Curso 3

Versión no interactiva

Introducción al muestreo

La versión interactiva de este curso está disponible gratuitamente en la siguiente dirección
https://elearning.fao.org/

En este curso

Acerca del curso

Este curso abarca los aspectos generales del muestreo en los inventarios forestales, y pretende introducir los conceptos básicos y las características de un estudio de muestreo, así como proporcionar una visión general de los componentes más importantes de un inventario forestal nacional (IFN).

Descargo de responsabilidad: Este curso no tiene por objeto formar adecuadamente a expertos en las estadísticas de muestreo necesarias para planificar, analizar, informar e interpretar correctamente las estimaciones basadas en muestras de un IFN.

¿A quién va dirigido este curso?

El curso está dirigido principalmente a quienes participan en las fases de muestreo y análisis de un IFN, pero puede realizarlo cualquier persona interesada en el tema. Específicamente, este curso está dirigido a:

1. Técnicos forestales responsables de la ejecución de los IFN de su país.
2. Equipos de monitoreo forestal nacional.
3. Estudiantes e investigadores, como parte del material curricular en escuelas de silvicultura y cursos académicos.
4. Jóvenes y nuevas generaciones de silvicultores.

Estructura del curso

Este curso consta de tres lecciones.

Lección 1: Acerca del muestreo	Esta lección presenta los conceptos y términos básicos asociados al muestreo estadístico. Ofrece una visión general de las características relevantes de un estudio de muestreo y explica los fundamentos del muestreo para un público no experto.
Lección 2: Elementos de diseño de un estudio de muestreo	Esta lección presenta los fundamentos de los elementos de diseño de los estudios de muestreo, ya que son pertinentes para los IFN, y los conceptos que se deben considerar al preparar una estrategia de muestreo. También

	explica cómo calcular el tamaño de la muestra asociada.
Lección 3: Diseño de estimaciones	En esta lección se examinan los métodos y fórmulas necesarios para obtener estimaciones no sesgadas a partir de los datos recogidos siguiendo una determinada estrategia de muestreo.

Acerca de la serie

Este curso concluye la serie de ocho cursos a su propio ritmo que cubren diversos aspectos de un IFN. Aquí puede ver la serie completa:

Curso	Aprenderá sobre el curso
Curso 1: ¿Por qué un inventario forestal nacional (IFN)?	Objetivos y propósito de un IFN, y cómo los IFN contribuyen a la formulación de las políticas y a la toma de decisiones en el sector forestal.
Curso 2: Preparación de un inventario forestal nacional (IFN)	La planificación y el trabajo necesarios para establecer un IFN eficiente o un Sistema nacional de monitoreo forestal (SNMF).
☞ **Curso 3: Introducción al muestreo**	(Este es el curso que está estudiando actualmente).
Curso 4: Introducción al trabajo de campo	Consideraciones para el trabajo de campo, variables a nivel de parcela y mediciones a nivel de árbol.
Curso 5: Gestión de datos en un inventario forestal nacional	Recopilación de información y gestión de datos para los IFN.
Curso 6: Garantía de calidad y control de calidad en un inventario forestal nacional	Procedimientos de GC y CC en la recopilación y gestión de datos de inventarios forestales.
Curso 7: Elementos del análisis de datos	Enfoques/cálculos típicos en los análisis de datos y temas relacionados.

	Curso 8: Resultados de los inventarios forestales nacionales Presentación de informes y difusión	Presentación de informes de los IFN y la importancia de la presentación de informes en el contexto de las acciones de REDD+.

Lección 1: Acerca del muestreo

Introducción de la lección

Esta lección presenta los conceptos y términos básicos asociados al muestreo estadístico. Además, ofrece una visión general de las características relevantes de un estudio de muestreo y explica los fundamentos del muestreo para un público no experto.

Objetivos

Al final de esta lección, usted podrá:

1. Describir la importancia del muestreo en los inventarios forestales
2. Definir los fundamentos del muestreo estadístico.
3. Explicar los conceptos básicos y la terminología asociada al muestreo.
4. Explicar la importancia de la exactitud y la precisión durante el proceso de estimación.

¿Por qué es necesario el muestreo?

Antes de que empecemos a conocer algunos de los aspectos relevantes del muestreo estadístico, retrocedamos un poco y reflexionemos brevemente sobre los fundamentos de los estudios de muestreo en general.

¿Por qué el muestreo es un concepto tan fundamental en el contexto de los inventarios forestales y del monitoreo forestal?

