الفصل الخامس

المعاينة وتوزيعات المعاينة

Sampling and Sampling Distributions

في علم الإحصاء، يتم عموماً عمل الاستدلال عن متغيرات المجتمع موضع الدراسة بناء على تحليل بيانات العينة. من هنا، يتضح أهمية العينات في علم الإحصاء. وبالطبع فإن العينة يجب أن تكون ممثلة للمجتمع إذا تم الاستدلال بناء على إحصاءات تم حسابها من عينة تم اختيارها بدقة.

وهناك نوعان من أساليب المعاينة يمكن استخدامها لاختيار العينات، العينات المبنية على الاحتمال والعينات غير الاحتمالية والتي سيتم التطرق لها بالتفصيل.

العينات المبنية على الاحتمال (العينات الاحتمالية)

Probability-Based Samples

تكون العينة المختارة مبنية على الاحتمال إذا كان أسلوب المعاينة المستخدم يعطي فرصة معلومة لكل مفردة من مفردات المجتمع بالاختيار في العينة. وباستخدام هذا الأسلوب في المعاينة فإنه يمكننا استخدام التعبيرات الاحتمالية في تحليل بيانات العينة. وعند اختيار العينة نفسها فإنه يوجد العديد من الطرق تعطي كل طريقه نوع مختلف من العينات ويعتمد اختيار طريقه المعاينة على مجموعة من العوامل تتمثل فيما يلي:

١- مدى تجانس المجتمع

نظراً لاشتمال المعاينة على أخطاء، فإن طريقة اختيار العينة تعتمد على خصائص المتغير المراد دراسته في المجتمع. فإذا كان المتغير أو الصفة توجد تقريباً في جميع مفردات المجتمع فإن لدينا مجتمعاً متجانساً ويمكن استخدام أسلوب معاينه بسيط مثل العينة العشوائية أو المنتظمة. من جهة أخرى، إذا كانت الصفة متركزة في مستويات مختلفة من خلال مجموعات مختلفة من المجتمع فإنه يجب تحديد تلك المجموعات ثم أخذ عينة من كل مجموعة على حدة. وهذه الحالة تعني استخدام أساليب للمعاينة أكثر تعقيداً مثل العينات الطبقية أو العنقودية والتي تساعد في قياس الاختلافات في المجتمع التي يمكن أن تتسبب في إعطاء قيم مختلفة للمتغير موضع الدراسة.

٢- درجة الدقة المطلوبة

بافتراض تساوي العوامل الأخرى، نرغب في اختيار أسلوب المعاينة الذي يعطي أقل تباين ممكن لإحصاءة العينة لحجم العينة المختارة. من ناحية أخرى، يجب أن نزيد حجم العينة لتقليل خطأ المعاينة والذي يعبّر عن الفرق بين قيمة الإحصاءة أو متغير العينة وقيمة معلمة المجتمع أو متغير المجتمع. وتزيد قيمة هذا الفرق؛ نتيجة للصدفة أو التغيرات العشوائية في اختيار الوحدات الأولية للعينة. وبناء على ذلك، فإن كبر تباين إحصاءة العينة تعني كبر خطأ المعاينة ومن ثمَّ انخفاض الثقة في العينة المختارة.

٣- تكاليف المعاينة

على الرغم من إمكانية زيادة درجة الثقة في العينة بزيادة حجمها إلا أن التكاليف المصاحبة لذلك سوف تزيد بزيادة الحجم. حيث ترتبط تكاليف المعاينة مباشرة بحجم العينة بينما الموثوقية عادة لا ترتبط بشكل مباشر.

وفي الغالب فإن حجم العينة يجب أن يزيد للحصول على زيادة قليله في الموثوقية وبذلك نستطيع تحقيق الكفاءة والذي يعني أن أسلوب المعاينة أكثر كفاءة

مقارنة بأسلوب آخر إذا تم الحصول على نفس درجه الموثوقية بتكاليف أقل. وعليه فإن أسلوب العينة العنقودية عادة أكثر كفاءة من أسلوب العينة العشوائية البسيطة عندما يكون المجتمع المراد أخذ عينة منه منتشر في مساحة واسعة مما يعني تحمل تكاليف سفر عند استخدام أسلوب المعاينة العشوائية البسيطة.

وتصميم العينة يعني تصميم العينة الاحتمالية وبذلك سوف يتم التطرق لأنواع العينات الأكثر استخداماً وتقييم كل منها على حدة بناء على معياري الكفاءة والموثوقية.

المعاينة العشوائية البسطة Simple Random Sampling

في العينة العشوائية البسيطة تتميز جميع عناصر مجتمع المعاينة بتساوي الفرصة في الاختيار. وكذلك في حالة اختيار عينه عشوائياً حجمها n فإن كل عينة محتمله بذلك الحجم لها أيضاً نفس الفرصة في الاختيار.

