جامعة البعث

كلية التمريض

المعهد التقاني لخدمات الطوارئ

الإحصاء الحيوي

المحاضرة الخامسة

الدكتور ياسر العمر

المعاينة وتوزيع العينات

SAMPLING AND SAMPLES DISTRIBUTION

التميز بين مصطلح العينة والمجتمع

The Distinction between the Sample and the Population

عندما نرغب بجمع معلومات من مجتمع ما لا يمكن – إلا ما ندر – أن نفحص كل عضو في المجتمع عند إجراء تقييم إحصائي وبدلاً من ذلك يمكن أن نأخذ مجموعة صغيرة عينةً مختارة من المجتمع ومن ثم نفحص أعضاء هذه المجموعة. وتستخدم النتائج لتقدير بعض مميزات هذا المجتمع .

وهكذا فإن الأهداف الرئيسة لأخذ العينات هي عبارة عن الحصول على البيانات التي تسمح لنا بإجراء الاستنتاجات المتعلقة بدراسة أعداد كبيرة من الأفراد اعتمادا على فحص العينة ودراسة علاقتها على سبيل المثال بوجود أو عدم وجود المرض . وربما تتعلق استنتاجات لإثبات أن المرض غير موجود في هذه المنطقة ، وبالعكس يمكن أن تهدف لإثبات وجود المرض أو تقييم مستوى حدوث المرض .

الاستنتاج الإحصائي Statistical Inference

يتم الحصول على الاستنتاج الإحصائي من المجتمع بأخذ العينة وهذا الاستنتاج يمكّننا من أن نرسم الاستنتاجات حول جوانب معينة من المجتمع عندما تكون هناك مجاميع كبيرة مقسمة إلى مجاميع أخرى أصغر عندئذ فإن العينة ضرورية للتقصي عن حالة معينة فمن المهم جداً أن نعي أن التميز بين العينة والمجتمع المأخوذ منه العينات كمكون أساسي للنظرية الإحصائية يتمثل بالاستنتاج الإحصائي هناك نقطتان أساسيتان في الاستنتاج الإحصائي تؤثران جداً في التحليل الإحصائي : وهما طريقة التقدير estimation واختبار النظرية hypothesis testing وسنناقش

في هذه المحاضرة طريقة التقدير بينما يركز اختبار النظرية في قرارنا في اعتماد النتائج التي حصلنا عليها بواسطة العينات المأخوذة أو أن النتائج المحصول عليها لا يمكن اعتمادها نظراً لوجود بعض العوامل أو الأسباب المتداخلة مع المسبب المدروس.

تقدير حدود المجتمع الإحصائي باستخدام المعاينة الإحصائية

Estimation of Population Parameters Using the Statistical Sampling

إن الغرض من أخذ العينة هو التعرف على جوانب عدة للمجتمع الإحصائي المدروس وهذه الجوانب تتلخص بعبارة حدود المجتمع الإحصائي population parameters وهي التي تميز توزيع المجتمع الإحصائي المدروس ويمكن وصف توزيع بيانات المجتمع الإحصائي إذا ما عرفنا طبيعة القيم التي حصلنا عليها في هذه البيانات. ومن غير الممكن أن نحدد الوسط الحسابي للمجتمع الإحصائي تماماً عندما نختار فقط العينة من المشاهدات الملاحظة في مجتمع ما . فمثلاً نحسب العينة الإحصائية والتي لها قيمة قريبة من القيمة الحقيقية ما أمكن لحد المجتمع إلا أن تقدير وسط العينة هو أفضل من طريقة تقدير وسط المجتمع ففي حين أن وسط العينة هو عبارة عن مجموع أعداد كافة المشاهدات في العينة مقسوماً على أعداد المشاهدات في العينة يكون وسط المجتمع هو مجموع كافة المشاهدات في المجتمع مقسوماً على أعداد المشاهدات في هذه الأفراد.