La respuesta a esta pregunta es muy sencilla: Cuando examinamos la evaluación de campo de las variables básicas, no es **factible ni eficaz observar todos los elementos** en la superficie forestal de un país. En lugar de ello, los expertos deben hacer inferencias sobre el estado actual y el cambio de las

variables objetivo mediante observaciones en subconjuntos relativamente pequeños o "muestras" de la superficie forestal total, denominadas parcelas de muestreo.

Podemos imaginar que el muestreo es similar a abrir pequeñas ventanas que nos permiten observar partes de la población para hacernos una idea del conjunto.

Un examen más detallado de estas muestras de la superficie total revela que la credibilidad de los resultados de un estudio de muestreo se ve influida por la forma en que **seleccionamos estas muestras, los métodos que utilizamos para obtener las observaciones únicas y las técnicas de estimación y los cálculos aplicados**. Estos son también los tres elementos de diseño de un estudio de muestreo, que deben planificarse teniendo en cuenta consideraciones estadísticas, algo que analizaremos con más detalle en la próxima lección.

¿Qué es el muestreo estadístico?

El proceso de selección garantiza que los elementos del muestreo puedan considerarse representativos de la población. Cuando el muestreo sigue las reglas de la estadística, lo llamamos **muestreo estadístico**. El muestreo estadístico viene determinado en gran medida por la aleatorización (y por la ausencia de consideraciones subjetivas o arbitrarias), lo que significa que, al aplicar una selección aleatoria, garantizamos que cada elemento de la población tenga una probabilidad definida y conocida de ser seleccionado.

Otros criterios de selección, como la **imparcialidad** o la **objetividad**, no son suficientes. Dado que las probabilidades de selección desempeñan un papel fundamental en el muestreo estadístico, estas técnicas también se denominan **muestreo probabilístico**. De este modo, se garantiza la representatividad de la muestra y se dispone de estimadores no sesgados (es decir, enfoques de estimación estadísticamente correctos) para la mayoría de los diseños de muestreo y observación comunes.

La selección subjetiva de los elementos "más representativos" de la población no es un muestreo estadístico y no permite realizar estimaciones ni inferencias estadísticas.

Imagínese que se envía a expertos con la tarea de encontrar la parcela "más representativa" en una superficie forestal (en cuanto a densidad de árboles, mezcla de especies, pendiente, condiciones del suelo, etc.). Resulta evidente que una estimación obtenida a partir de una parcela de este tipo se referiría exclusivamente a la elección del experto (mientras que otro experto probablemente elegiría una opción diferente).

Si bien una estimación basada en la opinión de un experto puede ser buena y aproximarse al valor de la población objetivo, todo depende del experto y no se ha definido un enfoque metodológico objetivo que pueda ser repetido por otra persona. El muestreo estadístico, por el contrario, es transparente en todos sus pasos metodológicos.

¿Sabía qué?

En el contexto de los inventarios forestales se inventaron y presentaron numerosas técnicas estadísticas de muestreo. Aunque las parcelas de muestreo ya se utilizaban ampliamente en silvicultura en el siglo XIX, en torno a 1900 se desarrolló -y se aceptó gradualmente- una técnica más formalizada de muestreo estadístico para grandes poblaciones como metodología para producir resultados válidos: en 1895, el estadístico noruego A.N. Kiaer presentó un enfoque de muestreo que entonces se denominó "**método representativo**", en el que la "**representatividad**" desempeñaba un papel central.

Los estadísticos de inventarios forestales de la época hicieron importantes contribuciones al análisis del muestreo sistemático por líneas. El primer IFN basado en el muestreo estadístico se implementó en Noruega entre 1919 y 1930. Le siguieron otros países nórdicos europeos a principios de los años veinte: Finlandia entre 1921-1924 y Suecia entre 1923-1929.

La congruencia estadística es una de las principales características del muestreo estadístico aplicado al monitoreo forestal. Sólo si se respetan los principios del muestreo estadístico podrá defenderse de forma convincente el diseño del inventario elegido cuando, por ejemplo, se planteen dudas sobre los resultados

Deducir inferencias a partir de una muestra

La estadística descriptiva se ocupa de la caracterización cuantitativa de una población de interés, o del dominio sobre el que se deberán producir tales afirmaciones descriptivas. El muestreo pretende deducir inferencias/conclusiones sobre la población total a partir de un número limitado de elementos de muestreo seleccionados. En los inventarios forestales, estos elementos suelen ser parcelas de muestreo, que son subconjuntos de la superficie total del bosque.

