الإحصاء والاحتمالات – المحاضرة الثانية

Statistics and probabilities-Lecture 2

Dr. Fadi Khalil, Dr. Ali Ahmed

Dr Soummaya Abdul-Hak

Doctor lecturer in statistics and programing

2025

المقدمة

الإحصاء هو لغة الأرقام التي تترجم الظواهر إلى معانٍ قابلة للفهم والتحليل من خلاله نستطيع وصف البيانات، فهم خصائصها، واستخلاص النتائج منها. في هذه المحاضرة سنتناول أنواع البيانات الإحصائية ثم نتدرج إلى أهم المقاييس الإحصائية، وهي مقاييس النزعة المركزية، مع التطرق إلى التشتت والالتواء، لنفهم كيف نقرأ الأرقام قراءة علمية دقيقة.

أولًا: أنواع البيانات الإحصائية

يمكن تصنيف البيانات بحسب طريقة عرضها إلى ثلاثة أنواع رئيسية:

١. البيانات المفردة (Ungrouped Data)

هي بيانات تُعرض في شكل سردي مباشر للقيم دون أي تنظيم في جداول تكرارية. بمعنى آخر، كل قيمة تُذكر كما هي، حتى وإن تكررت.

- **مثال 1:**

درجات خمسة طلاب في مادة الإحصاء:

$$12,\ 14,\ 16,\ 14,\ 18$$

هنا لدينا خمس قيم فقط، تُذكر كما جُمعت .ولا نذكر التكرار أو الفئات.

- **استخدامها:**

تُستخدم البيانات المفردة في العينات الصغيرة أو في المراحل الأولية من التحليل.

٢. البيانات المرتبة (Discrete or Arrayed Data)

هي البيانات التي تُنظَّم في جدول تكراري بسيط بحيث نضع كل قيمة أمام تكرارها (عدد مرات ظهورها).

جَامعة المَنارة
MANARA UNIVERSITY

◆ مثال 2:

القيمة x_i	التكرار (f_i)
12	1
14	2
16	1
18	1

هنا نفس البيانات السابقة، لكننا رتبناها وعددنا تكرار كل قيمة. يساعد هذا الشكل في تسهيل العمليات الحسابية مثل إيجاد الوسط الحسابي أو الوسيط

٣. البيانات المبوّبة (Grouped Data)

هي البيانات التي تُقسَّم إلى فئات أو مجالات (Class Intervals)، بحيث نضع أمام كل فئة عدد القيم الواقعة ضمنها (التكرار).

◆ **مثال 3:**

الفئة	التكرار f_i
$[0-9[$	3
$[9-18[$	5
$[18-27[$	8
$[27-36[$	4

في هذا النوع لا نهتم بالقيم الفردية، بل بالمجالات التي تنتمي إليها تلك القيم. ونستخدم **مركز الفئة** لتمثيلها حسابيًا في المعادلات.

◆ **ثانيًا :أهم المقاييس الإحصائية**

الإحصاء الوصفي يعتمد على ثلاثة أنواع من المقاييس:

١. مقاييس النزعة المركزية (Measures of Central Tendency)

وهي التي تحدد **النقطة التي تتركز حولها البيانات**، مثل:

- الوسط الحسابي
- الوسيط
- المنوال
- الوسط الهندسي
- الوسط التوافقي

٢. **مقاييس التشتت (Measures of Dispersion)**

وتُستخدم لمعرفة **مدى انتشار البيانات حول مركزها** مثل:

- المدى
- التباين
- الانحراف المعياري

<u>مقاييس النزعة المركزية</u>

١.الوسط الحسابي (Arithmetic Mean)

هو أكثر المقاييس استخدامًا في الإحصاء، ويُعرف بأنه: مجموع القيم مقسومًا على عددها الكلي.

