدعى بالمعدل Average إن عبارة المعدل تشير إلى قياس من القياسات العديدة للميول

المركزية للبيانات . وهناك العديد من هذه المقاييس إلا أن أهمها هي : الوسط الحسابي والوسط

الهندسي والوسيط والمنوال .

وهناك مقياس غير مركزي يسمى بالمئين ويعبر حقيقة مقياس موضعي يمثل الوسيط في

إحدى حالاته الهامة .

١ – الوسط الحسابي The Arithmetic Mean

وهذا المقياس أكثر شيوعاً لدى عموم الناس يحسب هذا المقياس يجمع كافة قيم المشاهدات

الموجودة في البيانات وتقسيمها على عدد هذه المشاهدات في هذه البيانات .

فإذا رمزنا للمتغير المستمر بالرمز (X) وكان هناك عدد من المشاهدات (n) في العينة

فعندئذ يكون الوسط الحسابي في العينة (X) كما يأتي :

$$\overline{X} = \frac{\Sigma \times}{n}$$

وفي جداول التوزيع التكرارية غير المجمعة يؤخذ التكرار بالحسبان فتصبح الصيغة كما يأتي

:

$$\overline{X} = \frac{\Sigma fx}{\Sigma f = n}$$

مثال : إذا اعتبرنا أن تكلفة المجموعات التشخيصية لخمس أمراض كل مجموعة سجلت القيم المالية التالية بالليرة السورية (القيمة مضروبة بالألف) و هي كالتالي:

٢٣ ،١٤، ٣٢ ،٩، ٢٧

فإن الوسط الحسابي هو :

$$\overline{X} = \frac{\Sigma \times}{n} = \frac{23+14+32+9+27}{5} = \frac{105}{5} = 21$$

بمعنى أخر، فإن المتوسط الحسابي لأسعار هذه الكيتات هي ٢١٠٠٠ ليرة سورية.

مثال : لنأخذ بيانات ناتجة من دراسةِ تركيز النطاف في السائل المنوي لواحد وأربعين عينة (القيم مضروبة بالقيمة ١٠٠٠).

٧٥ – ١١٠ – ١٢٣ – ١٣٩ – ١١٠– ٨٦ – ١٢٤ – ١٤٢ – ٩٠ – ١١٠ – ١٢٥ – ١٤٢ – ١١٣ – ١٢٦ – ١٤٢ – ٩٥ – ١١٧ – ١٢٧ – ١٤٣ – ٩٧ – ١١٨ – ١٢٨ – ١٤٣ – ٩٧ – ١٢٠ – ١٢٩ – ١٤٨ – ١٠٢ – ١٢١ – ١٣٣ – ١٥٩ – ١٠٤ – ١٢٢ – ١٣٨ – ١٠٤ – ١٢٢ – ١٣٨ – ١٠٥ – ١٢٣ – ١٣٩ .

الحل:

الوسط الحسابي = مجموع المشاهدات / عدد المشاهدات =٤١/٤٨١٣ = ١١٧.٣٩٠

٢– الوسيط The Median

وهو عبارة عن أحد المقاييس الأخرى للميول المركزية الشائعة التكرار باستخدامها . ويعرف الوسيط بأنه القيمة المركزية (الوسيطة) لمجموعة من المشاهدات (n) في عينة ما والتي تكون مرتبة، فعلى سبيل المثال ترتب المشاهدات تصاعدياً أو تنازلياً بمعنى آخر هو القيمة التي تفصل مجموعة بيانات إحصائية إلى جزئين متساويين فصلاً يجعل قيم أحدهما أقل من قيمة الوسيط بينما تكون قيم الجزء الثاني أكبر منها . وبهذا يكون الوسيط هو القيمة المنصفة لذلك الترتيب التصاعدي والتنازلي لهذه المشاهدات .

نرمز للوسيط بالرمز (Me) ويحسب الوسيط على النحو الآتي :

١– إذا كان عدد المشاهدات (n) وهي فردية فإن الوسيط يساوي :

$$Me = \frac{n+1}{2}$$

إذ إن :

n : تمثل عدد الحدود

وبهذا تكون قيمة الحد ذي الترتيب المقدر هي قيمة الوسيط .

٢– إذا كانت المشاهدات (n) زوجية فإن الوسيط يقع في القيم المنصفة لقسمي المشاهدات.

