

الفصل الثالث عشر

تقدير حجم المجتمع

Estimation of the Population Size

1.13 مقدمة

لقد تناولنا في الفصول السابقة تقدير الوسط الحسابي والمجموع الكلي والنسبة للمجتمع مع افتراض أن حجم المجتمع N معلوم، وإذا كان غير معلوم فعلياً أن نقوم بإهمال معامل التصحيح ونفترض أن حجم المجتمع كبير جداً، لكن في كثير من الأحيان يكون من الأهمية بمكان تقدير حجم المجتمع، فعلى سبيل المثال تقدير عدد الأسماك الموجودة في خليج العقبة أو الخليج العربي، أو تقدير عدد المتظاهرين الذين خرجوا في مظاهرة معينة، أو تقدير عدد الحيوانات البرية الموجودة في إحدى الصحاري ... إلخ.

هنالك أكثر من طريقة يمكن استخدامها لتقدير حجم المجتمع سوف نتناول ثلاثة منها بالتفصيل في هذا الفصل، وهي: الطريق المباشرة، والطريقة المعكوسة، والطريقة التي تعتمد على تقدير كثافة الوحدات في المجتمع.

2.13 تقدير حجم المجتمع باستخدام المعاينة المباشرة

نقوم في هذه الطريقة بسحب عينة عشوائية من أفراد المجتمع (الحيواني)، ونضع علامات مميزة على جميع الوحدات (الحيوانات) التي سحبت في العينة، ونقوم بإعادتها إلى المجتمع، وتترك لها الفرصة الكافية لكي تختلط بالمجتمع مرة أخرى، ثم نقوم في وقت لاحق بسحب عينة ثانية محددة الحجم من نفس

المجتمع، ونقوم بحصر عدد الوحدات (الحيوانات) التي تحمل العلامة المميزة والتي وضعت على الوحدات (الحيوانات) التي سحبت في العينة الأولى. إذا كانت N تمثل عدد أفراد المجتمع و r عدد الوحدات التي تحمل العلامة المميزة و $P = r / N$ يمثل نسبة الوحدات التي تحمل العلامة المميزة. لذا فإننا نستطيع تقدير N باستخدام $\hat{p} = \frac{s}{n}$ التي تمثل نسبة الوحدات والتي تحمل العلامة المميزة في العينة الثانية حيث إن s تمثل عدد الوحدات التي تحمل العلامة المميزة في العينة الثانية و n يمثل حجم العينة الثانية.

$$\hat{N} = \frac{r}{\hat{p}} = \frac{rn}{s}$$

مع تقدير للتباين

$$\hat{V}(\hat{N}) = s^2(\hat{N}) = \frac{r^2 n(n-s)}{s^3}; s > 0$$

نلاحظ أن \hat{N} تقدير متحيز إلى N حيث إن

$$E(\hat{N}) = N + \frac{N(N-r)}{nr}$$

إذا كان حجم r و n كبيراً فإن مقدار التحيز $\frac{N(N-r)}{nr}$ سيكون صغيراً ويمكن إهماله، كذلك يمكن أن نلاحظ أن \hat{N} يبالغ بزيادة تقدير N . ولزيد من المعلومات، ولتقديرات أخرى لـ N يراجع Chapman (1952).

مثال (1): رغب مدير دائرة الحيوانات البرية في إحدى المحافظات بتقدير عدد الغزلان في الصحراء التابعة لمحافظةه فنُصح بأن يسحب عينة عشوائية أولية بحجم $r = 100$ من الغزلان ويقوم بوضع علامة مميزة على جميع هذه الغزلان ومن ثم إرجاعها إلى الصحراء التي كانت تعيش فيها لمدة شهر، ومن ثم يقوم بسحب عينة ثانية بحجم $n = 50$ غزال، فقام بما نُصح بعمله، ومن ثم قام

بحصر عدد الغزلان التي تحمل الصفة المميزة في العينة الثانية فوجدها $s = 10$.
قدّر عدد الغزلان في هذه الصحراء، وأوجد حداً أعلى لمقدار الخطأ لهذا
التقدير.

الحل: يمكن تقدير عدد الغزلان باستخدام

$$\hat{N} = \frac{nr}{s} = \frac{100(50)}{10} = 500$$

يمكن إيجاد الحد الأعلى لمقدار الخطأ كما يأتي

$$2\sqrt{s^2(\hat{N})} = 2\sqrt{\frac{r^2 n(n-s)}{s^3}} = 2\sqrt{\frac{(100)^2 50(50-10)}{(10)^3}} = 282.84$$

لذلك فإننا نقدر عدد الغزلان بنحو 500 غزال وبحد أعلى للخطأ مقداره
282.84.

3.13 تقدير حجم المجتمع باستخدام العينة المعكوسة

هذه الطريقة مماثلة للطريق الأولى (العينة المباشرة)، ولكن هنا العينة
الثانية لا يكون حجمها ثابتاً، بل نسحب وحدات من المجتمع إلى أن تصل إلى
عدد معين (s) من الوحدات التي تحمل العلامة المميزة. وباستخدام هذه
الطريقة نستطيع تقدير حجم المجتمع N باستخدام

$$\hat{N} = \frac{r}{\hat{p}} = \frac{rn}{s}$$

ولكن s هنا ثابتة و n متغيرة على خلاف ما كان عليه الحال في طريقة العينة
المباشرة، وأما تقدير التباين إلى \hat{N} فيكون

$$s^2(\hat{N}) = \frac{r^2 n(n-s)}{s^2(s+1)}$$

لا بد من ملاحظة أن الشرط $s > 0$ لن يكون مشكلة في هذه الطريقة لأن S يجري تثبيتها من قبل الباحث على خلاف الطريقة السابقة، ولا بد من الإشارة هنا إلى أن التقدير \hat{N} وهو تقدير غير متحيز إلى N على خلاف الطريقة السابقة، وكذلك تقدير التباين فهو تقدير غير متحيز إلى التباين الحقيقي إلى \hat{N} .

مثال (2): لنفرض أننا في المثال السابق حددنا عدد الغزلان التي نريد أن نحصل عليها في العينة الثانية لتكون $s = 10$ وقمنا بسحب غزلان من المجتمع وبعد سحب 80 غزال حصلنا على 10 غزلان تحمل العلامة المتميزة، أي أن حجم العينة الثانية $n = 80$. قَدِّر عدد الغزلان في هذه الصحراء، وأوجد حداً أعلى لمقدار الخطأ لهذا التقدير.

الحل: يمكن تقدير عدد الغزلان باستخدام

$$\hat{N} = \frac{nr}{s} = \frac{100(80)}{10} = 800$$

يمكن إيجاد الحد الأعلى لمقدار الخطأ كما يأتي

$$2\sqrt{s^2(\hat{N})} = 2\sqrt{\frac{r^2 n(n-s)}{s^2(s+1)}} = 2\sqrt{\frac{(100)^2 80(80-5)}{(10)^2(10+1)}} = 467.1$$

لذلك فإننا نقدر عدد الغزلان بنحو 800 غزال ويحد أعلى للخطأ مقداره

467.1.

4.13 تقدير حجم العينة باستخدام المعاينة المباشرة والمعكوسة

لقد تكلمنا أعلاه عن استخدام العينة المباشرة والمعكوسة لتقدير حجم المجتمع، وربما يتساءل القارئ أيهما أفضل أو أكثر دقة في تقدير حجم المجتمع؟ إن العينة المعكوسة تعطي معلومات أكثر دقة في تقدير حجم المجتمع من العينة المباشرة إذا كان حجم العينة الثانية n الذي يتطلب لإعادة إمساك s من الوحدات التي تحمل العلامة المميزة صغيراً بالمقارنة إلى حجم

المجتمع. ولكن إذا لم تكن هنالك أي معلومات متوافرة عن حجم المجتمع N فإن سوء اختيار أو تحديد حجم العينة الأولية r سيؤدي إلى أن يكون حجم العينة الثانية n كبيراً جداً إذا تم استخدام المعاينة المعكوسة. على سبيل المثال إذا كان حجم المجتمع $N = 10000$ وحدة وحجم العينة العشوائية الأولية $r = 50$ ، فإن ذلك يتطلب أن يكون حجم العينة الثانية n كبيراً جداً للحصول على $s = 10$ وحدة تحمل العلامة المميزة في العينة الثانية.

أولاً: تحديد حجم العينة الأولى r والثانية n للمعاينة المباشرة

يمكن استخدام جدول (1) لتحديد حجم العينيتين r و n المطلوبتين لتقدير حجم المجتمع N وبحد أعلى ثابت لمقدار خطأ التقدير. ولكن استخدام هذا الجدول يتطلب بعض المعلومات السابقة عن حجم المجتمع N ، ولا بد من الإشارة إلى أن محتويات الجدول رقم (1) هي $V(\hat{N})/N$ للمعاينة المباشرة. إذا كانت لدينا معلومات تقريبية عن حجم المجتمع N ، يمكن أن نحدد تباين التقدير إذا كانت قيمتا r و n ثابتتين. يحتوي الجدول رقم (1) على قيمتي r و n من خلال النسب $p_1 = r/N$ و $p_2 = n/N$. لمزيد من المعلومات يراجع (Scheaffer et al. 2005).

جدول (1): قيم $V(\hat{N})/N$ للمعاينة المباشرة

$p_2 = n/N$	$p_1 = r/N$					
	0.001	0.01	0.1	0.25	0.5	1.0
0.001	999,000	99,000	9000	3000	1000	0
0.01	99,000	9,900	900	300	100	0
0.1	9,000	990	90	30	10	0
0.25	3,996	396	36	12	4	0
0.5	1,998	198	18	6	2	0
1.0	999	99	9	3	1	0

مثال (3): يعتقد مدير دائرة الحيوانات البرية في المثال رقم (1) أن عدد الغزلان في الصحراء التابعة لمحافظة لهذا العام هو تقريباً نفس العدد في السنة الماضية وهو بين 500 إلى 700 غزال. أوجد حد الخطأ في التقدير المصاحب للنسب

$p_1 = 0.25$ و $p_2 = 0.1$. حدد حجم العينيتين r و n المطلوبتين إذا رغبت دائرة الحيوانات البرية استخدامهما لإجراء مسح جديد لتقدير حجم المجتمع N .

الحل: سوف نأخذ الحد الأعلى لعدد الغزلان وهو 700 حتى نحصل على تقدير محافظ للتباين $V(\hat{N})$. نستخدم الجدول رقم (1) مع النسب $p_1 = 0.25$ و $p_2 = 0.1$ لنحصل على

$$\frac{V(\hat{N})}{N} = 30$$

نأخذ $N=700$ للحصول على

$$V(\hat{N}) = 700(30) = 21000$$

$$\sqrt{V(\hat{N})} = \sqrt{21000} = 144.914$$

وعليه سيكون حد الخطأ في التقدير هو

$$2\sqrt{V(\hat{N})} = 2(144.914) = 289.828$$

إذا رغبت دائرة الحيوانات البرية بإجراء مسح بالعينه جديد فإن حجم العينه r يجب أن يكون

$$r = p_1 N = 0.25(700) = 175$$

أما حجم العينه n فيجب أن يكون

$$n = p_2 N = 0.1(700) = 70$$

ثانياً: تحديد حجم العينه الأولى r والثانية n للمعاينة المعكوسة

يمكننا أن نوجد التباين $V(\hat{N})$ للعينه المعكوسة بنفس الطريقة التي استخدمناها في العينه المباشرة أعلاه وباستخدام جدول رقم (2). ولابد من الإشارة إلى أن محتويات الجدول رقم (2) هي $V(\hat{N})/N$ للمعاينة المعكوسة. ويمكننا أن نستخدم الجدول رقم (2) وذلك باستخدام النسب $p_1 = r/N$ و

$p_2 = n/N$. ولزيد من المعلومات يراجع (Scheaffer et al (2005).

جدول (2): قيم $V(\hat{N})/N$ للمعاينة المعكوسة

$p_2 = n/N$	$p_1 = r/N$					
	0.001	0.01	0.1	0.25	0.5	1.0
0.001	999	990	900	750	550	0
0.01		99	90	75	50	0
0.1			9	7.5	5	0
0.25				3	2	0
0.5					1	0
1.0						0

5.13 تقدير كثافة وحجم المجتمع باستخدام معاينة المربعات

إن تقدير عدد الوحدات في منطقة معينة (المجتمع) يمكن تحقيقه أولاً بتقدير عدد الوحدات لكل وحدة مساحة (أي كثافة الوحدات) ومن ثم ضربه بمساحة المنطقة (المجتمع) تحت الدراسة. على سبيل المثال إذا كان كاتب الطباعة يقوم بالمعدل بطباعة 3 كلمات خاطئة في الصفحة الواحدة، فإنه سيطبّع تقريباً 600 كلمة خاطئة عند طباعته لكتاب يحتوى على 200 صفحة. سوف نناقش تقدير الكثافة وعدد الوحدات في منطقة معينة، ولكن هذه الطريقة لا تقتصر فقط على المساحة لأنه يمكن تطبيقها على الحجم مثل عدد البكتيريا في سائل محدد الحجم، أو عدد المكالمات الهاتفية التي تتلقاها إلى دائرة معينة خلال مدة زمنية ثابتة، أو عدد الزائرين لموقع على الشبكة العنكبوتية (الإنترنت) خلال مدة زمنية محددة. ولكننا سنستخدم المساحة ووحدة المساحة المربعة من باب الملاءمة فقط.

لنفرض أن منطقة مساحتها A يجري مسحها بسحب عينة عشوائية حجمها n من القطع مساحة كل قطعة تساوي a . لنفرض أن $A = Na$ ، كل قطعة تسمى مربع على الرغم من أنها قد لا تكون مربعة المساحة. لنفرض أن m_i تمثل عدد الوحدات في المربع i ونفرض أن M تمثل عدد الوحدات في المجتمع ذو المساحة A حيث إن

$$M = \sum_{i=1}^N m_i$$

كذلك نفرض أن

$$\lambda = \frac{M}{A}$$

تمثل كثافة الوحدات أو كثافة الوحدات لكل وحدة مساحة. إن هدفنا هو تقدير λ ومن ثم $M = \lambda A$. لابد من الإشارة هنا إلى أن m_i عبارة عن متغير عشوائي لأنها تمثل عدد الوحدات في مربع ثابت المساحة جرى سحبه بطريقة عشوائية.

لنفرض أننا سحبنا عينة عشوائية حجمها n من المربعات، مساحة كل مربع a وقمنا بمشاهدة عدد الوحدات m_i في كل مربع، لذا فإن الوسط الحسابي

$$\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$$

يُعدُّ تقديراً لمعدل عدد الوحدات في المربع، وكذلك فإن

$$\hat{\lambda} = \frac{\bar{m}}{a}$$

يمثل تقدير لعدد الوحدات لكل وحدة مساحة، كما نلاحظ أن العينة هنا عبارة عن عينة عشوائية بسيطة؛ لذا يمكننا أن نقدر تباين $\hat{\lambda}$ بسهولة، وهو

$$\hat{V}(\hat{\lambda}) = \frac{1}{a^2} \hat{V}(\bar{m}) = \frac{1}{a^2} \frac{s_m^2}{n}; s_m^2 = \frac{1}{n-1} \sum_{i=1}^n (m_i - \bar{m})^2$$

ومن هذه النتيجة نستطيع بسهولة أن نوجد $\hat{M} = \hat{\lambda} A$ ومن ثم نوجد تقديراً لتباين \hat{M} .

لقد افترضنا للحصول على النتائج أعلاه أن توزيع الوحدات داخل كل مربع موزعة توزيعاً ثابتاً - أي لا توجد عشوائية في التوزيع - ولكن في كثير

من الحالات الوحدات داخل كل مربع قد تتوزع عشوائياً، على سبيل المثال عدد الأشجار المريضة في وحدة المساحة، عدد البكتريا في وحدة الحجم، عدد الحوادث المرورية في مدة زمنية؛ لذا يمكننا أن نبسط النتائج أعلاه على افتراض أن m_i تتبع توزيع بواسون، لنحصل على تقديرات إلى λ و M .
تقدير للكثافة λ هو

$$\hat{\lambda} = \frac{\bar{m}}{a}$$

وتقدير لتباين $\hat{\lambda}$ يكون

$$\hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}}{an}$$

و تقدير لحجم المجتمع M

$$\hat{M} = \hat{\lambda} A$$

و أخيراً تقدير لتباين \hat{M} هو

$$\hat{V}(\hat{M}) = A^2 \frac{\hat{\lambda}}{an}$$

مثال (4): تعاني إحدى المناطق من انتشار نوع من النمل المؤذي للمزروعات والإنسان، لذا تقدير كثافة انتشار هذا النوع من النمل يعد من أولويات الدوائر الصحية والزراعية في المنطقة. وكذلك لمعرفة الزيادة في حجم مجتمع النمل (أعداد النمل). جري سحب عينة عشوائية تحتوي على 50 مربع، مساحة كل مربع نحو 16 متراً مربعاً، وجرى حصر عدد قرى النمل في كل قطعة وحصلنا على النتائج الآتية:

5	4	3	2	1	0	عدد القرى
2	5	10	12	8	13	التكرار (عدد القطع)

استخدم هذه البيانات لتقدير كثافة قري النمل في المنطقة.

الحل:

$$\sum_{i=1}^{50} m_i = 0(13) + 1(8) + 2(12) + \dots + 5(2) = 92$$

و

$$\sum_{i=1}^{50} m_i^2 = 0(13) + 1(8) + 4(12) + \dots + 25(2) = 276$$

لذا فإن كثافة قري النمل لكل متر مربع هي

$$\hat{\lambda} = \frac{\bar{m}}{a} = \frac{1}{16} \left(\frac{92}{50} \right) = 0.11$$

ويمكن أن نقدر تباين $\hat{\lambda}$ ليكون

$$\hat{V}(\hat{\lambda}) = \frac{1}{a^2} \frac{s_m^2}{n} = \frac{1}{(16)^2} \frac{1}{50} \left[276 - \frac{(92)^2}{50} \right] = 0.0017$$

وكذلك يمكن إيجاد الحد الأعلى لمقدار الخطأ كما يأتي

$$2\sqrt{\hat{V}(\hat{\lambda})} = 2\sqrt{0.00017} = 0.026$$

وإذا كنا نعتقد أن قري النمل تتوزع بشكل عشوائي في المنطقة تحت

الدراسة، لذا سيكون تقديرنا لتباين $\hat{\lambda}$ بما يلي

$$\hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}}{an} = \frac{0.11}{(16)(50)} = 0.00014$$

و يحد أعلى لمقدار الخطأ هو

$$2\sqrt{\hat{V}(\hat{\lambda})} = 2\sqrt{0.00014} = 0.023$$

ونلاحظ أن تقديري التباين متقاربين جداً وهذا يؤشر إلى أن افتراض عدم العشوائية في توزيع قرى النمل في المنطقة معقول، ولا بد من الإشارة هنا إلى أن افتراضنا عدم العشوائية يعني أن $\hat{\lambda}$ و $\hat{V}(\hat{\lambda})$ يعتمدون على البيانات من خلال \bar{m} فقط.

لمزيد من المعلومات يراجع (Scheaffer et al (2005).

تمارين

1. ناقش الفرق بين المعاينة المباشرة والمعكوسة.
2. لنفرض أن تكاليف المعاينة غير مهمة، كيف يمكننا أن نخفض حد الخطأ باستعمال المعاينة المباشرة والمعكوسة.
3. تخشى دائرة الأسماك النهرية من تناقص نوع معين من السمك الذي يعيش في النهر الذي يمر بالولاية التابعة لها، لذا قامت باصطياد $r=100$ سمكة من هذا النوع من السمك خلال عدة أيام وقامت بوضع علامة مميزة عليها وأعادتها على الفور إلى النهر. بعد عدة أسابيع قامت بسحب عينة ثانية حجمها $n=120$ فوجدت أن 27 سمكة تحمل العلامة المميزة. قدر عدد الأسماك N الموجودة في النهر من النوع المعني بالدراسة، ومن ثم أوجد حداً أعلى لمقدار الخطأ.
4. ينتاب المسؤولين في إحدى الدوائر قلقاً من ازدياد عدد الحمام الموجود حول بنايتهم وعلى جدرانها الخارجية وسقوفها ومما يسببه من الإزعاج والأوساخ للبنية وساكنيها. لذا قاموا بتأجير فريق يتولى تقدير عدد الحمام الموجود حول البناية. قام الفريق باستخدام عدة طرق لمسك مجموعة من الحمام $r = 60$ ووضع علامة مميزة على كل حمامة ومن ثم إطلاق سراحها بأسرع وقت ممكن. بعد شهر أعادوا العملية من جديد ولكن مع تحديد أنه لا يمكن إيقاف العملية إلا بعد مسك 18 حمامة تحمل العلامة المميزة. استطاعوا الحصول على العدد المطلوب أي 18 حمامة تحمل العلامة المميزة ولكن بعد مسك 80 حمامة. قدر عدد الحمام الموجود حول البناية ومن ثم أوجد حداً أعلى لمقدار الخطأ.

