محاضرة (1)

مبادئ نظرية العينات

1-1 مقدمة

تهتم نظرية العينات بدراسة العلاقة بين المجتمع والعينات المسحوبة منه فيما يسمى بالاستدلال الإحصائي statistical inference . هناك عدة طرق لأخذ العينات من المجتمع ، لاستخدامها في الاستدلال الإحصائي ومن أشهر هذه الطرق هي العينة العشوائية وهي العينة التي تكون لكل مفرده من مفردات المجتمع نفس فرصة الاختيار في العيينة . فمثلاً نستعين بعينه مسحوبة من المجتمع لتقدير معالم هذا المجتمع مثل متوسطه أو تباينة أو غير ذلك . أو أعطاء عينه من المرضى بارتفاع الضغط، مثلاً دواء معين ثم قياس ضغطهم قبل وبعد تناولهم لهذا الدواء لمعرفة ما إذا كان هذا الدواء مفيد في خفض الضغط أم لا.

أي مجموعات من المفردات تشترك في صفه أصفات وتكون موضوع دراسة أو بحث فإن هذه المجموعة يطلق عليها إحصائيا مجتمع الدراسة (أو اختصاراً المجتمع Population). والمجتمع قد يكون مجموعة ما من البشر أو أشجار أنواع معينه من الفاكهة أو الحيوانات الزراعية أو إنتاج دولة ما لسلع معينه خلال فترة زمنية محدده...الخ.

والمجتمع قد يكون محدوداً إذا كان يمكن حصر عدد أفراده مثل سكان مدينة ما أو طلاب مرحلة تعليمية معينة ، وقد يكو ن المجتمع غير محدود (لانهائي) إذا كان لا يمكن حصر عدد أفراده مثل النجوم والكواكب أو الكائنات الحية بمياه المحيطات والأنهار وعند دراسة صفة ما أو صفات معينه لمجتمع ما ، فإن البيانات الإحصائية عن تلك الصفة أو الصفات تجمع بأحد أسلوبين

أولاً: أسلوب الحصر الشامل (census) وفيه تجمع البيانات عن كل مفرده من مفردات المجتمع، وهذا الأسلوب يتطلب وفرة في الوقت والمال والمجهود الفني وتزداد هذه المتطلبات وتتضاعف كلما ازداد حجم المجتمع (عدد أفراد المجتمع). وهذا الأسلوب لا يتبع عادة إلا في حالة التعدادات التي تجريها الدول و تدعمها بإمكانيات ضخمه مثل تعدادات السكان والتعدادات الصناعية والتعدادات الزراعية.

الثاني : أسلوب المعاينة (Sampling method) وفيه يتم جمع البيانات عن جزء من مفردات المجتمع يختار بطريقة أو بأخرى ويطلق عليه عينه (Sample) ثم بعد ذلك يتم تعميم نتائج الدراسة على المجتمع بأكمله. أي أن إسلوب المعينة يقصد به دراسة خصائص المجتمع من خلال دراسة عينه مسحوبة منه ، ونجاح هذا الإسلوب يعتمد على أن تحمل العينة أقصى درجة من دقة التمثيل للمجتمع المسحوبة منه.

من الأفضل في بعض الحالات الحصول على معلومات دقيقة عن طريق التعداد ال تام أو الحصر الشامل لجميع عناصر المجتمع، لكن لاستخدام أسلوب المعاينة فوائد جمة مقارنة بالتعداد الشامل يرد بيانها في الفقرة التالية.

2-1 بعض مزايا أسلوب المعاينة

يتميز أسلوب المعاينة عن أسلوب الحصر الشامل بمزايا عديده منها:

1 - **يؤدي استخدام العينات العشوائية إلى خفض تكاليف الدراسات الميدانية بسبب صغر حجم العينة بالنسبة إلى حجم المجتمع وهو ما يؤدي إلى تخفيض الأعباء الإدارية والفنية التي تتطلبها أي دراسة ميدانية.**