A partir del análisis de las observaciones recogidas sobre las variables objetivo de estas parcelas de muestreo, podemos obtener una estimación estadística del verdadero parámetro desconocido de la población. Por ejemplo: a partir de las biomasas por parcela de las n parcelas de muestreo podemos

elaborar una estimación de la biomasa por hectárea de toda la población. Intuitivamente está claro que no podemos esperar que esa estimación sea igual al valor verdadero: es una aproximación y variará cada vez que tomemos otra muestra siguiendo el mismo diseño de inventario.

Nota

Los valores verdaderos de una población se denominan **parámetros**, mientras que las estimaciones obtenidas a partir de estudios de muestreo se denominan **estadísticas**. El valor verdadero medio de una población, la media paramétrica, se estima a partir del valor medio de la muestra.

Es importante tener clara esta distinción**: los verdaderos parámetros nunca serán conocidos, sino estimados por la estadística de muestreo**. El valor verdadero es una constante, un valor fijo. La estadística de muestreo (= el valor estimado) es una variable aleatoria que puede tomar muchos valores distintos -dependiendo de la muestra que se haya seleccionado- y sigue una distribución determinada.

Veamos algunos ejemplos de cómo se podrían aplicar las definiciones anteriores a un inventario forestal para biomasa.

1. La población de árboles, por ejemplo, está determinada por una superficie, representada por un número infinito de centros de puntos sin dimensiones en los que podrían seleccionarse parcelas de muestreo.
2. La muestra consta de un número determinado de parcelas (tamaño de la muestra) que se ha seleccionado siguiendo el diseño de muestreo.
3. El valor verdadero -o parámetro poblacional- de, por ejemplo, la biomasa media, sería la biomasa media estimada sobre todas las posibles ubicaciones infinitas de muestreo en el área. Dado que sólo nos ocupamos del diseño actual de la parcela, el valor verdadero sigue siendo desconocido.
4. Utilizando un gráfico y un diseño de estimación adecuados, podemos obtener una estimación no sesgada a partir de la muestra que tenemos a mano.

Estimador y estimación

Cuando hablamos de un estimador en el muestreo estadístico, nos referimos al algoritmo o fórmula de cálculo que utilizamos para producir una estimación. Con el fin de producir estimaciones estadísticas, el estimador debe reflejar: el proceso de selección subyacente de los elementos de muestreo; y la forma en que se obtuvieron las observaciones individuales del elemento de muestreo.

¿Cuál es el concepto de población subyacente en los inventarios forestales*?*

Cuando las parcelas de muestreo son los "elementos de muestreo" que se seleccionan, la siguiente pregunta a la que llegamos es "¿cuál es entonces la población?

En términos generales, una población se define como el conjunto de todos los elementos de muestreo que teóricamente se pueden seleccionar. En el inventario forestal, se suelen utilizar parcelas de muestreo cuya ubicación se determina seleccionando un punto muestral. A continuación, la población queda definida por todos los puntos muestrales posibles dentro del área de interés. ¿Cuántos son?

El número de puntos en cualquier superficie es infinito. La población es entonces "el número total de parcelas de muestreo posibles en el área definida objeto de estudio", donde estas parcelas de muestreo se instalan alrededor de los puntos muestrales seleccionados.

Sin embargo, dado que muchas variables de interés son agregados de mediciones en árboles individuales que se encuentran en estas parcelas de muestreo, sólo variarán cuando cambie la composición de los árboles incluidos. Por lo tanto, para tales variables, podemos refinar el concepto de población y decir: "la población está compuesta por todos los grupos de árboles mutuamente excluyentes que tienen una probabilidad positiva de ser incluidos conjuntamente por el diseño de parcela definido". El tamaño de esta población no es infinito, sólo hay un número finito de opciones para las inclusiones conjuntas de árboles espacialmente dispersos.

Limitaciones para las conclusiones

De una muestra sólo podemos extraer conclusiones/inferencias sobre la parte de la población que tiene una probabilidad positiva de formar parte de la muestra. A esta parte de la población la denominamos **marco de muestreo**. En el mejor de los casos, el marco de muestreo comprende toda la población de interés, pero en realidad suele excluir algunas partes de la superficie forestal por diversos motivos, como la falta de accesibilidad o el riesgo de ingreso. Todas nuestras estimaciones se refieren

exclusivamente al conjunto de elementos de muestreo que están en el marco de muestreo y tenemos que asegurarnos de que el marco de muestreo cubra el máximo posible de toda la población.