وتعتبر العشوائية العامل الأساسي المحدّد في حالة العينة العشوائية البسيطة. ولكن ليس بالسهولة تحقيق شرط العشوائية دائماً فعلى سبيل المثال عند الرغبة في استبيان آراء الطلاب في جامعة كبيره حول الزراعة لا نستطيع اختيار عينة عشوائية بسهولة عن طريق مقابلة الطلاب عند باب اتحاد الطلاب حتى نحصل على مجموعة مشاهدات كافية. حيث إن معظم الطلاب ربما لن يذهب لذلك المكان إطلاقاً ولذلك فإن احتمال مقابلتهم سيكون مساوياً للصفر. وعليه فإن العينة المتحصل عليها بهذه الطريقة ليست عينة عشوائية.

وفي الغالب يتم اختيار طريقه المزج الكامل Thorough mixing وذلك لتحقيق العشوائية في العينة، وذلك بترقيم جميع الوحدات في المجتمع أو تعريفها وبذلك لا

يكون هناك فرق بين الوحدة التي تم اختيارها والوحدات الأخرى. حيث يتم وضع الأرقام في سلة كبيره ثم خلطها معاً، يلي ذلك تحديد حجم العينة ثم اختيار الوحدة الأولى ثم الثانية وهكذا حتى نحصل على العدد المطلوب. أما في حالة الاختيار بالإرجاع فإنه يتم خلط الأرقام معاً بعد كل عمليه إرجاع قبل الاختيار حتى نحافظ على نفس الاحتمالية لجميع الوحدات.

تجدر الإشارة إلى أن السحب بدون إرجاع يؤدي لزيادة احتمال الظهور للوحدات المتبقية بعد السحب ؛ نظراً لانخفاض العدد الإجمالي للوحدات بعد عملية السحب. ولكن جميع الوحدات المتبقية لها نفس الفرصة في الاختيار مما يحافظ على بقاء احتمالية الحصول على عينه متساوية حجمها n لأي مزيج يتم اختياره من الوحدات والتي يمكن سحبها من مجتمع الدراسة.

ويمكن تبسيط عمليه اختيار العينة باستخدام جدول الأرقام العشوائية والموضح في الملحق (الجدول رقم ١). يحتوي جدول الأرقام العشوائية على أرقام صحيحة موجبة من رقم صفر إلى رقم ٩ ويمكن تركيبها باستخدام طريقة المزج الكامل. حيث يتم إعداد عشرة كروت مرقمه من الصفر إلى ٩ في سلة معينة وسحب واحدة وتسجيلها كأول رقم في الخانة الأولى في جدول الأرقام العشوائية ثم إعادة الكرت ثم السحب مرة أخرى وتسجيل الرقم في الخانة الثانية في الجدول وهكذا حتى نكمل الجدول. وبالطبع فإن الطريقة الأكثر كفاءة هي استخدام برامج الحاسب الآلي لإنشاء الأرقام العشوائية.

باستخدام جدول الأرقام العشوائية نستطيع توليف أي رقمين للحصول على أرقام عشوائية من ٠٠ إلى ٩٩ وثلاثة أرقام للحصول على أرقام من ٠٠٠ إلى ٩٩٩ وهكذا. وبصفة عامة فإننا نختار عدد من الخانات المناسبة للعينة من المجتمع الذي

نرغب دراسته. فعلى سبيل المثال، إذا كان لدينا مجتمع يحتوي على ١٠٠٠٠ وحدة ونرغب في اختيار عينه حجمها ١٠٠ ، فإنه يجب أولاً وضع قائمة لوحدات المجتمع وترقيمها من ٠٠٠٠ إلى ٩٩٩٩. بعد ذلك نختار مكان للبداية عشوائياً في الجدول ثم نقرأ أفقياً أو للأسفل أو على القطر ونختار أرقام بالتتابع حتى نحصل على ١٠٠. في حالة تكرار القيم يتم حذفها والاستمرار في الاختيار حتى نحصل على العدد المطلوب. وأخيراً نختار من مجتمع الدراسة الأرقام التي تم الحصول عليها من الجدول. ويمكن استخدام العينة العشوائية البسيطة إذا كان من ا لسهولة تحديد وحدات المعاينة وكان المجتمع موضع الدراسة صغيراً ومتجانساً. أما في حالة المجتمعات الكبيرة فإن هذه الطريقة مكلفة ومستهلكة للوقت لأنه يجب ترقيم جميع الوحدات. أيضاً إذا كانت الوحدات القريبة من بعضها أكثر تجانساً مقارنة بالوحدات البعيدة كما في استبيانات دخل الأسرة فإن طريقه العينة العشوائية البسيطة قد تعطي عينة غير ممثلة للمجتمع.

