محاضرات مبادئ الاحصاء
الفصل الاول
للعام الدراسي 2015-2016

م.م. ورود باسم نور

المرحلة الاولى
قسم دارة الاعمال

المصدر

كتاب الاحصاء

تأليف د.محمود حسن المشهداني د.امير حنا هرمز

نظام المحاضرات الالكتروني

الفصل الاول

أولاً: تعريف علم الإحصاء

هنالك تعاريف عديدة للإحصاء اختلفت وتباينت من حيث المضمون والشمول باختلاف مراحل تطور هذا العلم ويمكن إجمالها في نوعين رئيسيين الأول منها اعتبر الإحصاء بأنه جمع لبيانات إحصائية أي جمع لجمل عددية للحقائق والظواهر في حين ان الثاني اعتبر الإحصاء بأنه جمع لطرق إحصائية أي جمع متكامل لمبادئ وأساليب تستخدم في تجميع وتحليل البيانات والمعلومات الإحصائية وفيما يلي التعريف الاخير:

الإحصاء: وهو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن الظاهرة معينة وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك.

وينقسم علم الإحصاء بذلك إلى فرعين:

1. **الإحصاء الوصفي**: ويتضمن هذا الفرع الطرق والأساليب المستخدمة في جميع البيانات والمعلومات عن الظاهرة معينة أو مجموعة من الظواهر وكيفية تنظيم وتصنيف وتبويب هذه البيانات مع إمكانية عرضها في جداول ورسوم بيانية وحساب بعض المؤشرات الإحصائية منها.
2. **الإحصاء الاستدلالي**: وهو الذي يهتم بموضوعي التقدير واختبار الفرضيات.

ثانياً: أهمية علم الإحصاء ومجالات تطبيقه

إن مجالات استخدام وتطبيق علم الإحصاء يمكن شملها بما يلي:

1. البحوث الرزاعية التطبيقية.
2. البحوث الصناعية التطبيقية.
3. البحوث النفسية والتربوية.
4. بحوث الرياضيات والشباب.
5. البحوث الطبية التطبيقية.
6. البحوث الاقتصادية التطبيقية.
7. البحوث الإدارية التطبيقية.
8. البحوث الهندسية التطبيقية.

ثالثاً: الطريقة الإحصائية في البحث العلمي:

1. تحديد مشكلة البحث أو فرضية البحث أو الدراسة.
2. جمع البيانات والمعلومات عن الظاهرة أو الظواهر ذات العلاقة بالبحث والدراسة.
3. تصنيف البيانات وتبويبها وعرضها.
4. حساب المؤشرات الإحصائية كتقديرات لمعالم مجتمع البحث أو الدراسة.
5. تحليل معطيات الدراسة والتوصل إلى النتائج على ضوء فرضية أو فرضيات البحث والدراسة.
6. تفسير النتائج وعملية اتخاذ القرارات بشأن فرضيات البحث.