5. قام الباحثون في إحدى الشركات بسحب عينات بحجم 100 سم³ من الهواء في المنطقة التي تعمل بها الشركة. الهدف من الدراسة هو معرفة كثافة نوع من البكتريا غير المضرة والموجودة في الهواء. لنفترض أن 15 عينة أعطتنا معدل 210 لكل عينة. قدر كثافة هذا النوع من البكتريا لكل سنتيمتر مكعب ومن ثم أوجد الحد الأعلى للخطأ.
6. ترغب إدارة الشرطة في إحدى المدن بتقدير عدد المتظاهرين الذي خرجوا في مظاهرة منددة بسياسة حاكم المدينة. قامت بتوزيع 80 قميصاً بلون مميز على عينة من المتظاهرين، وطلبت من كل شخص أن يلبس القميص، ومن ثم يعود إلى المظاهرة. بعدة مدة مناسبة قامت إدارة الشرطة بسحب عينة عشوائية ثانية حجمها 60 متظاهراً، فوجدت أنه لا يوجد إلا 5 أشخاص يرتدون القميص ذا اللون المميز في هذه العينة. قدر عدد المتظاهرين، ومن ثم أوجد حداً أعلى لمقدار الخطأ.
7. لنفترض أنه في السؤال رقم (5) من الصعوبة بمكان إيجاد عدد البكتريا في العينة ولكن يمكن بسهولة تمييز فيما إذا كانت البكتريا موجودة أم لا، في عينة حجمها 500 وجدنا أن 410 من هذه العينات تحتوي على هذا النوع من البكتريا. قدر الكثافة لكل سنتيمتر مكعب، ومن ثم أوجد الحد الأعلى للخطأ.
8. يجري عد السيارات التي تمر من خلال نقطة تقاطع بين شارعين رئيسيين في إحدى المدن خلال مُدَّةٍ زمنية محددة، طول كل مدة عشر دقائق، جرى تحديدها عشوائياً خلال ساعات العمل الرسمية، وجرى سحب عينة عشوائية حجمها 20 فترة زمنية، ووجدنا أن معدل عدد السيارات المارة خلال هذه المدة الزمنية ذات العشر دقائق هو 40 سيارة. قدر عدد السيارات المارة خلال ساعات الدوام الرسمي، وأوجد حداً أعلى للخطأ.

References

1. Chapman, D. J. (1952). Inverse, Multiple and Sequential Sample Censuses, *Biometrics*, 8, 286-306.
2. Darling, D. A. and Robbins, H. (1967). Finding the Size of a Finite Population, *Ann. Math. Statist.*, 38, 1392-1398.
3. Scheaffer, R.L., Mendenhall, W. and Ott, L. (2005). *Elementary Sampling Survey*, 6th ed., Duxbury, New York.
4. Sen, P. K. (1988). Asymptotics in Finite Population Sampling, *Handbook of Statistics Vol. 6*, pp 291-331, P.R. Krishnaiah and C.R. Rao eds.

الفصل الرابع عشر

المعاينة بطريقة الإمساك وإعادة الإمساك

Capture-Recapture Sampling

1.14 مقدمة

لقد استخدمت طريقة المعاينة بالإمساك وإعادة الإمساك لتقدير أعداد المجتمعات الحيوانية بما فيها الطيور، والأسماك، والحيوانات الثديية، والزواحف، والحشرات وغيرها من المخلوقات. وكذلك لتقدير قابلية اكتشاف الحيوانات في مجتمعاتها لاستخدامها في مسوحات أخرى. وكذلك لتقدير قابلية البقاء ولتقدير معلمات المجتمع الأخرى، كذلك جرى استخدام طريقة الإمساك وإعادة الإمساك لتقدير أعداد المجتمعات الإنسانية المراوغة أو غير المنضبطة مثل أعداد المشردين. وأخيراً تقدير عدد الإصابات المميتة في حوادث الطرق وغيرها.

يمكن تلخيص المعاينة بطريقة الإمساك وإعادة الإمساك كما يأتي: نقوم بسحب عينة عشوائية أولية من أفراد المجتمع، ونضع علامات مميزة على جميع الحيوانات أو نميز هذه الوحدات (الحيوانات) بطريقة أو بأخرى، ونعيدها إلى المجتمع بأسرع وقت، ومن ثم نقوم في وقت لاحق بسحب عينة ثانية مستقلة عن العينة الأولى، ونقوم بحصر عدد الحيوانات التي تحمل العلامة المميزة في هذه العينة، فإذا كانت العينة الثانية ممثلة للمجتمع فإن نسبة الوحدات التي تحمل العلامة المميزة فيها تساوي نسبة الوحدات التي

تحمل العلامة المميزة في المجتمع، وباستخدام هذه العلاقة نستطيع أن نحدد مجموع الوحدات في المجتمع كما سترى لاحقاً.

لا بد من الإشارة هنا إلى أنه ليس بالضرورة أن تُمسك الحيوانات أو يعاد إمساكها بشكل فعلي، فإذا كان يمكن أن تميز هذه الحيوانات عن طريق بعض العلامات الطبيعية الموجودة عليها فيمكن أن نقوم بالمسح بسحب عينتين مستقلتين، ونستطيع أن نميز عدد الحيوانات المعلمة التي تم مشاهدتها في العينتين، وهذا العدد يمثل إعادة الإمساك، ولكن في كثير من الحالات لا بد من مسك الحيوان باستخدام شبكة أو مصيدة أو أي طريقة يمكن أن يمسك بها، ومن ثم توضع علامات مميزة على الحيوانات مثل قطعة معدنية، أو صبغ، أو شريط، أو علم أو غيرها من العلامات التي نستطيع من خلالها تمييز الحيوانات. أما فيما يخص المجتمعات الإنسانية فالعينتان يمكن أن تكونا عبارة عن كشافين، فمثلاً الكشف الأول يمكن الحصول عليه من التعدادات السكانية الشاملة، والكشف الثاني يمكن أن يكون من دراسة لاحقة لنفس المجتمع، وكذلك يمكن أن يكون الكشف الأول عبارة عن الكشوفات عن الحوادث المرورية لدى مديريات شرطة المرور، والكشف الثاني يمكن أن يكون البيانات المتوافرة لدى شركات التأمين. في كلا المثالين يمكن أن يكون إعادة الإمساك عبارة عن الأشخاص الذين تمت مشاهدتهم في كلا الكشافين.

هناك كثير من البحوث المنشورة حول هذا النوع من العينات والتي يمكن الرجوع إليها للتوسع في هذا الموضوع، نذكر من هذه البحوث Cormack (1979) و Otiset al. (1978) و Pollock (1981, 1991) و Pollock al (1990) و Seber (1973, 1982, 1986, 1992) و Seber (1973) ومن الجدير بالذكر أن Seber (1973) صنف الطرق التي يمكن استخدامها للإمساك وإعادة الإمساك حسب نوع المجتمع. ومن المراجع التي تتناول استخدام طريقة الإمساك وإعادة الإمساك في

المجتمعات الإنسانية المراوغة أو غير المنضبطة (Cowan and Malec (1986) و (1991) Freedman و (1988) Sudman et al. و (1986,1991) Wolter.

الرموز والمصطلحات الآتية سوف يتم استخدامها في هذا الفصل
 τ : عدد الوحدات (الحيوانات) في المجتمع.

X : عدد الوحدات (الحيوانات) التي تحمل العلامات المميزة في المجتمع، وهو كذلك يمثل حجم العينة الأولية.

y : عدد الوحدات (الحيوانات) التي سحبت في العينة الثانية.

x : عدد الوحدات (الحيوانات) التي تحمل العلامات المميزة في العينة الثانية. وأخيراً لا بد من الإشارة إلى عينة بحجم n ترمز إلى مجموعة من الوحدات تم سحبها من المجتمع، وهذه الوحدات إما أن تكون قطع أرض، أو شرائط أرضية، أو خطوطاً مستقيمة، أو مناطق لإمساك الحيوانات أو ... إلخ.

2.14 الإمساك مرة واحدة

يمكن تلخيص طريقة الإمساك وإعادة الإمساك بما يأتي: نقوم بسحب عينة عشوائية أولية من المجتمع بحجم X من الحيوانات، ونقوم بوضع علامات مميزة على جميع هذه الحيوانات ثم نعيدها بأسرع وقت إلى المجتمع، ومن ثم نقوم في وقت لاحق بسحب عينة عشوائية جديدة وبصورة مستقلة عن العينة الأولى وبحجم y من الحيوانات، ونقوم بحصر عدد الحيوانات التي تحمل العلامة المميزة في هذه العينة وليكن x . إذا كانت العينة الثانية تمثل المجتمع بصورة كاملة، فإن نسبة الحيوانات التي تحمل العلامة المميزة في العينة الثانية يكون مساوياً تقريباً لنسبة الحيوانات التي تحمل العلامة في المجتمع أو بعبارة أخرى:

$$\frac{x}{y} = \frac{X}{\tau}$$

وبصورة مماثلة يمكننا القول: إن نسبة الحيوانات التي تحمل العلامة المميزة في المجتمع والتي مسكت في العينة الثانية تساوي تقريباً نسبة العينة الثانية إلى المجتمع، أي:

$$\frac{x}{X} = \frac{y}{\tau}$$

يمكننا حل أي من المعادلتين أعلاه للحصول على تقدير لحجم المجتمع غير المعلوم وهو

$$\hat{\tau} = \frac{y}{x} X$$

هذا التقدير يعرف بتقدير بيترسن، أما تقدير التباين إلى $\hat{\tau}$ فهو

$$s^2(\hat{\tau}) = \frac{Xy(X-x)(y-x)}{x^3}$$

وبفترة 95% ثقة تقريبه لتقدير حجم المجتمع

$$\hat{\tau} \pm 2\sqrt{s^2(\hat{\tau})}$$

بما أن عدد الحيوانات التي تحمل العلامة المميزة في العينة الثانية يمكن أن تكون صفراً، فإنه في مثل هذه الحالة لا يمكن تقدير حجم المجتمع؛ لأن تباين $\hat{\tau}$ يكون غير محدد. لذا جرى تعديل تقدير بيترسن $\hat{\tau}$ وتقدير تباينه ليكونا على الوجه الآتي:

$$\tilde{\tau} = \frac{(y+1)(X+1)}{x+1} - 1$$

$$s^2(\tilde{\tau}) = \frac{(X+1)(y+1)(X-x)(y-x)}{(x+1)^2(x+2)}$$

لمزيد من المعلومات يراجع (Seber (1982). وبفترة 95% ثقة تقريبه لتقدير حجم المجتمع

$$\bar{\tau} \pm 2\sqrt{s^2(\bar{\tau})}$$

ويمكننا إيجاد فترة ثقة دقيقة وليست بتقريبية وذلك بافتراض التوزيع الاحتمالي، أو باستخدام طريق المحاكاة يراجع (Buckland (1980, 1984). وأيضاً لمزيد من المعلومات حول إيجاد فترة ثقة لتقدير حجم المجتمع يراجع (Seber(1973, 1982, 1986, 1992).

لابد من الإشارة هنا إلى أن هذا الجزء من الفصل مشابه إلى حد كبير ما ورد في الفصل الثالث عشر وبالتحديد في 2.13. مثال (1): قام الباحثون في أحد الحقول بإمساك $X=300$ فأرة، ووضعت علامة مميزة عليها، ومن ثم أطلق سراحها، وبعد عدة أيام عاد الباحثون إلى حقل الدراسة، وقاموا وبصورة مستقلة عن المرة السابقة بمسك $y=200$ فأرة، فوجدوا أن من بينها $x=50$ تحمل العلامة المميزة. استخدم كلا التقديرين أعلاه لتقدير عدد الفئران في الحقل، ومن ثم أوجد 95% فترة ثقة تقريبه لتقدير عدد الفئران في حقل الدراسة.

الحل:

$$\hat{\tau} = \frac{y}{x}X = \frac{200}{50}(300) = 1,200$$

وبتقدير للتباين

$$s^2(\hat{\tau}) = \frac{Xy(X-x)(y-x)}{x^3} = \frac{300(200)(300-50)(200-50)}{50^3} = 18,000$$

وبفترة 95% ثقة تقريبيه لتقدير حجم المجتمع

$$\hat{t} \pm 2\sqrt{s^2(\hat{t})} = 1200 \pm 2\sqrt{18000} = (932, 1468)$$

أما إذا استخدمنا التقدير الثاني فنحصل على

$$\bar{t} = \frac{(y+1)(X+1)}{x+1} - 1 = \frac{(200+1)(300+1)}{50+1} - 1 \approx 1185$$

$$s^2(\bar{t}) = \frac{(X+1)(y+1)(X-x)(y-x)}{(x+1)^2(x+2)} = \frac{(300+1)(200+1)(300-50)(200-50)}{(50+1)^2(50+2)} = 16,774.5$$

وبفترة 95% ثقة تقريبيه لتقدير حجم المجتمع

$$\bar{t} \pm 2\sqrt{s^2(\bar{t})} = 1185.3 \pm 2\sqrt{16774.5} = (926, 1444)$$

3.14 نماذج للإمساك لمرة واحدة

هنالك عدة نماذج رياضية يكمن استخدامها في حالة الإمساك لمرة واحدة

سنتطرق إلى أهمها.

أولاً: النموذج المتعدد الحدود Multinomial Model

إن تاريخ مسك أو كشف الحيوان في الصيغة العامة للنموذج المتعدد الحدود لحالة الإمساك أو كشف الحيوان لمرة واحدة في المجتمعات المغلقة يمكن تصنيفه ضمن أربع حالات، هي: الحالة الأولى أن يمسخ بالحيوان في العينة الأولية والثانية، الحالة الثانية أن يمسخ في العينة الأولى ولا يمسخ في الثانية، الحالة الثالثة ألا يمسخ في العينة الأولى ولكن يمسخ في الثانية، وأخيراً الحالة الرابعة ألا يمسخ الحيوان في العينتين. بعد هذا نرى أن النموذج المتعدد الحدود له أربع احتمالات يكون مجموعها واحداً، ويمكن تطبيقه أو استخدامه لتاريخ إمساك كل حيوان، بالإضافة إلى ذلك إذا كان احتمال

مسك الحيوان في العينتين مستقلاً فإن النموذج لعدد الحيوانات التي لكل واحد منها تاريخ إمساك معين سيكون مضروب النموذج المتعدد الحدود Product Multinomials Model. يمكن أن يكون احتمال كشف أو مسك الحيوان مختلفاً من حيوان لآخر ومن عينة لأخرى، في مثله هذه الحالة يحتوي النموذج العام المتعدد الحدود على عدد كبير من المعالم Parameters، ولا يوجد عدد كافٍ من المشاهدات لتقدير هذه المعلمات؛ لذا لا بد من وضع بعض القيود حتى نستطيع تقدير حجم المجتمع، ومن أهم هذه القيود أن نفترض أن احتمال كشف أو مسك الحيوان سيكون متساوياً لجميع الحيوانات خلال مدة الدراسة، ولكن يمكن أن يكون الاحتمال مختلفاً بين العينتين، بالإضافة إلى افتراض أن العينتين جرى سحبهما بصور مستقلة. تحت هذه القيود أو الافتراضات نستطيع أن نحصل على تقدير القيمة العظمى لحجم المجتمع Maximum Likelihood Estimator (MLE) $\hat{\tau} = Xy/x$ وهو نفس تقدير بيترسن الذي مر معنا أعلاه.

ثانياً: نموذج التوزيع الهندسي الزائد Hypergeometric Model

إذا كان عدد الحيوانات التي أمسكت في العينة الأولى والثانية هو على النحو الآتي X و y ثابتين، والعينة الثانية عبارة عن عينة عشوائية بسيطة دون إرجاع وحدات العينة إلى المجتمع، لذا فإن عدد الوحدات في العينة الثانية التي تحمل العلامة المميزة x سيكون متغيراً عشوائياً ويتبع نموذج التوزيع الهندسي الزائد، مرة أخرى إن تقدير القيمة العظمى لحجم المجتمع سيكون نفس تقدير بيترسن وهو $\hat{\tau} = Xy/x$.

ثالثاً: نموذج بواسون Poisson Model

يمكن استخدام نموذج بواسون إذا كان عدد الحيوانات في الحالات الأربعة لتاريخ مسك الحيوان هي متغيرات عشوائية مستقلة، وتبعاً لذلك فإن حجم المجتمع τ يكون متغيراً عشوائياً.

لمزيد من المعلومات حول نماذج الإمساك وإعادة الإمساك يراجع Cormack (1979) و Otis et al. (1978) و Pollock (1981) و Seber (1982, 1986, 1992) و Pollock et al. (1990) و Wolter (1986).

4.14 تصميم المعاينة للإمساك وإعادة الإمساك

لقد افترضنا في النماذج أعلاه أن مسك أو كشف الحيوان يكون مستقلاً عن مسك الحيوانات الأخرى، أو أن العينة الثانية هي عبارة عن عينة عشوائية بسيطة مع عدم الإرجاع كما هو الحال مع النموذج الهندسي الزائد. عندما يتم الحصول على المشاهدات باستخدام عينة من الوحدات مثل قطع أراض، أو مواضع مصائد للحيوانات، أو شبكات لمسك الأسماك (الطيور) أو خطوط عرضية أو ... إلخ لا تكون العينة عينة عشوائية بسيطة، ولا تتم عملية كشف أو مسك الحيوانات مستقلة كما افترضنا في النماذج أعلاه. لمثل هذه الحالات (النماذج السابقة) تكون مفيدة لمعرفة التغيرات التي لا تعتمد على العينة، ولكن تغيرات المعاينة التي تعتمد على الطريقة التي حصلنا بها على العينة تتأثر ببعض العوامل مثل عدم تجانس المكان، وكذلك الفروقات بين المجموعات، والتي من الممكن أن تكون مسؤولة عن جزء كبير من تباين تقدير بيترسن $\hat{\tau}$.

من الناحية المثالية نرغب بأن تكون الحيوانات في منطقة الدراسة مثل حبات الفاصوليا الموجودة في الصحن. لتقدير عدد حبات الفاصوليا في

الصحن، نقوم بأخذ مجموعة من هذه الحبات ونضع علامة على كل حبة باستخدام القلم أو نلونها بلون مميز، ومن ثم نعيدها إلى الصحن ونخلطها مع الحبات الأخرى حتى تكون الحبات متجانسة داخل الصحن. ومن ثم نقوم بسحب عينة ثانية وهي التي يمكن اعتبارها عينة عشوائية، أي كل حبة لديها نفس الاحتمال بالظهور في العينة الثانية، وكل مجموعة حبات (عينة) لها نفس الاحتمال بالسحب. إذا كان المجتمع الحيواني الذي نرغب دراسته يمتاز بنفس الصفة أعلاه أي نستطيع أن نسحب عينة عشوائية بسيطة منه، فإن النموذج الهندسي الزائد يمكن استخدامه في مثل هذه الحالة، ويكون المتغير العشوائي هو عدد الحيوانات المعلمة (أي تحمل العلامة المميزة) في العينة الثانية. يمكن أن نستخدم النموذج المتعدد الحدود في مسوحات الحياة البرية التي يتم سحب العينة فيها بالطيران فوق منطقة الدراسة باستخدام مروحية، التي تؤدي إلى وضع علامة مميزة على جميع الحيوانات التي تم اكتشافها في العينة الأولى ولا يتم تعليم أي حيوان في العينة الثانية، لذلك نرى أن كلا العينتين يكون حجمها غير ثابت. في مثل هذه المسوحات يمكن أن نفترض أن سحب أي حيوان يكون مستقلاً عن سحب الحيوانات الأخرى، لهذا فإن استخدام نموذج مضروب متعدد الحدود يكون مناسباً.