المعاينة المنتظمة Systematic Sampling

يتم استخدام العينة المنتظمة بكثرة بدلاً من العينة العشوائية البسيطة إذا كان لدينا طريقه للوصول إلى قائمه بمفردات المجتمع. بهذه الطريقة يمكننا الحصول على عينة بأخذ كل K وحدة في المجتمع، حيث تشير إلى رقم صحيح يساوي تقريباً نسبه المعاينة $\frac{N}{n}$. وبالتالي إذا كان لدينا مجتمع حجمه ١٠,٠٠٠ والعينة n= 500 فإن قيمه k تساوي $\frac{10000}{500} = 20$ وعليه يمكننا اختيار العينة بأخذ كل وحدة عشرين في المجتمع. ولكي يكون لكل وحدة في المجتمع نفس الفرصة في الاختيار يجب اختيار نقطه البداية عشوائياً.

ومن ثم فإن المثال السابق سيتم اختيار رقم عشوائياً بين ١ و ٢٠. فإذا كان على

سبيل المثال الرقم المختار هو ٨ فإن ا لعينه ستحتوي على الأرقام ٨، ٢٨، ٤٨ ، إلى آخره. حتى نختار جميع الـ ٥٠٠ قيمه. وجدير بالذكر أن استخدام المعاينة المنتظمة ينجح في حالة وجود قائمة بمفردات المجتمع متاحة، مثل دليل العضوية، قائمة بعدد المزارعين المقترضين من بنك معين، إلى آخره. وأحياناً تستخدم العينات المنتظمة في التحكم الإحصائي بالجودة ، وخاصة عندما نختار العينة من جدول الإنتاج الحالي في الفترات المعتادة.

باستخدام طريقة العينة العشوائية المنتظمة، نحصل على عينة أكثر تمثيلاً مقارنة بطريقه المعاينة العشوائية البسيطة إذا كانت الوحدات في المجتمع المتقاربة أكثر تشابه من الوحدات المتباعدة مثل بيانات الدخل.

من جهة أخرى إذا كانت البيانات تحتوي على فترات أو دورات مخفية بنفس المدى، كما في فترات المعاينة فإن طريقه المعاينة المنتظمة قد لا تكون الاختيار المناسب. فعلى سبيل المثال، لا نستخدم طريقة المعاينة المنتظمة لبحث مبيعات محلات المواد الغذائية اليومية وخاصة إذا كنا نأخذ العينة من المبيعات كل سبعة أيام لأن مبيعات المواد الغذائية لا تتوزع على أيام الأسبوع. ولذلك، لو تم اختيار يوم الثلاثاء سنفقد المتسوقين خلال إجازة نهاية الأسبوع الذين يستفيدون من العروض أو الإجازة من العمل وبالتالي يكون تقديرنا غير صحيح.

المعاينة الطبقية Stratified Sampling

إذا كان لدينا معرفة مسبقة بالمجتمع ويلزمنا تقسيم المجتمع إلى طبقات مختلفة فإن طريقه المعاينة الطبقية أكثر كفاءة من طريقة المعاينة العشوائية البسيطة. ويتم اختيار الطبقات بحيث تحتوي كل طبقة على وحدات لها خصائص متجانسة ومتشابهة أكثر

من المجتمع ككل. بعد ذلك يتم أخذ عينه من كل طبقة باستخدام طريقه العينة العشوائية البسيطة وتكون العينة الإجمالية للمجتمع عبارة عن العينات التي تم أخذها من الطبقات المختلفة. وهناك العديد من الطرق الممكن استخدامها لاختيار عدد الوحدات المختارة في كل عينة فرعية. إحدى هذه الطرق هو استخدام العينة الطبقية المتناسبة بحيث يكون توزيع العينة على كل طبقة متناسب مع عدد وحدات المعاينة منها إلى حجم المجتمع الكلي أي أن عدد الوحدات في العينة المختارة من الطبقة يكون مساوٍ لنفس نسبة عدد وحدات هذه الطبقة في المجتمع. فعلى سبيل المثال، يمكننا أخذ عينة لإنتاجية القطن باستخدام أنواع التربة كطبقات.

الطريقة الأخرى تتم حسب التجانس داخل الطبقات بحيث يتم التوزيع بناء على الانحراف المعياري للطبقة فكلما كانت الطبقة أكثر تجانساً كان الانحراف المعياري أقل ولذلك قل عدد الوحدات المختارة في العينة من تلك الطبقة. وهذا يجعل التناسب عينة طبقية. ومن الأمثلة على ذلك عند الرغبة في أخذ عينة للتكاليف باستخدام حجم المزرعة كطبقات. في هذه الحالة يتم اختيار نسبة صغيره من المزارع الصغيرة مقارنة بالمزارع الكبيرة. وباستخدام الطبقات فإنه يتم تعيين فئات بحيث تكون وحدات المعاينة في كل فئة متماثلة ولكن الفئات مختلفة. ويتم التحكم بنسبة العينة من كل طبقة ولا ندع ذلك للصدفة ولذلك نحقق عينة ممثلة. وإذا كان التباين للخصائص المشاهدة لكل طبقه أصغر مقارنة بكامل المجتمع، كما هو في العادة صحيح، فإن درجه الموثوقيه لحجم عينه محدد سوف تكون عاليه أو الكفاءة لدرجة معينه من الموثوقيه سوف تكون عاليه بسبب اختيار طريقه المعاينة الطبقية. ولذلك فإن طريقة الطبقات مناسبة للتعامل مع المجتمعات غير المتجانسة. في بعض الحالات، يجب أن تقوم بتعيين طبقات متعددة أو طبقات فرعيه وذلك للحصول على أقصى فائدة من الطبقات. فمثلاً يتم تقسيم إنتاجيه