مثال : إن المعايير المناعية لمرض من الأمراض مأخوذة من ١٩ عينةً هي الآتي (القيم مضروبة بالقيمة ١٠٠):

٣٥٠ و ٤٩٩ و ٤١٢ و ٥٥٦ و ٧٨٩ و ٩٩١ و ٣١٤ و ٦٤٢ و ٣٨٨ و ٥١٠ و٥٩٨ و ٢٩٧

و ٤٥٥ و ٨٦٣ و ٥٣٦ و ٦٨٥ و ٤٢١ و ٣٣٣ و ٤٧٤

المطلوب : أوجد قيمة الوسيط للمعايير المذكورة أعلاه .

الحل : لحساب قيمة الوسيط نرتب القيم أولاً وليكن تصاعدياً فنحصل على :

٢٩٧ و٣١٤ و ٣٣٣ و ٣٥٠ و ٣٨٨ و ٤١٢ و ٤٢١ و ٤٥٥ و ٤٧٤ و ٤٩٩ و٥١٠ و ٥٣٦

و ٥٥٦ و ٥٩٨ و ٦٤٢ و ٦٨٥ و ٧٨٩ و ٨٦٣ و ٩٩١ .

فالملاحظ أن عدد المشاهدات هو فردي (١٩) قيمة ولذلك نحسب ترتيب الوسيط :

$$\frac{n+1}{2} = \frac{19+1}{2} = 10$$

ولذلك فإن الوسيط هو القيمة المعلمة بصندوق وهي القيمة العاشرة أي تكون قيمة الوسيط ممثلة بالحد العاشر وتكتب Me = 499 وهذه القيمة تتوسط مجموعة بيانات المثال فتقسمها إلى جزئين متساويين عَشْر عن يسارها أصغر منها بالقيمة وتسعة قيم عن يمينها أكبر منها .

وفي الحالة التي يكون فيها جدول n أي عدد قيمها زوجياً فإن ترتيب الوسيط 2/ (n + 1) سيعتبر عدداً كسرياً وسيقع الوسيط عندها بين قيمتين في التوزيع ويكون الوسيط في هذه الحالة عبارة عن الوسط الحسابي لهاتين القيمتين .

مثال : أوجد قيمة الوسيط للمشاهدات الآتية :

١، ٢، ، ٣، ٤

الحل :

بما أن هذه القيم مرتبة تصاعدياً نحسب ترتيب الوسيط مباشرة :

$$\frac{n+1}{2} = \frac{4+1}{2} = 2.5$$

ومن المشاهدات المذكورة أعلاه نجد أن الحد 2.5 غير موجود ويقع بين الترتيبين الثاني والثالث المقابلين للقيمتين ٢ و ٣ فإننا نعتبر أن قيمة الوسيط هي قيمة الوسط الحسابي لهاتين القيمتين :

$$Me = \frac{2+3}{2} = 2.5$$

ونجد فعلاً أن قيمة الوسيط 2.5 تقسم البيانات الأربعة المعطاة إلى جزئين متساويين .

٣- المتوسط الهندسي The Geometric Mean :

إذا أخذ لغاريتم كل قيمة في البيانات (إما للأس y أو للأس e) والتي تكون فيها تكراراتها موزعة بكثرة إلى الناحية اليمنى من المحور السيني فإن هذه البيانات تتحول بتناسق نظامي إلى شقي محور السينات (اليميني واليساري) إذا ما أخذنا القيم بشكلها اللغاريتمي .

وفي هذه الحالة فإن الوسط الحسابي للقيم المحولة لغاريتمياً هو أحد مقاييس الميول المركزية المفيدة . إلا أن مساوئ هذا المقياس هو أننا نتعامل مع مقياس لغاريتمي ولذلك نحول البيانات إلى قياسها الأصلي بحساب الرقم الصحيح المضاد للقيم اللغاريتمية anti logarithm وهذا ما يدعى الوسطَ الهندسي . وفي الدراسات الحيوية عندما تكون البيانات موزعة إلى اليمين أو اليسار من إحدى المحاور البيانية نحسب المتوسط الهندسي لنحصل على معدل القيم المراد قياسها وتصبح أكثر تناسقاً يميناً ويساراً .

يُحسب المتوسط الهندسي رياضياً كما يلي:

$$G = \sqrt[n]{X_1 x\ X_2\ x\ X_3 ......Xn}$$

إذ إن :

χ n .....χ١ : قيم المشاهدات ($\chi > 0$)

n : عدد قيم المشاهدات وهو الجذر النوني لجداءات مجموعة من القيم

وهكذا يُستخدم الوسط الهندسي وسطاً للمعدلات والنسب في كثير من الحالات (وليس في جميعها) كحالة معدل حدوث المرض وهو يمثل الوسط الحسابي في تفسيرها . إذ إن الوسط

الحسابي لا يجوز استخدامه في حالات المعدلات والنسب (إلا في حالات خاصة لن نتعرض لها هنا) .