الآن لتتصور أن عملية الإمساك وإعادة الإمساك في المجتمعات الحيوانية- التي يتم الحصول فيها على العينة الأولية بواسطة مصائد وضعت في أماكن مختلفة من منطقة الدراسة. ومع كل مصيدة هنالك مناطق بمساحات مختلفة لجلب انتباه الحيوان لدخولها ومن ثم مسكه، وكذلك حركة الرياح وحركة الحيوانات كلها تؤثر على عملية مسك الحيوان، ولكن الحيوانات المعادة إلى المجتمع بعد مسكها ووضع العلامة المميزة عليها يمكن أن تتوزع في المجتمع بصورة غير متجانسة، وربما له علاقة بالمنطقة التي مسكت بها أو المنطقة التي جرى إطلاق سراحها فيها، والذي قد يؤدي

إلى أن يكون توزيع الحيوانات المعلمة مخالفاً لتوزيع الحيوانات في المجتمع. ونتيجة لذلك سيكون نسبة الحيوانات المعلمة في منطقة معينة يختلف كثيراً عن المناطق الأخرى.

حسب النماذج المعتمدة التي تكلمنا عنها أعلاه لا بد أن تكون العينة الثانية عبارة عن عينة عشوائية بسيطة من الحيوانات، أو أن عملية مسك أو كشف الحيوان يكون مستقلاً عن الحيوانات الأخرى، ولكن في كثير من الحالات يكون هذا غير ممكن. بدلاً من ذلك يمكن أن نحصل على عينة ثانية من الحيوانات من خلال عينة من وحدات تسمى وحدات الكشف وتكون عبارة عن وحدات من الأرض، أو مصائد، أو قطع أرض، أو شبكات صيد أو غيرها من وحدات الكشف، في مثل هذه الحالات تكون الوحدات موزعة بشكل عشوائي على المجتمع، ولكن الحيوانات التي تم مسكها لا تكون مستقلة عن بعضها الآخر، وبما أن الحيوانات المعلمة غير موزعة بصورة متساوية في المجتمع؛ لذا سنرى أن بعض الوحدات ستحتوي على نسبة كبيرة من الحيوانات المعلمة بينما وحدات أخرى تكون نسبة الحيوانات المعلمة قليلة جداً أو صفراً.

فيما يأتي سنقوم باستخدام عينة عشوائية من وحدات الكشف والتي يتم سحبها مع الإرجاع أو دون إرجاع مع عملية الإمساك وإعادة الإمساك لتقدير حجم المجتمع.

أولاً: العينة العشوائية مع الإرجاع لوحدات الكشف

لنفترض أننا سحبنا عينة عشوائية (مع الإرجاع) حجمها n من وحدات الكشف على سبيل المثال قطع أرض أو مصائد أو شبكات صيد أو ... إلخ. لنفترض أن للوحدة i من وحدات الكشف y_i والتي تمثل عدد الحيوانات التي شوهدت في الوحدة x_i عدد الحيوانات المعلمة (أي تحمل العلامة المميزة) من

بين y_i ؛ لذا فإن عدد الوحدات التي تم مشاهدتها في العينة الثانية هو $\sum_{i=1}^n y_i$ من بينهم $\sum_{i=1}^n X_i$ حيوان يحمل العلامة المميزة. لنفترض أن \bar{X} و \bar{Y} تمثلان الوسط الحسابي للمتغيرين X_i و Y_i على التوالي. وكما نعلم فإن عدد الوحدات المعلمة في المجتمع X معلومة، إذا افترضنا أن المجتمع مغلق، أي عدد الحيوانات المعلمة بقي كما هو منذ أن تم تنفيذ العينة الأولى إلى أن تنفذ العينة الثانية؛ لذا فإن تقدير حجم المجتمع يكون

$$\hat{\tau} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n X_i} X = \frac{\bar{y}}{\bar{X}} X$$

وهو نفس تقدير النسبة Ratio المعروف، أما تقدير التباين إلى $\hat{\tau}$ فإن نفس تقدير تباين النسبة المعدل والمعروف أيضاً

$$s^2(\hat{\tau}) = \left(\frac{X}{\bar{X}}\right)^2 \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - rX_i)^2$$

حيث إن $r = \frac{\bar{y}}{\bar{X}}$. وأخيراً فترة $100(1-a)\%$ ثقة تقريبية لتقدير حجم المجتمع يمكن إيجادها باستخدام

$$\hat{\tau} \pm t_{\alpha/2} \sqrt{s^2(\hat{\tau})}$$

حيث إن $t_{\alpha/2}$ يمكن إيجادها من جداول توزيع t مع درجة حرية $n-1$. لمزيد من المعلومات ولصيغ أخرى لتقدير التباين يراجع (Seber (1982).

سيكون تباين تقدير النسبة صغيراً عندما تتوزع الحيوانات المعلمة داخل منطقة الدراسة بصورة متساوية، وعدم توزيع الحيوانات المعلمة بصورة متساوية سيؤدي إلى زيادة التباين. أخيراً إذا كان توزيع الحيوانات غير متساوٍ لدرجة كبيرة قد يؤدي هذا إلى جعل تباين $\hat{\tau}$ غير محدد.

مثال (2): قام الباحثون بمسك 382 طير في مجتمع للطيور بواسطة شباك شفافة، ومن ثم قاموا بوضع علامة براقعة على جميع الطيور التي مُسِكت، يستطيع المشاهد أن يرى هذه العلامات بواسطة منظار من مسافة بعيدة. العينة الثانية تم تنفيذها باستخدام ثماني خطوط محددة الأطوال، جرى وضعها بطريقة عشوائية في مجتمع الدراسة. يقوم المشاهد أو الباحث بالتحرك على كل خط ويحصى عدد الطيور التي يشاهدها بواسطة المنظار في أثناء سيره على الخط ويحصى من بينها عدد الطيور التي تحمل الإشارة المميزة. الجدول الآتي يبين عدد الطيور التي تم مشاهدتها y وعدد الطيور المعلمة x لكل خط

الخط	1	2	3	4	5	6	7	8
Y	33	13	1	45	21	82	14	0
X	20	2	0	15	5	39	4	0

استخدم تقدير النسبة لتقدير حجم المجتمع ومن ثم أوجد 95% فترة ثقة لتقدير حجم المجتمع.

الحل: نسبة العينة هي

$$r = \frac{\bar{y}}{\bar{x}} = \frac{26.125}{10.625} = 2.459$$

تقدير عدد الطيور

$$\hat{\tau} = \frac{\bar{y}}{\bar{x}} X = r X = 2.459(382) \approx 939$$

تقدير التباين

$$\sum_{i=1}^n (y_i - rx_i)^2 = [33 - 2.459(20)]^2 + L + [0 - 0(2.459)]^2 = 680.318$$

$$s^2(\hat{\tau}) = \left(\frac{X}{\bar{x}}\right)^2 \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - rx_i)^2 = \left(\frac{382}{10.625}\right)^2 \frac{1}{8(7)} (680.318) = 15,703.37$$

أما 95% فترة ثقة تقريبية لتقدير حجم المجتمع فتكون

$$\hat{\tau} \pm t_{\alpha/2} \sqrt{s^2(\hat{\tau})} = 939 \pm 2.365 \sqrt{15,703.37} = (643, 1235)$$

ولابد من الإشارة إلى أن تقدير التباين لتقدير بيترسن سيكون

$$s^2(\hat{\tau}) = \frac{Xy(X-x)(y-x)}{x^3} = \frac{382(209)(382-85)(209-85)}{85^3} = 4,789$$

وهو أقل بكثير من تقدير قيمة التباين التي حصلنا عليها هنا، والسبب يعود إلى أن تقدير بيترسن لم يأخذ في الحسبان عدم التجانس المكاني داخل المجتمع، وكذلك لا توجد استقلالية بين الطيور كما يفترضها التقدير. على هذا يكون التقدير هنا أكثر دقة لتقدير التباين الحقيقي لتقدير حجم المجتمع.

ثانياً: العينة العشوائية مع عدم الإرجاع لوحدات الكشف

يمكن استخدام العينة العشوائية من دون إرجاع الوحدات وذلك عندما تكون الوحدات في الكشف عبارة عن مجموعات غير متداخلة أو عندما تكون منطقة الدراسة مقسمة إلى مجموعة من المناطق الصغيرة، كل منطقة تمثل وحدة مستقلة مع كون احتمالات كشف جميع الحيوانات متساوية في جميع الوحدات التي تم سحبها في العينة. إن طريقة الإمساك وإعادة الإمساك تتضمن سحب عينة ثانية بحجم n من وحدات الكشف التي يبلغ عددها بالمجتمع N من دون إرجاع مع احتمال كشف أو مسك العناصر (الحيوانات) ثابتة لجميع العناصر الموجودة في وحدة الكشف كما هو موصوف من قبل Seber (1982, pp.111-114) و Wolter (1986). إن تقدير حجم المجتمع τ سيكون نفس تقدير بيترسن وهو

$$\bar{\tau} = \frac{y}{x} X$$

و مع تقدير للتباين كما اقترح من قبل Wolter (1986) وهو

$$s^2(\bar{\tau}) = \left(\frac{X}{x}\right)^2 \frac{N-n}{Nn(n-1)} \sum_{i=1}^n (y_i - rx_i)^2 + \frac{X\bar{y}(m_x - \bar{x})(\bar{y} - \bar{x})}{\bar{x}^3}$$

حيث إن $\mu_x = \frac{X}{N}$. يراجع Wolter (1986) لمزيد من المعلومات.

5.14 تقدير قابلية الاكتشاف باستخدام نماذج الإمساك وإعادة الإمساك

إن استخدام طريقة الإمساك وإعادة الإمساك لتقدير حجم المجتمع تقوم بصورة مباشرة أو غير مباشرة بتقدير احتمالات الإمساك أو الاكتشاف، على سبيل المثال فإن تقدير بيترسن يشترك مع تقديرات أخرى تستعمل لتقدير حجم المجتمع الحيواني بالعلاقة نفسها التي يمكن كتابتها بالشكل الآتي

$$\hat{\tau} = \frac{y}{\hat{p}}$$

حيث إن $\hat{p} = \frac{X}{X}$ هو عبارة عن تقدير احتمال الاكتشاف أو الإمساك في العينة الثانية، كذلك يمكن استخدام طريقة الإمساك وإعادة الإمساك لتقدير الاكتشاف لمسوحات أخرى مستقلة يمكن تنفيذها بنفس طريقة المعاينة المتبعة في العينة الثانية.

إذا افترضنا التوزيع الهندسي الزائد إلى x ، فإن التباين لتقدير الاكتشاف إذا عرفنا X و y سيكون

$$\text{var}(\hat{p}) = \frac{y}{X^2} \left(\frac{X}{\tau} \right) \left(1 - \frac{X}{\tau} \right) \frac{\tau - y}{\tau - 1}$$

إذا تم اختيار n من وحدات الكشف (مع الإرجاع) مثل قطع أراض، أو خطوط مستقيمة، أو مصائد أو شبكات صيد ... إلخ، ولنفترض أن x_i تمثل عدد الحيوانات التي تحمل العلامة المميزة في الوحدة i ، لذا فإن تقدير احتمال الاكتشاف لكل وحدة سيكون

$$\tilde{p} = \frac{x'}{X}$$

حيث إن $x' = \sum_{i=1}^n x_i$. أما تقدير التباين إلى \tilde{p} فيكون

$$s^2(\tilde{p}) = \frac{s_x^2}{nX^2}$$

حيث إن

$$s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)$$

مثال (3): بالعودة إلى المثال 1 قدر احتمال الإمساك أو الاكتشاف للفئران في العينة الثانية، ومن ثم أوجد تقدير الانحراف المعياري لهذا التقدير.

الحل: احتمال إعادة الإمساك أو الاكتشاف هو

$$\hat{p} = \frac{x}{X} = \frac{50}{300} = 0.17$$

أما تقدير التباين لهذا التقدير فيمكن إيجاده باستخدام $\hat{\tau}$ بدلاً من τ لنحصل على

$$s^2(\hat{p}) = \frac{y}{X^2} \left(\frac{X}{\hat{\tau}} \right) \left(1 - \frac{X}{\hat{\tau}} \right) \frac{\hat{\tau} - y}{\hat{\tau} - 1} = \frac{200}{300^2} \left(\frac{300}{1200} \right) \left(1 - \frac{300}{1200} \right) \frac{1200 - 200}{1200 - 1} = 0.00035$$

وهذا يعطينا تقديراً للانحراف المعياري مقداره 0.02.

قام عدد من الباحثين مثل (Ahlo 1990) و (Pollock and Otto 1983) و (Burnham and Oberton 1969) و (Overton 1969) باقتراح تقديرات يمكن استخدامها لتقدير حجم المجتمع في الشكل الآتي

$$\bar{\tau} = \sum_{i=1}^{\tau} z_i / \hat{\pi}_i$$

حيث إن $z_i = 1$ إذا كانت الوحدة i في المجتمع جرى سحبها في عينة أو أكثر و $z_i = 0$ خلاف ذلك و $\hat{\pi}_i$ تمثل تقدير احتمال اكتشاف أو مسك الحيوان في العينة الأخيرة. لقد تم تقدير الاحتمالات $\hat{\pi}_i$ بطريق مختلفة منها مباشرة وأخرى غير مباشرة، وهذا قد يؤدي إلى تحيز كبير بتقدير الاحتمالات، لمزيد من المعلومات يراجع (Pollock and Otiset al 1978)

Otto(1983). كذلك يراجع (Pollock et al. (1990 لمعرفة كيفية تقدير حجم المجتمع إذا كان احتمال كشف أو مسك الحيوانات مختلف.

6.14 نماذج أخرى

باستخدام عينات متتابعة من مجتمع يتعرض أفراده للموت، أو الولادة لعناصر جديدة، أو الهجرة خارج المجتمع، أو العودة إلى المجتمع، هذا يعني أن المجتمع لن يكون مغلقاً ومن ثم لا بد من الأخذ في الحسبان هذه التغيرات التي قد تحدث خلال مدة سحب العينات المتتابعة، كل ذلك يتطلب نماذج قد تحتوي على عدد كبير من المعالم. أحد النماذج الذي يسمح بموت بعض أعضاء المجتمع والذي سببه عملية الإمساك وإعادة الإمساك أُقترح من قبل Jolly (1965) وSeber(1965) ولقد تم توسيع هذا النموذج بطرق مختلفة من قبل عدد كبير من الباحثين، يراجع (Seber(1986. في هذا النموذج الاحتمالات يمكن أن تتغير من عينة إلى أخرى ولكنها تبقى ثابتة لجميع الحيوانات لكل مسك، لقد تم استخدام النموذج الهندسي الزائد المتعدد المتغيرات. لقد قام Seber(1982,1986,1992 بمراجعة جميع البحوث المنشورة حول هذا النموذج وأضاف تعديلاته المختلفة.

هناك طريقة جديدة نوعاً ما لتحليل طريق الإمساك وإعادة الإمساك للمجموعات المفتوحة باستخدام نماذج log-linear، لمزيد من التفاصيل يراجع (Cormack (1980, 1981, 1985,1989 وAgresti(1990. في هذه النماذج لوغاريتم توقع عدد الحيوانات مع تاريخ إمساك يكون عبارة عن دالة خطية لمجموعة من المعالم التي يمكن تفسيرها.

References

1. Agresti, A. (1990). *Categorical Data Analysis*, New York, Wiley.
2. Ahlo, J. M. (1990). Logistic Regression in Capture-recapture Models. *J. Amer. Statist. Assoc.* 46, 623-635.
3. Buckland, S. T. (1980). A Modify Analysis of the Jolly-Seber Capture-recapture Model. *Biometrics*, 36, 419-435.
4. Buckland, S. T. (1984). Monte Carlo Confidence Intervals. *Biometrics*, 40, 811-817.
5. Burnham, K. P. and Overton, W. S. (1978). Estimation of the Size of a Closed Population when Capture Probability Vary Among Animals. *Biometrika*, 65, 625-633.
6. Cormack, R. M. (1979). Models for Capture-recapture. In R. M. Cormack, G. P. Patil, and D. S. Robson (eds.), *Sampling Biological Populations*. Fairland, MD: International Co-operative Publishing House, 217-255.
7. Cormack, R. M. (1980). Model Selection in Capture-recapture Experiments. *GLIM Newsletter*, Dec., 27-29.
8. Cormack, R. M. (1981). Loglinear Models for Capture-recapture Experiments on Open Populations. In R.W. Hiorns and D. Cooke (eds.). *The Mathematical Theory of the Dynamics of Biological Populations II*. London, Academic Press, 243-273.
9. Cormack, R. M. (1985). Examples of the use of GLIM to Analyse Model Capture-recapture Studies. In B.J.T. Morgan and P. M. North (eds.) *Statistics in Ornithology* > Lecture Notes in Statistics, No. 29. New York, Springer-Verlag, 243-273.
10. Cormack, R. M. (1989). Log-linear Models for Capture-recapture. *Biometrics*, 45, 395-413.
11. Cowan, C. D. and Malec, D. (1986). Capture-recapture Models when both Sources have Clustered Observations. *J. Amer. Statist. Assoc.* 81, 347-353.
12. Freedman, D. A. (1991). Adjusting the 1990 Census. *Science*, 252, 1233-1236.
13. Jolly, G. M. (1965). Explicit Estimates from Capture-recapture with both Death and Immigration-stochastic Model. *Biometrika*, 52, 225-247.
14. Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D.R. (1978). Statistical for Capture Data from Closed Population. *Wildlife Monographs*, No. 2.
15. Otto, M. C. and Pollock, K. H. (1990). Size Bias in Transect Sampling: A field Test. *Biometrics*, 239-245.
16. Overton, W. S. (1969). Estimating the Number of Animals in Wildlife Population. In R. H. Giles (ed). *Wildlife Management Techniques*, 3rd ed. Washington, DC, Wildlife Society, 403-455.
17. Pollock, K. H. (1981). Capture-recapture Models: A Review of Current Methods, Assumptions and Experimental Design. In C. J. Ralph and J. M. Scott (eds.), *Estimating Numbers of Terrestrial Birds*, Studies in Avian Biology, No. 6 Oxford, Pergamon Press, 426-435.
18. Pollock, K. H. (1991). Modeling Capture-recapture, and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: Past Present and Future. *J. Amer. Statist. Assoc.* 86, 225-238.
19. Pollock, K. H., Nichols, J. D., Hines, J. E. and Brownie, C. (1990). Statistical Inference for Capture-recapture Experiments, *Wildlife Monograph*, 107, 1-97.

20. Pollock, K. H. and Otto, M. C. (1983). Robust Estimation of a Population Size in Closed Animal Population from Capture-recapture Experiments. *Biometrics*, 39, 1035-1049.
21. Seber, G. A. F. (1965). A Note on the Multiple-recapture Census. *Biometrika*, 49, 339-349.
22. Seber, G. A. F. (1973). *The Estimation of the Animal Abundance*, London, Griffin.
23. Seber, G. A. F. (1982). *The Estimation of the Animal Abundance*, 2nd ed. London, Griffin.
24. Seber, G. A. F. (1986). A Review of Estimating Animal Abundance. *Biometrics*, 42, 267-292.
25. Seber, G. A. F. (1992). A Review of Estimating Animal Abundance. *II International Statistical Review*, 60, 129-166.
26. Sekar, C. C. and Deming, W. E. (1949). On a Method of Estimating Birth and Death Rates and the Extent of Regression. *J. Amer. Statist. Assoc.* 5, 119-127.
27. Sudman, S., Sirken, M. G. and Cowan, C. D. (1988). Sampling Rare and Elusive Populations. *Science*, 240, 991-996.
28. Thompson, S. K (2002). *Sampling*, 2nd Wiley, New York.
29. Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *J. Amer. Statist. Assoc.* 81, 338-346.
30. Wolter, K. M. (1991). Accounting for America's Uncounted and Miscoounted. *Science*, 253, 12-15.

الفصل الخامس عشر

المعاينة بالخط القاطع

Line Transect Sampling

1.15 مقدمة

إن المسح بطريقة الخط القاطع للحيوانات أو النباتات الموجودة في منطقة الدراسة، يجرى عادة بأن يقوم العّاد أو الباحث بالتحرك على الخط الذي يجري اختياره ليقطع منطقة الدراسة، ويقوم بمشاهدة الحيوانات أو النباتات على جانبي الخط، ويحدد موقعها، وبعد ذلك يقوم بقياس المسافة العمودية بين موضع الحيوان أو النبات والخط القاطع، عادة ما تتم مشاهدة أعداد كبيرة من الحيوانات أو النباتات القريبة من الخط ليس لأن كثافتها قرب الخط أكبر من المناطق البعيدة، ولكن احتمال اكتشاف أو مشاهدة الحيوانات أو النباتات القريبة من الخط تكون أكبر، ويتناقص بشكل تدريجي كلما ابتعدنا عن الخط القاطع، ولا بد من الأخذ في الحسبان حالة عدم ثبات اكتشاف الحيوانات أو النباتات عندما نقوم بتقدير عدد أو كثافة الحيوانات في منطقة الدراسة.