رابعاً: تصميم البحوث

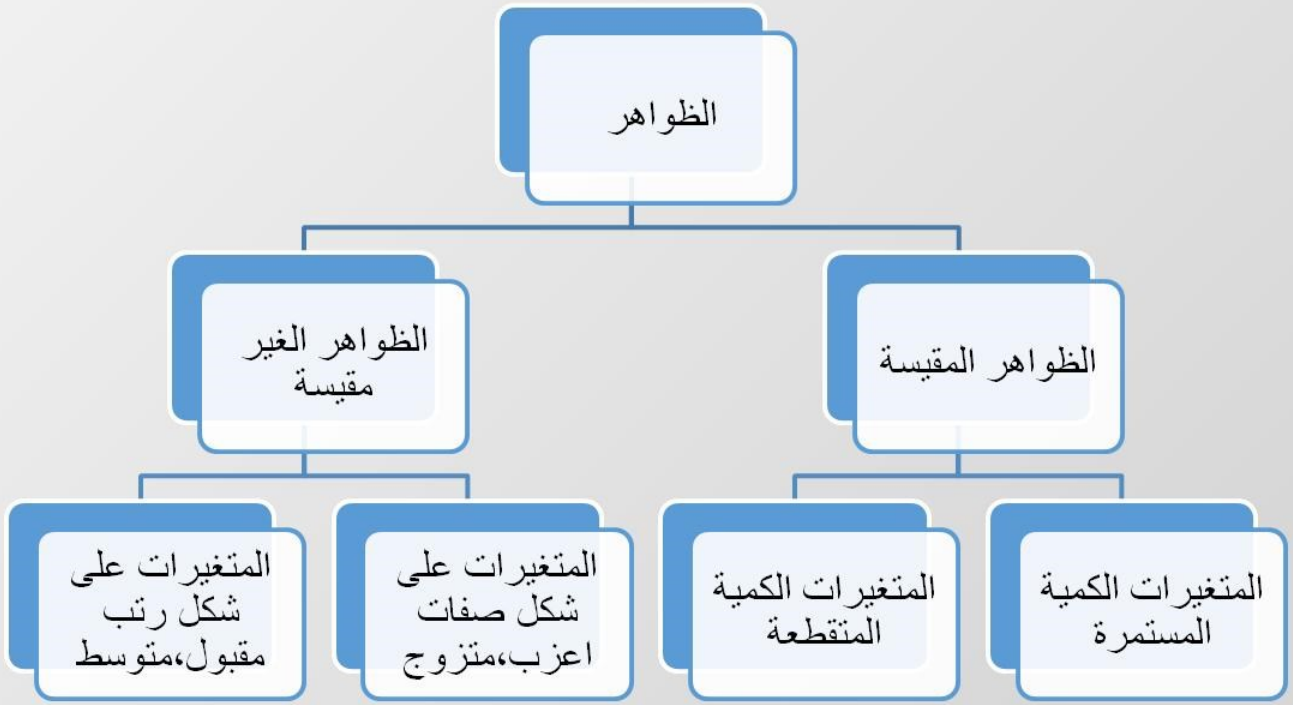
1. تحديد الغرض من البحث:
2. تحديد إمكانية التنبؤ الفعلي:
3. تحديد إطار البحث:

والمجتمع الإحصائي: هو عبارة عن مجموعة من المفردات أو الوحدات الإحصائية التي تشترك بصفة أو أكثر من الصفات المعنية وغالباً ما يتم الحصول عليها من البيانات والمعلومات المطلوبة. وأن المجتمع الإحصائي يمكن تقسيمه إلى قسمين بصورة عامة:

1. **المجتمع الإحصائي المحدد:**
2. **المجتمع الإحصائي الغير محدد:**

خامساً: الظواهر المقيسة

1. **المتغيرات الكمية المستمرة**: يمكن أن تأخذ أي قيمة بين قيمتين متتاليتين مثلا إن العمر الافتراضي لشخص ما يمكن أن يكون سنوات وعدد من الأشهر 18.5 وان وزن الطفل يمكن أن يكون كيلوات وبضع من الغرامات 5.26.
2. **المتغيرات الكمية المتقطعة**: وهي المتغيرات التي لايمكن أن تأخذ أي قيمة بين قيمتين متتاليتين مثلا أن عدد أولاد عائلة معينة لا يمكن أن يكون خمس أولاد ونصف بأخذ قيمة واحدة وهو خمس أولاد وهكذا.


الظواهر
الظواهر المقيسة
الظواهر الغير مقيسة
المتغيرات الكمية المستمرة
المتغيرات الكمية المتقطعة
المتغيرات على شكل صفات اعزب،متزوج
المتغيرات على شكل رتب مقبول،متوسط

الفصل الثاني
جمع وتصنيف وتبويب البيانات

اساليب جمع البيانات

1. اسلوب التسجيل الشامل:
2. اسلوب العينات:

ويمكن تعريف العينة: هي مجموعة مختارة من مفردات المجتمع الإحصائي قيد الدراسة مثل دراسة فاعلية دواء معين على بعض الأشخاص المصابين بمرض السرطان.