2 - **يتحقق وفر واضح في الوقت الذي ينفق في دراسة ميدانية على أساس عينة بدلاً من الحصر الشامل وتتضح أهمية الوقت عندما نقوم بدراسة ظاهرة تتغير بمرور الوقت ، وعينة قد يترتب على دراسة تلك الظاهرة في المجتمع كله بجمع البيانات من جميع مفردات المجتمع أن يمر وقت بديل فتكون البيانات والنتائج وقت ظهورها غير مطابقة لواقع المجتمع وتصبح النتائج ذات قيمه محدودة بعد أن فقدت عنصر المطابقة مع واقع الظاهرة وتوزيعها الحالي لمجتمع ، والتعدادات الدورية للسكان وبسبب ضخامة حجم العمل بها تستغرق وقتاً طويلاً حتى تصبح نتائجها جاهزة ومنشورة وقد يطول هذا الوقت إلى أكثر من ثلاث أو أربع سنوات حتى مع استخدام أحدث أجهزة الحاسبات الأليه الضخم ة ، ويكون على الباحثين مستخدمي هذه النتائج مراعاة الوقت الذي ينقض بين تاريخ إجراء التعداد وتاريخ نشر نتائجه وتعديل هذه النتائج في حدود ذلك .. وهذا دفع الكثير من الدول إلى تعزيز نتائج التعدادات الدورية للسكان بنتائج تعدادات تجري بين كل تعدادين متتاليين على أساس العينة.**

3 - **في المجتمعات غير المحدودة (اللانهائية) مثل مجتمع الكائنات الحية في البحار والمحيطات لا يمكن أن تتم الدراسة على أساس الحصر الشامل ولكن لابد وأن تتم الدراسة بأسلوب المعاينة.**

4 - **أيضاً هناك بعض الاختبارات لابد وأن تتم بأسلوب المعاينة لأن إجراء مثل هذ ه الاختبارات على أساس الحصر الشامل يؤدي إلى تلف المادة المختبرة أو هلاكها .. فاختبار صلاحية شحنه من المفرقعات مثلاً لابد وأن يتم على أساس العينة وبالمثل تحليل دم المرضى يتم على أساس عينه.**

3-1 أقسام العينات

تنقسم العينات عادة إلى قسمين رئيسين وهما عينات عشوائية وعينات غير عشوائية، وفيما يلي تفصيل لكل قسم منها.

1-3-1 العينات العشوائية

هي تلك العينات التي يتم اختيار مفرداتها حسب خطه إحصائية لا يكون فيها للباحث أو لمفردات العينة دخل في اختيار أي مفرده فيها ، حيث يتم الاختيار باستخدام أساليب معينة تلعب الصدفة خلالها الدور الأول في اختيار المفردة ولكن بشرط أن يتحقق لجميع المفردات احتمال ثابت ومحدد للاختيار. والعينات العشوائية إذا ما تم اختيارها بالطريقة العلمية السليمة والمناسبة يمكن أن تكفل درجه عالية من دقة التمثيل للمجتمعات المسحوبة منها لذلك فهي الوسيلة الأساسية في حالة البحوث العلمية الدقيقة .. من أهم أنواع العينات العشوائية مايلي.

(أ) العينة العشوائية البسيطة Simple random sample:

ويلجأ إليها الباحث في حالة ما إذا كان مجتمع الدراسة ليس كبيراً ويحمل قدراً من التجانس بين المفردا ت للصفة أو الصفات موضع الدراسة . والعينة العشوائية البسيطة تستغل فرص متكافئة لمفردات المجتمع للدخول في العينة ولكن المفردات التي تدخل في العينة تكون عن طريق الصدفة البحتة. والاختيار العشوائي يتم يدوياً عن طريق بطاقات متماثلة في الحجم واللون أو عن طريق جداول الأعداد العشوائية أو عن طريق الحاسب الآلي. ولكي يتحقق ذلك فإن الأمر يتطلب تحديد مفردات المجتمع تحديداً كاملاً ويكون هذا التحديد على شكل قائمة (أو خريطة) تضم كل مفردات المجتمع وهذه القائمة تسمى الإطار (Frame) ولا يجوز الاختيار العشوائي إلا من المفردات التي يضمها الإطار.