مثال : احسب معدل حدوث مرض داء السكري في عدد من البلدان لأربع سنوات حسب الأرقام المدرجة أدناه (لاحظ أننا عندما نقول معدل حدوث مرض في العلوم الطبية لا تعني المتوسط الحسابي و أنما قيمة عدد الحالات الجديدة للمرض مقسوما على المتوسط الحسابي للحيوانات الواقعة تحت خطر الإصابة خلال فترة زمنية محددة).

0.030 , 0.053 , 0.025 , 0.061

الحل : المتوسط الهندسي لهذه المعدلات المرضية:

$$G = \sqrt[5]{0.030 \times 0.053 \times 0.025 \times 0.06} = 0.25$$

أي معدل حدوث مرض داء السكري للفترة الزمنية المعتبرة هو 0.25 سنوياً .

٤ – المنوال The Mode :

يرمز للمنوال بالرمز Mo وهو مقياس معروف إلا أن هذا المقياس لا يستخدم دائماً مقياسً للميول المركزية ويعرف بأنه هو القيمة الأكبر تكراراً في مشاهدات العينة.

وتختلف قيمة المنوال عن قيمة كل من الوسط الحسابي والوسيط وعادة يكون هناك منوال واحد إذا كان حجم العينة كبيراً ولكن يمكن أن يكون منوالان في مجموعة قيم عينة واحدة إذا تساوى تكرار القيمتين الأكبر تكراراً وعندئذ يقال عن توزيع مشاهدات العينة بأنه ثنائي المنوال Bimodal .

إن مسوغ اعتبار المنوال من مقاييس النزعة المركزية يرتكز إلى حقيقة كونه التكرار الأكثر لإحدى القيم ، ولتكن درجته ٦٠٠٠ خلية / مليمتر لقيم الخلايا البيضاء في الدم في إحدى المشافي يعني أن قيم هذه الخلايا تدور حول هذه القيمة في المشفى المأخوذة منها العينة فلتحديد المنوال في مجموعة بيانات محصول عليها نبحث عن القيمة الأكثر تكراراً فيها وتكون هي المنوال .

مثال : لنأخذ قيم تعداد الكريات البيضاء مقدرة خلية في المليمتر الواحد في العينات المدرجة في الجدول التالي:

٤٥٠٠	٦٥٢٢	٧٩٢٢	٦٨٣٦	٧٠٠٠
١٠٠٠٠	٨٣٠٢	٨٣٢٣	٦٢٠٣	7300
٦٥٠٠	٧٤٠١	٩٣٣٢	٦٣٨٠	٥٤٠٠
٤٧٠٠	٧٨٠٠	٧٢٥٢	٨٧٣٢	٦٠٠٠
٦٢٠٠	٨٠٠٠	٦٢٩٢	٦٣٨٩	٥٠٠
٦١٠٠	٩١٠٠	٥٨٣٢	٩١٢٢	٧٢٠٠
٥٦٠٠	٩٠٠٠	٢٧٣٣	٨٣٦٣	٤٠٠٠
٥٩٠٠	٥٥٠٠	٧٣٦٣	٨٣٦٢	١٢٠٠٠
٦٣٠٠	٧١٩٩	٥٢٧٢	٦٣٢٨	٦٠٠٠
٦٢٠٠	٦٢١٠	٦٣٣٩	١٤١٦٢	٥٠٠٠

الحل :

من القيم المدرجة أعلاه نجد أن قيمة ٦٠٠٠ هي القيمة الأكثر تكراراً في هذه البيانات ولذلك فهي قيمة المنوال .

ويجب أن نذكر أن الإحصائيين لا يفضلون استخدام المنوال لتخليص البيانات للأسباب التالية :

١- تعتمد قيمة المنوال على دقة قيم البيانات المقيسة .

٢- بعض التوزيعات ليس فيها قيمة منوال بينما يكون لبعض توزيعات البيانات أكثر من منوال .

مثال : أوجد قيمة المنوال من مجموعة البيانات الآتية ؟

$$6 , 8 , 11 , 6 , 8 , 10 , 9 , 7$$

الحل : نجد أن لدينا منوالين $M_{o1} = 8$, $M_{o2} = 6$ لأن لهما نفس التكرار وهو التكرار الأكثر في التوزيع .

وفي بعض البيانات لا نجد منوالاً مثل $9 , 10 , 11 , 8 , 2 , 1 , 6$