المحاضرة الأولى

أولا: طبيعة علم الإحصاء The Statistics Science

أن كلمة إحصاء في الماضي كان يقصد منها العد والحصر، حتى سمي الإحصاء بعلم العد The science of counting. وان لفظة الإحصاء في الانكليزية (statistics) كانت تستعمل في أوربا للدلالة على أعمال وحسابات الدولة في شؤون الحرب و الضرائب وعدد السكان و المواليد و الوفيات و الإنتاج الزراعي .....

تعريف علم الإحصاء :-

هو العلم الذي يهتم بجمع وتنظيم وتلخيص البيانات وعرضها وتحليلها بأسلوب علمي للحصول منها على المعلومات اللازمة لاتخاذ القرارات المناسبة.

ويمكن تقسيم علم الإحصاء إلى قسمين رئيسيين:

1- الإحصاء الوصفي : Descriptive statistics ويشمل الطرق الإحصائية المستعملة في وصف مجموعة معينة من البيانات وتتضمن هذه الطرق الإحصائية على أساليب جمع البيانات data بصورة قياسية ورقمية numerical measurements يتم جمعها وتبويبها وتلخيصها وعرضها باستخدام المقاييس الإحصائية المناسبة لها .

2- الإحصاء ألاستنتاجي (الاستدلالي) Statistical Inference ويشمل الطرق الإحصائية التي تهدف الى عمل استنتاجات حول المصدر الذي جمعت منه البيانات أي يهتم باستخلاص النتائج من العينة وتعميمها على المجتمع واستخدامها في اتخاذ القرارات المناسبة .

البيانات و الرموز الإحصائية :

عند تجميع البيانات (Data) عن ظاهرة معينة فإننا نرمز لها بالرمز (y) وان كل مفردة او مشاهدة لهذه الظاهرة نرمز لها بالرمز (y_i) . فمثلا عند دراسة أطول الطلبة في احد الجامعات فأننا نرمز لصفة الطول بالرمز(y) وطول أي طالب بالرمز (y_i) وتسمى بالمفردة او المشاهدة Observation تختلف قيمة المشاهدة (y_i) من طالب لأخر ولهذا نقول ان (y) متغير Variable ولذلك يمكن ان نقسم الصفات من الناحية الإحصائية الى نوعين :

1- الصفات المتغيرة : وهي الصفات التي تتغير من شخص لأخر او من نبات لنبات او من جماد لأخر كما في صفة طول الطلاب السابقة الذكر وصفة العمر واللون والحرارة وحاصل النبات .... الخ.

2- الصفات الثابتة :وهي صفات غير متغيرة ضمن الزمان والمكان المحددين لها.

لها . فمثلاً طلبة الدراسات العليا لكلية التربية في جامعة ديالى للعام الدراسي 2006-2007 م .

وتنقسم المتغيرات الى قسمين :

أ- المتغيرات النوعية (الوصفية) Qualitative Variables

وهي تلك الصفات او الظواهر التي لا يمكن قياسها مباشرة بالأرقام العددية مثل صفة لون العيون (اسود ، ازرق ، بني) والحالة الاجتماعية (اعزب او متزوج) او (غني وفقير) (ذكر وانثى).....الخ .

أ.د عقيل خالد حمدان المفرجي - جامعة ديالى - كلية الزراعة

ب-المتغيرات الكمية : Quantative Variaables

وهي تلك الصفات او الظواهر التي يمكن قياسها مباشرة بأرقام عددية او بواسطة آلات القياس مناسبة فبيانات الطول يمكن ان تقاس بوحدات الطول وبيانات الوزن تقاس بوحدات الوزن .

وتنقسم المتغيرات الكمية من حيث القياس الى مجموعتين :-

1- متغيرات مستمرة (متصلة) : Continous Variables

المتغير المستمر هو المتغير الذي يأخذ المشاهدة أو المفردة فيه أي قيمة رقمية فلوفرضنا بأن أطوال طلبة الصف تتراوح بين 150.5 و 170 فسنقول أن اطوال الطلبة هي (170 أكبر من y أكبر من 150.5 أي أن y ممكن أن يأخذ أي قيمة بينهما .

2- متغيرات متقطعة او منفصلة (غير مستمرة): Discrete Variables

المتغير المتقطع هو المتغير الذي يأخذ قيماً متباعدة أو متقطعة فمثلا لدينا (20) نبات ونريد معرفة عدد النباتات المصابة فيها ، فمثلاً المصابة هي (5) ولا نعطي (5.5) او(4.6) او (3.7) نبات بل يجب ان نعطي نقلات كاملة أي من 4 مثلاً الى 5 او ان نرجع الى 3 وهذا هو المتغير المتقطع .