بفرض لدينا عينة عشوائية $x_1, x_2, \dots, x_n$ عندها:

📘 **القانون العام:**

$$\bar{X} = \frac{\sum x_i}{n}$$

حيث:

- $\bar{X}$: الوسط الحسابي

- $\sum x_i$: مجموع القيم
- n : عدد القيم

◆ **في البيانات المفردة:**

نحسب المجموع ثم نقسم على عدد القيم.

مثال4:

القيم: $10, 20, 30, 40$

$$\bar{X} = \frac{10+20+30+40}{4} = 25$$

◆ **في البيانات المرتبة:**

نستخدم القانون:

$$\bar{X} = \frac{\sum(f_i x_i)}{\sum f_i}$$

f_i : التكرار المقابل للقيمة x_i .

مثال 5:

x_i	f_i	$f_i x_i$
10	2	20
20	3	60
30	1	30
المجموع	**6**	**110**

$$\bar{X} = \frac{110}{6} \approx 18.33$$

في البيانات المبوّبة:

نستخدم **مراكز الفئات** (X_{mid}) أو الرمز x'_i

$$\bar{X} = \frac{\sum(f_i x'_i)}{\sum f_i}$$

مثال6 :

الفئة	x'_i	f_i	$f_i x'_i$
$[0 - 9[$	4.5	3	13.5
$[9 - 18[$	13.5	5	67.5
$[18 - 27[$	22.5	8	180
$[27 - 36[$	31.5	4	126
المجموع	—	**20**	**387**

$$\bar{X} = \frac{387}{20} = 19.35$$

مزايا الوسط الحسابي:

١. سهل الفهم والحساب.

٢. يعتمد على جميع القيم.

٣. يُستخدم في معظم التحليلات الإحصائية المتقدمة.

عيوبه:

١. يتأثر بشدة بالقيم المتطرفة (Outliers)

٢. لا يصلح للبيانات الاسمية أو الترتيبية.

٣. قد لا يمثل الواقع في حالة التوزيع المائل.

٢.الوسيط (Median)

هو القيمة التي **تقسم مجموعة البيانات إلى نصفين متساويين** بعد ترتيبها تصاعديًا. و نرمز له بـ Me

📘 **القانون العام لبيانات مفردة:**

- إذا كان عدد القيم فرديًا ← الوسيط Me هو القيمة الوسطى أو القيمة التي ترتيبها $X_{\frac{n+1}{2}}$
- إذا كان زوجيًا ← الوسيط Me هو متوسط القيمتين الوسطيتين أو القيمة $\frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$

مثال 7 :

القيم: $10,12,14,18,20$ الوسيط $14 = X_3 = X_{\frac{n+1}{2}} = Me$

مثال 8 :

القيم: $10,12,14,16,18,20$ الوسيط هو: $Me = \frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2} = \frac{X_3+X_4}{2} = \frac{14+16}{2} = 15$

◆ **في البيانات المرتبة:**

نستخدم التكرارات لتحديد موضع الوسيط حيث نبحث في التكرار التجميعي الصاعد عن أول عدد يساوي أو يتجاوز قيمة ترتيب الوسيط وهي $\frac{n}{2} = \frac{\sum f_i}{2}$ أي نصف مجموع التكرارات فتكون القيمة المقابلة له هي قيمة الوسيط.

مثال 9 :

x_i	f_i	التكرار التجميعي K_i
10	2	2
12	3	5
14	4	9
16	1	10

مجموع التكرارات $= 10 \leftarrow$ الوسيط هو القيمة رقم $5 = \frac{10}{2}$ تقع عند $x = 12$

- **في البيانات المبوّبة:**

نتبع الخطوات التالية:

- نوجد التكرار التجميعي الصاعد.
- نحدد ترتيب الوسيط بالعلاقة: $\frac{n}{2} = \frac{\sum f_i}{2}$ (أي 50% من البيانات) .
- نبحث في التكرارات التجميعية الصاعدة عن أول عدد يساوي أو يتجاوز قيمة ترتيب الوسيط بالتالي الفئة المقابلة له هي فئة الوسيط ، ونطبق العلاقة التالية لإيجاد قيمة الوسيط :

$$\mathrm{Me} = a + \left(\frac{\frac{\sum f_i}{2} - K_{i-1}}{f_i} \right) * L$$

حيث:

- a : الحد الأدنى للفئة الوسيطية
- $\sum f_i$: مجموع التكرارات
- K_{i-1} : التكرار التراكمي قبل الفئة الوسيطية
- f_i : تكرار الفئة الوسيطية
- L : طول الفئة

مثال10:

فيما يلي توزيع الدخول الشهرية لـ 65 أسرة بالدولار:

الدخل الشهري	[50 – 60[	[60 – 70[	[70 – 80[	[80 – 90[	[90 – 100[	[100 – 110[	[110 – 120[
عدد الأسر	8	10	16	14	10	5	2

الوسيط : نتبع الخطوات التالية :

- نوجد التكرار التجميعي الصاعد K_i (موضح بالجدول المساعد).
- نحدد ترتيب الوسيط بالعلاقة : $\frac{\sum f_i}{2} = \frac{65}{2} = 32.5$
- نبحث في التكرارات التجميعية الصاعدة عن أول عدد يساوي أو يتجاوز قيمة ترتيب الوسيط فنجد أنه 34 بالتالي الفئة المقابلة له هي فئة الوسيط أي $[70 - 80[$ ونطبق العلاقة التالية لإيجاد قيمة الوسيط :

$$Me = a + L * \frac{\frac{\sum f_i}{2} - K_{i-1}}{f_i}$$

ومنه :

$$Me = 70 + 10 * \frac{32.5 - 18}{16} = 79.06$$

الفئات	f_i	x'_i	K_i
$[50 - 60[$	8	55	8
$[60 - 70[$	10	65	18
$[70 - 80[$	16	75	$34 \leftarrow Me$
$[80 - 90[$	14	85	48
$[90 - 100[$	10	95	58
$[100 - 110[$	5	105	63
$[110 - 120[$	2	115	65
المجموع	6.5	—	

✅ **مزاياه:**

١. لا يتأثر بالقيم المتطرفة.

٢. يعبر بدقة عن المركز في التوزيع المائل.

٣. سهل الفهم في العينات الصغيرة.

⚠️ **عيوبه:**

١. لا يعتمد على جميع القيم.

٢. يصعب حسابه في البيانات المبوّبة إذا كان الجدول غير دقيق.

٣.المنوال (Mode)

هو القيمة **الأكثر تكرارًا** في مجموعة البيانات. و نرمز له بالرمز Mo

- **في البيانات المفردة:**

مثال11:

القيم: 2,3,3,5,6 ← المنوال $Mo = 3$

- **في البيانات المرتبة:**

نختار القيمة ذات التكرار الأعلى.

مثال12:

x_i	f_i
10	2
12	5
14	3

المنوال $Mo = 12$

- **في البيانات المبوّبة:**

- نحدد الفئة المنوالية وهي صاحبة أكبر تكرار.
- نطبق القانون التالي:

$$Mo = a + \frac{\Delta_1}{\Delta_1 + \Delta_2} * \mathrm{L}$$

حيث أن a : الحد الأدنى للفئة المنوالية.

L : طول الفئة المنوالية .

Δ_1الفرق بين تكرار الفئة المنوالية وتكرار الفئة التي قبلها.

Δ_2 الفرق

بين تكرار الفئة المنوالية وتكرار الفئة التي بعدها.

بالعودة للمثال10:

نحدد الفئة الأكثر تكراراً وهي الفئة [70,80[ونطبق العلاقة:

$$Mo = a + \frac{\Delta_1}{\Delta_1 + \Delta_2} * \mathrm{L}$$

وعليه:

$$Mo = 70 + \frac{6}{6+2} * 10 = 77.5$$

✅ **مزاياه:**

١. سهل الفهم، ولا يحتاج حسابات معقدة.