لقد جرى استخدام طريقة الخط القاطع لمجتمعات متعددة من ضمنها مجتمعات الطيور والحيوانات وأصناف كثيرة من النباتات، كذلك للمجتمعات التي تكون عملية اكتشاف أفرادها يعتمد على موضع أو مكان العّاد أو الباحث، وكذلك جرى استخدام الخط القاطع للمسوحات الجوية باستخدام الطائرات، والمسوحات البحرية باستخدام القوارب، والمسوحات

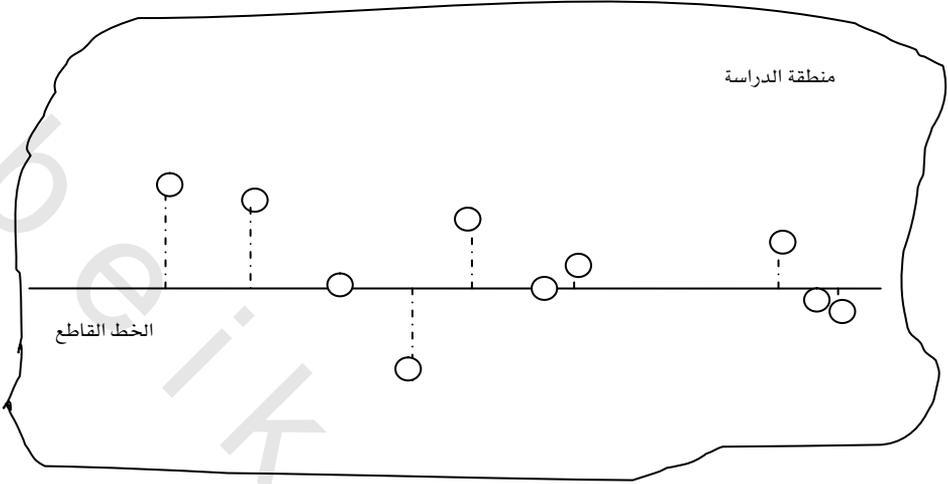
البرية باستخدام السيارات. وللتسهيل سوف نستخدم عبارة "حيوان أو حيوانات" للتعبير عن وحدات المجتمع الذي يجري دارسته، وقد تكون هذه الوحدات غير حيوانات.

عادة ما يوصف المسح بطريقة الخط أو الخطوط القاطعة بدالة الاكتشاف التي تعطي احتمال أن تتم عملية اكتشاف الحيوان (النبات) وهو في موضع معين من الخط القاطع، وعادة ما يتناقص احتمال اكتشاف الحيوان كلما كان موقعه من الخط أبعد، ويكون احتمال اكتشاف الحيوان (النبات) مساوياً لواحد إذا كان على الخط في أغلب الأحيان.

لمزيد من المعلومات ولتفاصيل أكثر يراجع: (1979) Gates و(1992) Buckland et al. و(1980) Burnham et al. و(1978) Eberhardt و(1992) Seber و(1982, 1986, 1992) Ramsey et al.

2.15 طرق تقدير الكثافة

الشكل 1: يوضح الحيوانات أو أي عناصر أخرى جرى مشاهدتها من على الخط القاطع، المسافة العمودية من العناصر إلى الخط القاطع جرى توضيحها بالخطوط المتقطعة. السؤال المهم هو كيف يمكننا أن نقدر كثافة العناصر في منطقة الدراسة إذا كانت لدينا المسافات العمودية من العناصر إلى الخط أو الخطوط القاطعة وكانت قابلية الاكتشاف للعناصر على الخط أو قربه كاملة، أي نستطيع أن نكتشف جميع الحيوانات أو العناصر، ولكن قابلية الاكتشاف تتناقص كلما ابتعدنا من الخط. سوف نبدأ باستخدام هذه المعلومات لتقدير كثافة الحيوانات أو العناصر وذلك باستخدام بعض الطرق البسيطة أو البدائية، ومن ثم ننتقل إلى الطرق الأكثر تقدماً.



الشكل 1: مشاهدة الحيوانات (العناصر) من الخط القاطع

إن الهدف الرئيس من الطرق التي سنتكلم عنها هنا هو تقدير كثافة الحيوانات (العناصر) في منطقة الدراسة. فيما يأتي بعض الرموز التي سنحتاجها في هذا الفصل:

A : مساحة منطقة الدراسة.

y_i : عدد الحيوانات (العناصر) التي تم مشاهدتها من الخط القاطع i .

n : حجم العينة ويكون هنا عدد الخطوط القاطعة التي جرى اختيارها في منطقة الدراسة.

T : عدد الحيوانات (العناصر) في منطقة الدراسة.

$D=T/A$: كثافة الحيوانات (العناصر) في منطقة الدراسة.

من الجدير بالذكر هنا أن Burnham et al. (1980) اقترح أن يكون عدد الحيوانات (العناصر) التي جرى مشاهدتها لا يقل عن 40 من أجل الحصول

على تقدير يمكن الاعتماد عليه لكثافة الحيوانات أو أي عناصر أخرى في منطقة الدراسة.

1.2.15 طريقة الشريط الضيق (Narrow-Strip Method)

على الرغم من كون قابلية اكتشاف الحيوانات البعيدة من الخط تكون غير كاملة، فإنه لا بد من وجود شريط ولو ضيق حول الخط تكون فيه قابلية الاكتشاف كاملة. باستخدام العناصر الموجودة داخل الشريط وبإهمال العناصر الموجودة خارجه يمكننا أن نقدر عدد عناصر المجتمع أو الكثافة لهذه العناصر في المجتمع باستخدام الطريقة الآتية:

لنفترض أن L تمثل طول الخط و w_0 تمثل المسافة القصوى من الخط والتي تكون قابلية الاكتشاف فيها كاملة. إذاً سيكون عرض الشريط $2w_0$ ومساحته $2w_0L$. لنفترض أن y_0 يمثل عدد الحيوانات أو العناصر التي تم مشاهدتها في الشريط. لتقدير الكثافة D في الشريط الضيق يمكننا أن نستخدم

$$\hat{D} = \frac{y_0}{2w_0L}$$

إذا كانت مساحة منطقة الدراسة A ، فإن عدد الحيوانات في المجتمع يمكن تقديره باستخدام

$$\hat{T} = A\hat{D} = \frac{Ay_0}{2w_0L}$$

بصورة عامة تكون المسافة w_0 أقل من المسافة القصوى التي يمكن أن تكتشف فيها العناصر، لذلك فإن y_0 يكون أقل من عدد العناصر المكتشفة، هنالك طرق عديدة جرى إتباعها لاختيار المسافة w_0 يكون فيها

الاكتشاف كاملاً، أحد هذه الطرق تكون بدراسة المدرج التكراري لمسافة الاكتشاف، وتجري ملاحظة المسافة التي ينحدر عندها التكرار النسبي للمشاهدات بصورة حادة.

مثال 1: لتقدير كثافة الطيور في إحدى المناطق قام الباحث باستخدام الخط القاطع وبطول $L=100$ متر، وتم اكتشاف $y=18$ طير بالمسافات الآتية من الخط:

0, 0, 1, 3, 7, 11, 11, 12, 15, 15, 18, 19, 21, 23, 28, 33, 34, 44

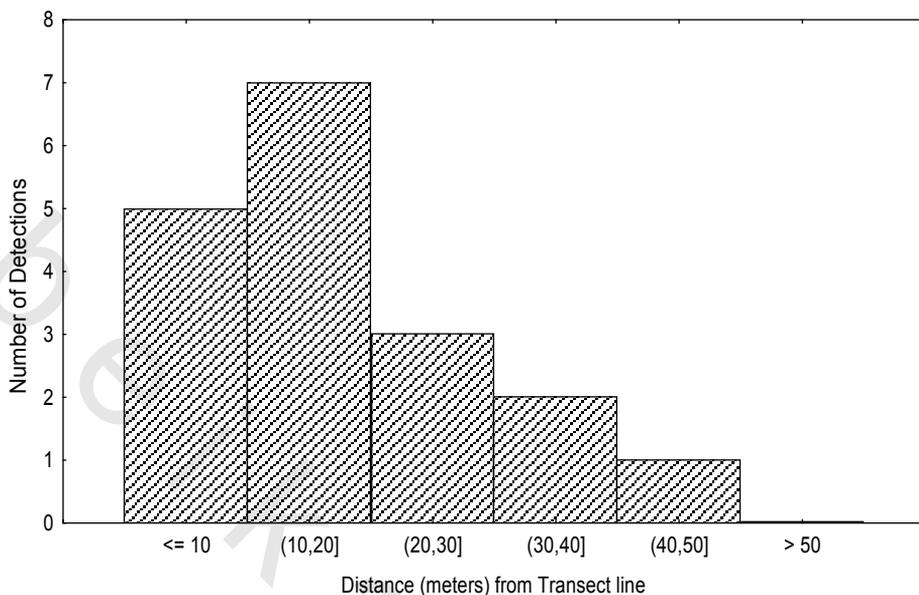
الحل: إذا قمنا برسم المدرج التكراري شكل 2 بفئات أطوالها 10 متر، نلاحظ أن 5 طيور تم مشاهدتها في الفئة الأولى، و 7 طيور في الفئة الثانية ما بين 10 إلى 20 متر، و 3 طيور في الفئة من 20 إلى 30 متر، وهكذا، نلاحظ أن عدد الطيور المشاهدة بعد الفئة الثانية قد انخفض بشكل ملحوظ؛ لذا اخترنا $w_0 = 20$ ، وعليه سيكون عرض الشريط الضيق $2 w_0$ يساوي 40م. لذا فإن عدد الطيور التي يمكن مشاهدتها في الشريط الضيق هي $y_0 = 12$. وباستخدام طريقة الشريط الضيق يكون تقدير كثافة الطيور

$$\hat{D} = \frac{y_0}{2 w_0 L} = \frac{12}{2(20)(100)} = 0.003$$

لذا فإن تقدير كثافة الطيور بالمتر المربع هي 0.003 طير.

على الرغم من أن طريقة الشريط الضيق سهلة جداً، إلا أنها غير مرضية لعدة أسباب منها:

1. لم يتم استخدام جميع المشاهدات التي جرى جمعها.
2. إن تحديد عرض الشريط الضيق جرى بشكل اعتباطي.
3. وأخيراً لأن قابلية اكتشاف الحيوانات لن تتناقص بشكل منتظم مع المسافة، لذا فإن عرض الشريط الضيق مع قابلية اكتشاف كاملة قد يكون صفرًا.



الشكل 2: المدرج التكراري لعدد الطيور التي تم اكتشافها لكل 10م من الخط القاطع

2.2.15 الطرق المعلمانية (Parametric Methods)

عندما تتم مشاهدة الحيوانات في شريط من الأرض، هذا يعني أن جميع الحيوانات في الشريط الذي نصف عرضه يساوي w وطوله يساوي L قد تمت مشاهدتها، لذا فإن تقدير الكثافة يكون $\hat{D} = y/2Lw$ أي عدد الحيوانات مقسوماً على مساحة الشريط، عندما تتم مشاهدة الحيوانات من الخط القاطع بدالة اكتشاف $g(x)$ لديها قابلية اكتشاف كاملة على الخط وتتناقص بالبعد عن الخط بمسافة x ، فإن المسافة لمشاهدة الحيوانات من الخطوط التي جرى وضعها في منطقة الدراسة بشكل عشوائي سيكون لديها دالة كثافة الاحتمال $f(X)$ وبنفس الشكل لدالة الاكتشاف $g(x)$ ولكن

المساحة تحت دالة كثافة الاحتمال $f(X)$ تساوي واحداً، وتقدير الكثافة سيكون

$$\hat{D} = \frac{y\hat{f}(0)}{2L}$$

والمشكلة التي تواجهنا تمكن بتقدير $f(0)$ الكثافة عند المسافة صفر من الخط القاطع.

يمكننا أن نتصور شريطاً مماثلاً من الأرض بعرض w تكون فيه قابلية الاكتشاف كاملةً، ويكون عدد الحيوانات المكتشفة في هذا الشريط متساوية بالمعدل لعدد الحيوانات المكتشفة من الخط القاطع، وتكون قابلية الاكتشاف متناقصة مع المسافة، العلاقة بين الخط القاطع والشريط المماثل من الأرض تكون

$$f(0) = \frac{1}{w}$$

و w يدعى نصف العرض الفاعل للخط القاطع. وبدلالة w يمكن تقدير الكثافة بالاعتماد على w الذي هو تقدير إلى w لتكون

$$\hat{D} = \frac{y}{2Lw}$$

يمكننا إما تقدير w أو $f(0)$ ومن ثم نستطيع تقدير الكثافة.

عندما يمكننا افتراض توزيع احتمالي معين يعتمد على معالم غير معلومة لدالة الاكتشاف $g(x)$ ، نستطيع أن نستخدم بعض الطرق الإحصائية مثل طريقة التقدير الاحتمالي العظمى (mle) لتقدير المعالم غير المعلومة من ثم نقدر $f(0)$ أو w . لقد تم دراسة بعض التوزيعات الاحتمالية التي تعتمد على معالم غير معلومة من قبل مجموعة من الباحثين منهم: Ramsey(1979)

وBuckland(1985) وBurnham et al. (1980) وQiunn and Gallucci(1980) وPollock(1978). سوف نستخدم توزيعان من أبسط التوزيعات كأمثلة في هذا الفصل.

لابد من الإشارة إلى أن من الفوائد لافتراض توزيع بسيط لدالة الاكتشاف سيؤدي إلى تقدير بسيط لكثافة المجتمع إذا كان الافتراض صحيحاً، ولكن من عيوبه أنه قد لا يكون لهذا التوزيع البسيط المرونة الكافية لتمثيل دالة الاكتشاف الحقيقية. سنتناول هنا توزيعين لدالة الاكتشاف هما التوزيع الأسي والتوزيع الطبيعي النصفى؛ وذلك لكونهما يؤديان إلى تقديرات بسيطة للكثافة.

في حالة التوزيع الأسي تكون دالة الاكتشاف $g(x)=\exp(-x/w)$. كلما كانت قيمة المعلمة w كبيرة أدى إلى أن تكون قابلية الاكتشاف من الخط كبيرة. باستخدام طريقة التقدير الاحتمالي العظمى (mle) نحصل على تقدير إلى $w = \bar{x}$ أي معدل مسافات الاكتشاف. لمزيد من المعلومات يراجع Ramsey(1979).

مثال 2: لنفترض أن دالة الاكتشاف للبيانات في المثال 1 تتبع التوزيع الأسي أي $g(x)=\exp(-x/w)$. قدر كثافة الطيور في منطقة الدراسة.

الحل: باستخدام البيانات في المثال السابق نستطيع أن نحصل على $w = \bar{x} = 16.39$ وعليه فسيكون تقدير الكثافة

$$\hat{D} = \frac{y}{2Lw} = \frac{18}{2(100)(16.39)} = 0.0055$$

بالرغم من كون استخدام التوزيع الأسي يؤدي إلى تقدير بسيط جداً لكثافة المجتمع، ولكنه غير واقعي لكثير من المجتمعات ولا يؤدي إلى تقدير دقيق للكثافة. أكثر من باحث مثل Buckland (1985) و Burnham et al.

(1980) و(1978) Eberhardt بحثوا هذا الموضوع وقالوا: أن دالة الاكتشاف يجب أن يكون لها مشتقة تساوي صفراً قريباً من الخط القاطع. أبسط التوزيعات التي ينطبق عليه هذا الشرط هو التوزيع الطبيعي النصفى. وبدالة احتمال هي

$$g(x) = \exp\left(\frac{-\pi x^2}{4w^2}\right)$$

ويتقدير إلى w

$$w = \sqrt{\frac{\pi}{2y} \sum_{i=1}^y x_i^2}$$

مثال 3: لنفترض أن دالة الاكتشاف للبيانات في المثال 1 تتبع التوزيع الطبيعي النصفى. قدر كثافة الطيور في منطقة الدراسة.
الحل: باستخدام البيانات في المثال 1 يمكننا حساب

$$\frac{1}{y} \sum_{i=1}^y x_i^2 = \frac{1}{18} (0^2 + 0^2 + \dots + 44^2) = 417.5$$

لذا فإن تقدير w سيكون

$$w = \sqrt{\frac{\pi}{2y} \sum_{i=1}^y x_i^2} = \sqrt{\left(\frac{3.1417}{2}\right)(417.5)} = 25.61$$

وعليه فإن تقدير الكثافة سيكون

$$\hat{D} = \frac{y}{2Lw} = \frac{18}{2(100)(25.6)} = 0.0035$$

3.2.15 الطرق غير المعلمانية (Nonparametric Methods)

لتفادي افتراض أن دالة الاكتشاف تتبع توزيعاً احتمالياً معيناً يمكننا أن نستخدم الطرق غير المعلمانية للتقدير، باستخدام البيانات لمتغير عشوائي من

دالة كثافة الاحتمال f ، الطرق غير المعلماتية تستخدم طرق تمهيد Smoothing معينة لتقدير قيمة $f(x)$ دالة كثافة الاحتمال لأي قيمة للمتغير x . وللخط القاطع فإن دالة كثافة الاحتمال التي تهمنها هي التي الدالة تقيس وتكتشف المسافات من الخط. سنتناول طريقتين من الطرق غير المعلماتية في هذا الفصل، ولمزيد من المعلومات يراجع (1985) Buckland و Burnham et al. و (1980) و Johnson and Routledge (1985). قابلية الاكتشاف على الخط تكون كاملة، لذا فإن تقدير الكثافة يكون $\hat{D} = y\hat{f}(0)/2L$.

تقدير $f(0)$ باستخدام طريقة الكرنل (Kernel Method)

هناك كم كبير من البحوث الإحصائية التي تتناول تقدير دالة كثافة الاحتمال معظمها استخدم طريقة الكرنل للتقدير. يُعدُّ (1986) Seber أول من اقترح استخدام طريقة الكرنل لتقدير الكثافة باستخدام الخط القاطع، وقد استخدمت الطريقة من قبل (1993) Quang لمسائل قريبة.

تستخدم الطريقة دالة الكرنل $K(x)$ التي يكون تكاملها 1

$$\int_{-\infty}^{\infty} K(x) dx = 1$$

إن تقدير دالة كثافة الاحتمال f عند القيمة x باستخدام طريقة الكرنل

هو

$$\hat{f}(x) = \frac{2}{yh} \sum_{j=1}^y K\left(\frac{x-x_j}{h}\right)$$

حيث إن h تمثل عرض النافذة و x_j قيمة المشاهدة z (التي تمثل المسافة العمودية من الخط القاطع إلى الحيوان z) و y تمثل عدد الحيوانات التي تم مشاهدتها، أما الرقم 2 فيستخدم عندما تحسب الكثافة بغض النظر عن أي جانب من الخط الذي تمت مشاهدة الحيوان عنده.

يمكننا تقدير $f(0)$ مع كون دالة الكرنل متماثلة بما يأتي

$$\hat{f}(0) = \frac{2}{yh} \sum_{j=1}^y K\left(\frac{x_j}{h}\right)$$

فإذا كانت دالة الكرنل تتبع التوزيع الطبيعي فإن

$$k\left(\frac{x_j}{h}\right) = \frac{1}{\sqrt{2p}} e^{-(1/2)(x_j/h)^2}$$

لقد أعطى Silverman (1986) قاعدة بسيطة للاختيار عرض النافذة h وهي

$$h = 0.9 a y^{-1/5}$$

حيث إن $a = \min(s, Q/1.34)$ ، أما s فتمثل الانحراف المعياري للمشاهدات التي هي عبارة عن المسافة العمودية للحيوانات المكتشفة من الخط، و Q فتمثل المدى الربيعي، ولكن إذا كنا نتعامل مع مسافة موجبة فقط فيجب أن نستخدم الوسيط بدلاً من المدى الربيعي.