(ب) العينة المنتظمة Systematic sample:

اختيار هذه العينة يتطلب وجود إطار للمجتمع كما في حالة العينة العشوائية البسيطة بحيث يعطى لكل مفرده من مفردات المجتمع رقماً متسلسلاً داخل الإطار ، ثم نختار مفردات العينة من الإطار بحيث يكون الرقم المتسلسل لكل مفردة يبعد بعداً ثابتاً منتظماً عن رقم المفردة السابقة لها وكذلك رقم المفردة اللاحقة لها. فمثلاً إذا كان لدينا مجتمعاً حجمه 2000 مفرده ونريد اختيار 100 عينه منتظمة حجمها مفرده فإننا نقسم الإطار إلى فترات منتظمة طول كل فترة = $\frac{2000}{100}=20$ مفردة ومن داخل مفردات الفترة الأولى (1 – 20) يختار مفرده واحدة عشوائياً ولتكن رقم 14 مثلاً وبناء على رقم تلك المفردة يتحدد باقي مفردات العينة المنتظمة فتكون هي المفردات ذات الأرقام 34 ، 54 ، .... ، 1974 ،1994.

والعينة المنتظمة كثيرة الاستعمال في التطبيقا ت العملية لقلة تكاليفها وقلة الأخطاء التي ترتكب في اختيار مفردات العينة فضلاً عن سهولة إجرائها. ولكن أهم عيوب المعاينة المنتظمة هو عدم

صلاحيتها إذا ما وجدت علاقة دورية مع ترتيب العناصر في القائمة وكان طول الفترة بين عناصر العينة مساوياً لطول الدورة أو إحدى مضاعفاتها.

(جـ) العينة العشوائية الطبقية Stratified random sample:

ويلجأ إليها الباحث في حالة ما إذا كان مجتمع الدراسة واضحاً به فئات (طبقات) بحيث أن التجانس أو التقارب داخل كل طبقة من طبقات مجتمع الدراسة أكبر من التجانس داخل المجتمع ككل (أي أن التشتت داخل المجتمع ككل أكبر من التشتت داخل كل فئة من فئاته على حده). في هذه الحالة يجب على الباحث مراعاة أن الطبقة داخل العينة بنفس نسبة و جودها داخل المجتمع (وأحياناً يوضع في الاعتبار عناصر أخرى مثل التشتت داخل الطبقة أو عنصر التكلفة لجمع البيانات عن الطبقة). بعد أن يتم تحديد عدد المفردات التي يجب سحبها من كل طبقة للدخول في العينة فإن هذه المفردات يتم سحبها عشوائياً من داخل الطبقة ومجموع هذه المفردات تكون العينة الطبقية العشوائية.

(د) العينة متعددة المراحل أو العنقودية clustered sample:

يلجأ إليها الباحث عندما يكون مجتمع الدراسة كبير جداً ومتناثراً على مساحات شاسعة تكلف الكثير من الوقت والجهد في التنقل بينها عند جمع البيانات، أيضاً في حالة عدم وجود إطار يضم جميع مفردات المجتمع فيستحيل الاختيار العشوائي مباشر من المجتمع . لهذا يلجأ الباحث إلى أخذ العينة على مراحل متعددة متتالية. في المرحلة الأولى يتم تقسيم المجتمع إلى عدد محدد من وحدات المعاينة الكبيرة الحجم ومنها يختار بعضها عشوائياً ثم يتلو ذلك كمرحلة ثانية تقسيم الوحدات المختارة عشوائياً من المرحلة الأولى إلى وحدات أقل منها في الحجم ثم يختار بعضها عشوائياً.. وهكذا تتابع مراحل التقسيم والاختيار العشوائي ، وعدد هذه المراحل ليس ثابت بل يتوقف على طبيعة مجتمع الدراسة وإمكانيات الباحث .. في المرحلة الأخيرة يصل الباحث إلى وحدات المعاينة التي سيجمع عنها بيانات البحث ويطلق عليها وحدات المعاينة الأولية.

1-3-2 العينات غير العشوائية:

هي تلك العينات التي لا تكفل لجميع مفردات المجتمع احتمال ثابت ومحدد للاختيار، وغالباً يتدخل الباحث في عملية الاختيار بصو رة أو بأخرى ... ومن أ هم أنواع العينات غير العشوائية:

(أ) العينة العمدية أو المقصودة Purposive sample:

يلجأ الباحث إلى هذه الطريقة فيما إذا كان مجتمع الدراسة كبير جداً وكانت إمكانياته لا تسمح له إلا بدراسة عينة حجمها صغير جداً بالنسبة لمجتمع الدراسة، في هذ ه الحالة يتعمد الباحث اختيار مفردات معينة كعينة لمجتمع الدراسة يرى بخبرته السابقة أن هذه العينة يمكن أن تعطي تمثيلاً مقبولاً لمجتمع الدراسة.