أنواع العينات:

تنقسم العينات بشكل عام إلى قسمين هما:-

أولاً: العينات العشوائية:ويقصد بالعينة العشوائية بأنها تلك المجموعة من المفردات المختارة من مجتمع الدراسة بحيث أنه ليس للباحث أي دخل في اختيار هذه المفردة دون تلك أي أن هناك مبدأ تساوي الفرصة لظهور أي مفردة من مفردات المجتمع ضمن هذه العينة وتنقسم العينات العشوائية إلى عدة أنواع منها:-

1. **العينة العشوائية البسيطة:**-وعلى فرض ان مجتمع الدراسة متجانس ومحدد وأن عدد مفردات المجتمع N وافرض ان هذه الدراسة تتطلب اختيار عينة حجمها n فان

- احتمال ظهور كل مفردة ضمن العينة $\frac{1}{N}$.
- ان عدد العينات الممكنة من هذا المجتمع r بحيث أن:

$$r = \binom{n}{N} = \frac{N!}{n!(N-n)!}$$

- عدد مرات ظهور أي مفردة من المفردات في جميع العينات:

$$\binom{n-1}{N-1} = \frac{(N-1)!}{(n-1)!(N-n)!}$$

2. العينة الطبقية العشوائية:-

تعتبر العينات المختارة وفق هذا الأسلوب من أفضل أنواع العينات وأكثرها دقة في تمثيل المجتمع فكثيراً ما نلاحظ مفردات المجتمع الإحصائي غير متجانس من حيث الصفة ذات العلاقة بموضوع البحث فمثلاً لو كنا بصدد دراسة المستوى العلمي لطلبة كلية الإدارة والاقتصاد فنلاحظ إن هذا المجتمع غير متجانس من حيث التخصص العلمي فهنالك مجموعة من الطلبة من اختصاص الإدارة وأخرى من الاقتصاد وأخرى من المحاسبة وأخرى من الإحصاء .وعلى فرض ان مجتمع الدراسة مؤلف من N مفردة وان هذا المجتمع من الممكن تجزئته إلى L من الطبقات وفق معيار متجانس ملائم لتلك الدراسة بحيث أن في كل طبقة متجانسة من المفردات وكما يلي:-

1 2 3 … L ,$N = N_1+N_2+\ldots+N_l$

وزن الطبقة

$$W_h=\frac{N_h}{N}$$

حيث h= 1,2, …., L ، يمثل h تسلسل الطبقة.

حجم العينة المطلوب اختيارها من الطبقة h هي $n_h=W_h.n$

3. الطبقة العشوائية المنتظمة:

على فرض ان مفردات مجتمع الدراسة البالغ عددها N مفردة مرتبة وفق ترتيب معين كأن يكون ترتيب تصاعدي أو تنازلي أو وفق أي معيار آخر للترتيب. وأفرض ان دراسة معينة تتطلب اختيار عينة من المفردات قوامها n مفردة. يتم تقسيم مفردات المجتمع المرتبة إلى عدد من المجاميع كل مجموعة منها تضم المجتمع مفردات إن وبافتراض .مفردة $k = \frac{N}{n}$ متسلسلة في الترتيب فأن هذه المجاميع ستكون:-

مجموعة الاولى	مجموعة ثانية	مجموعة الاخيرة
1 2 … k	k+1,k+2,…,2k	(n-1)k,…,nk

4. العينة المتعددة المراحل:

في هذه العينة يتم تقسيم المجتمع الإحصائي إلى وحدات تدعى بالوحدات الأولية ثم يتم اختيار عينة عشوائية من هذه الوحدات الأولية كمرحلة أولى يتم تقسيم كل وحدة أولية مختارة إلى وحدات اصغر تدعى بالوحدات الثانوية ويتم اختيار عينة عشوائية من الوحدات الثانوية لكل وحدة أولية كمرحلة ثانية. تقسم الوحدات الثانوية المختارة إلى وحدات أصغر ويختار من كل منها عينة عشوائية كمرحلة ثالثة. ويستمر التقسيم والاختيار على نفس الطريقة لحين الوصول إلى المفردات التي يتم جمع البيانات منها والتي تؤلف عينة البحث.