تقدير كثافة الحيوانات في المجتمع ستكون

$$\hat{D} = \frac{y\hat{f}(0)}{2L}$$

مثال 4: إذا كانت دالة الكرنل تتبع التوزيع الطبيعي، استخدمت البيانات في المثال 1 لتقدير كثافة الطيور في منطقة الدراسة.

الحل: الوسيط للقيمة المطلقة للبيانات في المثال 1 هو 15 و $15/1.34=11.19$. بما أن 11.19 أقل من قيمة الانحراف المعياري $s = 12.56$ وباستخدام الطريقة التي اقترحها Silverman (1986) يمكننا إيجاد عرض النافذة

$$h = 0.9(11.19)(18)^{-1/5} = 5.56$$

لذا سيكون تقدير $f(0)$ باستخدام طريقة الكرنل مع دالة تتبع التوزيع

الطبيعي هو

$$\hat{f}(0) = \frac{2}{yh} \sum_{j=1}^y K\left(\frac{x_j}{h}\right) = \frac{2}{yh} \sum_{j=1}^y \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x_j/h)^2}$$

الآن باستخدام البيانات في المثال 1 نستطيع تقدير $f(0)$

$$\hat{f}(0) = \frac{2}{yh} \sum_{j=1}^y \frac{1}{\sqrt{2\pi}} e^{-(1/2)(x_j/h)^2} = \frac{2}{18(5.65)\sqrt{2\pi}} [e^{-(1/2)(0/5.65)^2} + L + e^{-(1/2)(44/5.65)^2}] = 0.0376$$

يمكننا استخدام $\hat{f}(0)$ لتقدير كثافة الطيور في منطقة الدراسة

$$\hat{D} = \frac{y\hat{f}(0)}{2L} = \frac{18(0.0376)}{2(100)} = 0.0034$$

تقدير $f(0)$ باستخدام طريقة سلاسل فوريير (Fourier Series Method)

إن تقدير $f(0)$ باستخدام طريقة سلاسل فوريير هو

$$\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^m \hat{A}_k$$

حيث إن w^* يمثل المسافة العظمى التي يمكن أن يشاهد فيها الحيوان

و \hat{A}_k يمكن تعريفه

$$\hat{A}_k = \frac{2}{yw^*} \left[\sum_{i=1}^y \cos(k\pi x_i/w^*) \right]$$

أما m فتمثل عدد الحدود التي سنستخدمها لإيجاد القيم التقريبية إلى

$\hat{f}(0)$ ، ويمكننا استخدام القاعدة التي اقترحها (Burnham et al. 1980) التي

تقول ابدأ بـ $m=1$ ومن ثم اختر أول عدد صحيح m بحيث إن

$$\frac{1}{w^*} \sqrt{2/(y+1)} \geq |\hat{A}_{m+1}|$$

أما فيما يخص تحديد w^* المسافة العظمى التي يمكن أن يشاهد فيها الحيوان فقد اقترح كلُّ من (Burnham et al. (1980) و (Crain et al. (1979) استخدام مسافة أقل من المسافة العظمى التي تمت مشاهدة أحد الحيوانات فيها، وذلك من خلال إهمال 1 إلى 3% من المشاهدات الكبيرة على أنها شاذة. مزيد من المعلومات يراجع (Burnham et al. (1981) و (Quang (1990). مثال 5: استخدم طريقة سلاسل فورير لتقدير $f(0)$ واستخدمها لتقدير كثافة الطيور في منطقة الدراسة.

الحل: إذا استخدمنا البيانات في المثال 1 فإن القيمة العظمى هي 44 متراً، وإذا أهملنا هذه القيمة على أنها قيمة شاذة فإننا سوف نستخدم $w^* = 34$ والقاعدة أعلاه تكون متحققة عندما تكون $m=1$ ، لذلك نحتاج إلى حد واحد فقط، لذلك نحتاج حساب \hat{A}_1 فقط. وهو الذي يحتوي على 17 حداً، تمثل عدد المشاهدات بعد إهمال القيمة العظمى.

$$\hat{A}_1 = \frac{2}{17(34)} \left[\cos\left(\frac{1(3.1417)(0)}{34}\right) + L + \cos\left(\frac{1(3.1417)(34)}{34}\right) \right] = 0.0091$$

لذا فإن تقدير $f(0)$ سيكون

$$\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^m \hat{A}_k = \frac{1}{34} + 0.0091 = 0.0385$$

أخيراً فإن تقدير الكثافة يكون

$$\hat{D} = \frac{y\hat{f}(0)}{2L} = \frac{18(0.0385)}{2(100)} = 0.0033$$

3.15 تصاميم لاختيار الخطوط

إن تصميم المعاينة للمسح بالخط القاطع عبارة عن الطريقة التي يتم فيها اختيار مواقع الخطوط، تعتمد بعض الخواص المهمة مثل عدم الانحياز لتقدير

الكثافة أو تقدير التباين على تصميم الخطوط وليس على الفروض حول مجتمع الدراسة.

يمكن أن نشاهد مجموعة كبيرة من الحيوانات باستخدام خط واحد، ولكن لا يزال حجم العينة 1. وللحصول على تقدير دقيق للكثافة لابد من حجم عينة مساوٍ إلى n من الخطوط، بصورة خاصة إذا كان توزيع الحيوانات في المجتمع غير متماثل في مناطقه المختلفة.

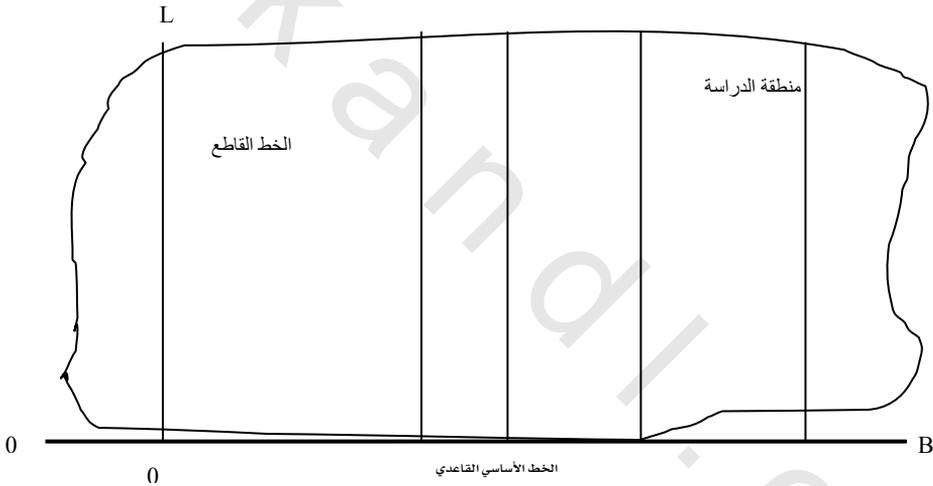
إن تقدير التباين بالاعتماد على عينة من عدة خطوط يفضل على التقدير الذي يعتمد على خط واحد. ولقد أكد وأوصى عدد كثير من الباحثين على ضرورة أن يعتمد تقدير التباين على أكثر من خط مثل: Burnham and Anderson (1976) و Eberhardt (1978) و Overton (1969) و Seber (1982). تعتمد الطرق التي تقدر التباين باستخدام خط واحد على وضع فروض على توزيع الحيوانات بالمجتمع منها: أن الحيوانات تتوزع بشكل متماثل وبصورة مستقلة في منطقة الدراسة، ولا بد من الإشارة هنا إلى أن مثل هذه الفروض غير عملية لأن الحيوانات غالباً ما تعيش في مجتمعات مع بعضها بعضاً في أجزاء معينة من مجتمعات وجودها.

4.15 اختيار الخطوط بطريقة عشوائية

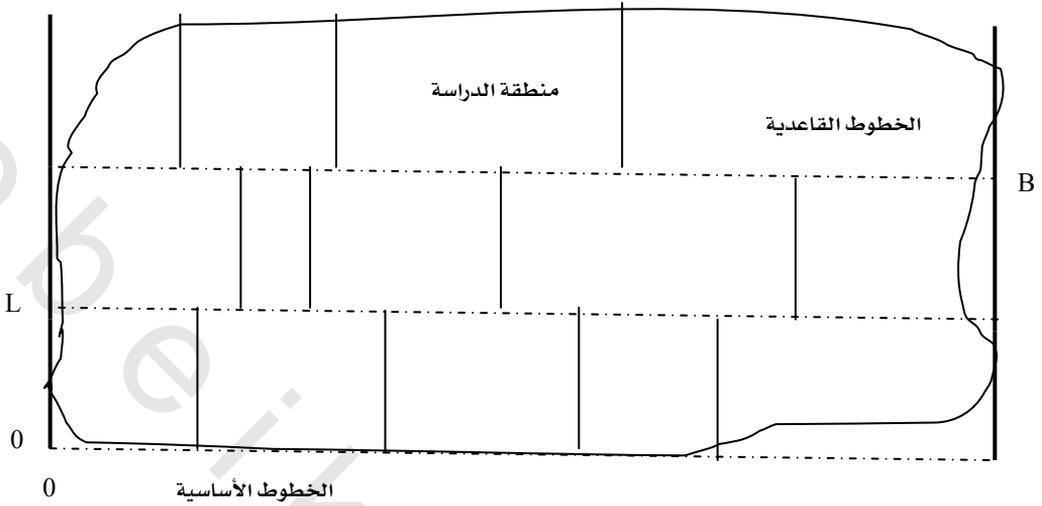
يمكن اختيار عينة عشوائية بحجم n من الخطوط في منطقة الدراسة وذلك برسم الخط الأساسي بطول B على خريطة منطقة الدراسة خلال منطقة الدراسة أو تحتها كما هو موضح في الشكل 3. ليس بالضرورة أن تكون منطقة الدراسة منتظمة الشكل. بعد ذلك نقوم بسحب n من الأعداد العشوائية من التوزيع المنتظم التي تقع في الفترة بين 0 و B ولتكن V_1, V_2, \dots, V_n ومن ثم نقوم برسم n من خطوط التقاطع التي تكون مواضعها V_1, V_2, \dots, V_n وتكون عمودية على الخط الأساسي. الشكل 3

يحتوي على خمسة خطوط عمودية على الخط الأساسي بطول L متر جرى تحديد مواضعها بسحب V_1, V_2, \dots, V_5 من الأعداد العشوائية في الفترة بين 0 و B .

إذا كانت منطقة الدراسة عرضها أكبر من طول الخط القاطع، نقوم برسم خطوط أساسية متوازية تبعد عن بعضها بعضاً مسافة L متر، ومن ثم نقوم بسحب نقاط البداية V_1, V_2, \dots, V_n للخطوط القاطعة في الفترة بين 0 و B كما هو موضح بالشكل 4، تمثل الخطوط المتقطعة الخطوط الأساسية، وهناك 11 خطاً قاطعاً جرى تحديد مواضعها في الفترة 0 و B بسحب نقاط البداية V_1, V_2, \dots, V_{11} من الأعداد العشوائية في الفترة بين 0 و B .



الشكل 3: عينة عشوائية من 5 خطوط بطول L متر لكل خط في منطقة الدراسة. إذا كانت منطقة الدراسة عرضها أكبر من طول الخط القاطع، فعلياً أن نقوم برسم خطوط أساسية متوازية تبعد عن بعضها الآخر بمسافة L متر. ومن ثم نقوم بسحب نقاط البداية V_1, V_2, \dots, V_n للخطوط القاطعة في الفترة بين 0 و B كما هو موضح بالشكل 4. تمثل الخطوط المتقطعة الخطوط الأساسية، وهناك 11 خطاً قاطعاً جرى تحديد مواضعها في الفترة 0 و B بسحب نقاط البداية V_1, V_2, \dots, V_{11} من الأعداد العشوائية في الفترة بين 0 و B .



الشكل 4: عينة عشوائية من 11 خطوط بطول L متر لكل خط في منطقة الدراسة

قد تكون منطقة الدراسة غير منتظمة لهذا لا يمكن أن نحصل على خطوط متساوية الأطوال، أي أن طول الخط القاطع يكون متغيراً عشوائياً L_i . لمزيد من المعلومات يراجع (Thompson 2002).

5.15 اختيار الخطوط باحتمالات متناسبة مع أطوالها

يمكننا اختيار الخطوط باحتمالات متناسبة مع أطوالها وذلك عن طريق اختيار n من النقاط بصورة مستقلة من التوزيع المنتظم من منطقة الدراسة. ويمكن تحقيق ذلك من خلال وضع منطقة الدراسة داخل مستطيل، من ثم نقوم باختيار إحداثيات النقاط بطريقة عشوائية حتى نحصل على n من النقاط داخل منطقة الدراسة. لكل نقطة جرى سحبها داخل منطقة الدراسة نقوم برسم خط يمر بهذه النقطة، ويكون عمودياً على الخط الأساسي ويقطع منطقة الدراسة. نلاحظ أن الخط الذي يقع في جزء واسع من منطقة الدراسة يكون طويلاً، لذا فإن احتمال اختياره يكون أكبر؛ لأن هنالك عدد أكبر

من النقاط في منطقة الدراسة تؤدي إلى اختياره؛ لذا فإن دالة كثافة الاحتمال للخط القاطع الذي بدأ من الموضع v على الخط الأساسي ستكون $L(v)/A$.

إذا كان لدينا خط واحد جرى سحبه بالطريق أعلاه، يمكننا أن نقدر الكثافة بما يأتي

$$\hat{D} = \frac{y}{2wL(v)}$$

حيث إن w تأثير نصف العرض للخط القاطع، و v نقطة التقاطع للخط القاطع مع الخط الأساسي، و $L(v)$ عرض منطقة الدراسة أو طول الخط العمودي من نقطة التقاطع الواقعة على الخط الأساسي إلى نهاية منطقة الدراسة. بعد هذا يمكننا إثبات أن \hat{D} تقدير غير متحيز إلى D ، لمزيد من المعلومات يراجع (Thompson 2002).

لنفترض أن \hat{D}_i تمثل تقدير الكثافة من الخط i في العينة بالحجم n ، لذا فإن تقدير الكثافة سيكون

$$\hat{D}_p = \frac{1}{n} \sum_{i=1}^n \hat{D}_i$$

هو تقدير غير متحيز إلى D . وبما أن الخطوط سحبت بطريقة عشوائية بالاعتماد على نقاط البداية، فإن \hat{D}_i تُعدُّ متغيرات عشوائية ومستقلة ولديها التوزيع نفسه. ويمكننا أن نحصل على تقدير غير متحيز لتقدير تباين \hat{D}_p باستخدام \hat{D}_i وهو

$$s^2(\hat{D}_p) = \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \hat{D}_p)^2$$

References

1. Barry, S. C. and Welsh, A. H. (2001). Distance Sampling Methodology, *Journal of the Royal Statistical Society B.* 63, 31-51.
2. Buckland, S. T. (1982). A Note on the Fourier Series Model for Analysing Line Transect Data, *Biometrics.* 38, 469-477.
3. Buckland, S. T. (1985). Perpendicular Distance Models for Line Transect Sampling, *Biometrics.* 41, 177-195.
4. Buckland, S. T. (1987). One the Variable Circular Plot Method of Estimating Animal Density, *Biometrics.* 43, 363-384
5. Buckland, S. T., Anderson, D. R., Burnham, K.P. and Laake, J. L. (1992). Distance Sampling Estimating of Abundance of Biological Population, Chapman & Hall, London.
6. Burnham, K.P. (1979). A Parametric Generalization of the Hayne Estimator for Line Transect Sampling, *Biometrics.* 35, 587-595.
7. Burnham, K.P. and Anderson, D. R (1976). Mathematical Models foe Nonparametric Inferences from Line Transect Data, *Biometrics.* 32, 325-336.
8. Burnham, K.P., Anderson, D. R. and Laake, J. L. (1980). Estimation of Density from Line Transect Sampling of Biological Populations, *Wildlife Monograph.* 72, Supplement to the Journal of Wildlife Management.
9. Burnham, K.P., Anderson, D. R. and Laake, J. L. (1981). Line Transect Estimation of Bird Population Density using Fourier Series, *Studies in Avian Biology.* 6, 466-482.
10. Crain B. R., Burnham, K.P., Anderson, D. R. and Laake, J. L. (1979). Nonparametric Estimation of Population Density for Line Transect Sampling using Fourier Series, *Biometrical Journal.* 21, 731-748.
11. Drummer, T. D. and McDonald, L. L. (1987). Size Biased in Line Transect Sampling, *Biometrics.* 43, 13-21.
12. Eberhardt, L. L. (1978). Transect Methods for Population Studies, *Journal of Wildlife Management.* 42, 1-31.
13. Gates, C. E. (1979). Line Transect and Related Issues. In R. M. Cormack, G. P. Patil and D. S. Robson (eds), Sampling Biological Population. Fairland, MD: International Co-operative Publishing House, 71-154.
14. Johnson, E. G. and Routledge, R. D. (1985). The Line Transect Method: A Nonparametric Estimator based on Shape Restrictions, *Biometrics.* 41, 669-679.
15. Otto, M. C. and Pollock, K. H. (1990). Size Biased in Line Transect Sampling: A Field Test. *Biometrics.* 46, 239-245.
16. Overton, W. S. (1969). Estimating the number of Animals in Wildlife Populations: In R. H. Giles (eds.) *Wildlife Management Techniques*, 3rd ed. Washington, DC: Wildlife Society, 403-455.
17. Pollock, K. H. (1978). A Family of Density Estimators for Line Transect Sampling, *Biometrics.* 34, 475-478.
18. Quang, P. X. (1990). Confidence Intervals for Densities in Line Transect Sampling, *Biometrics.* 46, 497-472.

19. Quang, P. X. (1991). A Nonparametric Approach to Size-biased Line Transect Model for Aerial Surveys, *Biometrics*. 47, 269-279.
20. Quang, P. X. and Lanctot, R. B. (1991). A Line Transect Sampling, *Biometrics*. 47, 1089-1102.
21. Quinn, T. J., II and Gallucci, V. F. (1980). Parametric Models for Line Transect Estimators of Abundance, *Ecology*. 61, 293-302.
22. Ramsey, F. L. (1979). Parametric Models for Line Transect Surveys, *Biometrika*. 66, 505-512.
23. Ramsey, F. L., Gates, C. E., Patil, G. P. and Taillie, C. (1988). On Transect Sampling to Assess Wildlife Populations and Marine Resources. In P. R. Krishnaiah and C. R. Rao (eds.), *Handbook of Statistics, Vol. 6 Sampling*. Amsterdam: Elsevier Sciences Publishers, 515-532.
24. Schweder, T. (1977). Point Process Models for Line Transect Experiments. In J. R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem (eds.), *Recent Development in Statistics*. Amsterdam: North-Holland, 221-242.
25. Seber, G. A. F. (1982). *The Estimation of the Animal Abundance*, 2nd ed. London, Griffin.
26. Seber, G. A. F. (1986). A Review of Estimating Animal Abundance. *Biometrics*, 42, 267-292.
27. Seber, G. A. F. (1992). A Review of Estimating Animal Abundance. *II International Statistical Review*, 60, 129-166.
28. Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.

obeikandi.com

الفصل السادس عشر

المعاينة بخط التقاطع

Line Intercept Sampling

1.16 مقدمة

إن المعاينة بطريقة خط التقاطع عبارة عن عينة من الخطوط يجري اختيارها في منطقة الدراسة، وعندما يقطع خط أو أكثر أحد عناصر المجتمع نقوم بقياس متغير أو أكثر في هذا العنصر. على سبيل المثال ممكن أن تكون الشجيرات المنتشرة في منطقة الدراسة هي العناصر التي نريد دراستها، فعندما يقطع خط أو أكثر إحدى هذه الشجيرات نقوم بقياس -على سبيل المثال- عدد الأغصان بهذه الشجيرة أو ارتفاعها أو عرضها أو عدد الثمرات الموجودة فيها ... إلخ.

إن الشجيرات الكبيرة في منطقة الدراسة يكون احتمال أن يقطعها خط أو أكثر أكبر من الشجيرات الصغيرة، لذا فإن الحصول على تقديرات غير متحيزة لمعالم المجتمع تعتمد على تحديد هذه الاحتمالات التي تعتمد على حجم الوحدات.

فيما يأتي بعض الرموز والمصطلحات التي سوف نستخدمها في هذا

الفصل

A : مساحة منطقة الدراسة.

a_i : مساحة العنصر i .

y_i : صفة ما نريد قياسها في العنصر i (إنتاج العنصر أو ارتفاعه أو عدد الأغصان).

