الاحصاء الحياتي :- Biostatistics

مقدمة :- Introduction

ان علم الاحصاء يعتبر من اهم الركائز التي ترتكز عليها عملية البحث العلمي في ميادينه المختلفة ويمكن القول انه لا يوجد مجال من مجالات الفكر والعمل الا واستعمل الاحصاء فيه بأساليبه المختلفة ومن اهم المجالات العلوم الحياتية.

تعـريف : علم الإحصاء هو ذلك الفرع من العلوم الذي يختص بالطرق العلمية لجمع البيانات و تنظيمها و تلخيصها و عرضها وتحليلها؛ وذلك للوصول إلى نتائج مقبولة و قرارات سليمة على ضوء هذا التحليل .

من التعريف السابق لعلم الإحصاء يتضح لنا أن العمليات الإحصائية تتم في صورة أربع خطوات تتمثل في :

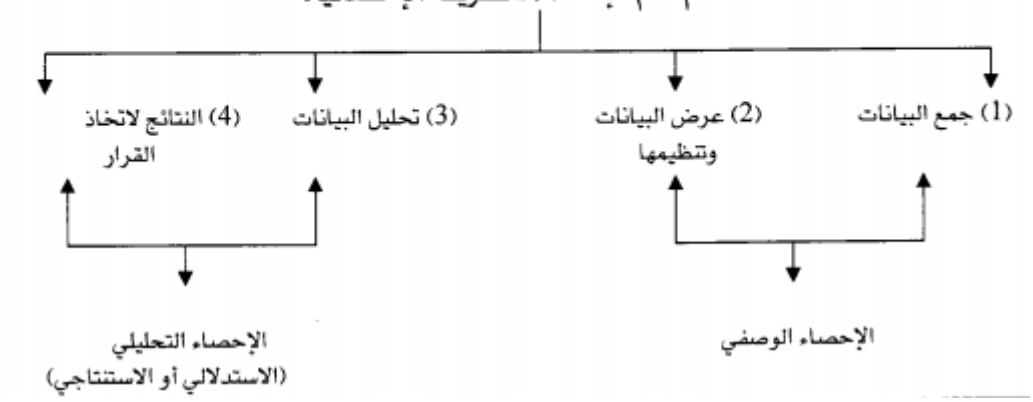
١ ـ جمع البيانات الرقمية أو العددية .

٢ ـ تنظيم البيانات في صورة جداول (العرض الجدولي) أو رسوم بيانية (العرض البياني) ، أو الاثنين معاً .

٣ ـ وصف البيانات باستخدام مفاهيم إحصائية معينة .

٤ ـ الاستدلال من البيانات على نتائج معينة يراد الوصول إليها .

تعريف علم الإحصاء: مجموعة النظريات والطرق العلمية التي تبحث في جمع البيانات وعرضها وتحليلها واستخدام النتائج في التنبؤ أو التقرير واتخاذ القرار.



ويقسم علم الاحصاء الى قسمين هما :-

1- الاحصاء الوصفي (Descriptive statistic)

يتضمن هذا القسم الطرق والاساليب المستخدمة لجمع البيانات وتصنيفها وتبويبها مع امكانية عرضها في جداول ورسوم بياتية وحساب بعض المؤشرات الاحصائية .

2- الاحصاء الاستدلالي (Inferential statistic)

يهتم هذا القسم بموضوع التقدير او التخمين (Estimation) واختيار الفرضيات .

تعريف الاحصاء الحياتي (Biostatistics) :-

يعني الاحصاء الحياتي اشياء مختلفة للاشخاص المختلفين فهو للعامة جداول واعداد عن البيانات الحياتية اما المعنى الاصطلاحي للاحصاء فهو رياضيات جمع البيانات للظواهر البايلوجية وتنظيمها وتحليلها وتفسيرها والتعميم من الخاص الى العام عن طريق استدلال خواص المجتمع من خواص العينة .

اهمية علم الاحصاء :-

يحتل الاحصاء مكاناً بين العلوم لما له من استعمالات واسعة للوصول الى قرارات صائبة لوصف او تفسير الظواهر المختلفة في جميع العلوم وهو المستعمل من قبل للافراد والجماعات المختلفة والدول على حد السواء وفي الحقيقة ان الانتصار العظيم في نزول الانسان على القمر ماكان يحدث لولا مساعدة علم الاحصاء ، واستخدم الاحصاء في مجالات كثيرة ونركز على اهمية علم الاحصاء في العلوم البايلوجية والطبية والصحة العامة والكيمياء .

1- في علم الاحياء (البايلوجي) :-تستخدم الطرق الاحصائية في دراسة الاجناس والفصائل المختلفة للحيوان والنبات ومعرفة خواص كل جنس بما يتميز عن غيره واختلاف مفردات الجنس الواحد في اية

خاصية معينة من الناحية الاحصائية ، فمثلاً نرى الذكور في الجنس البشري اطول قامة من الاناث مع ان الذكور فيما بينهم يختلفون في الطول الى درجة ما وكذلك الاناث ، كل ذلك يتم عن طريق جمع البيانات وتبويبها ودراستها دراسة احصائية والخروج بنتائج من هذه الصفات .

2- في الطب يستخدم الاحصاء لدراسة العلاقة بين متغيرات كثيررة منها على سبيل المثال العلاقة بين العمر وضغط الدم وكذلك العلاقة بين الوراثة والبيئة وتأثيراتهما على تكوين الفرد .