Población vs. marco de muestreo

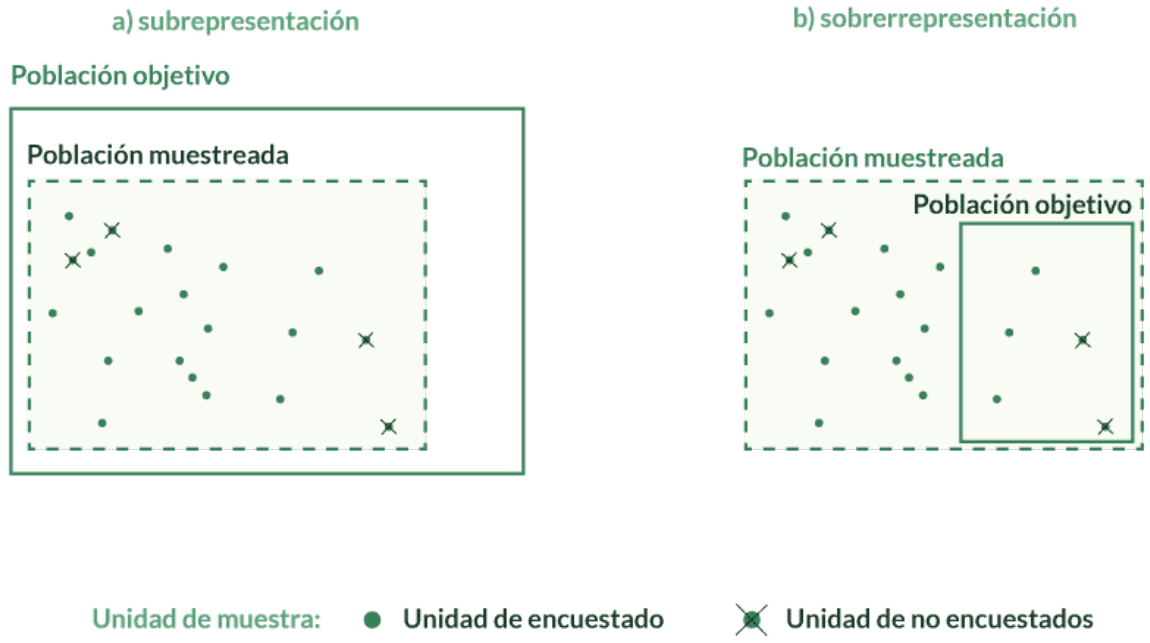
Imaginemos que un país no utiliza una definición biofísica de bosque -basada en criterios cuantitativos y cualitativos-, sino una definición administrativa o jurídica de "tierras forestales". Si la selección de los elementos de muestreo se limita a este marco de muestreo, no podemos extraer ninguna conclusión sobre los árboles y los bosques biofísicos que se dan fuera de las tierras forestales definidas. Todas las conclusiones se referirían exclusivamente a la superficie forestal situada dentro del terreno forestal delimitado.

Por consiguiente, tanto la población como el marco de muestreo deben definirse claramente y mencionarse durante la elaboración del informe y la interpretación de los resultados.

Además, puede haber algunos puntos en el marco de muestreo a los que no se pueda acceder por estar denegado el acceso o por motivos de seguridad. Estas observaciones omitidas se denominan **falta de respuesta.** La diferencia entre el **marco de muestreo** y la **falta de respuesta** es la siguiente: el marco de muestreo define los elementos de muestreo que supuestamente pueden seleccionarse y medirse. Pero puede ocurrir que algunos puntos muestrales resulten inaccesibles, que son los puntos de falta de respuesta, y existen diferentes técnicas para hacer frente a este problema. Por lo general, los índices de falta de respuesta son relativamente bajos en los IFN. Si bien existen técnicas de imputación para realizar predicciones basadas en modelos de las observaciones que podrían corresponder a dichas parcelas sin respuesta, en los IFN se suelen ignorar y se reduce el tamaño de la muestra.

El diagrama a continuación muestra que una población objetivo (área dentro de la línea continua) a menudo puede no coincidir con la población muestreada (área dentro de la línea de puntos). El ejemplo A es la subrepresentación, muy típica en los IFN, en los que determinadas áreas de la población se han clasificado previamente como, por ejemplo, no accesibles. El ejemplo B es la sobrerrepresentación, más rara en contextos del IFN, pero posible si la población objetivo de interés es una subpoblación particular del país, considerando que el muestreo se diseñó originalmente para todo el país.

En ambos casos, algunas unidades de muestreo eran accesibles (encuestados) y otras inaccesibles (no encuestados).



