المحاضرة الاولى

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.

تعريف علم الاحصاء (Statistics):-

هو العلم الذي يهتم بجمع البيانات وتصنيف وتبويب وتحليل البيانات واستخلاص النتائج والاستنتاجات منها .

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 للأشخاص المختلفين فهو للعامة جداول واعداد عن البيانات الحياتية اما المعنى الاصطلاحي للإحصاء 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 للأفراد والجماعات المختلفة والدول على حد السواء وفي الحقيقة ان الانتصار العظيم في نزول الانسان على القمر ما كان 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 كثيرة منها على سبيل المثال العلاقة بين العمر وضغط الدم وكذلك العلاقة بين الوراثة والبيئة وتأثيراتهما على تكوين الفرد .

3- في الصحة العامة :- يستخدم الاحصاء لدراسة الامراض السارية ونسبة زيادتها ونقصها في المجتمع وكذلك دراسة حالة المعوقين والوفيات ونسبة الزيادة في السكان .

4- في الكيمياء :- يستخدم الاحصاء لتحليل البيانات المتعلقة بتكرير النفط ومعرفة نسبة مكوناته وكذلك دراسة العلاقة بين الغازات او الفلزات او العمليات الكيمياوية من ناحية تحليل البيانات المتعلقة بها وكذلك التجارب الكيمياوية في اعداد بحوث الماجستير والدكتوراه والبحوث العلمية الاخرى وغيرها من التجارب في مجال النفط والمعادن وجمع البيانات المتعلقة بها ودراستها دراسة احصائية لغرض الاستفادة منها في اعداد خطط التنمية الصناعية والبتروكيمياوية .

المعالم والرموز الاحصائية Statistical parameter and notation

$<$ اكبر $\geq$ اكبر او يساوي

$>$ اصغر $\leq$ اصغر او يساوي

$\sum$: sigma تقرأ Sum وهي دلالة للجمع

$\overline{y}$: الوسط الحسابي للعينة

M : الوسط الحسابي للمجتمع

S^2 : تباين العينة S الانحراف المعياري للعينة

6^2 : تباين المجتمع 6 الانحراف المعياري للمجتمع

$S\overline{y}$: الانحراف المعياري لمتوسط العينة او الخطأ القياس

S^2p : التباين المشترك

Sp : الانحراف المعياري المشترك

$C.V$: معامل الانحراف

$d.f$: درجة الحرية

$S.S$: مجموع المربعات

$H0$: فرضية العدم

$H1$: الفرضية البديلة

α : مستوى المعنوية

$M.D$:الانحراف المتوسط

nPr تباديل r من n

nCr توافيق r من n

r = معامل الارتباط

B = معامل الانحدار

F $Fcal$ المحسوبة

F $Ftabl$ الجدولية

t $t\ cal$ المحسوبة

t $t\ tabl$ الجدولية

x^2 كاي مربع كاي

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 (g, pg , mg) 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 تكريت 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 تكريت هي متغير تأخذ مدى معين لمجتمع طلبة جامعة تكريت ، 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 تكريت او عدد ردهات المرضى في مستشفى تكريت التعليمي .

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 10 $= 5\,C2 = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1\ (3 \times 2 \times 1)}$

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 = $\frac{\text{مجموع المرضى}}{\text{مجموع العينة}}$

$$10 = \frac{100}{10} + \quad \text{93, 83,73, 63, 53 , 43, 33, 23 , 13,3 ,}$$

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 الطب البيطري ونحتاج اخذ عينة من مجتمع كلية طب الاسنان عددها (20 عنصراً) علما ان مجتمع كلية الطب البيطري 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

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

	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

المحاضرة الثانية

عرض وتلخيص البيانات Data presentation and Summarization

بعد ان تجمع البيانات وفق الاساليب التي ذكرناها تبدأ مرحلة عرض وتلخيص البيانات مستندة على طبيعة البيانات والهدف الاساسي من جمعها وهناك ثلاث طرق اساسية لعرض وتلخيص البيانات وهي :-

1- طريقة العرض الجدولي Tabular Presentation
2- طريقة العرض البياني Graphic Presentation
3- حساب المقاييس الاحصائية

1- طريقة العرض الجدولي Tabular Presentation

تعرض البيانات على شكل جداول

A- الجداول البسيطه :- وهي الجداول التي توزع فيها البيانات حسب صفة واحدة ويتألف الجدول من عمودين الاول يمثل تقسيمات الصفة او الظاهرة الى فئات او مجموعات والثاني يبين عدد المفردات التابعة لكل فئة او مجموعة ، فالجدول التالي يبين توزيع (100 طالب) من طلبة كلية الطب البيطري حسب صفة الوزن .

الفئات الوزن (كغم)	عدد التكرار الطلبة
60 – 62	5
63 – 65	15
66 – 68	45
69 – 71	27
72 – 74	8
	100

B- الجداول المركبة :- وهي الجداول التي توزع فيها البيانات حسب صفتين او ظاهرتين او اكثر في نفس الوقت والجداول لصفتين تتألف من الصفوف وتمثل فئات او مجاميع احد الصفتين والاعمدة تمثل فئات او مجاميع الصفة الاخرى اما المربعات التي تقابل الصفوف والاعمدة فتحوي على المفردات او التكرارات المشتركة ، والجدول التالي يبين توزيع (100) طالب م طلاب كلية طب الاسنان حسب صفة الوزن والطول :

الوزن (كغم) / الطول(سم)	51-60	61-70	71-80	Total
121 – 140	20	6	4	30
141 – 160	2	40	10	52
161 – 180	2	6	10	18
Total	24	52	24	100

ان هذا النوع من الجداول تسمى بجداول التوزيع التكراري Frequency Distribution وهي عبارة عن تلخيص وترتيب بيانات المتغير التي سبق وان جمعت ووضعت مقسمة الى عدد من المجاميع كل منها تسمى بالفئة (class) وهذه الفئات قد تكون مرتبة تصاعدياً او تنازلياً حسب طبيعة البيانات ويسمى توزيع قيم الظاهرة العددية حسب الفئات بالتوزيع التكراري وقد تكون فئات التوزيع التكراري متساوية في الطول اوغير متساوية حسب طبيعة الدراسة ومتطلباتها

ملاحظات على 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 الحدود 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 :①- C = الحد الاعلى – الحد الادنى + 1

= 62 – 60 + 1 = 3

②- طول الفئة يساوي الفرق بين الحدود الدنيا لفئتين متتاليتين = 63 – 60 =3

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

④- طول الفئة يساوي الفرق بين مركز فئتين متتاليتين = 64 – 61 =3

⑤- طول الفئة الفرق بين الحدود الحقيقية الدنيا لفئتين متتاليتين = 62.5 – 59.5 =3

⑥- طول الفئة الفرق بين الحدود الحقيقية العليا لفئتين متتاليتين = 65.5 – .62 =3

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

$$\text{yi} = \frac{60+62}{2} + = 61$$

الحدود الحقيقية :

الحد الادنى الحقيقي ① الحد الادنى للفئة – 0.5

الحد الادنى الحقيقي = 60 – 0.5 = 59.5

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

الحد الادنى الحقيقي = 61 - $\frac{1}{2}$ × 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 + $\frac{1}{2}$ طول تلك الفئة

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