بعض المفاهيم الاحصائية :-

1- المتغير Variable :- يقصد به اي صفة او عنصر قابل للتغير في النوع والكم من فرد الى أخر في نفس المجتمع ويكون المتغير اما :-

A- متغيرات وصفية او نوعية Qualitative Variable

وهي الصفة التي لا يمكن قياسها مباشرة بأرقام عددية لان الفرق بين المفردات تكون في النوع وليس في الكم ومن الامثلة على ذلك (الصحة ، اللون ، الذكاء ، والجنس ، والحالة الاجتماعية)

B- صفة كمية Quantitative Variable : وهي الصفة التي يمكن قياسها مباشرة بأرقام عددية كالاختلاف بين الافراد في الطول والوزن ومستوى الهيموكلوبين والهرمونات وعدد خلايا الدم الحمراء ومستوى الدهون في مصل الدم Lipid profile (TC ، TG، HDL-C ، LDL-C ، VLDL) ويمكن قياسها بوحدات القياس المختلفة كالسنتيمتر والكيلوغرام (mg , pg ,g) وتنقسم المتغيرات الكمية الى :-

1- متغيرات متصلة او مستمرة Continuous variable

المتغير المتصل هو المتغير الذي تأخذ كل مفردة قيمة رقمية او كسر بين حدي المتغير الكلي فلو فرضنا اطوال الطلبة يتراوح بين (130.5 و 170 سم) ،كمية الهيموكلوبين (12.5 – 14 ملغم لكل لتر من الدم)

2- متغيرات غير متصلة او مستمرة Discontinuous Variable
هي المتغيرات التي تأخذ المشاهدة او المفردة فيها قيم متباعدة او متقطعة غير مستمرة اي هو الذي لا تأخذ كل مفردة فيه قيمة كسرية بل لا تزيد قيمة المتغير او تنقص بأقل من واحد فعدد الطلاب عدد الكتب كلها متغيرات غير متصلة او مستمرة .

المشاهدة Observation :-

تعتبر المشاهدة ك بمثابة المواد الاولية التي يتعامل معها الباحث فأذا اراد باحث ان يقيس مستوى الكلوكوز في مصل دم احد الجرذان ولنفرض ان مستوى الكلوكوز في مصل دم هذا الجرذ هو (120 ملغم/ 1مل) فأن هذا العدد يمثل المشاهدة ، لذا فأن المشاهة هي سجل رقمي لحادثة وان مجموع المشاهدات تكون البيانات Data .

المجتمع Population :-

المجتمع من الناحية الاحصائية يمثل جميع الافراد او العناصر التي تشترك في صفة متغير واحد او اكثر تميزه تماماً عن بقية المجتمعات ويتعلق مفهوم المجتمع بالهدف المحدد للبحث الاحصائي فقد يشكل طلبة جامعة كربلاء مجتمعاً ، والمجتمع هو عبارة عن جميع القيم التي يمكن ان يأخذها المتغير ، فمثلاً عند دراسة مستوى الهيموكلوبين في دم طلبة جامعة كربلاء وصفة مستوى الهيموكلوبين في دم طلبة جامعة كربلاء هي متغير تأخذ مدى معين لمجتمع طلبة جامعة كربلاء ، والمجتمع اما ان يكون :-

A- مجتمع محدود Finite Population

وهو المجتمع الذي يمكن حصر مفرداته كما هو الحال في مستوى الهيموكلوبين في دم طلبة جامعة كربلاء او عدد ردهات المرضى في مستشفى الحسين .

B- مجتمع غير محدود Infinite Population

هو المجتمع الذي من الصعب او المستحيل حصر مفرداته مثل عدد البكتريا في مستعمرة بكترية او حقل معين .

العينة Sample :-

العينة هي جزء المجتمع وهي عبارة عن مجموعة من المشاهدات اختير بطريقة ما من المجتمع حيث ان دراسة المجتمع ككل قد يكون صعباً ويحتاج الى وقت وجهد ومال لذا فقد استعيض عن دراسة المجتمع بدراسة العينة ومنها نستطيع ان نستنتج خواص المجتمع الذي اخذت منه العينة ، فقد تكون العينة انسان او حيوان او نبات او جزء معلوم من نبات معين تجري عليه التجارب في المختبرات والعينة هي احدى ادوات البحث العلمي .

ومن اهم انواع العينات :-

1- العينة العشوائية البسيطة Simple Random Sample
وهي تلك العينة التي تسحب من مجتمع الدراسة بحيث يكون احتمال فرض ظهور اية مفردة من مفردات المجتمع الاحصائي في العينة متساوياً وبمعنى اخر تعني اعطاء كل فرد من المجتمع نفس الفرصة للظهور في العينة ويتم اختيارها كما يلي :
مثل استخدام طريقة البطاقات او القرعة
اذا كان لدينا (5) مرضى واردنا اختيار مريضين عشوائياً فما عدد الطرق الممكنة لأختيار مريضين لاجراء بعض الفحوصات .

ان عدد الطرق الممكنة $n\,Cr = \frac{n!}{r!(n-r)!}$

توافيق $5\,C2 = \frac{5!}{2!(5-2)!}$

طرق مختلفة $5\,C2 = \frac{5\times4\times3\times2\times1}{2\times1\,(3\times2\times1)} = 10$

بمعنى انه يوجد عشر بطاقات يكتب عليها اسم مريضين ويتم اختيار بطاقة من العشرة بطاقات عشوائياً .

فأذا كانت اسماء المرضى c,e,d,b,a فأن العشر بطاقات يكون مكتوب عليها de, ce ,cd , be ,bd , bc , ae , ad , ac , ab ويتم سحب اي بطاقة من العشرة .

2- العينة المنتظمة :- وهي اختيار العينات بشكل منتظم من قائمة المجتمع حيث يتم اختيارها من خلال ترقيم عناصر المجتمع الاحصائي بحيث يتم تحديد قاعدة للاختيار تستند على تحديد اختيار العنصر الاول ولتبسيط الشرح لو كان مجتمع الاصل (100 مريض) وتريد اختيار (10 مرضى) لأجراء بعض الفحوصات عليهم فمثلاً تأخذ الارقام العشرة الاولى وتوضع في صندوق ويتم السحب ، فمثلاً حصلنا على

الرقم (3) فيكون العينات العشرة المرضى هي كالآتي ، وتكون المسافة $= \frac{\text{مجموع المرضى}}{\text{مجموع العينة}}$ $10 = \frac{100}{10} +$

93, 83,73, 63, 53 , 43, 33, 23 , 13,3 ,

حيث ان الفاصلة = 10 بين مريض و آخر وتسمى هذه العينة المختارة (عينة منتظمة Systematic Sample) .