مثلاً إذا أراد باحث دراسة خصائص اقتصادية أو اجتماعية معينه عن ريف دولة ما ، وكانت إمكانياته المالية والإدارية لا تسمح له بعينة سوى سكان قرية واحدة ، فإنه في هذه الحالة إذا ما تم اختيار القرية عشوائياً من بين آلاف القرى بتلك الدولة فإن الصدفة قد تأتي بقرية بعيدة في خصائصها (من حيث الظاهرة موضوع الدراسة) عن خصائص معظم قرى تلك الدولة ... كأن تأتي بالصدفة قرية ساحلية معظم سكانها من الصيادين أو قرية قريبة من مشروع صناعي ضخم يستوعب في قواه العاملة معظم سكانها.. هذه القرية أو تلك قد يأخذ النمط المعيشي لسكناها طابعاً خاصاً – نابعاً عن ظروفها الخاصة – بعيداً عن النمط المعيشي المعتاد لبقية القرى، لذلك فأي منها لا يمكن أن يعطي تمث يلاً مقبولا لريف تلك الدولة . لهذا فإن الباحث وعلى ضوء خبراته السابقة يتعمد اختيار قرية معينة يرى أنها – من وجهة نظره الشخصية- يمكن أن تمثل الريف . وهذه الطريقة غير علميه وغالباً يتم اللجوء إليها في حالة البحوث التمهيدية.

(ب) العينة الحصصيه Quota sample:

وهي نوع خاص من العينات غير العشوائية وتستخدم كثيراً في معاينة الرأي العام (على سبيل المثال عمليات استطلاعية الرأي العام التي يقوم بها معهد جالوب قبل إجراء انتخابات الرئاسة في الولايات المتحدة الأمريكية)... في هذه الطريقة يقسم المجتمع موضوع الدراسة إلى طبقات بالنسبة إلى صفات أو خصائص معينة ويتم العمل على تمثيل كل طبقة منها في العينة بنسبة وجودها في المجتمع الأصلي (وعلى سبيل المثال في حالة دراسة الدخل لمنطقة ما ورؤى أن يكون حجم العينة المطلوبة 100 فرد مثلاً عندما يريد الباحث أن يقوم جامعو ا البيانات بالحصول على البيانات من 20 موظفاً، 45 من العمال الحرفيين ، 35 من ذوي الأعمال الحرة .. وتترك الحرية لجامعي البيانات في اختيار الأفراد المطلوبة فيها حدود المواصفات الموضوعة لكل طبقة من الطبقات المذكورة.

واضح أنه رغماً من أن هذه الطريقة في ظاهرها مماثلة للعينة الطبقية العشوائية.. إلا أنه في الحالة الأخيرة (العينة الطبقية العشوائية) يكون اختيار المفردات عشوائياً من داخل كل طبقة ولا يترك لجامع البيانات حرية اختيار المفردات من كل طبقة والذي قد يترتب عليه تميزاً كبيراً.

عموماً.. يلجأ الباحث إلى العينة الحصصيه إذا كان من الم رغوب فيها اظهار النتائج في وقت قصير مع التغاضي عن توافر درجة دقة عالية بتلك النتائج.

1-4 أخطاء البيانات الإحصائية

تتعرض البيانات الإحصائية التي يتم جمعها إلى نوعين من الأخطاء:

1-4-1 خطأ التميز

وهو ينتج عن مصادر متعددة، منها أخطاء في تصميم البحث أو التجربة أو أخطاء فنيه أثناء جمع البيانات أو خلال العمليات الحسابية التي تتم على البيانات المتجمعة .. أخطاء التميز تزداد بازدياد الفروق بين الإمكانيات (المادية والفنية) اللازم توافرها لضمان أقصى درجة دقة ممكنه وبين الإمكانيات الفعلية المتاحة للباحث.

أخطاء التميز قد توجد في البيانات التي يتم جمعها بأسلوب الحصر الشامل وقد توجد أيضاً في البيانات التي يتم جمعها بأسلوب المعاينة، ولكنها إن وجدت فهي غالباً أكبر في الحالة الأولى (الحصر الشامل) مما هي عليه في الحالة الثانية (المعاينة) باعتبار أن حجم العمل في تلك الحالة يكون أقل وبالتالي قد يسهل توفير الإمكانيات اللازمة وتجنب الأخطاء الفنية.