Conceptos básicos del muestreo

Aunque las estadísticas tienden a volverse complejas y a veces no son fáciles de digerir, mucho de lo que se expresa en fórmulas complejas es, de hecho, relativamente fácil de entender con algunas matemáticas básicas, y a menudo también es bastante intuitivo.

En la siguiente sección, nos centraremos en varios conceptos estadísticos y nos referiremos únicamente a aquellos que son relevantes para los IFN. Sin embargo, hay mucho más que aprender sobre los inventarios forestales que estos pocos conceptos.

Algunos conceptos y términos importantes

Cuando tomamos una muestra de una población (o del marco de muestreo) no hay un único resultado: cada selección de una nueva muestra alternativa arrojará una estimación diferente que es igual de válida que todas las demás.

Como no podemos determinar el único **valor verdadero** (llamado **parámetro**) de la población a partir de una muestra, la estimación que obtenemos siempre conlleva incertidumbre. Cuando tomamos una muestra de una población (o del marco de muestreo) no hay un único resultado: cada selección de una

nueva muestra alternativa arrojará una estimación diferente que es igual de válida que todas las demás.

De hecho, cuando determinamos este margen, también se trata de una estimación. Una medida típica de la incertidumbre es el **intervalo de confianza**, que define un intervalo en torno al valor estimado en el que esperamos el valor verdadero con una probabilidad definida.

¿Qué es el tamaño de la muestra?

El tamaño de la muestra se refiere al número de observaciones seleccionadas de forma independiente (elementos de muestreo observados) que se extraen del marco de muestreo. Aquí, el término "independiente" significa: la selección de un elemento no tiene ningún efecto sobre la selección de otro. Este proceso de selección ocurre si los elementos de la muestra se seleccionan aleatoriamente.

Sin embargo, en los inventarios forestales esto no suele ocurrir, ya que las muestras se recogen a intervalos fijos. Es importante señalar que la "selección independiente", tal como se describe aquí, no debe confundirse con "la independencia de las variables", que es un concepto completamente diferente.

En la Lección 2 se proporcionará más información sobre cómo determinar el tamaño del muestreo para diversos diseños de inventarios forestales.

¿Cuál es la diferencia entre intensidad de muestreo y tamaño de la muestra?

La intensidad de muestreo se refiere a la proporción del marco de muestreo que se observa. En cambio, el **tamaño de la muestra** se refiere al número absoluto de elementos de muestreo seleccionados (independientemente).

La intensidad de muestreo se define como la fracción de la población de elementos de muestreo que entran en la muestra. No obstante, como este concepto no es aplicable a la población infinita, definimos la intensidad de muestreo en los inventarios forestales mediante el área: es la fracción del área muestreada (= la suma de todas las áreas de las parcelas) del área total sobre la que se define la población.

¿Qué significa varianza poblacional?

La varianza poblacional **cuantifica la variabilidad de la población**. Es una característica de la población de los elementos de muestreo. Es decir, para cada elemento de población, hay un valor para una

variable objetivo específica, como la biomasa por hectárea. La varianza poblacional paramétrica es la verdadera varianza de todos estos valores. Y esta varianza verdadera (paramétrica) puede estimarse a partir de una muestra.

¿En qué se diferencia la varianza poblacional de la varianza del error?

Esta distinción es un elemento clave para comprender una gran parte de las estadísticas de muestreo relevantes para los IFN. Mientras que la varianza poblacional es una estimación de la variabilidad entre los elementos de la población (observaciones a partir de parcelas de inventario), la varianza del error es una propiedad de la muestra. Esto significa que cuantifica la variación esperada entre estimaciones repetidas de la misma variable objetivo (por ejemplo, la biomasa media por área).

Supongamos que un IFN se lleva a cabo repetidamente durante mil veces, cada vez con una nueva selección de parcelas. En este caso, la variación entre todas las medias individuales es una estimación de la varianza del error. Esta información es importante para juzgar la calidad de una muestra, ya que da respuesta a la pregunta "¿qué ocurriría si repitiéramos nuestra muestra una y otra vez? ¿obtendríamos siempre resultados bastante similares, o esperaríamos que los inventarios repetidos dieran lugar a resultados muy variables?

En el segundo caso, diríamos que nuestra estimación es menos precisa, y en el primero que nuestra estimación es precisa. Generalmente, no notificamos la **varianza del error**, sino su raíz cuadrada: el error estándar. Se trata de una de las estadísticas más relevantes estimadas a partir de una muestra, ya que cuantifica la precisión de la estimación. La razón por la que se notifica la raíz cuadrada y no la varianza del error es muy sencilla: el error estándar se presenta en las mismas unidades que la propia estimación y, por lo tanto, es mucho más fácil de entender.