خطأ المعاينة Sampling Error :

ليس من الإمكانية بمكان أن تكون قيمة العينة الإحصائية مساوية تماماً لقيمة حد المجتمع التي قدرناها . فيجب علينا أن نوصي أن هناك دائماً احتمالية بوجود خطأ في التقدير لأننا أخذنا العينات من الأفراد دون النظر إليها جميعا وهذا ما ندعوه بخطأ المعاينة Sampling error ونحتاج إلى تقييم الدقة للعينة الإحصائية كتقدير لحدود المجتمع، ومن أجل هذا الغرض نحسب ما

يدعى الخطأ المعياري للتقدير Standard error of the estimate سواء كان هذا التقدير المحسوب وسطاً لمتوسط قيم مكونات الدم أو نسبة كنسبة الأفراد المعرضة لداء البريميات سواء كانت تلك الأفراد لديها أو ليس لديها معيار إيجابي للبريميات (وهنا يكون شكل المتغير بيزي أي أن نقول إيجابي أو سلبي) ويمكننا حساب الخطأ المعياري للوسط (والذي يرمز له بـ SEM) .

بالقانون : $SEM = \frac{s}{\sqrt{n}}$

بإعتبار أن القيمة s تحسب بـ : $s = \sqrt{\frac{(x-\bar{x})^2}{n-1}}$

بإعتبار أن : s : الانحراف المعياري ، n : عدد المشاهدات

x: قيمة المشاهدة ، $\overline{X}$: متوسط قيمة المشاهدة .

مثال : إن تقدير الانحراف المعياري لتعداد كريات الدم البيضاء قيمته 2352 خلية بالميكروليتر الواحد لأعداد عينات دم ٢٥٦ ولذلك حسب القانون المدرج أعلاه يحسب الخطأ المعياري للوسط كما يأتي :

$$\frac{2352}{\sqrt{256}} = 147\, cell / ml$$

التفريق بين الانحراف المعياري والخطأ المعياري للوسط

The Distinction between the Standard Deviation and the Standard Error of the Mean

إن كل من الانحراف المعياري SD و الخطأ المعياري للوسط SEM هما قياسان لهما تطبيقات مختلفة عن بعضهما كثيراً ومن المهم هنا أن يكون لدينا فهم واضح للتمييز بين الانحراف المعياري للمشاهدات والخطأ المعياري للوسط وهذا التفريق يجنبنا سوء التفسير للبيانات المدروسة :

إن الانحراف المعياري هو قياس لتشتت المشاهدات . وهو يعطي مؤشراً إلى مدى تقارب المشاهدات لمعدلاتها ويمكن أن يستخدم لإنشاء حد مرجعي Reference Interval والذي يعرف بمدى التشتت للبيانات.

بينما الخطأ المعياري للوسط هو قياس دقة وسط العينة كتقدير لوسط المجتمع. وهو يقيم خطأ العينة بإعطاء مؤشر إلى مدى التقارب بين وسط العينة ووسط المجتمع المأخوذة منه العينات ويستخدم الخطأ المعياري المعدل لإنشاء ما يدعى بحد الثقة Confidence Interval والذي يسمح لنا بأن نحكم على دقة تقديرنا لمعدلات المجتمع المدروس .

حد الثقة للوسط الحسابي

The Confidence Interval for a Mean

مفهوم حدود الثقة Understanding Confidence Intervals

لقد ركزنا دائماً على أن توزيع العينات بالنسبة للوسط هو توزيع افتراضي Hypothetical Distribution . وفي ممارستنا العملية لا نأخذ عينات متكررة من الأفراد المدروسة وإنما نأخذ عادة عينة واحدة فقط وتستخدم هذه العينة كتقدير لوسط القيم المأخوذة من الأفراد المدروسة . وهذا يعطينا تصويراً لدرجة الدقة في تقديراتنا لهذه الأفراد أو المجتمعات المدروسة . يجب أن ينوه هنا بأن استخدام عينة واحدة من كل فرد عند التحليل تجنبنا مشكلة أساسية في التحليل الإحصائي والتي تسمى بمشكلة القياسات المتكررة Repeated Measures وهذا يعطي نتائج الدراسة دقة أكثر وهناك العديد من البرامج الإحصائية الحاسوبية التي تجنبنا هذه المشكلة مباشرة كبرنامج SAS، ، وغيره من البرامج الإحصائية وإذا كان لدينا قياسات متكررة فيجب أخذ معدل النتائج لدراسة مشكلة ما في هذا المجتمع أو ذاك .