٢. لا يتأثر بالقيم المتطرفة.

٣. مفيد للبيانات الاسمية (مثل الأكثر شيوعًا من فئة معينة).

⚠️ **عيوبه:**

١. قد لا يكون فريدًا (إذا كان هناك أكثر من منوال).

٢. لا يعتمد على كل القيم.

٣. لا يصلح في العينات الصغيرة جدًا.

ملاحظة : ١. إذا كان $\bar{X} = Me = Mo$ فالتوزيع متماثل .

٢. إذا كان $\bar{X} > Me > Mo$ التوزيع ملتوي نحو اليمين .

٣. إذا كان $\bar{X} < Me < Mo$ التوزيع ملتوي نحو اليسار .

4. الوسط الهندسي (Geometric Mean)

📘 المفهوم النظري:

الوسط الهندسي هو مقياس نزعة مركزية يُستخدم عندما تكون البيانات نسبية أو مضروبة وليست مجمّعة جمعًا، أي عندما نريد أن نعرف معدل التغير النسبي أو معدل النمو المركب على مدى فترات زمنية أو عمليات متتابعة.

بعبارة بسيطة:

"الوسط الهندسي (G) يعبّر عن القيمة المتوسطة لنسبة النمو أو التغير عبر الزمن، بحيث نحصل على نفس النتيجة لو حصل النمو بنفس النسبة في كل فترة".

🔹 الصيغة الرياضية:

إذا كانت لدينا n قيم موجبة:

$$G = \sqrt[n]{X_1 * X_2 * X_3 * \dots * X_n}$$

أو بطريقة أبسط باستخدام اللوغاريتمات (في الحسابات الكبيرة) :

$$\log(G) = \frac{\sum \log(X)}{n}$$

ثم نأخذ **اللوغاريتم العكسي** للنتيجة للحصول على G

أمثلة توضيحية:

مثال 13: معدل النمو الاقتصادي

افترض أن الناتج المحلي لبلدٍ ما نما خلال 3 سنوات على النحو التالي:

- السنة الأولى: $+5\%$
- السنة الثانية: $+10\%$
- السنة الثالثة: $+15\%$

نريد معرفة **معدل النمو السنوي المتوسط.**

أولاً: نحول النسب إلى معاملات نمو:

$$1.05\ ,\ \ 1.10\ ,1.15$$

الوسط الهندسي يساوي

$$G = \sqrt[3]{1.05 * 1.10 * 1.15} = \sqrt[3]{1.32825} = 1.099 \Rightarrow 9.9\%$$

النتيجة :معدل النمو السنوي المتوسط هو 9.9% وليس المتوسط الحسابي (الذي سيكون 10%)، لأن النمو يحدث بشكل تراكمي وليس خطيًا.

مثال 14: تحليل البكتيريا في المختبر

إذا تضاعف عدد البكتيريا بنسبة 200%، ثم 150%، ثم 100% على التوالي في ثلاث فترات، فإن استخدام الوسط الهندسي يعطينا معدل النمو الحقيقي المتراكم عبر الفترات الثلاث.

لنجد الوسط الهندسي

- نبدأ بتحويل النسب المئوية على معاملات نمو

$$(1 + 200\%) = 3$$
$$(1 + 150\%) = 2.5$$
$$(1 + 100\%) = 2$$

إذاً عوامل النمو هي $3, 2.5, 2$

- ومنه معدل النمو الحقيقي المتراكم عبر الفترات الثلاث (الوسط الهندسي) هو

$$G = \sqrt[3]{3 * 2.5 * 2} = 2.466$$

لتحويله إلى نسبة مئوية : نطرح 1 أي $1 - 2.466 = 1.466$. إذا الوسط الهندسي لمعدل النمو المتراكم هو 146.6%

الاستخدامات العملية:

١. **في الاقتصاد:**

- لحساب معدل النمو المركب للناتج المحلي، أو الاستثمار، أو الدخل.
- لتقدير العائد الوسطي السنوي على استثمار يتغير سنويًا.