3- العينة الطبقية Stratified Sample

يتم في هذا النوع من العينة تقسيم المجتمع الاحصائي اولاً الى مجموعات فرعية تسمى كل منها (طبقة Strate) ومن ثم تتم عملية المعاينة من كل طبقة ، وعادة تكون جميع عناصر الطبقة الواحدة متجانسة فيما يتعلق بالخصائص موضوع الدراسة فعلى سبيل المثال لو اريد اجراء دراسة معينة على مجتمع كلية طب الاسنان ونحتاج اخذ عينة من مجتمع كلية طب الاسنان عددها (20 عنصراً) علما ان مجتمع كلية طب الاسنان عدده (1000 فرد) حيث كان مجتمع كلية الطب مقسم الى الطبقات التالية

رقم الطبقة	اسم الطبقة	عدد افراد الطبقة
1	اساتذة	150
2	موظفين	250
3	طلبة	600

الحل :-

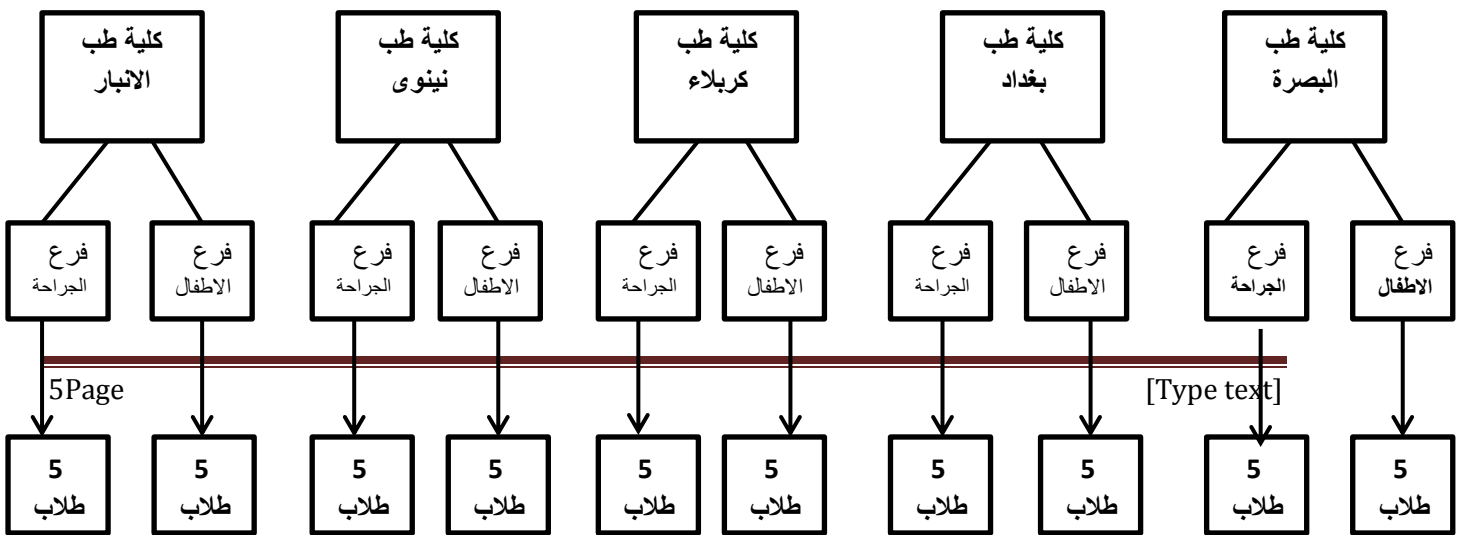
يتم الاختيار عدد مفردات كل طبقة حسب العلاقة

عدد افراد كل طبقة $= \frac{\text{حجم الطبقة}}{\text{حجم المجتمع}} \times$ حجم العينة

	العينة	عدد افراد الطبقة	رقم الطبقة
عدد افراد طبقة الاساتذة $= \frac{150}{1000} \times 20 = 3$	3	150	1 اساتذة
عدد افراد طبقة الموظفين $= \frac{250}{1000} \times 20 = 5$	5	250	2 موظفين
عدد افراد طبقة الطلاب $= \frac{600}{1000} \times 20 =$	$\frac{12}{20}$	$\frac{600}{1000}$	3 طلبة

4- العينة العنقودية متعددة المراحل Multi-stage cluster sample

تعتبر المعاينة العنقودية احد الآليات التي يمكن استخدامها لاختيار العينات من خلال تقسيم المجتمع الى مجموعات او عناقيد على سبيل المثال نريد التعرف على مستوى التعليم الطبي في العراق فنختار 5 كليات طب ومن كل كلية فرعين ومن كل فرع 5 طلاب



جداول التوزيع التكراري :-

1- الفئات Classes :- وهي المجاميع التي قسمت اليها قيم المتغير وكل فئة تأخذ مدى معين من قيم المتغير ولكل فئة دائما يكون لدينا حد اعلى وحد ادنى وكذلك حدان حقيقيان حد ادنى حقيقي وحد اعلى حقيقي وتدعى بالحدود الفعلية .

الحد الادنى Lower class limit : هي اقل قيمة من قيم المتغير يسمح لها بالدخول في هذه الفئة على سبيل المثال الفئة (50 -55) فالحد الادنى للفئة هو (50).

الحد الاعلى upper Class limit : وهو اكبر قيمة من قيم المتغير يسمح لها بالدخول في هذه الفئة مثل (50 – 55) فالحد الاعلى (55) ، ولكل فئة حدان حقيقيان حد ادنى حقيقي Lower class boundary وحد اعلى حقيقي Upper class Boundary وقد تسمى هذه الحود الحقيقية بالحدود الفعلية حيث يتم تحويل البيانات من بيانات منفصلة الى بيانات مستمرة او متصلة .

2- طول الفئة Class length or Class width

هو مقدار المدى بين حدي الفئة وهذا يستحسن ان تكون اطوال الفئة متساوية وسنرمز لطول الفئة بالرمز (C) .

3- مركز الفئة Class mark : هو عبارة عن منتصف المدى بين حدي الفئة وسنرمز بالرمز (yi)

4- تكرار الفئة Class frequency :هو عدد مفردات او القيم التي تقع في مدى تلك الفئة وسنرمز له بالرمز (fi) ويكون مجموع التكرارات مساوية للعدد الكلي لقيم الظاهرة .