N : عدد العناصر في منطقة الدراسة ($N=16$ في الشكل 1).

$$\tau_y = \sum_{i=1}^N y_i$$

: المجموع الكلي للمجتمع للمتغير y .

μ_y : الوسط الحسابي للمجتمع للمتغير y .

$$C = \sum_{i=1}^N a_i / A$$

: تغطية N من العناصر لمنطقة الدراسة.

B : طول الخط الأساس.

L : طول الخط القاطع.

V_i : طول التقاطع بين خط التقاطع والعنصر i كما هو موضح بالشكل 1.

W_i : تمثل ضل (عرض) العنصر i على خط الأساس كما هو موضح بالشكل 1

n : عدد العناصر التي قُطعت بخط التقاطع.

m : حجم العينة وهو يمثل عدد الخطوط التي تم اختيارها في منطقة الدراسة.

$D=N/A$ كثافة العناصر في منطقة الدراسة.

2.16 عينة عشوائية من خطوط ثابتة الاتجاه

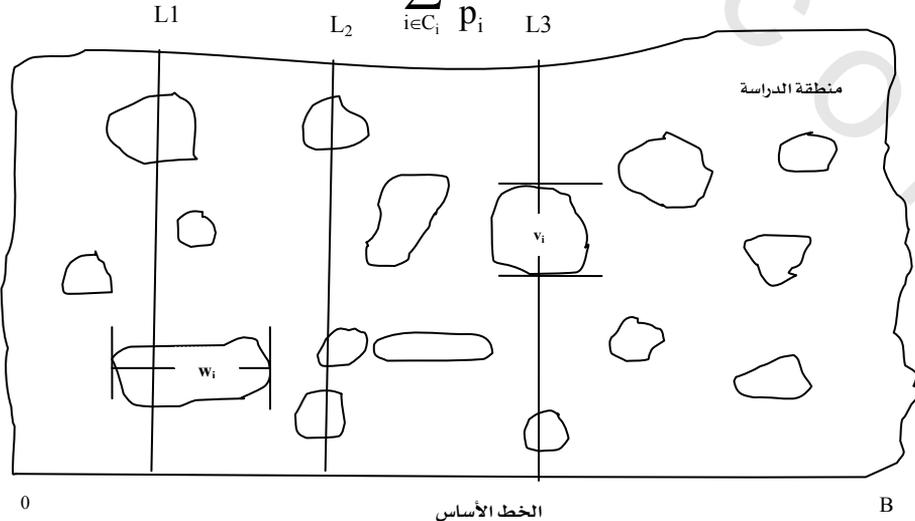
إن من أبسط التصاميم هو اختيار m من الخطوط بطريقة عشوائية وذلك باختيار m من المواضع بطريقة عشوائية على طول الخط الأساس لمنطقة الدراسة والذي يمثل عرض منطقة الدراسة. ومن ثم نقوم بمد الخطوط القاطعة التي تكون عمودية على الخط الأساس. الشكل 1 يوضح هذا التصميم. في الشكل 1 لدينا ثلاثة خطوط متساوية الطول.

لنفترض أن لدينا N من العناصر في منطقة الدراسة، مع كل عنصر لدينا متغير واحد على الأقل نرغب في دراسته وليكن y ، الهدف هو تقدير مجموع المجتمع الكلي $\tau = \sum_{i=1}^N y_i$ للمتغير y أو الكثافة لكل وحدة مساحة $D_y = \tau / A$. إن احتمال أن يقطع الخط k العنصر i يكون متناسباً مع عرض العنصر W_i كما هو موضح في الشكل 1؛ لذا فإن احتمال أن يقطع الخط k العنصر i هو

$$p_i = \frac{W_i}{B}$$

لنفترض أن C_i تمثل مجموعة العناصر التي قطعت من قبل الخط k في عينة حجمها m من الخطوط. لنعرف متغير جديداً هو

$$X_v = \sum_{i \in C_i} \frac{y_i}{P_i}$$



الشكل 1: عينة عشوائية من 3 خطوط ثابتة الاتجاه بطول L لكل خط في منطقة الدراسة

نلاحظ أن المتغير X_k عبارة عن تقدير غير متحيز إلى المجموع الكلي للمجتمع τ . وبما أن العينة العشوائية التي تتكون من m خط سوف تعطينا X_1, X_2, \dots, X_m - هذه التقديرات تتوزع بشكل مستقل ومشابه لبعضها الآخر - فإن وسطها الحسابي يمثل تقدير غير متحيز إلى τ

$$\hat{\tau} = \frac{1}{m} \sum_{k=1}^m X_k$$

وبما أن العناصر لم تُسحب بصورة مستقلة عن بعضها الآخر - حيث إن العنصر يمكن أن يُقطع من قبل أكثر من خط - فلا بد من تكرار الخطوط أو بعبارة أخرى سَحَبَ عينة عشوائية بحجم m من الخطوط لنستطيع أن نستخدمها للحصول على تقدير غير متحيز إلى τ باستخدام

$$s^2(\hat{\tau}) = \frac{S_x^2}{m}$$

حيث إن

$$s_x^2 = \frac{1}{m-1} \sum_{k=1}^m (X_k - \hat{\tau})^2$$

3.16 عينة من خطوط عشوائية الموضع والاتجاه

لنفترض أن كل خط جرى سحبه بطريقة عشوائية كاملة من منطقة الدراسة. يمكننا أن ننفذ هذه الطريقة وذلك بأن نسحب مجموعة من النقاط بعدد الخطوط التي ننوي سحبها بطريقة عشوائية من منطقة الدراسة، كل نقطة تمثل نقطة المنتصف لكل خط طوله L . ومن هنا نقوم بتحديد اتجاه

الخط باختيار قيمة الزاوية بطريقة مستقلة، ولتحديد قيمة الزاوية نقوم بسحب رقم بطريقة عشوائية من التوزيع المتماثل في الفترة $[0, \pi]$. وإذا حدث أن خرج قسم من الخط خارج منطقة الدراسة فعلينا أن نقوم بمد هذا القسم في منطقة أخرى داخل منطقة الدراسة. لمزيد من المعلومات راجع Kaiser (1983).

بمعرفة اتجاه الخط θ فإن احتمال أن يكون العنصر i قطع من قبل الخط هو

$$P_i(\theta) = L w_i(\theta) / A$$

حيث أن $w_i(\theta)$ عبارة عن عرض العنصر باتجاه عمودي على θ أي المسافة القصوى بين خطين باتجاه θ التي قطعت العنصر كما هو مبين بالشكل رقم 2.

أما الاحتمال غير المشروط لاختيار العنصر i فهو

$$p_i = LC_i / A$$

حيث إن

$$C_i = E[w_i(\theta)]$$

أي توقع $w_i(\theta)$ لتوزيع θ .

يمكننا أن نجد تقدير غير متحيز للمجموع الكلي للمجتمع τ باستخدام الاحتمال المشروط أو الاحتمال غير المشروط، يمكننا أن نعرف متغير جديد للخط k

$$x_k(\theta) = \sum_{i \in C_k} \frac{y_i}{P_i(\theta)}$$

أو

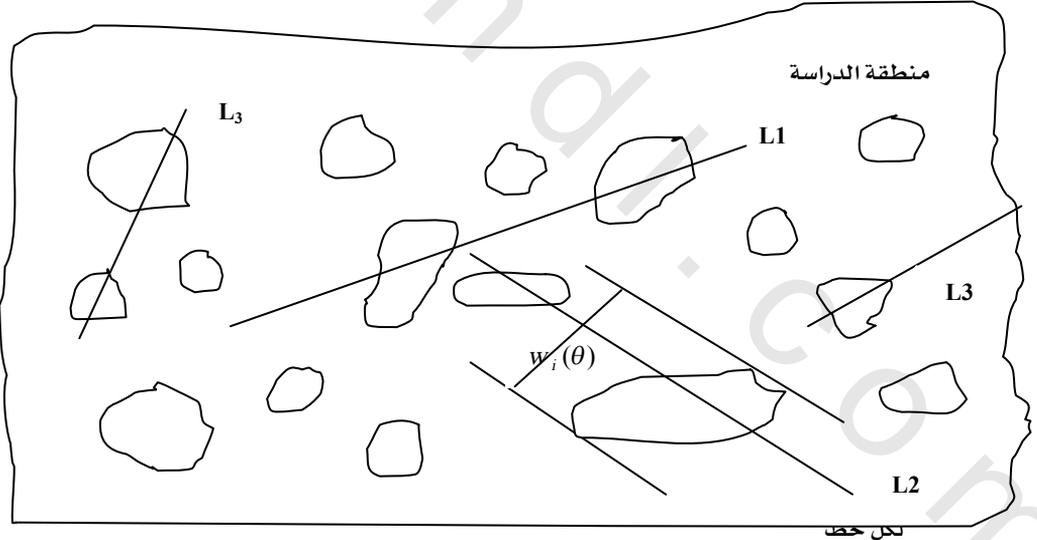
$$x_k = \sum_{i \in C_i} \frac{y_i}{p_i}$$

حيث أن كلاً من x_k و $x_k(\theta)$ تقديرين غير متحيزين إلى τ .
وإذا قمنا بسحب عينة بحجم m من الخطوط فسنحصل على تقديرين غير متحيزين إلى τ هما

$$\hat{\tau}_{p(\theta)} = \frac{1}{m} \sum_{k=1}^m x_k(\theta)$$

و

$$\hat{\tau}_p = \frac{1}{m} \sum_{k=1}^m x_k$$



الخط الأساس

الشكل 2: عينة عشوائية من 3 خطوط متغيرة الاتجاه والموضع بطول L

ويمكننا أن نحصل على تقدير غير متحيز لتقدير التباين لكلا التقدير باستخدام تباين العينة لكلٍ من X_k و $X_k(\theta)$ وهما على التوالي

$$s^2(\hat{\tau}_{p(\theta)}) = s_{x(\theta)}^2/m$$

حيث إن

$$s_{x(\theta)}^2 = \frac{1}{m-1} \sum_{k=1}^m (X_k(\theta) - \hat{\tau}_{p(\theta)})^2$$

و

$$s^2(\hat{\tau}_p) = s_x^2/m$$

حيث إن

$$s_x^2 = \frac{1}{m-1} \sum_{k=1}^m (X_k - \hat{\tau}_p)^2$$

ولمقارنة أي التقديرين $\hat{\tau}_p$ و $\hat{\tau}_{p(\theta)}$ أفضل يراجع (Kimura and Lemberg (1981).

ولابد من الإشارة هنا إلى أن استخدام الخطوط العشوائية (الموضع والاتجاه) يوفر لنا تقديرات غير متحيزة لتقدير المجموع الكلي (الكثافة والتغطية) ولكن عملية تنفيذها في حقل الدراسة قد تكون صعبة إن لم تكن مستحيلة إذا كان حقل الدراسة كبيراً جداً، لذا نوصي باستخدام طريقة الخطوط الثابتة الاتجاه التي تكون عمودية على خط الأساس، أو استخدام العينة المنتظمة من الخطوط العمودية على خط الأساس. لمزيد من

المعلومات يراجع (Lucsa and Seber (1977) و McDonald(1980).

4.16 تقدير التغطية والكثافة باستخدام خطوط التقاطع

إن من أهم استخدامات معاينة خطوط التقاطع هو تقدير التغطية والكثافة لعنصر أو أكثر في منطقة الدراسة. إن وحدة المعاينة هي خط التقاطع الذي يكون عمودياً على الخط الأساس ويقطع منطقة الدراسة، أي يكون طوله بعرض منطقة الدراسة. إذا كانت منطقة الدراسة مستطيلة فسيكون طول الخطوط متساوياً وبخلاف ذلك ستكون أطوال الخطوط غير متساوية. وبما أن العناصر في منطقة الدراسة لن تكون متساوية الحجم، لذا فإن احتمالات أن قطع هذه العناصر من قبل الخط القاطع غير متساوية، وستكون هذه الاحتمالات متناسبة مع حجم العناصر في منطقة الدراسة.

1.4.16 تقدير التغطية

لقد استُخدمت خطوط التقاطع لتقدير التغطية لعنصر أو عناصر معينة لمنطقة الدراسة أول مرة من قبل (Canfield 1941). من البدهي أن تقدر نسبة التغطية لعنصر معين في منطقة الدراسة بما يغطيه هذا العنصر لخط. إذا كان لدينا خط واحد بطول L فيمكننا أن نقدر التغطية باستخدام

$$\hat{c} = \sum_{i=1}^n v_i / L$$

يمكننا أن نثبت أن هذا التقدير هو تقدير غير متحيز إلى التغطية C ، يراجع (Lucas and Seber 1977). لتقدير التباين لا بد أن نضع افتراضات معينة على توزيع العناصر في المجتمع، وفي معظم الأحيان هذه الافتراضات غير واقعية، أو نقوم بسحب عينة عشوائية من الخطوط.

لنفترض أننا قمنا بسحب عينة بحجم m من الخطوط وبطول L نفسه. وقمنا بتقدير التغطية لكل خط وهي

$$\hat{c}_k = \sum_{i=1}^{n_k} v_i/L, \quad k = 1, 2, \dots, m$$

حيث إن n_k يمثل عدد العناصر التي تم قطعها من قبل الخط k . وبما أن العينة العشوائية التي تتكون من خط m سوف تعطينا تقديرات $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_m$ وهذه التقديرات تتوزع بشكل مستقل ومشابه لبعضها الآخر؛ لذا فإن وسطها الحسابي يمثل تقديراً غير متحيزاً إلى C وهو

$$\hat{c} = \frac{1}{m} \sum_{k=1}^m \hat{c}_k$$

يمكن أن نحصل على تقدير غير متحيز إلى تقدير تباين \hat{c} باستخدام

$$s^2(\hat{c}) = \frac{S_c^2}{m}$$

حيث إن

$$S_c^2 = \frac{1}{m-1} \sum_{k=1}^m (\hat{c}_k - \hat{c})^2$$

أما إذا كانت أطوال الخطوط غير متساوية فيجب أن نستخدم الوسط الحسابي المرجح لتقدير قيمة التغطية وهو

$$\hat{c} = \frac{\sum_{k=1}^m L_k \hat{c}_k}{\sum_{k=1}^m L_k}$$

2.4.16 تقدير الكثافة

لنفترض أن منطقة الدراسة عبارة عن مستطيل بمساحة $L \times B$ كما هو موضح بالشكل 1 أعلاه. وفي هذا المستطيل نقوم بمد خط التقاطع وذلك بعد

تحديد موضعه على خط الأساس باختيار نقطة عشوائية داخل الفترة (0, B) ومن ثم مد الخط ليقطع منطقة الدراسة ويكون عمودياً على الخط الأساس. إن احتمال أن يقطع الخط العنصر i يكون متناسباً مع عرض العنصر w_i وهو

$$p_i = \frac{w_i}{B}, \quad i = 1, 2, \dots, N$$

إن تقدير عدد العناصر في المجتمع يكون

$$\hat{N} = \sum_{i=1}^n 1/p_i = B \sum_{i=1}^n 1/w_i$$

وتقدير الكثافة للعناصر $D=N/A$ سيكون

$$\hat{D} = B \sum_{i=1}^n (1/w_i) / (BL) = (1/L) \sum_{i=1}^n 1/w_i$$

حيث إن n تمثل عدد العناصر التي قُطعت بخط التقاطع الذي طوله L .

وتقدير الكثافة $D_y = \tau_y / A$ للمتغير y يمكن حسابه باستخدام

$$\hat{D}_y = (1/L) \sum_{i=1}^n y_i / w_i$$

أما تقدير المجموع الكلي $\tau = \sum_{i=1}^N y_i$ للمتغير y الذي تم قياسه في العناصر

التي جرى قطعها من قبل خط القاطع فهو

$$\hat{\tau}_y = B \sum_{i=1}^n y_i / w_i$$

و تقدير الوسط الحسابي للمجتمع μ_y للمتغير y فيمكن الحصول عليه باستخدام

$$\hat{\mu}_y = \left(\sum_{i=1}^n y_i / w_i \right) \left(\sum_{i=1}^n 1/w_i \right)$$

للحصول على تقدير لتباين التقديرات أعلاه لابد من سحب عينة عشوائية بحجم m من الخطوط. لذا يمكننا الحصول على تقديرات مستقلة لكثافة العناصر $(\hat{D}_1, \hat{D}_2, \dots, \hat{D}_m)$ ولكثافة المتغير y ، $(\hat{D}_{y_1}, \hat{D}_{y_2}, \dots, \hat{D}_{y_m})$ في منطقة الدراسة. ومن ثم نجد الوسط الحسابي والتباين لهذه التقديرات لاستخدامها في التحليلات الإحصائية الأخرى مثل إيجاد فترة ثقة أو اختبار الفرضيات. هذا إذا كان أطوال الخطوط متساوية ولكن إذا كانت أطوال الخطوط غير متساوية فلا بد من استخدام الوسط الحسابي المرجح لتقدير كثافة العناصر أو تقدير الكثافة للمتغير y التي ستكون على النحو الآتي:

$$\hat{D} = \frac{\sum_{k=1}^m L_k \hat{D}_k}{\sum_{k=1}^m L_k}$$

و

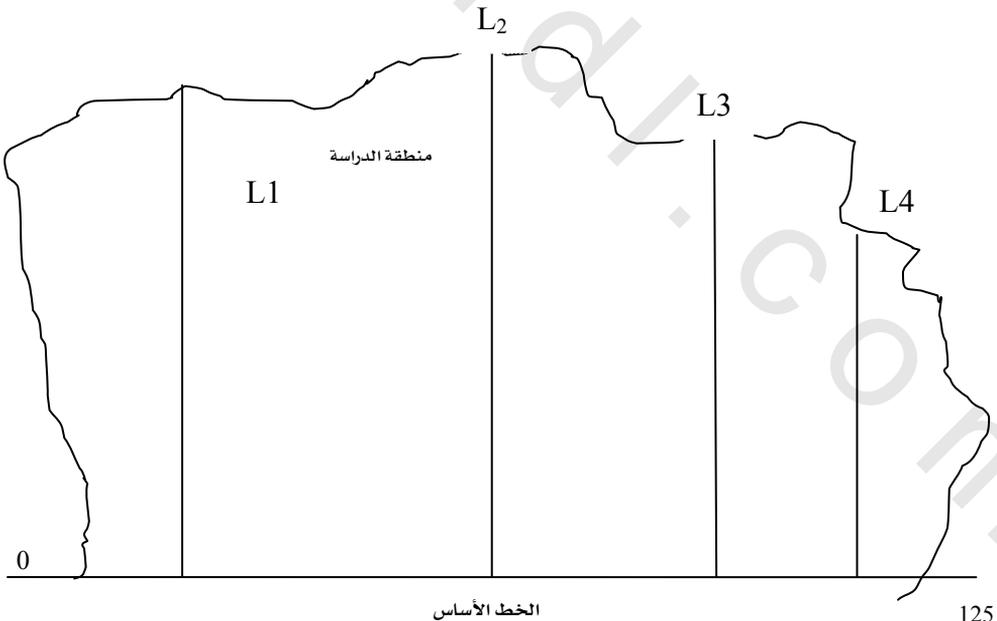
$$\hat{D}_y = \frac{\sum_{k=1}^m L_k \hat{D}_{y_k}}{\sum_{k=1}^m L_k}$$

إذا دققنا النظر في المعادلات الواردة في هذا الجزء سنجد أنها تعتمد على $w_i, i = 1, 2, \dots, n$ الموجود في مقام هذه المعادلات. وبصورة عامة إذا كان حقل الدراسة كبير فإن احتمال أن يقطع العنصر i يكون $p_i = \frac{w_i}{B}$ حتى ولو كررنا الخطوط، لذا فإن قيمة صغيرة جداً إلى w_i سيؤدي إلى قيمة احتمال قريبة من الصفر وهذا سيؤدي إلى أن تكون التقديرات غير مستقرة

ولا يمكن الاعتماد عليها. ولا بد في مثل هذه الحالات أن نحذف العناصر الصغيرة جداً التي تعطينا احتمالاً قريباً من الصفر من العينة، وعلينا أن نُعدّها عناصر شاذة.