Es intuitivamente evidente que la confianza y la credibilidad o certeza en los resultados dependen de esta varianza del error. Si no se facilita información sobre la varianza del error, el usuario de la información puede concluir que un único inventario por sí solo no es suficiente para utilizarlo como referencia (ya que el siguiente probablemente producirá una estimación diferente).

Exactitud y precisión

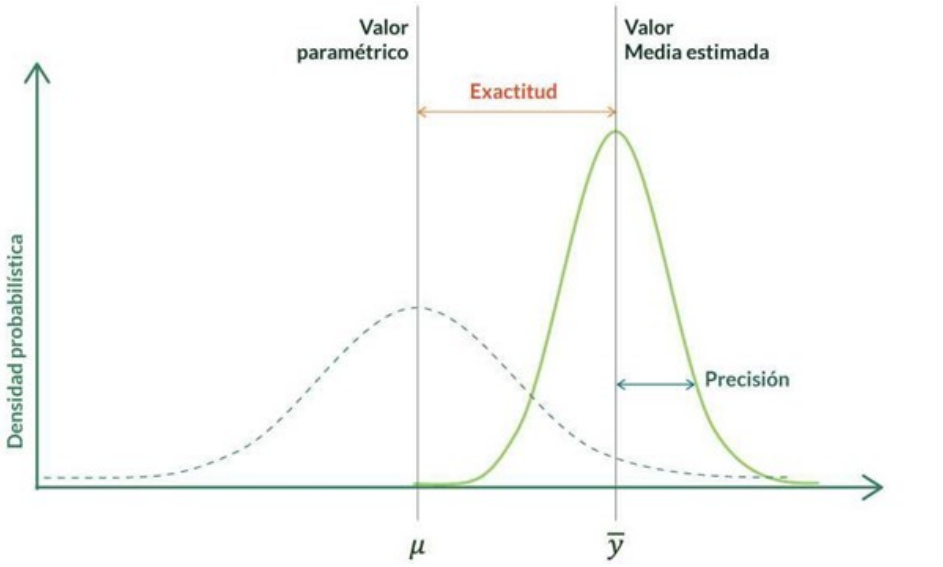
Ya nos hemos referido anteriormente al concepto de precisión, pero ahora vamos a destacar la pertinencia y el significado tanto de la exactitud como de la precisión tal y como las utilizamos en el

muestreo de inventarios forestales. Comprendamos mejor los conceptos con la ayuda de un ejemplo de un blanco de dardos.

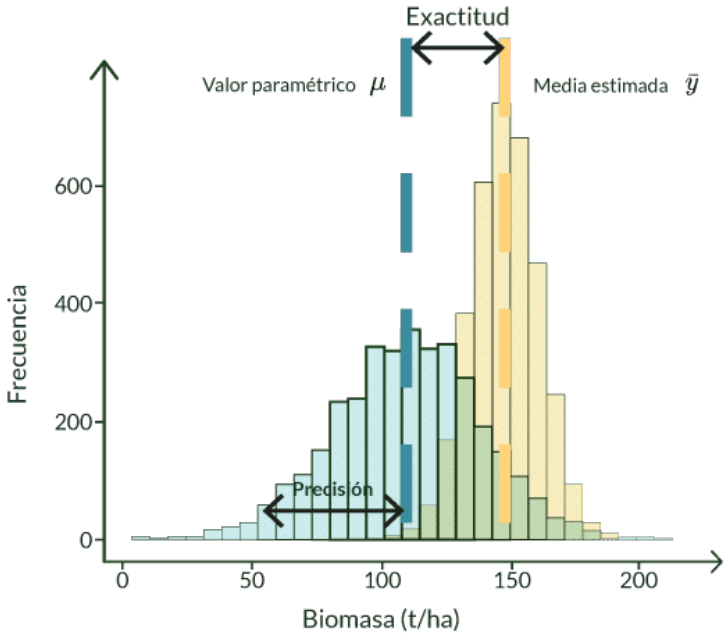
Baja precisión, alta exactitud	**Baja exactitud, alta precisión**
Imagine que lanza 4 dardos. La distribución de los aciertos respecto del centro es una expresión de su exactitud (el promedio de los aciertos dará como resultado una posición cercana al centro).	Siguiendo con el ejemplo del blanco, la dispersión de los aciertos es una expresión de su precisión (los lanzamientos repetidos caen muy próximos.
	