وسوف نوضح ما سبق شرحه بالتفصيل في الجدول التالي الذي يبين توزيع طلبة كلية طب الاسنان حسب صفة الوزن :

الوزن (كغم) الفئات class	عدد الطلبة التكرار (fi)	مركز الفئات Yi	الحدود الحقيقية	التكرار النسبي	التكرار المئوي
60 – 62	5	61	59.5 – 62.5	0.05	5
63 – 65	15	64	62.5 – 65.5	0.15	15
66 – 68	45	67	65.5 – 68.5	0.45	45
69 – 71	27	70	68.5 – 71.5	0.27	27
72 - 74	8	73	71.5 – 74.5	**0.08**	**8**
	100			1	100

طول الفئة : □- C = الحد الاعلى – الحد الادنى + 1

= 62 – 60 + 1 = 3

□- طول الفئة يساوي الفرق بين الحدود الدنيا لفئتين متتاليتين = 63 – 60 =3

□- طول الفئة يساوي الفرق بين الحدود العليا لفئتين متتاليتين = 65 – 62 =3

□- طول الفئة يساوي الفرق بين مركز فئتين متتاليتين = 64 – 61 =3

- طول الفئة الفرق بين الحدود الحقيقية الدنيا لفئتين متتاليتين = 62.5 – 59.5 =3

- طول الفئة الفرق بين الحدود الحقيقية العليا لفئتين متتاليتين = 65.5 – 62. =3

مركز الفئة yi = $\frac{\text{الحد الاعلى–الحد االادنى}}{\mathbf{2}}$

$$61 = + \frac{\mathbf{06-62}}{\mathbf{2}} = \text{yi}$$

الحدود الحقيقية :

الحد الادنى الحقيقي الحد الادنى للفئة – 0.5

الحد الادنى الحقيقي = 60 – 0.5 = 59.5

الحد الادنى الحقيقي = مركز تلك الفئة - $\frac{\mathbf{1}}{\mathbf{2}}$ طول تلك الفئة

الحد الادنى الحقيقي = 61 - $\frac{\mathbf{1}}{\mathbf{2}}$ × 3 = 61 – 1.5 = 59.5

الحد الاعلى الحقيقي = الحد الاعلى للفئة +0.5

الحد الاعلى الحقيقي = 62 +0.5 = 62.5

الحد الاعلى الحقيقي =مركز تلك الفئة + $\frac{\mathbf{1}}{\mathbf{2}}$ طول تلك الفئة

الحد الاعلى الحقيقي = 61 + $\frac{\mathbf{1}}{\mathbf{2}}$ × 3 = 61 +1.5 = 62.5

الرسوم البيانية

الرسم (العرض) البياني للبيانات، هو أحد طرق التي يمكن استخدامها في وصف البيانات، من حيث شكل التوزيع ومدى تمركز البيانات، وفي كثير من النواحي التطبيقية يكون العرض البياني أسهل وأسرع في وصف الظاهرة محل الدراسة، وتختلف طرق عرض البيانات بيانياً حسب نوع البيانات المبوبة في شكل جدول تكراري، وفيما يلي عرض للأشكال البيانية المختلفة .

١ - المدرج التكرار .

نرسم المدرج التكراري على محورين متعامدين إحداهما أفقي يمثل الفئات والثاني رأسي يمثل التكرار ، وتكون وحدة القياس على كل محور متناسقة مع بعضها البعض . نرسم مستطيلات متلاصقة على الفئات قاعدتها طول الفئة محسوباً من الحدود الحقيقية، وارتفاعاته عبارة عن تكرار هذه الفئات .

فمثلاً بالنسبة للفئة الأولى يكون المستطيل قاعدته بادئه من الحد الفئة الأولى، ومنتهية بالفئة الثانية، وارتفاع المستطيل هو تكرار الفئة الأولى . وهكذا لباقي المستطيلات التي تمثل باقي التكرارات .

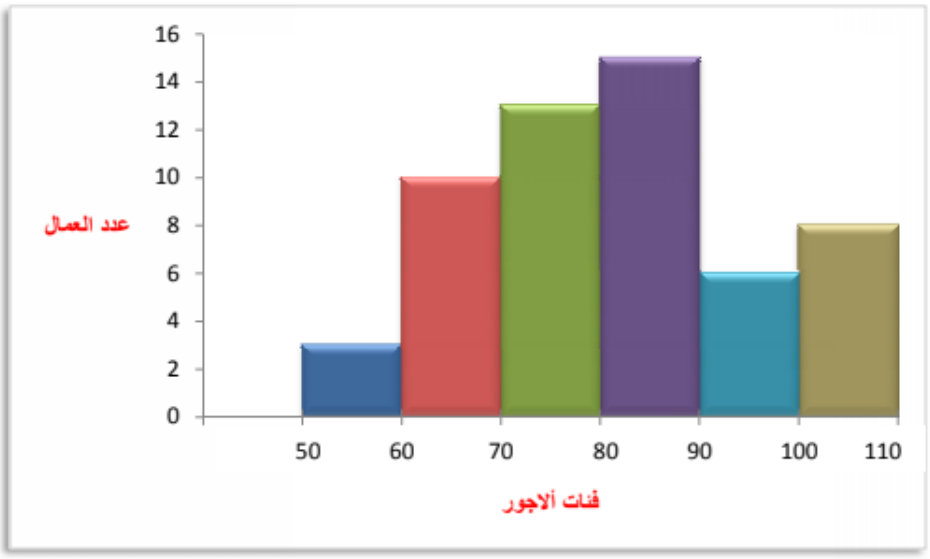
مثـال (2)

فيما يلى أجور 55 عاملاً في إحدى المؤسسات

فئات الأجور	50-	60-	70-	80-	90-	100-110	المجموع
عدد العمال	3	10	13	15	6	8	55

ارسم المدرج التكراري .

الحـل :



مثـال (3)

فيما يلي التوزيع التكراري لأوزان 100 عينة من الدواجن حجمها بالجرام، اختيرت من أحدى المزارع بعد 45 يوم .

الوزن	600-	620-	640-	660-	680-	700-720	المجموع
عدد الدجاج (التكرار)	10	15	20	25	20	10	100

ارسم المدرج التكراري .

٢ - الرسوم الدائرية .