وإن أفضل طريقة لتقييم ما إذا كان تقديرنا جيداً هو أن تعرف مدى القيم ضمن معدل القيم الحقيقية للأفراد المدروسة والتي تحدد ضمن احتمالية معينة ، هذا المدى للقيم يعرف بالحدود العليا والدنيا upper and lower limits

(وهو ما يسمى بحدود الثقة) ويدعى حد الثقة بالنسبة للوسط Confidence Interval for the mean

١- إذا كان حد الثقة كبيراً فإن وسط العينة يكون عندها تقديراً غير دقيق بالنسبة لمعدل المجتمع المدروس أو الأفراد المدروسة .

٢- إذا كان حد الثقة صغيراً أو ضيق المدى فيكون عندئذ وسط العينة تقديراً جيداً فهو بشكل آخر ذو تقدير دقيق بالنسبة لمعدل الأفراد المدروسة .

إن الثقة 95% تمثل أن معدل القيم في الأفراد المدروسة تشمل ما يدعى بحد الثقة 95% بالنسبة للوسط ونحسب نموذجياً حد الثقة 95% لأي حد من الحدود المراد معرفة دقة الحساب فيها. لكننا يمكن أن نجد أحياناً حدي الثقة 95 %أو 99% مستخدمة في بعض الدراسات . إذ إن حد الثقة 99% هو بالتأكيد أكثر شمولاً ودقة من حد الثقة 95% لأنها أكثر ثقة للحد الذي يشمله ذاك أو هذا الحد . ويعتمد مدى اتساع حد الثقة على :

١- درجة الثقة المطلوبة .

٢- حجم العينة (فالحجم الكبير للعينة يتطلب تقديراً أكثر دقة وبالتالي حد ثقة بمجال ضيق) .

٣- توفر الخاصية المتعلقة بالتقصي عن الحالة (فوجود متغيرات أكثر لعدد من المشاهدات المدروسة يتطلب تقدير أقل دقة وبالتالي حد ثقة بمجال أوسع)

حساب حد الثقة بالنسبة للوسط الحسابي

Calculating the Confidence Interval for the Mean

إن القانون العام لحد الثقة بالنسبة للوسط يحسب كما يلي :

$$\text{CI } 95\% : \ \overline{X} \mp 1.96 \times \text{SEM}$$

بإعتبار أن :

$\overline{X}$: الوسط .

1.96 = ثابت .

SEM: الانحراف المعياري للوسط وباعتبار أن : $SEM = \frac{Q}{\sqrt{n}}$

بإعتبار أن σ: الانحراف المعياري .

n: عدد المشاهدات .

فيصبح القانون :

$$\overline{X} \mp 1.96 \times \frac{Q}{\sqrt{n}}$$

$$\left(\overline{X} - 1.96 \times \frac{Q}{\sqrt{n}}, \overline{X} + 1.96 \times \frac{Q}{\sqrt{n}} \right)$$

ويتراوح مدى حد الثقة الأعلى والأدنى من خلال الحصول على القيم المحسوبة في القانون من القوسية المذكورة أعلاه P .

إن الثابت بالنسبة لحد الثقة 95% هو 1.96 (يقرب غالباً إلى القيمة 2)

بينما الثابت بالنسبة لحد الثقة 99% هو 2.58

والثابت بالنسبة لحد الثقة 90% هو 0.68

وعندما تكون قيمة الانحراف المعياري σ (الانحراف المعياري لمجتمع الدراسة) مجهولة نستعيض عنه بالانحراف المعياري الخاص بالعينة والذي يحسب بالقانون الآتي :

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{n - 1}}$$

إن هناك قوانين أخرى لحساب حد الثقة ولا سيما بالنسبة لتوزيع t ستودنت ولا مجال لتفصيلها وشرحها هنا لأن القانون العام المذكور أعلاه يمكن تطبيقه عندما يكون توزيع القيم للمشاهدات توزيعاً طبيعياً .

توزيع المعاينة للنسبة :

The Sampling Distribution of the Proportion

إن مفهوم توزيع المعاينة للنسبة هو نفس ذلك التوزيع المتعلق بمعدلات العينات المدروسة فهو توزيع افتراضي لخصائص مشاهدات تكون فيها الفوائد مقدرة عندما نريد أن نحصل على استنتاج إحصائي للنسب المئوية لمجتمع من المجتمعات أو منطقة من المناطق أو مجموعات من أفراد نريد دراستها .