القطن إلى طبقات وفقاً لنوعية الأرض هل هي جافة أم مروية ثم يلي ذلك تقسيم فئات التربة وفقاً لتلك الطبقات.

المعاينة العنقودية Cluster Sampling

المعاينة العنقودية معاكسه تماماً للمعاينة الطبقية. ففي المعاينة العنقودية يتم أولاً اختيار مجموعات من الوحدات الفردية تسمى عناقيد من المجتمع عشوائياً ثم بعد ذلك نختار جميع العينات أو العينات الفرعية من الوحدات وذلك من كل عنقود للحصول على العينة الكاملة. ولتحقيق نتائج جيده فإن الاختلاف بين العناقيد يجب أن يكون صغيراً بقدر المستطاع والاختلاف بين الوحدات داخل العنقود يجب أن يكون كبير. الهدف من ذلك هو تمييز العنقود بصفات مشابهه للمجتمع بحيث يكون أي عنقود عبارة عن عينه ممثلة. عملياً، هذه الشروط ربما يكون من الصعوبة تحقيقها ولكن يجب أخذها في الاعتبار عند تصميم المعاينة العنقودية.

العناقيد تسمى أحياناً وحدات المعاينة الأولية. فإذا كانت جميع الوحدات في العناقيد المختارة مشمولة في العينة، فإن لدينا معاينة من مرحلة واحدة. أما إذا أخذنا عينة فرعيه من وحدات العناقيد فإن لدينا معاينة من مرحلتين. وكذلك هناك معاينة متعددة المراحل. فمثلاً لإجراء دراسة أو مسح لمعرفة وجهة نظر طلاب الكليات الأمريكية تجاه الزراعة، يمكن اختيار الجامعات كعناقيد أو وحدات المعاينة الأولية ثم اختيار الكليات داخل تلك الجامعات كمرحلة ثانية ثم اختيار الطلاب داخل تلك الكليات كمرحلة ثالثة. والميزة الرئيسة للمعاينة العنقودية هو انخفاض تكاليفها مع درجه معينه من الموثوقية. ولذلك، عندما تكون العينة من مجتمع موّزع جغرافياً، كما في المثال السابق حول رأي الطلاب، فإنه يتم أخذ المناطق الجغرافية كوحدات معاينة أولية. وجدير بالذكر أن التباين في العينة العنقودية قد يكون أكبر مقارنة بطرق المعاينة

الأخرى التي تمت مناقشتها، ولكن انخفاض التكاليف يتيح لنا أخذ عينه بحجم أكبر وتكاليف أقل. وعليه فإن كبر حجم العينة سيؤدي لتقليل التباين بدرجه تجعل طريقه المعاينة العنقودية أكثر كفاءة لكل دولار يتم إنفاقه.

المعاينة التتابعية (التسلسلية) Sequential Sampling

تستخدم هذه الطريقة بشكل واسع في رقابة الجودة ويمكن أن تكون مفيدة في دراسات الميزانية الزراعية. حيث تشتمل على اختبار للعينات الصغيرة نسبياً في رقابة الجودة، بناء على قاعدة مخرجات العينة، واتخاذ القرار بقبول أو رفض كامل المنتج. وإذا لم تؤدِ العينة الصغيرة إلى قرار واضح يمكن زيادة حجم العينة. فمثلاً يتم سحب وحدات إضافية حتى نستطيع الوصول للقرار. هذه الطريقة تحافظ على انخفاض تكاليف المعاينة؛ نظراً لأنه في معظم الأحيان يمكننا صنع القرار بناء على النتائج المتحصل عليها من العينات الصغيرة جداً. وفي دراسات الميزانية، يتم مقابلة مجموعات صغيرة من المزارعين بخصوص تطبيقات الإنتاج. فإذا كانوا موافقين يتم تحديد الميزانية بناء على تلك التطبيقات أما إذا كان خلاف ذلك يتم أخذ عينه أكبر حجم حتى نصل إلى اتفاق.

العينات غير الاحتمالية Non-probability Samples

تتميز طريقه العينات غير الاحتمالية بأن جميع مفردات المجتمع ليس لها فرصة متساوية في الاختيار. ويتم استخدام هذه الطريقة بسبب الكفاءة والتكلفة ولكنها لا تعطي نتائج موضوعية؛ كما في طريقة العينات الاحتمالية.