وهي عبارة عن دائرة تقسم إلى قطاعات زواياها المركزية تتناسب مع القراءات ويمكن حساب الزاوية الخاصة بقطاع يمثل قراءة من القراءات من القانون التالي :

$$\text{الزاوية المركزية} = \frac{\text{القراءة نفسها}}{\text{مجموع القراءات}} \times 360^{o}$$

أو

$$\text{الزاوية المركزية} = \frac{\text{التكرار للفئة}}{\text{مجموع التكرارات}} \times 360^{o}$$

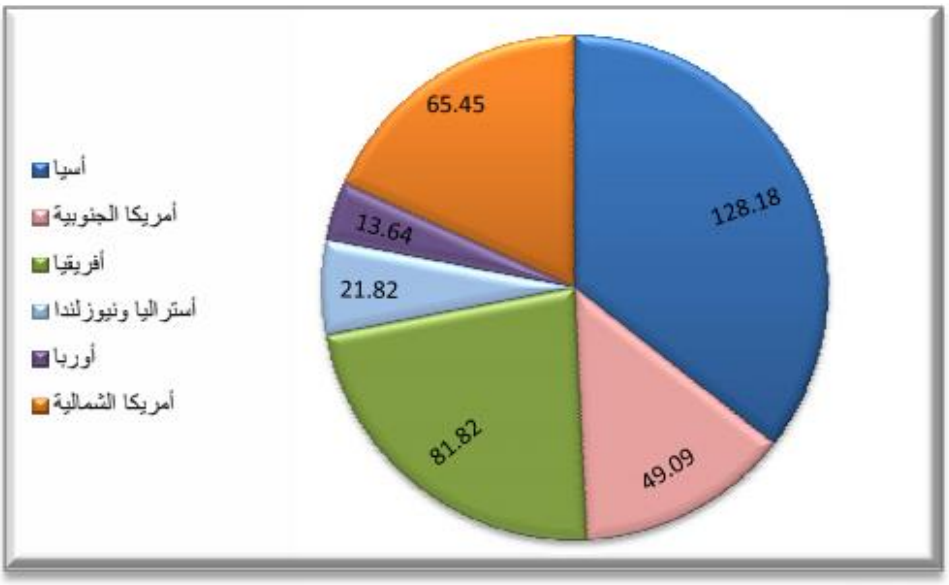
ولحساب النسبة المئوية :

$$\text{النسبة المئوية} = \frac{\text{التكرار للفئة}}{\text{مجموع التكرارات}} \times 100$$

مثـال (4)

الجدول التالي يمثل تقريباً مساحات القارات في العالم مثلها بالرسوم الدائرية :

القارة	المساحة بالمليون كم^2
أسيا	47
أمريكا الجنوبية	18
أفريقيا	30
أستراليا ونيوزلندا	8
أوربا	5
أمريكا الشمالية	24
المجموع	132



الجدول التالي يمثل عدد الأشخاص المتبرعين للدم مثلها بالرسوم الدائرية :

فصيلة الدم	A	AB	O	المجموع
العدد (التكرار)	20	40	60	120

التوزيع التكراري النسبي Relative Frequency distribution

هو توزيع تكراري يبين الاهمية النسبية لكل فئة ويحسب التكرار النسبي

$$\text{التكرار النسبي لأي فئة} = \frac{\text{تكرار تلك الفئة}}{\text{المجموع الكلي للتكرارات}} \qquad 0.05 \ \frac{\mathbf{5}}{\mathbf{100}} =$$

التكرار المئوي = التكرار النسبي × 100

التكرار المئوي = 0.05 × 100 = 5

التوزيعات التكرارية المتجمعة Cumulative Distribution :-

في جدول التوزيع التكراري العادي الذي سبق شرحه يبين توزيع قيم المتغير على الفئات المختلفة ولكن بعض الاحيان قد يكون هناك حاجة الى معرفة عدد القيم او المفردات التي تقل او تزيد عن قيمة معينة والجداول التي تحوي مثل هذه المعلومات تدعى بجداول التوزيع التكراري المتجمعة وهي نوعان من الجداول :-

A- جداول التوزيع التكراري التجمعي التصاعدي :وهذا التوزيع يعطينا عدد المفردات التي تقل قيمتها عن الحد الادنى لفئة معينة وهو الذي يبين تراكم التكرارات ابتداء من الفئة الاولى وانتهاء بالفئة الاخيرة ، يتم

احتساب التكرارات المتجمعة على اساس حدود الفئة العليا وتسمى less than cumulative distribution .

B- جداول التوزيع التكراري التجمعي التنازلي : وهو الجدول الذ يعطينا عدد المفردات التي تزيد عن الحد الادنى لفئة معينة وكذلك هو التوزيع الذي يبين تناقص التكرارات ابتداء من الفئة الاولى في التوزيع وانتهاء بالفئة الاخيرة ويتم حساب التكرارات على اساس الحدود الدنيا للفئات

مثال// اوجد التوزيع التكراري التجمعي التصاعدي والتنازلي لجدول التوزيع التكراري الذي يبين توزيع طلبة كلية طب الاسنان حسب صفة الوزن More than cumulative distribution

توزيع الطلبة حسب صفة الوزن

الفئات Class	التكرارات fi
60 – 62	5
63 – 65	15
66 – 68	45
69 – 71	27
72 - 74	**8**
	100

جدول التوزيع التكراري التجمعي التصاعدي

جدول الفئات	التكرار المتمجع الصاعد
اقل من 60	0
اقل من 63	5
اقل من 66	20
اقل من 69	65
اقل من 72	92
اقل من 74	100

جدول التوزيع التكراري:-

جدول الفئات	التكرار المتجمع النازل
60 فاكثر	100
63 فأكثر	95
66 فأكثر	80
69 فأكثر	35
72 فـأكثر	8
74 فأكثر	0

الخطوات العامة لتكوين جدول توزيع تكراري :-

1- استخراج المدى الكلي

يرمز له بالرمز R

$R = y\ max - y\ min + 1$

2- تحديد عدد الفئات ويرمز لعدد الفئات M

يفضل ان لا يقل عدد الفئات في التوزيع عن 5 ولا يزيد عن 15 فأذا قل عدد الفئات في التوزيع عن (5) فئات فأن عملية التبويب قد تؤدي الى عدم كشف الصفات الاساسية للمجتمع اي عدم اعطاء صورة واضحة لصفات المجتمع اما اذا زاد عدد الفئات عن (15) فئة فأن ذلك فيه صعوبات في اجراء العمليات الحسابية لبعض المؤشرات ويمكن حساب عدد الفئات حسب الصيغ التالية :-

A- صيغة يول $M = 2.5 \sqrt[4]{n}$

حيث n هي عدد المشاهدات

B- صيغة ستيرجس Steruges :- $M = 1 + 3.3 \log (n)$

3- ايجاد طول الفئة

طول الفئة = $\frac{\text{المدى}}{\text{عدد الفئات}}$ تقرب النتيجة الى اقرب عدد صحيح

4- كتابة حدود الفئات :-

بحيث ان جميع قيم المتغير عند كتابة حدود الفئات تضع بين الحد الادنى للفئة الاولى والحد الاعلى للفئة الاخيرة .