1-4-2 خطأ المعاينة العشوائية أو خطأ الصدفة

وهو الخطأ الناتج عن فروق الصدفة بين مفردات المجتمع التي دخلت العينة وبين تلك المفردات التي لم تشأ الصدفة أن تدخل العينة .. وخطأ الصدفة يمكن تقليل قيمته إذا ما تم اختيار العينة بالطريقة المناسبة وإذا ما كان حجم العينة مناسباً لحجم المجتمع وخصائصه.

1-5 المعالم والإحصاءات

المقاييس الإحصائية التي تحسب من بيانات مجتمع الدراسة بأكمله يطلق عليها معالم المجتمع(Parameters of population) ، أما المقاييس الإحصائية التي تحسب من بيانات عينه مسحوبة من مجتمع الدراسة فيطلق عليها إحصاءات(Statistics) ويعتبر كل إحصاء منها بمثابة تقدير أو قيمه تقديرية لمعلمة المجتمع المناظر، فيكون المتوسط الحسابي المحسوب من بيانات

العينة تقدير لمعلمة المجتمع المناظرة وهي المتوسط الحسابي المسحوب منه هذه العينة وهكذا .. ويجب ألا يغيب عن الأذهان بأن حساب قيمة المتوسط الحسابي من بيانات العينة ليس هدفاً في حد ذاته ولكن وسيلة للتعرف على المتوسط الحسابي للمجتمع موضوع الدراسة .. وهكذا بالحال بالنسبة لباقي المقاييس الإحصائية التي تحسب من العينة.

للتفرقة بين المعالم والإحصاءات يجب أن نرمز لكل منها برموز تختلف عن رموز الأخرى، على سبيل المثال يرمز للمتوسط الحسابي للمجتمع بالرمز μ بينما يرمز للمتوسط الحسابي للعينة بالرمز $\bar{x}$ أيضاً للانحراف المعياري للمجتمع بالرمز σ ، بينما يرمز للانحراف المعياري للعينة بالرمز s وهكذا.

1–5–1 توزيعات المعاينة: Sampling Distributions

نفرض أننا أخذنا عينه حجمها n من مجتمع ما ، ثم سحبنا منها بعض المقاييس الإحصائية مثل المتوسط الحسابي ، التباين ، ... فإن كل مقياس من هذه المقاييس يعتبر متغير عشوائي في ذاته يختلف من عينه إلى أخرى – هذا المتغير العشوائي يخضع لتوزيع معين – هذا التوزيع يسمى بتوزيع العينة. فمثلاً نقول أن توزيع المعاينة للمتوسط الحسابي وهو عبارة عن توزيع جميع المتوسطات الحسابية للعينات المأخوذة من نفس هذا المجتمع ذات الحجم n ، وكذلك فإن توزيع المعاينة للتباين هو توزيع جميع التباينات المحسوبة من عينات لها نفس الحجم n ومأخوذة من نفس المجتمع ، وهكذا ... .

2–5–1 توزيعات المعاينة للأوساط: Sampling Distributions of Means

نفرض أننا سحبنا عينه حجمها n من مجتمع لانهائي ، القيمة المتوقعة له تساوي μ والانحراف المعياري هو σ فإن المتوسط الحسابي $\bar{X}$ يخضع لتوزيع ما ، متوسط هذا التوزيع وانحرافه المعياري هما

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad , \quad \mu_{\bar{X}} = \mu \qquad (4\text{-}1)$$

وفي الحالة التي يكون فيها المجتمع الأصلي المسحوبة منه العينة مجتمع طبيعي (ويرمز له بالرمز $N(\mu , \sigma^2)$ فإن توزيع المعاينة للمتوسط الحسابي $\bar{X}$ يكون في هذه الحالة توزيع طبيعي أيضاً له نفس المتوسط الأصلي μ ولكن انحرافه المعياري يساوي $\sigma / \sqrt{n}$ أي بمعنى أن ،

$$X \sim N(\mu,\sigma^2) \quad \Rightarrow \quad \overline{X} \sim N(\mu,\frac{\sigma^2}{n}) \qquad (4\text{-}2)$$

ومن ثم يكون

$$z = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \sim N(0,1) \qquad (4\text{-}3)$$