المجتمع والعينة : Population and sample

1- **المجتمع** : **Population** هو جميع القيم أو المفردات التي يمكن أن يأخذها المتغير ،وتعبير أخر : هو عدد الوحدات التي تنطوي تحت صفة واحدة أو أكثر مشتركة بحيث تميز عن باقي المجتمعات تميز كامل . فمثلاً أذا كانت دارستنا متعلقة بأطول طلبة جامعة معينة فأن المجتمع في هذه الحالة هو اطوال جميع الطلبة في تلك الجامعة .و المجتمع أما ان يكون :

أ- محدودا: Finite Population وفيه يمكن حصر عدد الوحدات الانتاجية في مصنع ما ليوم معين .

ب- غير محدود: Infinite Population

وهو المجتمع الذي يصعب فيه او يستحيل حصر عدد مفرداته مثل مجتمع سمك معين في نهر ديالى . او عدد البكتريا في حقل ما، ويرمز للمجتمع بالرمز N .

2- العينة: Sample

هي جزء من المجتمع ، وهي عبارة عن مجموعة من القيم او المشاهدات اختيرت بطريقة ما من المجمع وهي اصغر وحدة نعمل عليها بحيث تمثل المجتمع .

حجم العينة : ويرمز لها بالرمزn وهو عدد مفردات العينة .

اسباب اخذ العينات :

1 ـ تقليل الكلفة لان الشمول الكامل للمجتمع يكون ذا كلفة باهضة .

2ـ دقة البيانات : اذ ان التعداد الشامل قد يقلل من دقة البيانات لان احتمال وقوع الاخطاء يزداد مع زيادة حجم المجتمع .

أ.د عقيل خالد عمران المدرس في جامعة ديالى - كلية الزراعة

3ـ تلف وحدات العد المدروسة بعد انتهاء التجربة ومن هنا لا يمكن استعمال كل وحدات المجتمع بل نأخذ عينات ممثلة للمجتمع .

4ـ الحاجة الانية : احيانا تجبرنا الظروف على اخذ العينة حتى لو كانت الإمكانية متوفرة ، مثلا لو طلب منا تقدير انتاج العراق من الحنطة لكي تستطيع وزارة التجارة تقدير ما تستورده منه عند ذاك نحن لا نستطيع ان نحصي كل ارض العراق لأنه يكلف جهد ومال ووقت وكذلك انتظار موسم حصاد المحصول لجميع المساحات المزروعة وعليه من الافضل اخذ عينة من مكان معين وبصورة عشوائية وبدون تحيز ونقدر الحاصل فيها ثم نعمم ذلك على جميع المساحة المزروعة ، اذن الحاجة الانية يجب ان تعطي فكرة معينة الآن وليس في أي وقت اخر.

طرق اخذ العينة :

1ـ الاجتهاد الشخصي : وعيبها انه لا يوجد ضمان في كونها ممثلة للمجتمع لأننا قد نخطئ بسبب الاعتماد على اجتهادات شخصية واحيانا يكون الاجتهاد الشخصي به تحيز مقصود او غير مقصود وبحسن نية .

2ـالعينة العشوائية البسيطة : Simple random sample

وهذه تتحقق بشرطين :

ا ـ وجود تجانس مقبول بين مفردات المجتمع فيما يتعلق بمتغير معين .

ب ـ الاختيار كيفما اتفق : وكمثال للشرط الاول نلاحظ ان تجانس اعمار الطلبة في المرحلة الاولى لكلية التربية في جامعة ديالى لا يوجد بينهم من هو في عمر عشر سنوات ولا في عمر خمسون سنة ،وعندها ناخذ عينة كيفما اتفق أي ان كل فرد في المجتمع يجب ان تكون له نفس الفرصة والاحتمال مع الذين شملوا بالعينة .

3 ـ العينة الطبقية :اذا كان المجتمع غير متجانس فيجب ان نقسمه الى طبقات ثم نأخذ عينة ممثلة لكل طبقة ومثال على ذلك لو كان مجتمع مكون من نوعين من الابقار (فريزيان و جيرسي) ففي هذه الحالة لا يمكن اخذ عينة واحدة من الاثنين وبصورة مجتمعة لانهما مختلفان بأدرار الحليب او العمر او الوزن فقد تظهر ابقار فريز بأن في العينة أكثر من ابقار الجيريسي لذلك علينا ان نقسم هذا المجتمع الى طبقات ، بحيث يوجد تجانس مقبول في كل طبقة والطبقات تختلف بعضها عن بعض . وهذا مثال اخر :

انواع الابقار	عدد الأبقار	
N_1 —— A	200	N = 1 000
N_2 ——B	300	
N_3 —— C	500	

حيث ان N تشمل المجتمع الكلي للأبقار المختلفة الغير متجانسة وعددها (1000) .