ثانياً: العينات غير العشوائية:-

ويقصد بالعينة غير العشوائية أنها تلك المجموعة من المفردات المختارة من مجتمع الدراسة بطريقة يكون للباحث دخل في اختيار هذه المفردة دون تلك وليس على أساس عشوائي وذلك لاعتبارات تتعلق بطبيعة تلك الدراسة. ومن هذه العينات:

1. العينة الحصصية.
2. العينة العمدية.
3. العينة العنقودية.
3. العينة المزدوجة .

وسائل جمع البيانات:

بعد تحديد حجم العينة وأسلوب المعاينة الملائم في اختيار مفردات هذه العينة من مجتمع الدراسة يتطلب الأمر اختيار الوسيلة الملائمة في جميع البيانات والمعلومات عن الظاهرة أو الظواهر المتعلقة بتلك الدراسة. وهناك وسائل عديدة في جمع البيانات أهما ما يلي:-

1- أسلوب الجمع المباشر:

2-الاستمارة الإحصائية:

الاستبيان: هو عبارة عن استمارة يتم من خلالها جمع البيانات والمعلومات من مفردات (أو بعض مفردات) مجتمع الدراسة وذلك عن طريق مواجهة الباحث الشخصية للمفردة الإحصائية ، أو ع طريق المراسلة كما هي الحال في التعدادات السكانية مثلاً حيث يتم مواجهة رب الأسرة (الأسرة تمثل المفردة الإحصائية) أو من ينوب عنه لغرض ملء استمارة التعداد بالبيانات والمعلومات اللازمة.

الأخطاء الشائعة في جمع البيانات:

يحدث أحياناً أن يقع الباحث في بعض الأخطاء لدى جمعه البيانات والمعلومات التي يتطلبها البحث. هذه الأخطاء تحدث نتيجة سوء استخدام الطريقة الإحصائية. وفيما يلي موجز لأهم هذه الأخطاء:-

أ- **خطأ التحيز**:

ب- **خطأ الصدفة**:

الجداول الإحصائية:-

فالجداول على نوعين:-

1. الجداول الأولية: وهي جداول تفصيلية توضح الحقائق وتعرض البيانات المتعلقة بظاهرة معينة.
2. الجداول الثانوية: وهي جداول مختصرة غرضها عرض نتائج ذات أهمية تتعلق ببحث معين.

إن الجداول تهدف إلى إبراز البيانات وتوضيحها في أضيق حيز.

تصنيف وتبويب البيانات:-

إن البيانات المستحصل عليها بخصوص الظاهرة المعينة تسمى بالبيانات الخام أو البيانات الأولية أو البيانات غير المصنفة. وهي غالباً ما تكون غير منظمة وفي هذه الحالة يتعذر على الباحث تكوين فكرة عن هذه الظاهرة أو تلك التي جمعت عنها البيانات كذلك يتعذر على الباحث الاعتماد بشكلها الغير منظم لأغراض التحليل الإحصائي للوصول إلى النتائج المطلوبة. لذلك فأن أولى الخطوات الهامة بعد عملية جمع البيانات هي عملية مراجعة وتصنيف وتبويب البيانات.