لا تخضع للتوزيع الطبيعي ولكنها تتوزع توزيع $\overline{X}$ أما إذا كان المجتمع غير طبيعي فإن حيث أن $(n \geq 30)$ الكبيرة n يكون قريباً من التوزيع الطبيعي لقيم

$$z = \frac{\sqrt{n}(\bar{x}-\mu)}{\sigma} \xrightarrow[n \to \infty]{as} N(0,1) \qquad (4\text{-}4)$$

وتعتبر النتيجة السابقة الهامة جداً في الإحصاء وخاصة في التطبيقات العلمية وتسمى نظرية والتي تنص على أنه في حالة العينات الكبيرة Central Limit Theorem النهاية المركزية ، حيث أن $\frac{\sigma^2}{n}$ و μ يخضع للتوزيع الطبيعي بالمعاملات $\overline{X}$ الحجم فإن المتوسط الحسابي هما متوسط وتباين المجتمع الأصلي بغض النظر عن شكل توزيع المجتمع الأصلي. ومن σ^2, μ بصرف النظر عن توزيع المجتمع الأصلي.(4-3) الكبيرة تتحقق العلاقة n ثم فإنه لقيم

هو المتوسط الحسابي لعينه عشوائية مسحوبة من مجتمع لانهائي $\overline{X}_1$ كذلك فإنه إذا كان هو المتوسط الحسابي لعينة عشوائية $\overline{X}_2$ ، وكان σ_1 وانحرافه المعياري هو μ_1 متوسطه هو وكانت العينتين مستقلتين σ_2 وانحرافه المعياري μ_2 مسحوبة من مجتمع لانهائي آخر متوسط فإن المجموع الجبري لمتوسط العينتين يخضع لتوزيع المعاينة بالمعاملات

$$\mu_{(\bar{x}\pm\bar{x}_2)} = \mu_1 \pm \mu_2 \quad \text{and} \quad \sigma^2_{(\bar{x}_1\pm\bar{x}_2)} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad (4\text{-}5)$$

هما حجم العينة الأولى والثانية. n_1, n_2 حيث

يخضع لتوزيع طبيعي أيضاً $(\bar{x}_1 \pm \bar{x}_2)$ وإذا كان المجتمعين الأصليين طبيعيين فإن وعليه فإنه في هذه الحالة(4-5)بالبارامترات المعطاة في

$$z = \frac{(\bar{x}_1 \pm \bar{x}_2)-(\mu_1 \pm \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad (4\text{-}6)$$

أما إذا كان أحد المجتمعين أو كليهما لا يتوزع توزيعاً طبيعياً فإن $(\bar{x}_1 \pm \bar{x}_2)$ لا يتوزع توزيعاً طبيعياً كذلك ، ولكن لقيم n_1, n_2 الكبيرة فإنه طبقاً لنظرية النهاية المركزية السابقة فإن $(\bar{x}_1 \pm \bar{x}_2)$ يتوزع توزيعاً قريباً من التوزيع الطبيعي وبذلك يمكننا استخدام نفس العلاقة في (4-6) حالة العينات الكبيرة.

3–5–1 توزيع المعاينة للتباين: Sampling Distribution of The Variance

إذا كان $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ هو تباين عينه عشوائية حجمها n مأخوذة من مجتمع متوسطه μ وتباينه σ^2 وعزمه الرابع حول المتوسط هو μ_4 فإن

$$\mu_{s^2} = \sigma^2 \quad \text{and} \quad \sigma^2{}_{s^2} = \frac{\mu_4 - \sigma^2}{n-1} \qquad (4\text{-}7)$$

وبالتالي فإن $\mu_4 = 3\sigma^2$ وإذا كان المجتمع طبيعي فإن

$$\sigma^2_{s^2} = \left(\frac{2}{n-1}\right)\sigma^2 \qquad (4\text{-}8)$$

نلاحظ هنا أن S^2 لا تتوزع طبيعي حتى ولو كان المجتمع طبيعي ، ولكنه يتوزع توزيع . أما إن كان المجتمع الأصلي يخضع $(n \geq 100)$ الكبيرة n قريب من التوزيع الطبيعي وذلك لقيم بعدد χ^2 يخضع لتوزيع يسمى توزيع مربع كاي $(n-1)s^2/\sigma^2$ للتوزيع الطبيعي فإن المتغير درجات حرية يساوي $n-1$ أي أن .