٢. **في الطب وعلم الأحياء:**

- لحساب المتوسط النسبي للتكاثر البكتيري أو معدلات انتشار الأدوية.
- في مقارنة الجرعات النسبية أو التراكيز الدوائية بين التجارب.

٣. **في البيئة:**

- لقياس التغير النسبي في معدلات التلوث أو النمو السكاني عبر الزمن.

٤. **في الكيمياء والفيزياء:**

- يستخدم لحساب متوسط معدلات التفاعل النسبي عندما تكون القيم مضروبة وليست مجموعات.

مزايا الوسط الهندسي:

١. يعطي صورة واقعية أكثر عندما تكون البيانات نسبية أو متغيرة تراكميًا.

٢. لا يتأثر كثيرًا بالقيم المتطرفة الكبيرة.

٣. يُستخدم في الحسابات اللوغاريتمية والاحتمالية.

٤. يمثل نقطة التوازن النسبي للبيانات.

⚠️ **عيوبه:**

١. لا يمكن استخدامه في وجود قيم صفرية أو سالبة.

٢. معقد قليلًا في الحساب اليدوي.

٣. لا يعطي نتائج منطقية عندما تكون البيانات جمعًا بسيطًا وليست نسبًا.

٥. الوسط التوافقي (Harmonic Mean)

المفهوم النظري:

الوسط التوافقي يُستخدم عندما تكون البيانات مرتبطة بالمقلوبات، أي عندما تكون العلاقة بين الكميات عكسية. فهو المقياس المناسب عندما يكون المعدل المطلوب نسبة بين مقدارين (مثل المسافة والزمن، أو الكمية والسعر).

بمعنى آخر:

"الوسط التوافقي يُستخدم عندما تكون القيمة الكلية مرتبطة بمعدل يعتمد على المقامات، وليس البسط".

◆ **الصيغة الرياضية:**

إذا كانت لدينا n قيم موجبة:

$$H = \frac{n}{\sum \frac{1}{x_i}}$$

أمثلة توضيحية:

مثال 15: حساب السرعة المتوسطة

سيارة تسير مسافة معينة ذهابًا وإيابًا:

- في الذهاب: 60 كم/سا
- في العودة: 40 كم/سا

نريد السرعة المتوسطة الكلية.

الوسط التوافقي:

$$H=\frac{2}{\frac{1}{60}+\frac{1}{40}}=\frac{2}{\frac{100}{2400}}=\frac{2}{0.0416}=48\text{كم/سا}$$

النتيجة :السرعة المتوسطة الفعلية = 48كم/سا وليس 50كم/سا كما قد يعطي الوسط الحسابي، لأن الزمن المقضي عند السرعة الأقل أكبر.

مثال 16: معدل السعر في السوق

تاجر يشتري نفس السلعة مرتين:

- المرة الأولى: السعر = 10 دولار
- المرة الثانية: السعر = 20 دولار

ما متوسط السعر الذي دفعه فعليًا لكل وحدة؟

$$H=\frac{2}{\frac{1}{10}+\frac{1}{20}}=\frac{2}{0.15}=13.33\text{ دولار}$$

ومنه المتوسط الحقيقي 13.33 دولار

مثال 17: عن تحليل أداة شبكة الانترنيت

لدينا ثلاثة مزودي خدمة انترنت (ISP) يقدّمون سرعات مختلفة للتحميل $(Mbps)$ لمستخدمين في منطقة معينة وهي كالتالي

مزود1 : $10\ Mbps$

مزود2: $20\ Mbps$

مزود3 : $30\ Mbps$

لحساب الوسط التوافقي

$$H = \frac{n}{\sum \frac{1}{x_i}} = \frac{3}{\frac{1}{10} + \frac{1}{20} + \frac{1}{30}} = 16.3636 \ \ Mbps$$

الوسط التوافقي هنا يُظهر أن السرعة المتوسطة للتحميل عبر مزودي الخدمة الثلاثة هو حوالي $16.3636 \ \ Mbps$ فهذا الرقم يعتبر أكثر دقة من المتوسط الحسابي (الذي سيكون $20 \ Mbps$) في حال وجود قيماً متباينة. حيث يعطي وزناً أكبر للقيم الأقل.