5- استخراج عدد التكرارات لكل فئة :-

مثال : البيانات التالية تمثل درجات 13 طالب من طلبة كلية الصيدلة في مادة الانسجة

50 , 51 , 53 , 59 , 61 , 61 , 62 , 63 , 68 , 69 , 71 , 74 , 79

الحل :

1- استخراج المدى $R = y\ max - y\ min + 1$

$R = 79 - 50 + 1 = 30$

2- تحديد عدد الفئات

طريقة يول $M = 2.5 \sqrt[4]{13}$

$M = 2.5 \times 1.898 = 4.75 \simeq 5$

3- طريقة سترج $M = 1 + 3.3 \log (13)$

$M = 1 + 3.3 \times 1.106 = 4.65 \simeq 5$

4- طول الفئة $= \frac{R}{M} = \frac{30}{5} = 6$

الحد الادنى للفئة الاولى (50)

طول الفئة = الحد الاعلى – الحد الادنى +1

6 = الحد الاعلى – 50 +1

6 = س – 50 +1

س = 56 – 1 = 55

الحد الاعلى = 55 الحد الاعلى للفئة الاولى

بأضافة طول الفئة للحد الادنى والحد الاعلى للفئة الاولى تحصل الفئات الاخرى التالية

الفئات	التكرار fi	مركز الفئات yi	احدود الحقيقية	التكرار النسبي	التكرار المئوي
50 – 55	3	52.5	49.5 – 55.5	0.23	23
56 – 61	3	58.5	55.5 – 61.5	0.23	23
62 – 67	2	64.5	61.5 – 67.5	0.15	15
68 – 73	3	70.5	67.5 – 73.5	0.23	23
74 - 79	**2**	76.5	73.5 – 79.5	0.15	15
	13				

مثال : البيانات التالية تمثل تركيز المونولديهايد في في اناث الارنب المزالة من المبايض والتي عددها 40 انثى علماً ان تركيز المونولديهايد في مصل الدم مقاس $m\ mol/L$ اعرض هذه البيانات في جدول توزيع تكراري:

2.6	2.2	4.1	3.5	2.0	3.2	3.7	3.0
3.7	3.4	1.6	3.1	3.3	3.8	3.1	2.4
3.1	2.5	4.3	3.4	3.6	2.9	3.3	3.9
3.4	3.3	3.1	3.7	4.4	3.2	4.1	1.9
3.5	2.3	3.8	3.2	2.6	3.9	3.0	4.2

الحل :-

1- استخراج المدى R = y max – y min+ 0.1

R = 4.4 – 1.0 + 0.1 = 2.9

2- تحديد عدد الفئات M= 2.0 $\sqrt[4]{\mathbf{40}}$

M= 2.5 × 2.51 = 6.28 ≃ 6

3- ايجاد طول الفئة = $\frac{\mathbf{2.9}}{\mathbf{6}} = \frac{\boldsymbol{R}}{\boldsymbol{M}}$ = 0.483 □ 0.5

4- كتابة حدود الفئات بما ان اقل قيمة (1.6) للمتغير تأخذ الحد الادنى للفئة الاولى 1.5

.: الحد الادنى للفئة الاولى = 1.5

الحد الاعلى للفئة الاولى

طول الفئة = الحد الاعلى – الحد الادنى + 0.1

0.5= س – 1.5 +0.1

س = 2.0 - 0.1

س = 1.9 الحد الاعلى للفئة الاولى

ثم نضيف طول الفئة للحد الادنى والحد الاعلى نحصل على الفئات الاخرى :-

حدود الفئات	التكرار fi	مركز الفئات	احدود الحقيقية
1.5 – 1.9	2	1.7	1.45 – 1.95
2.0 – 2.4	4	2.2	1.95 – 2.45
2.5 – 2.9	4	2.7	2.45 – 2.95
3.0 – 3.4	15	3.2	2.95 – 3.45
3.5 – 3.9	10	3.7	3.45 – 3.95
4.0 – 4.4	**5**	4.2	3.95 – 4.45
	40		

مقاييس النزعة المركزية

تُسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات، وهى القيم التي تتركز القيم حولها، ومن هذه المقاييس، الوسط الحسابي، والمنوال، والوسيط، والوسط الهندسي، والوسط التوافقي، والرباعيات.

سنكتفي في هذا المقرر بدراسة الوسط (المتوسط) الحسابي .

الوسط (المتوسط) الحسابي

المتوسط أو الوسط الحسابي يعتبر من أهم مقاييس النزعة المركزية و الأكثر شهرةً والأكثر استخداماً في الإحصاء والحياة العملية إذ يستخدم عادةً في الكثير من المقارنات بين الظواهر المختلفة. ويمكن حسابه للبيانات المبوبة وغير المبوبة ، كما يلي :

أولا : الوسط الحسابي للبيانات غير المبوبة :

تعريف : إذا كان لدينا مجموعة من المشاهدات للمتغير X وهي $x_1 , x_2 , x_3 , ... , x_n$ فان الوسط الحسابي يساوي حاصل جميع المشاهدات أو البيانات مقسوما على عددها . ويرمز له بالرمز $\overline{x}$ (ويقرأ $x\ bar$) .

فإن الوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\overline{x}$ يحسب بالعلاقة التالية :

$$\overline{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{الوسط الحسابي} = \frac{\text{مجموع القيم}}{\text{عدد القيم}}$$

فيما يلي درجات 8 طلاب في مقرر الرياضيات .

34 32 42 37 35 40 36 40

والمطلوب إيجاد الوسط الحسابي لدرجة الطالب في الامتحان .

الحـل :

$$\begin{aligned}\overline{x} &= \frac{x_1 + x_2 + ... + x_n}{n} \\ &= \frac{34+32+42+37+35+40+36+40}{8} \\ &= \frac{296}{8} \\ &= 37\end{aligned}$$

أي أن الوسط الحسابي لدرجة الطلاب في اختبار الرياضيات يساوي 37 درجة .