المعاينة الملائمة Convenience Sampling

باستخدام طريقه المعاينة الملائمة يتم اختيار العينة بناء على ما هو مناسب لنا بدون وضع اعتبار لتمثيل المجتمع بصورة ملائمة. وتتمثل ملائمتها في الوقت والتكاليف ورأي الإدارة. ونرى ذلك النوع من ا لمعاينة في التلفزيون المحلي عندما يكون التقرير المقدّم عبارة عن مقابلة مع شخص في الشارع للحصول على رأيه حيال بعض المواضيع في الأخبار. وهذا التصميم يجب أن لا يستخدم لعمل استدلال عن المجتمع.

المعاينة التحكمية (الاجتهادية) Judgment Sampling

في هذه الحالة يتم استخدام الاجتهاد الشخصي بدلاً من بعض الطرق الموضوعية مثل الأرقام العشوائية لاختيار العناصر من المجتمع الممثلة للعينة. فمثلاً، وكيل التعليم ا لمحلي ربما يتم اختياره بواسطة الباحث لاختيار المزارعين يدوياً لمقابلتهم لمعرفه رأيهم حيال التطبيقات الإدارية ؛ نظراً لأن الباحث يرغب في التحدث فقط لأفضل المديرين.

المعاينة الحصصية Quota Sampling

في المعاينة الحصصية يتم تصميم العينة لتكون مشابهة للمجتمع تبعاً لبعض الصفات الرئيسية. ويتم سحب العناصر من المجتمع حتى نصل لحصة تلك الصفة ثم يتم اختيار عناصر تحتوي على صفة مختلفة وهكذا. فمثلاً إذا كان لدينا معرفة بأن المحافظة أو المقاطعة تحتوي ٢٠ ٪ مزارعين قمح، ٦٠ ٪ مزارعين قطن و٢٠ ٪ مربين للثروة الحيوانية فإنه يمكن اختيار عينة حصصية تشتمل على نفس النسب. ونظراً لأن تقديرات هذا النوع من العينات معرض لاختلافات كبيره مقارنة بالمعاينة الاحتمالية فإنه يفضل استخدام طرق المعاينة الاحتمالية.

توزيعات المعاينة Sampling Distributions

يساعدنا توزيع المعاينة لفهم المشاكل المتعلقة بالمعاينة. حيث يمكننا باستخدام توزيع المعاينة معرفة اختلاف الإحصاءات مثل المتوسط والتباين من عينه إلى عينه عند سحب عينات عشوائية متتالية بنفس الحجم ومن نفس المجتمع ويسمى التوزيع الاحتمالي لمثل هذه الإحصاءات بتوزيع المعاينة. ولذلك فإنه سيكون هناك توزيع معاينه للمتوسط وللتباين وهكذا. وتعبّر هذه التوزيعات عن المفهوم الأساسي للاستدلال الإحصائي وهي كذلك مهمة في نظرية القرار الإحصائي.

توزيع المعاينة للمتوسط الحسابي Sampling Distribution of the Mean

تعتمد فكرة توزيع المعاينة على المعاينة المتكررة. ولذلك عند سحب مجموعة من العينات بنفس الحجم من مجتمع معين وحساب المتوسط الحسابي لكل عينه، فإنه يمكن معاملة تلك المتوسطات كمتغير عشوائي. ويعني ذلك أن المتوسط الحسابي يأخذ قيماً متغيرة بناء على العناصر التي تم سحبها من المجتمع بطريقة عشوائية كعينة. وعند معاملة جميع العينات العشوائية باحتمالات متساوية، فإنه يمكن وضع قيم مختلفة للمتوسط الحسابي في شكل توزيع احتمالي عن طريق جمع الاحتمالات المرتبطة بكل قيمة. التوزيع الاحتمالي الناتج يتميز بشكل معين بغض النظر عن شكل التوزيع الاحتمالي للمجتمع الذي تم سحب العينات منه. بالإضافة لذلك فإن القيمة المتوقعة للتوزيع الاحتمالي تساوي قيمة متوسط المجتمع وكذلك تباين التوزيع الاحتمالي ذو علاقة بتباين المجتمع. ويمكن إيضاح ذلك من خلال المثال التالي