3.4.16 مثال

قام بهذه الدراسة مجموعة من الطلبة الباحثين لتقدير الكثافة والتغطية لإحدى الشجيرات المنتشرة في منطقة الدراسة، حيث قاموا بمد خط الأساس بطول 125 متر في منطقة الدراسة كما هو موضح في الشكل 3 أدناه. ومن ثم قاموا بسحب أربعة أرقام عشوائية في الفترة (0.125) ومن موضع هذه النقاط قاموا بمد أربعة خطوط عمودية على الخط الأساس وقطعت منطقة الدراسة بأطوال مختلفة وهي على النحو الآتي $L_1 = 81$ و $L_2 = 106$ و $L_3 = 79$ و $L_4 = 50$ متر. الجدول 1 يعطينا v_i و w_i و y_i الارتفاع الأقصى للشجيرة لجميع الخطوط بالسنتيمترات.



الشكل 3: عينة عشوائية من 4 خطوط ثابتة الاتجاه بأطوال مختلفة

الجدول 1: يعطينا الجدول V_i طول التقاطع بين الخط والعنصر (الشجيرة) و W_i ظل الشجيرة (عرض) على خط الأساس و Y_i الارتفاع الأقصى للشجيرة للخطوط الأربعة.

V_{1i} : 268, 234, 193, 53, 50, 24, 13, 19, 43, 74, 72, 39, 40, 40 W_{1i} : 153, 87, 79, 78, 185, 145, 38, 52, 22, 38, 59, 20, 42, 102 Y_{1i} : 170, 83, 110, 104, 141, 65, 28, 42, 24, 42, 31, 34, 29, 66	الخط الأول L_1
V_{2i} : 52, 16, 95, 59, 112, 62, 44, 56, 86, 31, 15, 27, 53, 34, 38, 140, 142, 30 W_{2i} : 67, 31, 83, 195, 136, 145, 72, 115, 98, 129, 88, 26, 63, 112, 34, 21, 136, 95 Y_{2i} : 67, 26, 70, 120, 125, 119, 58, 81, 39, 114, 78, 83, 39, 87, 6, 41, 113, 68	الخط الثاني L_2
V_{3i} : 190, 263, 60, 80, 40, 132, 94, 50, 150, 39, 118 W_{3i} : 115, 87, 57, 97, 57, 197, 58, 254, 185, 35, 124 Y_{3i} : 120, 80, 40, 65, 50, 135, 93, 120, 95, 33, 96	الخط الثالث L_3
V_{4i} : 130, 175, 159, 182, 47, 4 W_{4i} : 96, 208, 68, 139, 50, 72 Y_{4i} : 93, 143, 79, 86, 58, 63	الخط الرابع L_4

استخدم البيانات لتقدير التغطية من قبل الشجيرات لهذه المنطقة، وكذلك قدر كثافة هذه الشجيرات في منطقة الدراسة، وأخيراً كثافة المتغير Y ، الذي يمثل الارتفاع الأقصى للشجيرات في منطقة الدراسة. كذلك أوجد تقدير التباين والخطأ المعياري للتقديرات الثلاث على افتراض أن أطوال الخطوط متساوية.

الحل: لقد حُسبت التغطية لكل خط من الخطوط باستخدام العلاقة أعلاه،
فلحساب التغطية للخط الأول على سبيل المثال استخدمنا

$$\hat{c}_1 = \sum_{i=1}^{14} v_i/L_1 = \frac{1}{81}(2.68+2.34+\dots+0.40) = 0.14346$$

الجدول 2 يعطينا التغطية لكل خط، أما التغطية لمنطقة الدراسة فقد
استخدمنا الوسط الحسابي المرجح لحساب التغطية وهو

$$\hat{c} = \sum_{k=1}^4 L_k \hat{c}_k / \sum_{k=1}^4 L_k = [0.14346(81) + \dots + 0.1394(50)] / (81 + \dots + 50) = 0.1319$$

أما فيما يخص حساب تقدير التباين والخطأ المعياري فقد افترضنا أن
أطوال الخطوط متساوية واستخدمنا

$$s_c^2 = \frac{1}{m-1} \sum_{k=1}^m (\hat{c}_k - \bar{\hat{c}})^2 = \frac{1}{4-1} [(0.1435 - 0.1350)^2 + \dots + (0.1394 - 0.1350)^2] = 0.00049284$$

حيث إن

$$\bar{\hat{c}} = \frac{1}{m} \sum_{k=1}^m \hat{c}_k$$

يمثل الوسط الحسابي غير المرجح، لذا فإن تقدير التباين والخطأ المعياري
يكونا على التوالي:

$$s^2(\hat{c}) = \frac{s_c^2}{m} = \frac{0.00049294}{4} = 0.00012312$$

و

$$s(\hat{c}) = \sqrt{\frac{s_c^2}{m}} = \sqrt{0.00012312} = 0.0111$$

الآن نقوم بحساب الكثافة لكل خط من الخطوط الأربعة، فعلى سبيل المثال سنحسب الكثافة للخط الأول

$$\hat{D}_1 = 1/L_1 \sum_{i=1}^{14} 1/w_{i1} = [(1/1.53) + (1/0.87) + \dots + (1/1.02)] / (81) = 0.33789$$

الجدول 2 يعطينا الكثافة لكل خط، أما الكثافة لمنطقة الدراسة فقد استخدمنا الوسط الحسابي المرجح لحسابها

$$\hat{D} = \sum_{k=1}^4 L_k \hat{D}_k / \sum_{k=1}^4 L_k = [81(0.337889) + \dots + 50(0.11425)] / (316) = 0.2384$$

ولحساب التباين والخطأ المعياري نفترض أن أطوال الخطوط متساوية ونستخدم العلاقة

$$s_D^2 = \frac{1}{m-1} \sum_{k=1}^m (\hat{D}_k - \bar{\hat{D}})^2 = \frac{1}{4-1} [(0.1435 - 0.2235)^2 + \dots + (0.1394 - 0.2235)^2] = 0.01010$$

حيث إن

$$\bar{\hat{D}} = \frac{1}{m} \sum_{k=1}^m \hat{D}_k$$

يمثل الوسط الحسابي غير المرجح. لذا فإن تقدير التباين والخطأ المعياري يكونا على التوالي:

$$s^2(\hat{D}) = \frac{s_D^2}{m} = \frac{0.01010}{4} = 0.002525$$

$$s(\hat{D}) = \sqrt{s^2(\hat{D})} = \sqrt{0.002525} = 0.0503$$

و أخيراً نقوم بحساب الكثافة للحد الأقصى لارتفاع الشجيرات للخطوط الأربعة، فعلى سبيل المثال سنحسبها للخط الأول

$$\hat{D}_{y_1} = 1/L_1 \sum_{i=1}^{14} y_{i1} / w_{i1} = [(1.7/1.53) + (0.83/0.87) + \dots + (0.66/1.02)] / (81) = 0.16425$$

الجدول 2 يعطينا الكثافة للحد الأقصى لارتفاع الشجيرات لكل خط، أما الكثافة لمنطقة الدراسة فقد استخدمنا الوسط الحسابي المرجح لحسابها وهو:

$$\hat{D}_y = \sum_{k=1}^4 L_k \hat{D}_{yk} / \sum_{k=1}^4 L_k = [81(0.16425) + \dots + 50(0.0919)] / (316) = 0.1395$$

ولحساب التباين والخطأ المعياري نفترض أن أطوال الخطوط متساوية ونستخدم العلاقة الآتية

$$s_{Dy}^2 = \frac{1}{m-1} \sum_{k=1}^m (\hat{D}_{yk} - \bar{D}_y)^2 = \frac{1}{4-1} [(0.16425 - 0.1332)^2 + \dots + (0.091934 - 0.1332)^2]$$

$$= 0.001225$$

حيث إن

$$\bar{D}_y = \frac{1}{m} \sum_{k=1}^m \hat{D}_k$$

والذي يمثل الوسط الحسابي غير المرجح؛ لذا فإن تقدير التباين والخطأ المعياري يكونا على التوالي

$$s^2(\hat{D}_y) = \frac{s_{Dy}^2}{m} = \frac{0.001225}{4} = 0.0003063$$

$$s(\hat{D}_y) = \sqrt{s^2(\hat{D}_y)} = \sqrt{0.0003063} = 0.0175$$

جدول 2: تقدير قيم التغطية والكثافة والكثافة للارتفاع الأقصى لكل خط، وتقدير التغطية والكثافة والكثافة للارتفاع الأقصى لمنطقة الدراسة باستخدام الوسط الحسابي المرجح والوسط الحسابي غير المرجح والخطأ المعياري للتقديرات المختلفة لمنطقة الدراسة.

تقدير الكثافة إلى: $y \hat{D}_y$	تقدير الكثافة: \hat{D}	تقدير التغطية: \hat{C}	
0.16425	0.33789	0.14346	الخط الأول L_1
0.16018	0.27235	0.10302	الخط الثاني L_2
0.11651	0.16947	0.15392	الخط الثالث L_3
0.09193	0.11425	0.1394	الخط الرابع L_4
0.1395	0.2384	0.1319	الوسط الحسابي المرجح
0.1332	0.2235	0.1350	الوسط الحسابي غير المرجح
0.0175	0.0503	0.0111	الخطأ المعياري

References

1. Butler, S. A. and McDonald, L. L. (1980). Unbiased Systematic Sampling Plans for the Line Intercept Method, *Journal of Range Management*, 36, 463-468
2. Canfield, R. H. (1941). Application of the Line Intercept Method in Sampling Range Vegetation, *Journal of Forestry*, 39, 388-394.
3. DeVries, P. G. (1979). Line Transect Sampling-statistical Theory, Applications and Suggestions for Extended use Ecological Inventory. In R. M. Cormack, G. P. Patil and D. S. Robson (eds), *Sampling Biological Population*, Vol. 5 Statistical Ecology. Fairland, MD: International Co-operative Publishing House, 1-70.
4. Eberhardt, L. L. (1978). Transect Methods for Population Studies. *Journal of Wildlife Management*. 42, 1-31.
5. Kaiser, L. (1983). Unbiased Estimation in Line-intercept Sampling. *Biometrics*, 39, 965-976.
6. Kimura, D. K. and Lemberg, N. A. (1981). Variability of Line Intercept Density Estimate: A Simulation Study of the Variance of Hydroacoustic Biomass Estimate. *Canadian Journal of Fisheries and Aquatic Sciences*, 38, 1141-1152.
7. Lucas, H. A. and Seber, G. F. (1997). Estimating Coverage and Particle Density using Line Intercept Sampling, *Biometrika*, 64, 618-622.
8. McDonald, L. L. (1980). Line-intercept Sampling for Attributes other than Coverage and Density. *Journal of Wildlife Management*, 44, 530-533.
9. Quzng, P. X. (1991). A Nonparametric Approach to Size-biased Line Transect Sampling, *Biometrics*, 39, 269-279.
10. Seber, G. A. F. (1979). Transect of random Length. In R. M. Cormack, G. P. Patil and D. S. Robson (eds), *Sampling Biological Population*, Vol. 5 Statistical Ecology. Fairland, MD: International Co-operative Publishing House, 182-192.
11. Thompson, S. K (2002). *Sampling*, 2nd Wiley, New York.

الفصل السابع عشر

المعاينة المكانية أو كريج انج

Spatial Sampling or Kriging

1.17 مقدمة

غالباً ما يكون مطلوباً في الدراسات الجيولوجية التنبؤ أو تقدير المخزون من المعادن أو النفط في موقع أو مكان معين، ويعتمد هذا التنبؤ على ما تمت مشاهدته في مواقع أو أماكن أخرى في المنطقة، وهذه المواقع غالباً ما تكون متباعدة عن بعضها بصورة غير منتظمة في المنطقة، وربما يكون مطلوباً التنبؤ أو تقدير المخزون من المعادن في المنطقة وليس فقط في موقع معين، فعند دراسة التلوث البيئي نقوم بقياس التلوث في موقع أو مواقع معينة، ومن ثم تستخدم هذه المواقع للتنبؤ أو تقدير مقدار التلوث البيئي في مواقع أخرى أو حتى منطقة أكبر. أخيراً يمكننا أن نستخدم المعاينة المكانية لتقدير أو التنبؤ بعدد الحيوانات أو النباتات الموجودة في منطقة الدراسة من خلال مشاهدة عينة من المواقع في هذه المنطقة.

لا بد لنا أن نذكر في القيمة Y_t من المعدن، أو النفط، أو التلوث، أو عدد الحيوانات في الموقع t بوصفها متغيراً عشوائياً، ونرغب في تقدير أو التنبؤ بقيمة متغير جديد Y_0 باستخدام القيم Y_1, Y_2, \dots, Y_n التي شوهدت في n من المواقع t_1, t_2, \dots, t_n . والمتغير Y_0 يمكن أن يكون قيمة المتغير في موقع جديد أو المجموع الكلي للمتغير في منطقة جغرافية أكبر. وبما أن Y_0 يُعدُّ متغيراً

عشوائياً وليس معلمة أو قيمة لصفة ثابتة للمجتمع، فإن الاستدلال الإحصائي يُعدُّ تنبؤاً وليس تقديراً بالرغم من أن التنبؤ يكون مكانياً وليس زمنياً.

إن مشكلة التنبؤ المكاني وحلها المسمى كريج انج (Kriging) يشبه إلى حد كبير طريقة التنبؤ أو التقدير باستخدام المعلومات الإضافية أو الخارجية كما مر معنا في الفصل السابع، ويمكن أن تكتب معادلات التنبؤ باستخدام التغاير (Covariance) أو باستخدام تباينات الفروق (Variogram)، علماً بأن التغاير يستخدم بصورة تقليدية مع السلاسل الزمنية، أما تباينات الفروق فتستخدم مع التنبؤ المكاني أو الإحصاء الجيولوجي (Geostatistics).

هناك مراجع كثيرة للتنبؤ المكاني، وكريج انج، والإحصاء الجيولوجي من أهمها: (Hohn (1988) و (Cressie (1986, 1989, 1991) و (Journel (1987, 1989) وأيضاً هناك مراجع في سلسلة العمليات المكانية من أهمها (Matern (1960, 1986) و (Diggle (1983) و (Cox and Isham (1991) و (Ripley (1980) و (1981).

2.17 التنبؤ المكاني

1.2.17 دالة التغاير المكانية (Spatial Covariance Function)

غالباً ما تكون قيمة المتغير الذي نرغب بدراسته غير مستقلة بين موقع وآخر في المسوحات البيئية والجيولوجية، بل -بالأحرى- تكون قيم المتغير في المواقع القريبة يعتمد بعضها على الآخر، ويمكننا أن نلخص العلاقة بين المتغيرين y_1 و y_2 المصاحبتين للموقعين t_1 و t_2 بالتغاير

$$\text{cov}(y_1, y_2) = E[y_1 - E(y_1)][y_2 - E(y_2)]$$

عندما يكون التغاير بين موقعين يعتمد على موضعهما فقط، وليس على مكانهما المحدد في منطقة الدراسة، هذه العلاقة يمكن أن نلخصها بدالة

التغاير $c(h)$ عندما يكون التغاير للمتغير y لموقعين متباعدين عن بعضهما بمقدار h بما يأتي:

$$c(h) = \text{cov}(y_{t+h}, y_t)$$

أما إذا كانت منطقة الدراسة ثنائية الأبعاد فالموقع t سيعطى الإحداثيات بعدين أي $t = (t_1, t_2)$. كذلك المسافة h بين المواقع ستكون $h = (h_1, h_2)$ ، لذا فإن دالة التغاير ستعتمد على الاتجاه والمسافة بين المواقع.

2.2.17 التنبؤ الخطي (كريج انج) Linear Prediction (Kriging)

لنفترض أنه تم مشاهدة عينة بحجم n من المواقع وقياسها، ونرغب في تقدير ما سنجده في موقع آخر، وللتسهيل سنكتب y_i لقيمة المتغير y الذي تمت مشاهدته في الموقع i في العينة n ، $i=1, 2, \dots, n$ ، وسنكتب c_{ij} بدلاً من $\text{cov}(y_i, y_j)$ للتغاير بين قيم y في الموقعين i و j . نلاحظ أن $c_{ii} = \text{var}(y_i)$. وباستخدام قيم المتغير y في المواقع t_1, t_2, \dots, t_n تجعلنا نرغب في التنبؤ بقيم المتغير y_0 في الموقع t_0 . وللتسهيل أيضاً سنفترض أن الأوساط الحسابية $E(y_i)$ $i=1, 2, \dots, n$ متساوية.

إن الهدف الأساس هو إيجاد \hat{y}_0 التي هي عبارة دالة لقيم المتغير y والتي شوهدت باستخدام عينة بحجم n ، والتي تكون تقديراً غير متحيز إلى y_0 ، أي

$$E(\hat{y}_0) = E(y_0)$$

و كذلك تقوم بتصغير متوسط مربع خطأ التنبؤ أي

$$E(y_0 - \hat{y}_0)^2$$

يكون أصغر ما يمكن. بصورة عامة إن أفضل تقدير هو التوقع المشروط إلى y_0 إذا علمنا قيم y_1, \dots, y_n ، ولتحديد قيمة $E(y_0 | y_1, \dots, y_n)$ لابد من معرفة التوزيع الاحتمالي المشترك إلى y_1, \dots, y_n وهذا سيكون صعباً.

بصورة عامة يمكن إيجاد دالة خطية لمشاهدات المتغير y تكون غير متحيزة، وتقوم بتصغير متوسط مربع خطأ التنبؤ، ويمكن كتابتها بالشكل الآتي

$$\hat{y}_0 = \sum_{i=1}^n a_i y_i$$

مشكلتنا الأساسية تكمن في كيفية إيجاد قيم a_1, a_2, \dots, a_n التي تقوم بتصغير متوسط مربع خطأ التنبؤ مع المحافظة على عدم التحيز. يمكننا أن نحصل على حل لهذه المشكلة باستخدام طريقة مضروب لكرانج (Lagrange multiplier method) والتي يمكن كتابتها بصيغة المصفوفات

$$f = G^{-1}h$$

حيث إن

$$G = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} & 1 \\ c_{21} & c_{22} & \cdots & c_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \text{ و } h = \begin{pmatrix} c_{10} \\ c_{20} \\ \vdots \\ c_{n0} \\ 1 \end{pmatrix} \text{ و } f = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ m \end{pmatrix}$$

علماً بأن الثابت m والذي تم الحصول عليه مع المعاملات a_i يمثل مضروب لكرانج الذي يستخدم لإيجاد متوسط مربع خطأ التنبؤ.

إن أفضل مُتنبئ خطي في الجيولوجيا يدعى مُتنبئ كريج انج (Kriging predictor)، ومتوسط مربع خطأ التنبؤ إلى \hat{y}_0 يدعى تباين كريج انج وهو

$$E(y_0 - \hat{y}_0)^2 = c_{00} - \sum_{i=1}^n a_i c_{i0} - m$$

إذا كان المجتمع يتبع التوزيع الطبيعي، أي أن التوزيع الاحتمالي المشترك إلى y_1, \dots, y_n يتبع التوزيع الطبيعي المتعدد، فإن $\hat{y}_0 = \sum_{i=1}^n a_i y_i$ هو أفضل مُتنبئ ولديه أصغر متوسط لمربع الخطأ بالمقارنة بأي مُتنبئ آخر ولذا فهو دالة للمشاهدات y_1, \dots, y_n .

إن الحصول على التنبؤ الأفضل يعتمد على معرفة دالة التغير، أي أن تكون معلومة. ولكن في الحياة العملية دالة التغير تكون غير معروفة ويجري تقديرها باستخدام البيانات التي تم الحصول عليها من المسح الحالي أو من المسوحات السابقة. ولكن إذا كانت السلسلة ثابتة ومتشابهة الخواص في جميع الاتجاهات، فإن التغير للمواقع التي تبعد عن بعضها بمسافة d يمكن تقديرها باستخدام التغير للعينة والتي تعتمد على n_d من الأزواج المتميزة للمواقع التي تبعد عن بعضها الآخر بمسافة مقدارها d تقريباً. لذا فإن تقدير التغير للعينة يكون

$$\hat{c}(d) = \frac{1}{n_d} \sum (y_{ti} - \bar{y})(y_{tj} - \bar{y})$$

والمجموع يكون لجميع الأزواج المتميزة من المشاهدات التي تبعد عن بعضها الآخر بمسافة d وعدد هذه الأزواج هو n_d ، ويمكننا أن نجد منحىً توفيقياً ممهداً أو مهذباً (smooth curve) باستخدام طريقة المربعات الصغرى غير الخطية (nonlinear least squares) لتقدير دالة التغير ومن ثم تقدير التغير لأي مسافة.