الجدول التالي يمثل عدد المخالفات في تجاوز الإشارة وهي حمراء عند إحدى الإشارات المرورية . أوجد المتوسط لعدد المخالفات :

اليوم	السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة
عدد المخالفات	120	90	80	85	100	150	200

ثانياً : الوسط الحسابي للبيانات المبوبة :

إذا كان لدينا عدد k من الفئات ذات المراكز $x_1, x_2, \ldots, x_k$ ولها تكرارات $f_1, f_2, \ldots, f_k$ على الترتيب ، فإن الوسط الحسابي يعطى بالعلاقة التالية:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots + f_k x_k}{f_1 + f_2 + f_3 + \ldots + f_k} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

الجدول التالي يعرض توزيع 40 متدرباً حسب أوزانهم :

فئات الوزن	40-	50-	60-	70-	80-	90-100	المجموع
عدد المتدربين	4	7	13	10	5	1	40

والمطلوب إيجاد الوسط الحسابي .

الحـل :

لحساب الوسط الحسابي يجب اتباع الخطوات التالية :

١- حساب مراكز الفئات x .

٢- ضرب مركز الفئة في التكرار المناظر له (xf) ، وحساب المجموع $\sum xf$

٣- حساب الوسط الحسابي بتطبيق العلاقة

$$\bar{x}=\frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$$

فئات الوزن	التكرارات f	مراكز الفئات x	$x\ f$
40-	4	$\frac{40+50}{2}=\frac{90}{2}=45$	$4\times45=180$
50-	7	$\frac{50+60}{2}=\frac{110}{2}=55$	$7\times55=385$
60-	13	$\frac{60+70}{2}=\frac{130}{2}=65$	$13\times65=845$
70-	10	$\frac{70+80}{2}=\frac{150}{2}=75$	$10\times75=750$
80-	5	$\frac{80+90}{2}=\frac{170}{2}=85$	$4\times85=340$
90-100	1	$\frac{90+100}{2}=\frac{190}{2}=95$	$1\times95=95$
المجموع	$\sum f=40$		$\sum x\cdot f=2595$

Page 17 / 20

إذن الوسط الحسابي لوزن المتدربين هو :

$$\bar{x} = \frac{\sum_{i=1}^{6} x_i f_i}{\sum_{i=1}^{6} f_i} = \frac{2595}{40} = 64.87 \ k.g$$

أي أن متوسط وزن المتدرب يساوي $64.87 \ k.g$

طريقة الوسط الفرضي : تستخدم هذه الطريقة عندما تكون قيم مفردات العينة اعداد كبيرة ويصعب التعامل معها وخصوصاً عند عدم توفر الحاسبة تفي هذه الطريقة بالغرض

$$\bar{y} = a + \frac{\sum di}{n}$$

حيث a الوسط الفرضي

$\sum$ di مجموع الانحرافات عن الوسط الفرضي

n= عدد المشاهدات

مثال // اذا كانت اوزان ستة طلاب من طلبة كلية الصيدلة كالآتي :-

yi= 85 , 67 , 80 , 75 , 60 , 55

اوجد الوسط الحسابي؟

الحل :

نختار وسط فرضي وليكن = 75

yi	di =yi -a
85	10
67	-8
80	5
75	0
60	15
55	$\frac{-20}{-28}$

$\bar{y}$ =a+$\sum$di

$\bar{y}$= 75 - $\frac{28}{6}$

$\bar{y}$= 75-4.67 =70.33

ملاحظة// لا يتغير الوسط الحسابي بتغير الوسط الفرضي

ب – الوسط الحسابي في حالة البيانات المبوبة

$$\bar{y} = \frac{\sum fiyi}{\sum fi}$$

مثال // اوجد الوسط الحسابي للبيانات التالية التي تبين توزيع (100) طالب من طلبة كلية الصيدلة حسب صفة الوزن ، اوجد الوسط الحسابي لوزن طلبة الكلية

$\bar{y}=\frac{\sum fiyi}{\sum fi}$

$\bar{y}=\frac{6754}{100}$ =67.54

الفئات الوزن كغم	التكرار fi عدد الطلبة	yi مركز الفئات	fiyi
60 – 62	5	61	305
63- 65	15	64	960
66 – 68	45	67	3015
69 – 71	27	70	1890
74 - 72	8	73	584
	100		6754

جـ - الوسط الحسابي المرجح أو الموزون Weighted Mean

من الناحية العملية هناك الكثير من الحالات تكون بعض المفردات اكثر اهمية من الاخرى مما يتوجب الامر اخذ ذلك بنظر الاعتبار لدى حساب الوسط الحسابي ، فمثلا عند حساب معدل درجات الطالب فأن الامر يستوجب الاخذ بنظر الاعتبار عدد الساعات الاسبوعية المخصصة لكل درس وهذا يعني ترجيح المفردات بأوزان معينة تمثل اهمية كل منها وعنده ادخال اهمية المفردات في حساب الوسط الحسابي فأن عندئذ يسمى الوسط الحسابي المرجح وبتعبير آخر لكل قيمة من المشاهدات (yi) وزن خاص يتناسب مع اهميتها (wi) فالوسط الحسابي لهذه القيم يحسب كما يلي:-

$$\bar{y}_w=\frac{\sum wiyi}{\sum wi}$$

حيث ان $\bar{y}_w$ الوسط الحسابي الموزون

wi اوزان وأهمية (المفردة)

yi قيمة المشاهدة

مثال // اذا كانت درجات احد الطلبة في الصف الاول في كلية الصيدلة في الدروس المقررة في هذه المرحلة حسب الساعات الاسبوعية المحدد لكل درس ، المطلوب حساب معدل الطالب ؟

الدرجة	عدد الساعات

2	62
2	80
2	75
3	88
3	84
3	84
3	86
3	90

الحل //

wiyi	الاهمية wi	الدرجة yi
124	2	62
160	2	80
150	2	75
264	3	88
252	3	84
258	3	84
270	3	86
1478	18	

$$\frac{\sum wiyi}{\sum wi} = \frac{1478}{18} = 80.714$$

الوسط الحسابي الموزون في حالة البيانات المبوبة :-

$$\bar{y}_w = \frac{\sum wifiyi}{\sum wifi}$$

yi= مركز الفئة
Fi= التكرار
wi= الاهمية

مثال // اوجد الوسط الحسابي الموزون للبيانات التالية التي تمثل انتاج معمل الادوية في سامراء من الادوية بالطن وعدد المكائن العاملة وعدد ساعات العمل ؟

wifiyi	wifi	yi	عدد ساعات العمل wi	عدد المكائن العاملة fi	فئات الانتاج بالطن
72	24	3	6	4	2 – 4