N_1,N_2,N_3 هي طبقات ويوجد داخل كل طبقة تجانس مقبول .ولو اردنا اختيار (100)

بقرة كعينة وهذا يساوي n من الـ (1000) بقرة وهي تساويN.فان الطبقة C أي (N3) تشكل نصف المجتمع (N) اذن يجب ان تكون العينة المأخوذة نصفها الطبقة N_3(C) وهكذا بالنسبة للطبقات الاخرى . بالنسبة للطبقة A (200) بقرة هي مجتمع قائم بحد ذاته ولكن لا يرمز لها بالرمز N بل N_1 أي هي مجتمع من اصل مجتمع اكبر وهكذا بالنسبة للطبقة B.

n_1,n_2,n_3 = حجم العينات (كنسبة) الذي يؤخذ من الطبقات الاولى والثانية والثالثة وكما يلي :

$$n_1 = \frac{N_1}{N} = \frac{200}{1000} = 20\%$$

$$n_2 = \frac{N_2}{N} = \frac{300}{1000} = 30\%$$

$$n_3 = \frac{N_3}{N} = \frac{500}{1000} = 50\%$$

إذن :

$$n = n1 + n2 + n3 = 0.20 + 0.30 + .050 = 100$$

وبعد ذلك يجب اختبار العينة داخل كل طبقة بطريقة المعاينة العشوائية البسيطة بسبب حصول التجانس داخل كل طبقة. وتوزيع المجتمع إلى طبقات يجب أن يكون الشكل الطبيعي لذلك أن C, B, A هي الطبقات من حيث متغير الحليب الذي تدرسه فقط – ولكن عندما تكون الطبقات الثلاثة متجانسة لمتغير أخر مثل العمر الذي لا ندرسه لذلك يهمل

الرموز الاحصائية Statistical no tattily.

كما مر بنا سبقاً سنرمز للمتغير بالرمز y ولكل قيمة له بالرمز yi ، فلو كانت أعمار 5 طلاب كاللاتي : 16، 22 ، 24 ، 18 ، 20 سنه فتكتب : yi = 20 ، 18 ، 24 ، 22 ، 16

أي ان : وهي قيمة المشاهدة الأولى للمتغير Y_1 20

و قيمة المشاهدة الثانية للمتغير $Y_2 = 18$

و قيمة المشاهدة الثالثة للمتغير $Y_3 = 24$

و قيمة المشاهدة الرابعة للمتغير $Y_4 = 22$

و قيمة المشاهدة الثانية للمتغير $Y_5 = 16$

وهكذا 1..... إلى أن نصل إلى القيمة الأخيرة:

تعني القيمة الخامسة

أي القيمة الأخيرة (n=5) للمتغير أو المشاهدة الأخيرة .

وعادة يرمز لمجموعة قيم المتغير بالرمز $\sum_{i=1}^{n} y_i$

فالرمز $\sum$ هو حرف اغريقي يسمى Sigma

أ.د عثمان خالد علوان الفرجي - جامعة ديالى - كلية الزراعة

أي مجموع او Summation of والرقمان i,n هما حدي المجموع وعلية فالرمز $\sum_{i=1}^{n} y_i$ يقرأ كالاتي :

مجموع قيم y مبتدأ من المشاهدة الاولى وحتى المشاهدة الاخيرة .

أي

$$\sum_{i=1}^{n} yi = y_1 + y_2 + y_3 + ... + yn$$

وللاختصار والسهولة فقد يكتب الرمز السابق بدون ذكر حدي المجموع أي فقط (∑yi) إذا لم يكن هناك خوفا من الالتباس .

وهناك مجموع جزئي مثل : $\sum_{i=3}^{5} yi$

أي مجموع المشاهدات الثالثة والرابعة والخامسة فقط : $\sum_{i=3}^{5} yi = y_3 + y_4 + y_5$

ويرمز لمجموع مربعات جميع المشاهدات بالرمز $\sum_{i=1}^{n} y_i^{\,2}$ ويساوي :

$$\sum_{i=1}^{n} y_i^{\,2} = y_1^{\,2} + y_2^{\,2} + y_3^{\,2} + ... + y_n^{\,2}$$

ويرمز لمربع مجموع المشاهدات بالرمز $\left(\sum_{i=1}^{n} y_i\right)^2$ ويساوي :

$$\left(\sum yi\right)^2 = \left(y_1 + y_2 + y_3 + ... + y_n\right)^2$$

كما يرمز لحاصل ضرب متغيرين مثل x وy بالرمز $\sum x_i y_i$ وهو يساوي :

$$\sum x_i\, y_i = y_1\, x_1 + y_2\, x_2 + y_3\, x_3 + ... + y_n\, x_n$$

ويرمز لحاصل ضرب مجموعتين لقيم متغيرين بالرمز $(\sum x_i)\,(\sum y_i)$ وهو يساوي :

$$(\sum xi)\,(\sum yi) = (x1 + x2 + x3 + ... + xn)(y1 + y2 + y3 + ... + yn)$$

سؤال : اذا علمت ان قيم كل من المتغيرين x و y هي كالاتي :

x_i= 2,6,3,1

y_i= 3,9,6,2

اوجد قيمة كل مما ياتي :ـ

1- $\sum_{i=1}^{n} y_i$ 2 – $\sum_{i=2}^{3} y_i$ 3 – $\sum y_i^{\,2}$ 4 – $(\sum y_i)^2$ 5 – $\sum x_i y_i$

6 – $(\sum x_i)(\sum y_i)$ 7 – $\sum (y_i - x_i)^2$ 8 – $\sum (x_i - 3)(y_i - 5)$