Como puede ver en el grafico a continuación, podemos representar gráficamente una distribución sobre todas las observaciones recopiladas (que aquí son valores individuales en el eje x). Los valores del eje y indican la frecuencia relativa de observaciones de los valores respectivos.

Aunque la distribución sólida conduce a una precisión relativamente alta (distribución estrecha) de la barra y media estimada, como en la figura, no es muy exacta (es decir, está sesgada) debido a su desviación del parámetro verdadero μ. Por el contrario, la distribución en línea discontinua da como resultado una estimación muy exacta de la media (aquí idéntica al valor paramétrico μ), pero una precisión relativamente baja.



Veamos ahora dos ejemplos en los que 3 500 parcelas de muestreo midieron la biomasa media por ha. El histograma amarillo representa una distribución con una precisión relativamente alta (distribución estrecha) de la media estimada: pero una baja exactitud (es decir, está sesgada) debido a su desviación del parámetro verdadero. Por el contrario, el histograma azul refleja una estimación muy exacta de la media (aquí idéntica al valor paramétrico μ), pero una precisión relativamente baja, debido a la mayor amplitud de la distribución.



Volviendo al grafico anterior, recuerde que la estadística de muestreo calculada a partir de una muestra no es más que una estimación de la biomasa real de la población, basada en el conjunto de parcelas seleccionadas. Si repetimos (imaginariamente) la estimación de la estadística de muestreo con una selección diferente bajo el mismo diseño, produciremos estimaciones diferentes de la biomasa poblacional. La distribución de estas medias estimadas representa la frecuencia relativa de estas diferentes estimaciones de la biomasa poblacional.

La amplitud de esta distribución es una expresión de la variabilidad (o dispersión) en torno al valor medio estimado (barra y). Si la dispersión de estos valores es baja y están relativamente próximos entre sí, podemos concluir que la repetición de muestras alternativas probablemente daría lugar a estimaciones similares. Por lo tanto, la amplitud de esta distribución también permite hacer una afirmación sobre la precisión (ver gráfico anterior).

Por otro lado, la **exactitud** -o corrección- es la desviación del valor esperado a partir de muestras repetidas con respecto al parámetro poblacional verdadero; esta desviación también se denomina **sesgo** o **sesgo del estimador**. Dado que el valor verdadero sigue siendo desconocido, la magnitud de esta desviación no puede cuantificarse a partir de la propia muestra. Es más bien una propiedad del estimador aplicado y la expresión de un error sistemático que no puede compensarse aumentando el tamaño de la muestra.

La única forma de garantizar estimaciones "no sesgadas" es una prueba matemática de que el diseño del muestreo y los métodos aplicados permiten realizar estimaciones correctas (diseño no sesgado) o simulaciones empíricas (en caso de que la estimación se base en la aplicación de modelos).

¡ Tenga en cuenta lo siguiente: En los estudios de muestreo no tenemos información sobre el valor verdadero de la población (objetivo), sólo disponemos de la muestra (dardos). Estamos ignorando la posición del centro, y la exactitud sólo puede garantizarse utilizando estimadores no sesgados.

¿Cuáles son las posibles razones de las estimaciones sesgadas? Vamos a averiguarlo.

Sesgo de selección	Se utilizó una selección no estadística y no se garantiza que la muestra sea representativa (por ejemplo, una selección subjetiva de parcelas cercanas a la carretera).
Sesgo del observado	Las observaciones o mediciones son sistemáticamente erróneas (por ejemplo, el DAP siempre se mide en 1 m de altura en lugar de 1,3 m).
Sesgo del estimador	Un cálculo sistemáticamente erróneo (por ejemplo, aplicando constantemente un factor de expansión de parcela erróneo, de forma que todas las observaciones de parcela sean demasiado altas).
Sesgo del modelo	En caso de técnicas de muestreo basadas en modelos o asistidas por modelos, pero también en caso de observación modelizada (por ejemplo, aplicación de modelos de biomasa erróneos), un posible sesgo del modelo afectará directamente al sesgo de la estimación.

Nota

El significado limitado de la intensidad de muestreo en relación con la precisión de la estimación

En las directrices del inventario o incluso en las normativas gubernamentales encontramos a veces umbrales para la intensidad de muestreo (proporción mínima de área) que debe muestrearse (por ejemplo, al menos el 3% del área de bosque). Sin embargo, esta intensidad de muestreo tiene muy poco significado para la precisión resultante de las estimaciones. La precisión depende del tamaño de la muestra. Examine detenidamente los estimadores presentados al final de esta lección y verá que la "intensidad de muestreo" no aparece en ninguna de las fórmulas.