125	25	5	5	5	4 – 6
252	36	7	6	6	6 – 8
108	12	9	4	3	8 – 10
88	8	11	4	2	10 - 12
645	105			20	

$$\overline{y}_w = \frac{\sum wifiyi}{\sum wifi} = \frac{645}{105} = 6.134 \quad \text{طن}$$

خصائص الوسط الحسابي :-

1- مجموع انحرافات القيم عن وسطها الحسابي = صفر

2- مجموع مربعات انحرافات القيم عن وسطها الحسابي = اقل ما يمكن

اقل ما يمكن = $\sum(yi - \overline{y})$

3- يأخذ الوسط الحسابي بعين الاعتبار جميع القيم في حسابه

4- يتأثر الوسط الحسابي بالقيم الشاذه او المتطرفة لان الوسط الحسابي يأخذ بنظر الاعتبار جميع القيم .

5- هناك صعوبة في حساب الوسط الحسابي في حالة الفئات المفتوحة لانه من الصعب تحديد مراكز الفئات وهذه المشكلة تحل بتحديد مراكز الفئات

الوسيط Median:-

يعرف الوسيط بأنه القيمة التي تمثل المرتبة الوسطى عندما ترتب القيم قيد الدرس تصاعدياً او تنازلياً وهذا يعني ان نصف القيم تقل عن قيمة الوسيط والنصف الاخر يزيد عنها

أ- ايجاد الوسيط لبيانات غير مبوبة

1- يتم ترتيب القيم تصاعدياً او تنازلياً

2- اذا كان عدد القيم فردي (n) فالوسيط يكون القيمة التي ترتيبها $\frac{n+1}{2}$ واذا كان عدد القيم زوجي (n) فالوسيط هو الوسط الحسابي للقيمتان اللتان ترتيبهما $\frac{n}{2}$ و $\frac{n}{2}+1$

مثال // اوجد الوسيط للبيانات التي تمثل مستوى الهيموغلوبين في دم (7) رجال ملغم / ديسلتر

yi=11 , 12 , 13 , 12 , 13 , 11 , 14

الحل :- 1- ترتب البيانات ترتيب تصاعدي 11 , 11 , 12 , 12 ,13 , 13 ,14

3- ايجاد رتبة الوسيط

بما ان عدد القيم (n) = فردي

رتبة الوسيط = $\frac{n+1}{2} = \frac{7+1}{2} = \frac{8}{2} = 4$

Me = 12

مثال // اوجد الوسيط للبيانات التي تمثل مستوى الهيموغلوبين في دم 8 رجال ملغم/ديلتر

yi= 11 , 12, 13 ,12 , 13 , 11 , 14 , 10

الحل :- 1- ترتيب البيانات تصاعدياً 10 , 11 , 11, 12 , 12 , 13 , 13 , 14

2- ايجاد رتبة الوسيط

بما ان عدد القيم زوجي = 8

فالوسيط هو الوسط الحسابي للقيمتان التي ترتيبهما $\frac{n}{2}$ ، $\frac{n}{2}+1$

$$4 = \frac{8}{2} = \frac{n}{2}$$

$$5 = \frac{8}{2}+1 = \frac{n}{2}+1$$

$$Me = \frac{12+12}{2} = \frac{24}{2} = 12$$

ب – ايجاد الوسيط لبيانات مبوبة :

$$Me = Li + [\frac{\frac{\sum fi}{2} - F}{fi}] \times C$$

Li = هي الحد الادنى الحقيقي لفئة الوسيط

$\frac{\sum fi}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد زوجي

$\frac{\sum fi+1}{2}$ = رتبة الوسيط في حالة مجموع التكرارات عدد فردي

F= التكرار المتجمع الصاعد عند بداية فئة الوسيط

C= طول الفئة (طول فئة الوسيط)

fi = التكرار المتجمع الصاعد عند نهاية فئة الوسيط - التكرار المتجمع الصاعد عند بداية فئة الوسيط

مثال // اوجد الوسيط للبيانات التالية التي تبين توزيع 100 طالب من طلبة كلية اصيدلة حسب صفة الوزن

الفئات	التكرار fi
60 – 62	5
63 – 65	15
66 – 68	45
69 – 71	27

72 - 74	8
	100

جدول الفئات	التكرار المتمجع الصاعد	
اقل من 60	0	
اقل من 63	5	
اقل من 66	20	الحد الادنى لفئة الوسيط←
اقل من 69	65	الحد الادنى لفئة الوسيط ←
اقل من 72	92	
اقل من 74	100	

50

1- ايجاد التكرار المتجمع الصاعد

2- ايجاد رتبة الوسيط رتبة الوسيط $= \frac{\sum fi}{2} = \frac{100}{2} = 2$

Li= الحد الادنى لفئة الوسيط = 66

الحد الادنى الحقيقي لفئة الوسيط = 65.5

طول الفئة C = 3

$$Me = Li + [\frac{\frac{\sum fi}{2} - F}{fi}] \times C$$

$$Me = 65.5 + [\frac{50 - 20}{45}] \times 3$$

$$Me = 65.5 + \frac{30}{45} \times 3$$

$$Me = 65.5 + 0.67 \times 3$$

$$Me = 65.5 + 2.01$$

$$Me = 67.51$$

ملاحظات عن الوسيط :-

1- يستخدم الوسيط كمقياس للنزعة المركزية بدلاً عن الوسط الحسابي عندما تكون هناك قيمة شاذة في التوزيع .

2- يستخدم في حالة الفئات المفتوحة

3- الوسيط قليل الحساسية للمتغيرات التي تحدث في قيم البيانات الاصلية لانه يهتم بالقيم الواقعة في المنتصف ويهمل الاطراف على عكس الوسط الحسابي الذي يعتبر شديد الحساسية لانه يأخذ بعين الاعتبار جميع القيم في حسابه .

4- يمكن استخدامه في حالة المتغيرات الوصفية التي لا تعبر عنها في الارقام كما هو الحال في ترتيب الاشخاص وفقاً لخصائصهم .