1- مراجعة البيانات:

2- تصنيف البيانات:

3- تبويب البيانات:

والهدف من عملية التبويب هو إبراز البيانات وتوضيحها في أضيق حيز ممكن وهناك أربعة أشكال رئيسية هي:-

أ- التبويب الزمني

ب- التبويب الجغرافي

ج- التبويب الكمي

د- التبويب على أساس صفة معينة

التوزيعات التكرارية: وهو عبارة عن تلخيص وترتيب البيانات المتغير العشوائي التي سبق أن صنفت وجمعت مقسمة عدد من المجاميع كل منها تسمى (Class) بالفئة هذه الفئات قد تكون مرتبة تصاعدياً أو تنازلياً حيث طبيعة البيانات ويسمى توزيع عدد قيم إلى فئات بالتوزيع التكراري.

ملاحظة: قد تكون فئات التوزيع التكراري متساوية في الطول أو غير متساوية وفيما يلي أهم مكونات التوزيع التكراري:-

1. **المدى الكلي للتوزيع (T.R)**: يعرف المدى الكلي بأنه الفرق ما بين اكبر قيمة وأصغر قيمة من المجموعة مضافاً إليها العدد 1 فإذا رمزت للمدى الكلي بالرمز T.R.

قانون المدى الكلي $\mathbf{T.R = X_L - X_S + 1}$

حيث أن X_L: اكبر قيمة من مجموعة القيم. X_S : اصغر قيمة من مجموعة القيم.

1. **عدد فئات التوزيع [M]**: يمثل عدد المجاميع التي يتألف منها التوزيع التكراري وهناك صيغ تقريبية يمكن من خلالها تحديد عدد فئات التوزيع حيث يرمز تعدد الفئات بالرمز ...واهم هذه الصيغ :

أ- صيغة يول (yole): $M = 2.5\sqrt[4]{n}$

حيث ان n تمثل عدد المفردات

ب- صيغة سترجس $M = 1 + 3.322log_{10}n$

ملاحظة : عند تطبيق تقريب النتائج الى اقرب عدد صحيح.

طول الفئة (L) ويسمى احيانا المدى الفئوي وبمثل مابين الحد الادنى للفئة وحدها الاعلى . ان طول الفئة يتناسب عكسياً مع عدد الفئات في التوزيع فكلما كبر طول الفئة قل عدد الفئات والعكس صحيح واذا رمزنا لطول الفئة بالرمز L فأنه يمكن تحديد قيمة (L) :

$$L = \frac{T.R}{M}$$

حيث ان : T.R = المدى الكلي M = عدد الفئات

ملاحظة : تستخدم هذه الصيغة في ايجاد طول الفئة عندما يكون اطوال الفئات متساوية في التوزيع التكراري .

مركز الفئة : يمثل مركز الفئة قيمة من قيم المتغير العشوائي x التي تتوسط المسافة بين الحد الادنى والحد الاعلى للفئة فأذا رمزنا للحد الادنى بالرمز (L.L) والحد الاعلى بالرمز (U.L) ولمركز الفئة بالرمز

$$X = \frac{U.L + L.L}{2}$$

ملاحظات مهمة حول التوزيع التكراري:-

1. ليس من الضروري عند تطبيق أن يكون الحد الأدنى للفئة الأولى مساوي تماماً لأصغر قيمة في المجموعة بل قد يكون أقل منها.
2. أن التوزيع التكراري قد يكون توزيع مغلق أو توزيع مفتوح ويقصد بالتوزيع المغلق بأنه ذلك التوزيع الذي يمتلك حداً أدنى للفئة الأولى وحدا أعلى للفئة الأخيرة .

1. **تكرار الفئة**: يمثل تكرار الفئة جزء من مفردات العينة التي تتصف بكونها تقع من القيمة العددية ما بين مدى الفئة بحسب ان مجموع هذه الأجزاء بشكل مفردات العينة n فإذا رمزنا لتكرارات الفئات بالرموز:

$[F_1, F_2, \ldots F_m]$

$\Leftarrow \Leftarrow \Leftarrow [F_1, F_2, \ldots F_m]$فأن

التوزيع التكراري المتجمع: وهو التوزيع الذي يبين كمية التكرار المتجمع عند قيمة معينة من قيم المتغير العشوائي، وهو على نوعين هما:-

أ-**التوزيع التكراري المتجمع الصاعد**: وهما التوزيع الذي يبين تراكم التكرارات ابتداءً من الفئة الأولى في التوزيع وانتهاء بالفئة الأخيرة.