مثال 1: أجريت بعض المسوحات لصيد الروبيان قرب إحدى الجزر القريبة من شواطئ ولاية ألاسكا الأمريكية لتقدير دالة التباين المكاني المشترك، التي سوف تستخدم للتنبؤ بكمية الروبيان التي يمكن صيدها في مواقع جديدة. لقد جرى رسم بيانات الصيد بالموضع على خريطة لمنطقة الدراسة، وأيضاً سُجلت كمية الصيد بالباون والمسافة بالميل البحري. وقد صُمم قارب البحث ليقوم برمي شبكة الصيد في مواضع تبعد عن بعضها مسافة ميل بحري، ومن ثم جرى حساب التباير للعينة باستخدام أزواج من المشاهدات جمعت لفترات من المسافة. ومن ثم حسبنا المنحنى التوفيقي $a \exp(-bx)$ باستخدام طريقة المربعات الصغرى غير الخطية لتقديرات التباير؛ لذا فإن الدالة التوفيقيّة للتباين المشترك ستكون

$$c(x) = 5.1 e^{-0.49x}$$

لنفترض أنه تم رمي الشبكة مرة واحدة وصدنا $y_1 = 5.526$ ألف باوند من الروبيان، ومن ثم تم رمي الشبكة مرة ثانية على بعد 6 أميال بحرية وصدنا $y_2 = 1.417$ ألف باوند من الروبيان. ما هو تقديرك لكمية الروبيان y_0 الذي سنصيده إذا ألقينا الشبكة على بعد ميل بحري واحد من الرمية الأولى و5.4 أميال من الرمية الثانية.

الحل: التباين سيكون $c(0)=5.1$ ، والتباير لرميتين يبعدان عن بعضهما 6 أميال بحرية $c_{12} = 5.1 e^{-0.49(6)} = 0.3$ ، التباير للمواضع الجديدة: $c_{10} = 5.1 e^{-0.49(1)} = 3.1$ ، و $c_{20} = 5.1 e^{-0.49(5.4)} = 0.4$. لذا فإن دالة التنبؤ الخطي

ستكون

$$\begin{pmatrix} a_1 \\ a_2 \\ m \end{pmatrix} = \begin{pmatrix} 5.1 & 0.3 & 1 \\ 0.3 & 5.1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 3.1 \\ 0.4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.22 \\ -0.95 \end{pmatrix}$$

لذا فإن $a_1 = 0.78$ ، و $a_2 = 0.22$ ، و $m = -0.95$. لذا فإن تنبؤنا للصيد في
الموضع الجديد سيكون

$$\hat{y}_0 = 0.78(5.526) + 0.22(1.1417) = 4.662$$

أو 4662 باوند مقدار الصيد في الموضع الجديد. أما تقديرنا لمتوسط مربع
خطأ التنبؤ فسيكون

$$\hat{E}(y_0 - \hat{y}_0)^2 = 5.1 - 0.78(3.1) - 0.22(0.4) + 0.95 = 3.5$$

وأما الخطأ المعياري لتقديرنا فسيكون $\sqrt{3.5} = 1.9$ أو 1900 باوند.

3.2.17 تباين الفروق (Variogram)

غالباً ما يستخدم تباين الفروق (variogram) لوصف التغيرات المكانية في
العلوم الجيولوجية بدلاً من التباين، ويمكن تعريف تباين الفرق لقيم المتغير y
في مواضع مختلفة كما يأتي:

$$\text{var}(y_{t+h} - y_t) = 2\gamma(h)$$

الدالة $\gamma(h)$ تسمى سيمافيريوجرام (Semivariogram). عندما تكون
العملية العشوائية مستقرة للدرجة الثانية فسيكون تباين الفروق والتباين
يحتوي على المعلومات نفسها ، وذلك لأن

$$\gamma(h) = c(0) - c(h)$$

حيث $c(0) = \text{var}(y_t)$ وهو تباين y في الموقع t . لقد
أشار (Cressie, 1986, 1989, 1991) إلى أن تباين الفروق يكون موجوداً لبعض
العمليات العشوائية حتى وإن لم تكن مستقرة للدرجة الثانية، لذا فإنها أكثر
عمومية من التباين.

لنفرض أن العملية كذلك مستقرة في الوسط الحسابي أي

$$E(y_t) = E(y_s)$$

لأي موقعين t و s في منطقة الدراسة. إذن

$$\text{var}(y_{t+h} - y_t) = E(y_{t+h} - y_t)^2$$

وبطريقة بسيطة يمكن أن نقدر $\gamma(h)$ باستخدام

$$2\hat{\gamma}(d) = \frac{1}{n_d} \sum (y_{t_i} - y_{t_j})^2$$

والمجموع يكون لجميع الأزواج المتميزة من المشاهدات التي تبعد عن بعضها بمسافة d وعدد هذه الأزواج هو n_d . للبيانات المتباعدة بشكل غير منتظم، يمكننا أن نجمع معاً الأزواج من المواقع التي تبعد عن بعضها تقريباً بنفس المسافة. وأخيراً يمكننا أن نجد منحني توفيقياً ممهداً لتقدير تباينات الفروق لجميع المسافات. ولقد أشار Cressie(1991) إلى ضرورة تقدير واستخدام γ بدلاً من c دالة التغاير؛ لأن $\hat{\gamma}(d)$ تقدير غير متحيز إلى $\gamma(d)$ ، بينما $\hat{c}(d)$ تقدير متحيز إلى $c(d)$.

يمكننا أن نكتب معادلات التنبؤ بدلالة السيمافيريوجرام. لنفترض أن المشاهدات تم قياسها في n من المواقع، ونرغب أن نتنبأ ماذا سنجد في موقع آخر. لتسهيل الرموز، سوف نستخدم الرمز y_i للدلالة على قيم المتغير y التي شوهدت في الموقع i في العينة، و $i=1,2,\dots,n$ ، كذلك سنكتب λ_{ij} لترمز لقيمة السيمافيريوجرام $\gamma(y_i - y_j)$ للفرق بين الموقعين i و j . نرغب بالتنبؤ لقيمة المتغير y_0 في الموقع t_0 ، باستخدام قيم المتغير y في المواقع t_1, t_2, \dots, t_n .

إن الهدف هو إيجاد الدالة \hat{y}_0 لمشاهدات المتغير y التي تكون تقدير غير

متحيز إلى y_0 أي

$$E(\hat{y}_0) = E(y_0)$$

وتقوم بتصغير متوسط مربع خطأ التنبؤ

$$E(y_0 - \hat{y}_0)^2$$

ويمكننا كتابة المقدّر الخطي كما يأتي:

$$\hat{y}_0 = \sum_{i=1}^n a_i y_i$$

مشكلتنا كيف يمكننا أن نجد قيم a_1, \dots, a_n والتي تقوم بتصغير متوسط مربع خطأ التنبؤ مع المحافظة على عدم التحيز. الحل يمكن كتابته بصيغة المصفوفات

$$a = \Gamma^{-1} \gamma$$

حيث إنّ

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \text{ و } \gamma = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix} \text{ و } a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ m^* \end{pmatrix}$$

ومتوسط مربع خطأ التنبؤ إلى \hat{y}_0 هو

$$E(y_0 - \hat{y}_0)^2 = \sum_{i=1}^n a_i \gamma_{i0} + m^*$$

ملاحظة: $m^* = -m$ حيث إنّ m يمثل الثابت الذي وجدناه باستخدام دالة التباين.

مثال 2: استخدم البيانات في المثال الأول لتقدير كمية الروبيان المتوقع صيده في الموقع الجديد باستخدام طريقة السيمافيريوجرام.

الحل: يمكننا إيجاد دالة السيمافيريوجرام باستخدام دالة التغير التي وجدناه في المثال 1 وهي

$$\gamma(x) = 5.1(1 - e^{-0.49x})$$

لذا فإن قيمة السيمافيريوجرام لرميتين يبعدان عن بعضهما 6 أميال بحرية هي: $\gamma_{12} = 5.1(1 - e^{-0.49(6)}) = 4.8$. وعليه فإن قيمة السيمافيريوجرام للموقعين الجديدين ستكون: $\gamma_{10} = 5.1(1 - e^{-0.49(1)}) = 2.0$ و $\gamma_{20} = 4.7$. لذا فإن دالة التنبؤ الخطي باستخدام السيمافيريوجرام ستكون

$$\begin{pmatrix} a_1 \\ a_2 \\ m^* \end{pmatrix} = \begin{pmatrix} 0 & 4.8 & 1 \\ 4.8 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2.0 \\ 4.7 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.22 \\ 0.95 \end{pmatrix}$$

لذا فإن $a_1 = 0.78$ ، و $a_2 = 0.22$ ، و $m^* = 0.95$ ؛ لذا فإن تنبؤنا للصيد في الموضع الجديد سيكون

$$\hat{y}_0 = 0.78(5.526) + 0.22(1.1417) = 4.662$$

أو باوند مقدار الصيد في الموضع الجديد، أما تقديرنا لمتوسط مربع خطأ التنبؤ فسيكون

$$\hat{E}(y_0 - \hat{y}_0)^2 = 0.78(2.0) - 0.22(4.7) + 0.95 = 3.5$$

وأما الخطأ المعياري لتقديرنا فسيكون $\sqrt{3.5} = 1.9$ أو 1900 باوند.

للتنبؤ بقيمة المتغير y لمجتمع أو منطقة الدراسة وليس لموقع أو لمواقع

محددة يراجع (Thompson 2002).

3.17 التصميم المكاني

يعتمد متوسط مربع خطأ التنبؤ للمتنبئ الخطي غير المتحيز \hat{y}_0 لقيمة y_0 سواء كانت y_0 تمثل قيمة المتغير في موضع معين أو المتوسط لمنطقة الدراسة، على قيم التغير بين y_0 وقيم العينة لكل موقع، كذلك على قيم التغير بين المواقع المختلفة. لذا يمكننا أن نستخدم المعلومات المتوافرة من دالة التغير أو تباينات الفروق لنقرر ما الذي نختاره من المواقع في العينة ليعطينا أفضل تنبؤ، بعبارة أخرى ما هو أفضل تصميم للمعاينة يمكن استخدامه للتنبؤ عن قيمة المتغير y_0 .

لنفترض أن c_{ij} ترمز إلى دالة التغير $\text{COV}(y_i, y_j)$ بين قيم y في الموقعين i و j ، ولنفترض أن c_{i0} تمثل التغير بين y_0 و y_i . بما أن $E(\hat{y}_0) = E(y_0)$ ، لذا

$$\text{فإن متوسط مربع خطأ التنبؤ إلى } \hat{y}_0 = \sum_{i=1}^n a_i y_i \text{ سيكون}$$

$$E(y_0 - \hat{y}_0)^2 = \text{var}(y_0 - \hat{y}_0) = \text{var}(y_0) + \text{var}(\hat{y}_0) - 2 \text{cov}(y_0, \hat{y}_0)$$

لذا

$$E(y_0 - \hat{y}_0)^2 = c_0 + \sum_{j=1}^n \sum_{i=1}^n a_i a_j c_{ij} - 2 \sum_{i=1}^n a_i c_{i0}$$

إذا دققنا النظر بالمعادلة الأخيرة يتضح لدينا أن أفضل التنبؤات التي سنحصل عليها من عينة بحجم n من المواقع يكون عندما تكون قيمة التغير قليلة بين المواقع وكبيرة بين المواقع و y_0 . لنفترض أن دالة التغير متماثلة بشكل دائري وتتناقص مع المسافة أو أن دالة تباين الفروق متماثلة وتتزايد مع المسافة. ويمكننا الحصول على قيمة كبيرة إلى c_{i0} وذلك عندما تكون المواقع المختارة بالعينة قريبة من الموضع 0 الذي نرغب في التنبؤ بقيمته. أما إذا كنا نرغب في التنبؤ عن قيمة الوسط الحسابي لمنطقة الدراسة فيفضل أن

تكون المواقع المختارة في العينة قريبة من مركز منطقة الدراسة. وللحصول على قيمة صغيرة إلى C_{ij} نقوم بتوزيع المواقع المختارة في العينة بصورة منتظمة على منطقة الدراسة أو بتقسيم منطقة الدراسة إلى عدد كبير من الطبقات الصغيرة.

للتنبؤ عن قيم y في مواضع منفردة في منطقة الدراسة، يمكننا أن نصغر قيمة أسوأ أخطاء التنبؤ باختيار المواقع التي عددها n لتصغير أقصى مسافة لأي نقطة في منطقة الدراسة من المواقع القريبة في العينة. لقد قام كل من $McBratney et al.(1981a,b)$ و $Y fanits et al.(1987)$ بدراسة هذه المشكلة، وتم المقارنة بين شبكات مختلفة من نقاط العينة المنتظمة على شكل مربع، ومستطيل، ومثلث، ومسدس. ولقد تم التوصل إلى أن الشبكة المثلثة والمتساوية الأضلاع تؤدي إلى تصغير أبعد مسافة من نقاط العينة إلى النقاط الأخرى في منطقة الدراسة التي ليست موجودة في العينة.

قام $Matern(1960,1986)$ بدراسة فاعلية مختلف تصاميم المعاينة للتنبؤ بقيمة الوسط الحسابي لمنطقة الدراسة. ولقد تم افتراض أن دالة التغيرات متماثلة دائرياً ومتناقصة مع المسافة. لقد وجد للعينات المنتظمة أن الشبكة المثلثة أكثر فاعلية من الشبكة المربعة، بينما الشبكة المستطيلة وُجد بأنها أقل فاعلية من الشبكتين المثلثة والمربعة.

كذلك قام $Matern$ بدراسة أفضل الطرق لتقسيم منطقة الدراسة إلى طبقات، وتوصل إلى أن أفضل التصميم أو التصميم الأكثر فاعلية هو الذي يقوم بتقسيم منطقة الدراسة إلى طبقات صغيرة ومضغوطة الشكل، ولقد تم التوصل إلى أفضل شكل، أو الشكل الأكثر فاعلية للطبقة تكون على الترتيب الآتي: دائرة، ومسدس، ومربع، ومثلث متساوي الأضلاع. ولا بد من الإشارة إلى أن معظم مناطق الدراسة لا يمكن تقسيمها إلى طبقات على

شكل دوائر. ولقد وجد أن الأشكال الثلاثة الأولى قريبة من بعضها من حيث الفاعلية. أما المثلث فهو أقل فاعلية من الأشكال الثلاثة الأولى. أما الطبقات المستطيلات فهي أقل فاعلية من المربعات، وتقل فاعليتها كلما ابتعدنا كثيراً من الشكل المربع.

إن أهم فوائد العينة المنتظمة تكمن في إمكانية نشر وحدات العينة بشكل متباعد إذا كانت الوحدات القريبة من بعضها مترابطة طردياً. ولكن العينة المنتظمة التي تكون وحداتها على شكل صف قد تكون غير فاعلة إذا كانت الوحدات المسحوبة بشكل منتظم (صف) تتوافق مع دورات متكررة في المجتمع، ويمكننا أن نسحب وحدات العينة المنتظمة بحيث لا تكون على شكل صف راجع (Bellhouse(1988) كما يأتي: نقوم بتقسيم منطقة الدراسة المستطيلة الشكل إلى $m_1 m_2$ من المربعات، يكون عدد الصفوف m_1 والأعمدة m_2 . يكون طول ضلع المربع k . ومن ثم نقوم بسحب m_1 من الأعداد العشوائية من التوزيع المنتظم في الفترة $(0, k)$ ، والتي ستكون عبارة عن الإحداثيات للمحور الأفقي. ونقوم بعد ذلك بسحب m_2 من الأعداد العشوائية من التوزيع المنتظم في الفترة $(0, k)$ ، وهذه ستكون عبارة عن الإحداثيات للمحور العمودي، لذا فإن نقاط العينة للصف الأول والعمود الأول قد تم اختيارها، أما بقية نقاط العينة فيجري تحديدها بإضافة مضاعفات العدد الثابت k لكل نقطة من إحداثيات المحور الأفقي ولكل نقطة من إحداثيات المحور العمودي.

لقد أثبت (Bellhouse(1981) أن العينة الطبقيّة التي تحتوي على وحدة واحدة في كل طبقة أكثر فاعلية من العينة المنتظمة التي جرى سحبها بالطريقة أعلاه. لقد قام (Olea(1982,1984a,b) بدراسة استخدام تصاميم العينة الطبقيّة والمنتظمة واستخداماتهما في المعاينة الجيولوجية، كما قام

Barnes(1988) بالنظر إلى تحديد حجم العينة لتقدير القيم المتطرفة بدلاً من القيم المتوسطة في منطقة الدراسة، وقام Gilbert(1987) بدراسة وتلخيص بعض سمات المعاينة المكانية لمراقبة تلوث البيئة، وأخيراً قام McArthur(1987) بمقارنة مجموعة من تصاميم المعاينة والمقدّرات لتقدير متوسط تركيز التلوث في نقطة المصدر في منطقة الدراسة. ويُعدُّ Muller(2001) أحدث مرجع للتصاميم المكانية.

References

1. Barnes, R. (1988). Bounding the Required Sample Size for Geologic Site Characterization, *Mathematical Geology*, 20, 447-490.
2. Bellhouse, D. R. (1981). Area Estimation by Point-Counting Techniques, *Biometrics*, 37, 303-312.
3. Bellhouse, D. R. (1988). Systematic Sampling, In P. R. Krishnaiah and C. R. Rao (eds.), *Handbook of Statistics*, Vol. 6, Sampling, Amsterdam, Elsevier Science Publisher, 125-145.
4. Cox, D. R. and Isham, V. (1980). *Point Processes*, Chapman & Hall, London.
5. Cerssie, N. (1986). Kriging Nonstationary Data, *Journal of the American Statistical Association*, 81, 625-634.
6. Cerssie, N. (1989). Geostatistics. *The American Statistician*, 43, 197-202.
7. Cerssie, N. (1991). *Statistics for Spatial Data*, Wiley, New York.
8. Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*, Academic Press, New York.
9. Gilbert, R. O. (1987). *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York.
10. Hohn, M. E. (1988). *Geostatistics and Petroleum Geology*, Van Nostrand Reinhold, New York.
11. Isaaki, E. H. and Srivaslava, R. M. (1989). *Introduction to Applied Geostatistics*, Oxford University Press, Oxford.
12. Journel, A. G. (1987). *Geostatistics for Environmental Sciences: An Introduction*, Project CR 811893, Las Vegas, NV: U. S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory.
13. Journel, A. G. (1988). Non-Parametric Geostatistics for Risk and Additional Sampling Assessment. In L. Keith (ed.), *Principles of Environmental Sampling*, Washington, DC: American Chemical Society, 45-72.
14. Journel, A. G. (1989). *Short Course in Geology*, Vol. 8, *Fundamentals of Geostatistics in Five Lessons*. Washington, DC: American Geophysical Union, 40 pp.
15. Matern, B. (1960). *Spatial Variation*. *Meddelanden fran Statens Skogsforskningsinstitut*, 45 (5).
16. Matern, B. (1986). *Spatial Variation*, 2nd ed., Springer-Verlag, Berlin.
17. McArthur, R. D. (1987). An Evaluation of Sample Design for Estimating a Locally Concentrated Pollutant, *Communications in Statistics – Simulation and Computation*, 16, 735-759.
18. McBratney, R. W., Webster, R. and Burgess, T. M. (1981a). The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables I, *Computer and Geostatistics*, 7, 331-334.
19. McBratney, R. W., Webster, R. and Burgess, T. M. (1981b). The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables II, *Computer and Geostatistics*, 7, 335-336.
20. Muller, W. G. (2001). *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*, (2nd ed.), Physica-Verlag.

21. Olea, R. A. (1982). Optimization of the High Plains Aquifer Observation Network, Groundwater Series, No. 7 Lawrence, KS: Kansas Geological Survey, University Kansas, 73 pp.
22. Olea, R. A. (1984a). Sampling Design Optimization for Spatial Functions, *Mathematical Geology*, 16, 369-392.
23. Olea, R. A. (1982b). Systematic Sampling of Spatial Functions, Series on Spatial Analysis, No. 7 Lawrence, KS: Kansas Geological Survey, University Kansas, 57 pp.
24. Ripley, B. D. (1981). Spatial Statistics, Wiley, New York.
25. Thompson, S. K. (2002). Sampling, 2nd Wiley, New York.
26. Yfantis, E. A., Flatman, G. T. and Behar, J. V. (1987). Efficiency of Kriging Estimation for Square, Triangular and Hexagonal Grids, *Mathematical Geology*, 19, 183-205