نفرض أن لدى مزارع مجموعة عجول صغيرة عمرها سنة واحدة وعددها ستة وكانت أوزانها عند الولادة كالتالي (رطل)

$$X = \{90, 80, 100, 80, 90, 100\}$$

وعليه فإن متوسط المجتمع آنفاً يساوي ٩٠ رطلاً وتباينه σ^2 يساوي ٦٦.٦٧.
وفي حالة الرغبة في اختيار عينة من ذلك المجتمع حجمها ٢ عجل بدون إرجاع فإن عدد العينات الممكن الحصول عليها يمكن حسابه كالتالي :

$$_6C_2 = 6!/(6-2)!(2!) = 15$$

و يمكن حساب المتوسط لكل عينه واستخدام ذلك المتوسط كمتغير عشوائي لإيجاد توزيع المعاينة للمتوسط (الجدول رقم (٥,١).
ويتضح من الجدول أن القيم المحتملة للمتوسط قريبة جداً من متوسط المجتمع. ونظراً لوجود تكرار في قيم المتوسط للعينات المختارة فإنه يمكن إيجاد التوزيع الاحتمالي لقيم المتوسط كما في الجدول رقم (٥,٢). حيث يمكن حساب متوسط التوزيع الاحتمالي في الجدول باستخدام صيغة القيمة المتوقعة كما يلي :

$$E(\overline{X}) = \Sigma \overline{X} \cdot P(\overline{X}) = (80)(1/15) + (85)(4/15) + (90)(15/15) + (95)(4/15) + (100)(1/15) = 1350/15 = 90$$

والتي تساوي نفس القيمة التي تم حسابها لمتوسط المجتمع. وهذه العلاقة متحققة دائماً طبقاً لمفهوم نظرية النهاية المركزية.

نظرية النهاية المركزية Central Limit Theorem

إذا تم اختيار عينات عشوائية بحجم n مشاهدات من مجتمع له متوسط منتهي μ وانحراف معياري σ ، فإن توزيع متوسط العينة $\overline{X}$ يقترب من التوزيع الطبيعي بمتوسط μ وانحراف معياري $\frac{\sigma}{\sqrt{n}}$ عندما يكون حجم العينة n كبير. وكلما زاد حجم العينة n كان التقريب أكثر دقة.

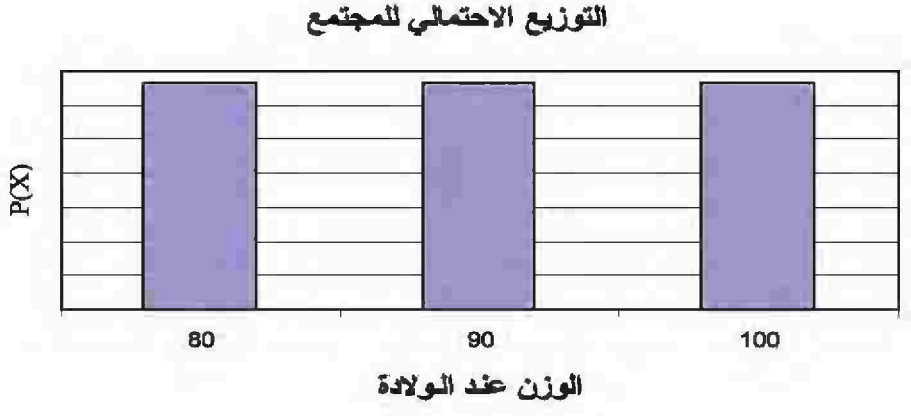
وعند رسم التوزيع الاحتمالي لمجتمع صغير من أوزان مواليد العجول فإنه يأخذ شكل المستطيل كما في الشكل رقم (٥,١)، ولكن توزيع العينة للمتوسط والذي تم إيجاده برسم جميع قيم المتوسطات الممكنة واحتمالاتها المصاحبة يأخذ شكل التوزيع الطبيعي، كما في الشكل رقم (٥,٢).

الجدول رقم (٥,١). متوسطات العينة لجميع التوليفات الممكنة لعينة حجمها اثنين.

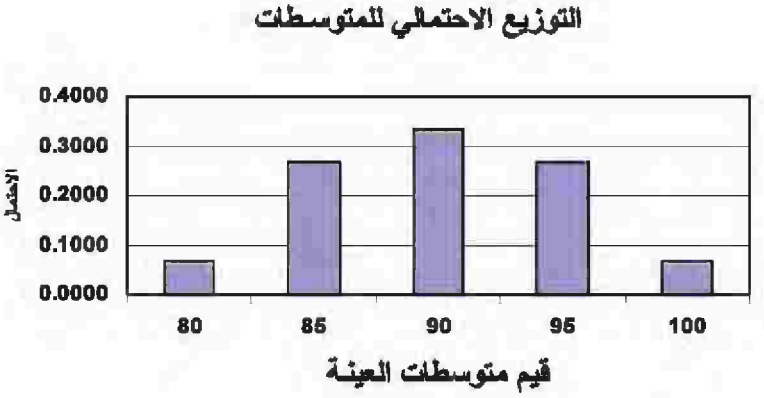
توليفة العينة	قياسات العينة	المجموع (ΣX)	متوسط العينة ($\overline{X}$)
X_1,X_2	٨٠ ،٩٠	١٧٠	٨٥
X_1,X_3	١٠٠ ،٩٠	١٩٠	٩٥
X_1,X_4	٨٠ ،٩٠	١٧٠	٨٥
X_1,X_5	٩٠ ،٩٠	١٨٠	٩٠
X_1,X_6	١٠٠ ،٩٠	١٩٠	٩٥
X_2,X_3	١٠٠ ،٨٠	١٨٠	٩٠
X_2,X_4	٨٠ ،٨٠	١٦٠	٨٠
X_2,X_5	٩٠ ،٨٠	١٧٠	٨٥
X_2,X_6	١٠٠ ،٨٠	١٨٠	٩٠
X_3,X_4	٨٠ ،١٠٠	١٨٠	٩٠
X_3,X_5	٩٠ ،١٠٠	١٩٠	٩٥
X_3,X_6	١٠٠ ،١٠٠	٢٠٠	١٠٠
X_4,X_5	٩٠ ،٨٠	١٧٠	٨٥
X_4,X_6	١٠٠ ،٨٠	١٨٠	٩٠
X_5,X_6	١٠٠ ،٩٠	١٩٠	٩٥