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1) \qquad (4\text{-}9)$$

ويعتبر توزيع مربع كاي من التوزيعات الهامة في الإحصاء التطبيقي ودالة كثافته هي

$$f(y) = y^{\frac{v-1}{2}} e^{-y/2} \quad , \quad y > 0 \qquad (4\text{-}10)$$

حيث v هي عدد درجات الحريه للتوزيع وتعتبر هي المعامل الوحيد له ويتضح من شكل الدالة أنها داله متصلة وتقع بأكملها فوق النصف الموجب لمحور السينات ، منحنى هذه الداله غير متماثل ويعتبر من المنحنيات موجبة الالتواء ويقل التواؤه (وبالتالي يقترب من التماثل) كما زادت درجات الحريه v. وتكون القيمة المتوقعة لهذا التوزيع هي v و تباينه هو $2v$ أي بمعنى أن

$$E(y) = \mu_y = v$$

$$V(y) = \sigma^2 = 2v \qquad (4\text{-}11)$$

فإذا كان s_1^2 هو تبلين عينه عشوائية حجمها n_1 مسحوبة من مجتمع طبيعي $N(\mu_1, \sigma_1^2)$ ، وكان s_2^2 هو تباين عينه عشوائية أخرى حجمها n_2 ومسحوبة من مجتمع طبيعي آخر $N(\mu_2, \sigma_2^2)$ وكانت العينتان مستقلتان فإن المتغير:

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1) \qquad (4\text{-}12)$$

حيث أن $F(n_1 - 1, n_2 - 1)$ تسمى بتوزيع F بدرجتي الحريه $n_1 - 1$ و $n_2 - 1$ و دالة الكثافة الإحتماليه للمتغير y الذي يخضع لتوزيع F بدرجتي الحريه v_1, v_2 تعطى بالصورة:

$$f(y) = \frac{y^{\frac{v_1}{2}-1}}{(v_1 y + v_2)^{\frac{v_1+v_2}{2}}} \quad , \quad y > 0 \qquad (4\text{-}13)$$

وكما يتضح من الداله في (4-13) أن المنحنى يقع بالكامل في النصف الموجب لمحور السينات ، وهو أيضاً غير متماثل وموجب الالتواء ولكن يقترب من التماثل كلما زادت درجات الحريه v_1, v_2 كما في حالة توزيع χ^2.

ذكرنا سابقاً أنه إذا كان $\bar{X}$ هو المتوسط الحسابي لعينه حجمها n مأخوذة من مجتمع طبيعي بالمعاملات μ,σ^2 فإن

$$z=\frac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma} \quad \sim \quad N(0,1)$$

هذا إذا كانت قيمة σ معلومة ، ولكن في حالة ما إذا كانت قيمة σ غير معلومة فإننا نستخدم بدلا منها S الانحراف المعياري للعينة ولكن في هذه الحالة يصبح المتغير $\frac{\sqrt{n}\,(\bar{x}-\mu)}{S}$ يخضع لتوزيع يعرف بتوزيع t ستيودنت $t-\text{student}$ بدرجات حريه $n-1$ ، أي أن

$$t=\frac{\sqrt{n}\,(\bar{x}-\mu)}{s} \quad \sim \quad t(n-1) \qquad (4\text{-}14)$$

دالة الكثافة لتوزيع t بدرجات حريه v تعطي بالصورة:

$$f(t)= \quad (1+\frac{t^2}{v})^{-\frac{v+1}{2}} \qquad (4\text{-}15)$$

وهو يشبه في ذلك المنحنى الطبيعي القياسي $N(0,1)$ وهو توزيع متماثل حول محور y ولكنه أقل تحدباً من التوزيع الطبيعي القياسي ولكنه يقترب من التوزيع الطبيعي كلما زادت درجات الحريه.

وإذا كان $\bar{X}_2$ و S_1^2 هما المتوسط الحسابي والتباين لعينه حجمها n_1 مأخوذة من مجتمع طبيعي متوسط هو μ_1 وكان $\bar{X}_2$ و S_2^2 هما المتوسط الحسابي والتباين لعينه أخرى حجمها n_2 ومأخوذة من مجتمع طبيعي آخر له المتوسط μ_2 وكانت العينتان مستقلتان فإن المتغير

$$t=\frac{(\bar{X}_2 \pm \bar{X}_2)-(\mu_1 \pm \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \sim \quad t(n_1+n_2-2) \qquad (4\text{-}16)$$

حيث أن $S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ يسمى بالتباين المشترك للعينتين The Pooled Variance.