ويمتاز توزيع المعاينة للنسبة بالخصائص الآتية :

١- يعتبر هذا التوزيع قريباً من التوزيع الطبيعي إذا كان حجم العينة كبيراً. وفي الواقع إن توزيع النسب المئوية هو في الحقيقة توزيع اسمي لكن هذا التوزيع كما شرحنا أقرب إلى التوزيع الطبيعي لعينة كبيرة عدد مشاهداتها n.

٢- إن معدل قيم توزيع المعاينة النسبي هو معدل النسبة لجميع الأفراد المدروسة وتقدر بالقيمة π وهكذا فإن النسب المئوية للعينة p تحدد بعينة وحيدة مستفردة هو تقدير غير منحرف بالنسبة للنسبة المئوية لجميع الأفراد المدروسة

٣- إن الانحـــراف المعيـــاري لتوزيـــع العينـــات هـــو النســـبة المئويـــة التـــي تحســـب بالقانون :

$$\sqrt{\pi(1-n)/n}$$

وهو ما يدعى بالخطأ المعياري للنسبة المئوية وعبارة عن قياس دقة p كتقدير لقيمة π وهذا الخطأ المعياري يحسب بقيم مشاهدات العينة كما يأتي :

$$SE(P)=\sqrt{\frac{P(1-P)}{n}}$$

بإعتبار أن قيمة P تحسب كما يأتي : $P=\frac{r}{n}$

n: حجم العينة .   r = عدد المشاهدات في العينة .

ومع أنه يمكننا تقدير قيمة π بحساب قيمة p فإن توزيع العينة المئوية تبقى توزيعاً قريباً من التوزيع الطبيعي لعينة كبيرة الحجم عدد مشاهداتها n .

ويجب أن نلاحظ أننا إذا استبدلنا تقدير قيمة p بنسبة مئوية p% فإن تقدير الخطأ المعياري يحسب كنسبة مئوية بالقانون :

$$S\,E(P\%)=\sqrt{\frac{P\%(100-P\%)}{n}}$$

حد الثقة للنسبة المئوية

The Confidence Interval for a Proportion

إن حد الثقة لنسبة أفراد الدراسة π يحسب بإضافة وطرح النسبة المئوية للعينة (p) مضروباً بالخطأ المعياري وهكذا فإن حد الثقة 95% لنسبة أفراد الدراسة يمكن أن يقدر بالقانون الآتي :

$$P \mp 1.96 \times SE(P) = \left\{P - 1.96\sqrt{\frac{P(1-P)}{n}}, P + 1.96\sqrt{\frac{P(1-P)}{n}}\right\}$$

ويجب أن نلاحظ أيضاً أننا يمكن أن نعدل القانون المذكور أعلاه إذا أردنا أن تكون القيمة المدروسة على شكل نسب مئوية وذلك باستبدال كل نسبة مئوية بالقيمة المئوية واستبدال القيمة (1) داخل كل جذر تربيعي بالقيمة 100 .

مثال : اختيرت عينة عشوائية مكونة من 115 فرد من مجتمع في منطقة ما في إحدى المناطق. اختبرت عينات الدم المحصول عليها من أفراد هذه العينة للبحث عن وجود أضداد جراثيم داء الدوران Leptospira وحسب معايير الأضداد المحصول عليها فقد صنفت إلى أضداد إيجابية وأخرى سلبية . ففي هذه العينة وجد أن هناك 36 فرد يحمل معايير أضداد إيجابية لهذا المرض . إن تقدير النسبة المئوية للأفراد المعرضة لداء الدوران يمكن أن تحسب كما يأتي:

$$36/115 = 0.31$$

وهكذا يمكن أن يقدر الخطأ المعياري في هذه النسبة كما يأتي :

$$SE_{(P)} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.313(1-0.313)}{115}} = 0.043$$

وبالتالي يمكن أن نحسب حد الثقة لنسبة الأفراد المعرضة للإصابة بداء الدوران كما يأتي :

$CI95\% = P \mp 1.96\ SE(P)$

$= (0.313 - 1.96 \times 0.043\ ,\ 0.313 + 1.96 \times 0.043)$

$= (0.228\ ,\ 0.398)$

وهكذا فإن ثقتنا 95% للنسبة الحقيقية للأفراد المعرضة للإصابة بداء الدوران تتراوح ما بين 0.23 إلى 0.40 .