الجدول رقم (٥,٢). التوزيع الاحتمالي لمتوسطات العينة.

متوسط العينة ($\overline{X}$)	التكرار (f)	الاحتمال، $P(\overline{X})$	$\overline{X}.P(\overline{X})$
٨٠	١	١٥/١	١٥/٨٠
٨٥	٤	١٥/٤	١٥/٣٤٠
٩٠	٥	١٥/٥	١٥/٤٥٠
٩٥	٤	١٥/٤	١٥/٣٨٠
١٠٠	١	١٥/١	١٥/١٠٠
الإجمالي	١٥	١=١٥/١٥	٩٠=١٥/١٣٥٠



الشكل رقم (٥,١). التوزيع الاحتمالي لمجتمع أوزان العجول المكتسبة.



الشكل رقم (٥,٢). توزيع المعاينة لمتوسطات العينة لعينات حجمها يساوي اثنين.

الخطأ المعياري للمتوسط Standard Error of the Mean

الانحراف المعياري لتوزيع المعاينة للمتوسط يسمى الخطأ المعياري للمتوسط وبذلك فإنه ليس هناك خلط بينه وبين الانحراف المعياري للمجتمع. الخطأ المعياري مرتبط بالانحراف المعياري للمجتمع ؛ كما في نظرية النهاية المركزية فهو يساوي حاصل قسمة الانحراف المعياري للمجتمع على الجذر التربيعي لحجم العينة.

ولكن عندما يكون لدينا مجتمعاً صغيراً كما في مثال أوزان مواليد العجول، فإنه يجب استخدام معامل تصحيح المجتمع المحدود. وتوضح المعادلة التالية رقم (5.1) طريقة حساب الخطأ المعياري للمتوسط باستخدام معامل التصحيح :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \qquad (5.1)$$

حيث :

σ = الانحراف المعياري للمجتمع.

N = حجم المجتمع.

n = حجم العينة.

$\sqrt{\frac{N-n}{n-1}}$ = معامل تصحيح المجتمع المحدود.

وبتطبيق الصيغة الرياضية (5.1) على مثال أوزان مواليد العجول نحصل على:

$$\sigma_{\bar{x}} = \frac{8 \cdot 167}{\sqrt{2}} \sqrt{\frac{6-2}{6-1}} = (5 \cdot 78)\ (0.8944) = 5.17$$

حيث إن معامل التصحيح يساوي تقريباً ٠,٨٩ ولذلك عمل على تخفيض $\sigma_{\bar{x}}$ نوعاً ما.

الجدول رقم (٥.٣). حساب الخطأ المعياري للمتوسط باستخدام التوزيعات الاحتمالية.

متوسط العينة، $\bar{X}$	$P(\bar{X})$	$\bar{X} - \mu$	$(\bar{X} - \mu)^2$	$(\bar{X} - \mu)^2 . P(\bar{X})$
٨٠	١٥/١	١٠-=٩٠-٨٠	١٠٠	١٥/١٠٠
٨٥	١٥/٤	٥-=٩٠-٨٥	٢٥	١٥/١٠٠
٩٠	١٥/٥	٠=٩٠-٩٠	٠	٠
٩٥	١٥/٤	٥=٩٠-٩٥	٢٥	١٥/١٠٠
١٠٠	١٥/١	١٠=٩٠-١٠٠	١٠٠	١٥/١٠٠
الإجمالي	١=١٥/١٥			$\sigma_{\bar{x}}^2$ =400/15=26.67
				$\sigma_{\bar{x}} = \sqrt{26.67} = 5.17$

بطريقه أخرى ، يمكن حساب $\sigma_{\bar{x}}$ مباشره من البيانات المعطاة في الجدول رقم (٥,٢). وذلك باستخدام الصيغة الرياضية لحساب التباين للتوزيع الاحتمالي للبيانات التالية:

$$\sigma^2 = \Sigma(X-\mu)^2 \cdot P(X) \qquad (5.2)$$

وتشير النتائج المتحصل عليها في الجدول رقم (٥,٣) إلى الوصول لنفس النتيجة التي حسبت سابقاً $\sigma_{\bar{x}} = 5.17$. ولكن في حالة معرفتنا لقيمة الانحراف المعياري للمجتمع فإننا نستخدم طريقه معامل التصحيح لسهولتها. وفي الحقيقة، فإنه لا يتم سحب جميع العينات الممكنة من أي توزيع حيث يتم في العادة سحب عينة واحدة. ولذلك فإنه ليس لدينا الخيار لحساب $\sigma_{\bar{x}}$ باستخدام متوسطات التوزيعات الاحتمالية كما في الجدول رقم (٥,٣) ، ولحسن الحظ فإن لدينا صيغة رياضيه معتمده على الانحراف المعياري للمجتمع σ .