وسنرمز للتكرار المتجمع الصاعد لأي فئة بالرمز F_i

∴ التكرار المتجمع الصاعد للفئة الأولى $F_1 = f_1$

التكرار المتجمع الصاعد للفئة الثانية $F_2 = f_1 + f_2$

التكرار المتجمع الصاعد للفئة الثالثة $F_3 = f_1 + f_2 + f_3$

⋮

التكرار المتجمع الصاعد للفئة الأخيرة $F_M = f_1 + f_2 + \ldots$ $F_M = n$

ب- **التوزيع التكراري المتجمع النازل**: وهو التوزيع الذي يبين تناقص التكرارات ابتداءً بالفئة الأولى في التوزيع وانتهاءً بالفئة الأخيرة منه وسنرمز للتكرار المتجمع النازل F'

التكرار المتجمع النازل للفئة الأولى	n
التكرار المتجمع النازل للفئة الثانية	$n - f_1$
التكرار المتجمع النازل للفئة الثالثة	$n - f_1 - f_2$
⋮	
التكرار المتجمع النازل للفئة الأخيرة	$n - f_1 - f_2 - \cdots - f_{m-1}$

التوزيع التكراري النسبي: أن التوزيع التكراري النسبي هو توزيع تكراري اعتيادي كما سبق توضيحه فيه التكرارات معبر عنها بنسبة مئوية يمكن الحصول عليها من قسمة تكرار كل فئة على مجموع التكرارات الكلية. فعلى افتراض أن هنالك توزيع تكراري عدد فئاته M وان التكرارات المقابلة لهذه الفئات وحسب تسلسلها f_1, f_2, f_m عندئذ فأن التكرارات النسبية المقابلة لهذه الفئات ستكون حيث أن:

$$f^* = \frac{f_i}{n} . 100$$

وأن n : تمثل مجموع التكرارات .

ويمكن حساب التوزيع التكراري المتجمع الصاعد النسبي والتوزيع التكراري التنازل السلبي وذلك من خلال قسمة التكرارات المتجمعة الصاعدة على مجموع التكرارات الكلية (أي n) فإذا رمزنا للتكرار المتجمع الصاعد النسبي بـ F* فأن :

إذا رمزنا للتكرار المتجمع الصاعد النسبي بالرمز $F^* = \frac{F_i}{n}$

ونرمز للتكرار المتجمع النازل النسبي بالرمز $F^* = \frac{F'}{n}$

العرض الهندسي للبيانات:

يتم عرض البيانات بهيئة رسوم بيانية وأشكال هندسية متعددة الأشكال والتصاميم والبعض منها بهيئة رسوم تصويرية. وهذه الأشكال هي:-

1- الأشرطة البيانية:

هي مجموعة من المستطيلات الرأسية أو الأفقية قواعدها متساوية وتمثل الصفة التي تم على أساسها التبويب (سنة، شهر ، محافظة...) وارتفاعاتها تمثل البيانات المقابلة لتلك الصفة (درجات الحرارة حسب الشهر ، السنة اكياس محصول الحنطة حسب المحافظات....) وتكون على نوعين:-

أ-الأشرطة البيانية المفردة:

وهي أشرطة بيانية تخص صنف واحد من البيانات المبوبة مثل عدد طلبة المقبول في التعليم العالي للسنوات 2000 – 2005

ب- الأشرطة البيانية المركبة: وهي أشرطة تخص صنفين أو اكثر من البيانات المئوية.