الاستخدامات العملية:

١. **في الفيزياء والهندسة:**

- لحساب السرعة المتوسطة عندما تكون المسافات متساوية.
- لحساب المقاومات الكهربائية عندما تكون المقاومات متصلة على التوازي.

٢. **في الاقتصاد والتمويل:**

- لتقدير متوسط السعر الفعلي للوحدات في حالة اختلاف الكميات أو الأسعار.
- في تحليل معدلات العائد العكسي أو الفوائد المتغيرة.

٣. **في الطب والإحصاء الحيوي:**

- لحساب المعدل الوسطي الزمني لتدفق الدم أو الاستجابة الدوائية عند فترات مختلفة.
- عند تحليل المتوسطات العكسية مثل متوسط معدل الوفيات بالنسبة للمدة.

٤. **في علم الحاسوب والبيانات:**

- يُستخدم في بعض خوارزميات التصنيف الإحصائي عندما يكون المقياس العكسي أفضل من المتوسط الحسابي.

- يستخدم في تحليل أداء الشبكات: لتحديد السرعة الفعلية التي يمكن أن يتوقعها المستخدمون
- تقييم أداء الخوارزميات: عند قياس زمن الاستجابة أو زمن التنفيذ
- تحليل البيانات المالية: مثل حساب العوائد على الاستثمار عندما تكون النسب متباينة

⚖️ مزايا الوسط التوافقي:

١. يعطي نتائج دقيقة عندما تكون البيانات مرتبطة بالمقلوبات.

٢. يوازن بين القيم الكبيرة والصغيرة بدقة عالية.

٣. مناسب للسرعات والمعدلات الزمنية والمقاومات الكهربائية.

⚠️ عيوبه:

١. لا يمكن استخدامه إذا احتوت البيانات على صفر أو قيمة سالبة.

٢. حساس جدًا للقيم الصغيرة (فالقيمة الصغيرة تُؤثر بشدة على النتيجة).

٣. أقل سهولة في الفهم للمتعلمين المبتدئين.

🔶 مقارنة بين الوسط الهندسي والتوافقي

العنصر	الوسط الهندسي	الوسط التوافقي
المفهوم	مقياس لمتوسط النمو النسبي	مقياس للمعدل العكسي أو الزمني
نوع البيانات	نسب أو معدلات نمو	معدلات تعتمد على المقامات (كالزمن أو السعر)
المعادلة	$G = \sqrt[n]{X_1 * X_2 * \ldots * X_n}$	$H = \dfrac{n}{\sum \frac{1}{x_i}}$
القيم الصفرية	لا يُستخدم معها	لا يُستخدم معها
القيم الصغيرة	تأثيرها ضعيف	تأثيرها كبير جدًا
الاستخدامات	الاقتصاد، الأحياء، البيئة	الفيزياء، السرعة، الاقتصاد، الطب

العلاقة بين الوسط الحسابي والهندسي والتوافقي

العلاقة الرياضية الأساسية بين هذه المقاييس الثلاثة هي:

$$H \le G \le \bar{X}$$

أي أن: **الوسط التوافقي ≥الوسط الهندسي≥ الوسط الحسابي**

مثال17:

القيم: 4,16

$$\bar{X} = \frac{4+16}{2} = 10$$

$$G = \sqrt{4 * 16} = 8$$

$$H = \frac{2}{\frac{1}{4} + \frac{1}{16}} = 6.4$$

وبذلك:

$$H = 6.4 \le G = 8 \le \bar{X} = 10$$

>> انتهت المحاضرة<<