تمارين Exercises

١- ما الفرق بين التعداد والعينة؟ اشرح مع إعطاء ثلاثة أسباب أو أكثر لماذا المعاينة أكثر ملائمة من التعداد في حالات معينة.

٢- أعط مثال للحالات التي يكون فيها المجتمع :

أ) غير محدود.

ب) يعتبر غير محدود، ولكنه في الحقيقة محدود.

ج) محدود.

٣- ما الهدف من وجود مجموعه ضابطة في الدراسة؟ أعط مثال لحالة تكون فيها المجموعة الضابطة ذات أهمية.

٤- ما الفرق بين الأخطاء العشوائية والمنتظمة؟ وما الذي يجب تقليله في المعاينة؟ اشرح.

٥- باستخدام جدول الأرقام العشوائية، الملحق جدول (١)، اختر عينة عشوائية بسيطة حجمها عشر مفردات من مجتمع حجمه ١٠٠ مع شرح خطوات الاختيار المستخدمة.

٦- باستخدام جدول الأرقام العشوائية، الملحق جدول (١)، اختر عينه عشوائية منتظمة حجمها ٥ ٪ من مجتمع حجمه ٣٠٠. وضح خطوات الاختيار المستخدمة.

٧- أعد حل تمرين ٥، ٦ باستخدام أسلوب المعاينة من قائمة أدوات : تحليل البيانات في برنامج الجداول الالكترونية Excel . أطبع النتائج التي تم الحصول عليها ثم اشرح خطوات الاختيار المستخدمة.

٨- لدى شركة الزراعة العالمية مندوبي مبيعات في سكرامنتو ومينابليس، هيوستن، وأتلانتا. ويحتوي كل مكتب على ١٦ مندوباً. فإذا كانت المبيعات الأسبوعية لأي مندوب مبيعات تتوزع طبيعياً بمتوسط ٣٠,٠٠٠ دولار وانحراف معياري ٦٤٠٠ دولار فما هو المدى حول المتوسط بنسبه ٩٥,٥٪ للحالات التالية:

أ) المبيعات الأسبوعية لأي مندوب مبيعات تم اختياره عشوائياً.

ب) متوسط المبيعات الأسبوعية للمندوب في مكتب هيوستن.

ج) متوسط المبيعات الأسبوعية للمندوب في الشركة.

٩- ما هو احتمال اختيار عينة عشوائية بسيطة بمتوسط ٥٠ أو أكثر من مجتمع متوسطه ٤٦ إذا كان حجم العينة ٨١ وتباين المجتمع ٣٢٤؟

١٠- إذا كان متوسط المشتريات الأسبوعية من الطعام لعدد ٥٠ مستهلك هي ١٢٠ دولاراً بانحراف معياري ٢٥ دولار. تم اختيار مجموعة عشوائية مكونة من ٥ أشخاص من هذا المجتمع لدراسة عاداتهم الشرائية. المطلوب إيجاد متوسط مشتريات هذه المجموعة من الطعام والانحراف المعياري.

١١- معمل للأصول الوراثية لديه ٥ مواقع مزروعة بأصناف حديثه من القطن لها إنتاجية مختلفة كالتالي:

الموقع	إنتاجية الايكر
V	٥٢٠
W	٥٠٠
X	٥٤٠
Y	٥٧٠
Z	٤٨٠

تم اختيار موقعين عشوائياً لتقدير متوسط الإنتاجية، المطلوب:

أ) أوجد جميع العينات الممكنة المكونة من موقعين ثم احسب متوسط الإنتاجية لكل عينه.

ب) احسب المتوسط والتباين للتوزيع الاحتمالي من البيانات في الفقرة أ وكذلك من البيانات الأصلية.

١٠- قررت مجموعة كبيرة من مزارعي البرتقال في الشاطئ الغربي - Cal co-op إنشاء مصنع للعصير. تم اختيار عدد ١٦ مصنعاً متشابهه بطريقه عشوائية وقدّر متوسط إنتاجها فكان ٣٠٠٠ جالون عصير في اليوم بانحراف معياري ٤٠٠ جالون. ما هو احتمال أن يكون للمصنع المختار عشوائياً متوسط يومي في حدود ٢٠٠ جالون من متوسط المجموعة.