المستطيل البياني:

عبارة عن شكل هندسي يستخدم في تمثيل بيانات ظاهرة معينة يمكن تجزئتها إلى عدد من الأصناف المقابلة للتجميع. ويتم باختيار مستطيل ذو قاعدة مناسبة وبعد ذلك يتم تمثيل كل صنف من البيانات يستطيل جزئي داخل المستطيل الكبير يحد أن مجموع مساحات المستطيلات الجزئية تمثل مساحة المستطيل أكبر وتتم عملية اختيار قواعد المستطيلات الجزئية وفق ما يلي:-

$$\text{طول قاعدة المستطيل الجزئي} = \frac{\text{عدد بيانات الصنف}}{\text{مجموع البيانات}} \times \text{طول قاعدة المستطيل}$$

الدائرة البيانية PIE-chart:

شكل هندسي مشابه للمستطيل البياني ولكن يتمثل ببيانات الصف بقطاعات داخل دائرة بحيث ان مجموع مساحات القطاعات تمثل مساحة الدائرة.

$$\text{زاوية القطاع} = \frac{\text{عدد بيانات الصنف}}{\text{مجموع البيانات}} \times 360^\circ$$

الخط البياني Line – Chart:

شكل يوضح التغيرات الحاصلة لظاهرة عبر فترة معينة من الزمن .

المدرج والمظلع والمنحنى التكراري:

عبارة عن اشكال هندسية الهدف منها عرض البيانات المبوبة في جدول توزيع تكراري.

أ-المدرج التكراري Histogram:

عبارة عن مجموعة من المستطيلات تمثل كل منها طول الفئة في التوزيع التكراري وارتفاعاتها تمثل قيمة التكرار المقابلة لتلك الفئات . تكون المستطيالت منفصلة في حالة البيانات المتقطعة ومتصلة في حالة البيانات المستمرة.

ب- المضلع التكراري Frequency polygon :

عبارة عن عدد من المستقيمات المتصلة مع بعضها على شكل سلسلة ونقطة اتصال المستقيم بالآخر تقليل مركز الفئة أي أنه ازدياد القيم (مركز الفئة والتكرار).

ج- المنحنى التكراري Frequency curve:

مشابه للمضلع ولكن بدل تمرر النقاط بمستقيم تمرر بمنحنى.

يغلق مع المحور السيني بمركزي فئة وهمي

ملاحظة 1: يمكن رسم المدرج والمنحني والمضلع التكراري معا فتكون نقاط المنحنى أو المضلع هو منتصف رووس المستطيلات.

ملاحظة 2: يرسم المنحنى التكراري.

منحنيات التوزيعات التكرارية المتجمعة

1. المنحنى التكراري المتجمع الصاعد

وهو الشكل الذي يرسم بحدود عليا والتكرار المتجمع الصاعد

أ- متغير متقطع يكون بشكل خطوط مستقيمة

ب- متغير مستمربشكل منحنى تكراري متجمع صاعد

2. المنحى التكراري المتجمع النازل

وهو الشكل الذي يرسم بحدود دنيا والتكرار المتجمع النازل

أ- متغير متقطع يكون بشكل خطوط مستقيمة

ب- متغير مستمربشكل منحنى تكراري متجمع نازل

الفصل الرابع
رمز الجمع والضرب

رمز الجمع Σ

1. مجموع قيم عناصر السلسلة
2. مجموع مربعات قيم عناصر السلسلة
3. مربع مجموع قيم عناصر السلسلة
4. مجموع مقلوب عناصر السلسلة
5. مقلوب مجموع عناصر السلسلة
6. مجموع لوغارتمات عناصر السلسلة
7. مجموع جذور عناصر السلسلة

مجموع الكمية الثابتة إلى n من الممرات

مجموع حاصل ضرب ا لثابت بقيم عناصر السلسلة x_i

رمز الضرب $\prod$

1. ضرب قيم عناصر السلسلة
2. ضرب مقلوب عناصر السلسلة
3. مقلوب ضرب عناصر السلسلة
4. ضرب لوغارتمات عناصر السلسلة
5. ضرب جذور عناصر السلسلة