Estimaciones puntuales y por intervalos

Por lo general, el valor estimado por sí solo no es información suficiente para una interpretación adecuada o para la elaboración de informes y la toma de decisiones. Recuerde que no hemos observado todo, sino que hemos obtenido una estimación a partir de una muestra. Si notificamos una media estimada (por ejemplo, el volumen medio o la biomasa por unidad de superficie), que denominamos **estimación puntual**, esta información por sí sola no permite emitir juicio alguno sobre la calidad (o fiabilidad, credibilidad o certeza) de dicha estimación.

También necesitaríamos información adicional sobre la precisión estimada de dichas estimaciones puntuales para poder informar sobre su calidad. Dicha información se da en términos de un intervalo alrededor de la media estimada, en el que esperaríamos el valor verdadero con una cierta probabilidad, y esto es lo que llamamos una **estimación de intervalo**.

Consejos prácticos

Notificación de estimaciones

Al Notificar estimaciones, es una buena práctica decir "*a partir de nuestro estudio de muestreo estimamos que las existencias en formación son de 200 m³/ha ± x" y no "a partir de nuestro estudio de muestreo concluimos que las existencias en formación son de 200 m³/ha.*

El hecho de que se trate exclusivamente de estimaciones también se desprende del hecho de que acompañamos nuestras estimaciones de valores medios (estimaciones puntuales) con estimaciones de la precisión de esta estimación (estimaciones de intervalo).

Las preguntas que inmediatamente pueden surgir aquí son:

- ¿A partir de cuántas observaciones independientes (parcelas) se estimó esta media?
- ¿Cuánto variaron estas observaciones individuales (= varianza poblacional)?
- ¿Cuál es la variación esperada de esta media si repitiéramos (virtualmente) la muestra muchas veces con el mismo diseño (= varianza del error)?

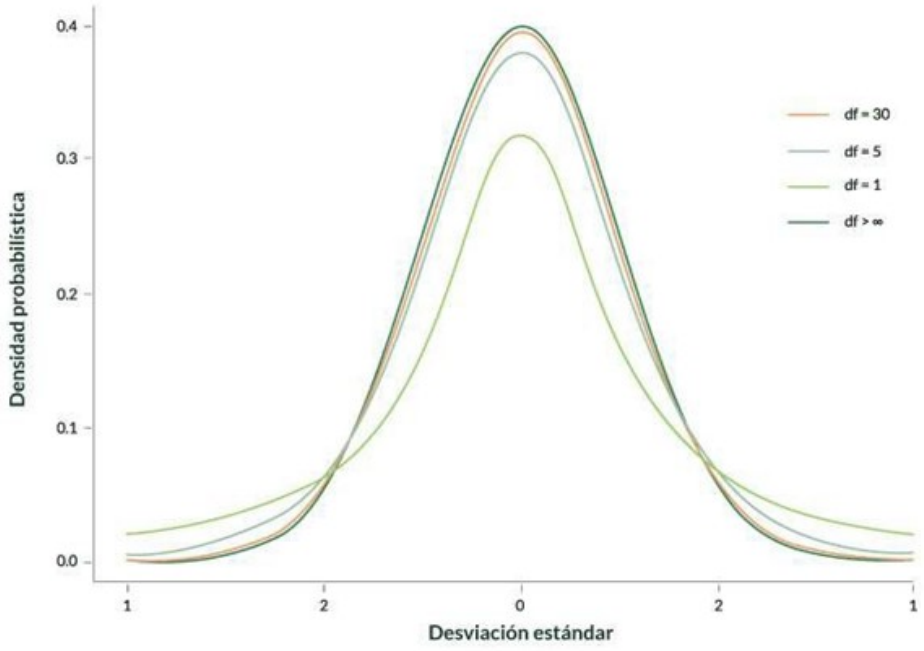
Todas las preguntas anteriores influyen en la amplitud del llamado **intervalo de confianza** en torno a una media estimada. Este intervalo de confianza es un enunciado probabilístico a partir del cual podemos saber en qué intervalo alrededor de la media estimada esperamos encontrar el parámetro poblacional verdadero (desconocido) con una probabilidad definida.

Sin embargo, esto sólo es posible si consideramos una determinada distribución de las estimaciones, y aquí es donde entra en juego una interesante propiedad de las muestras estadísticas.

Distribución de las muestras

Un carácter muy interesante de las muestras permite definir este intervalo: las estimaciones procedentes de muestreos repetidos tienden a seguir una distribución normal. Esto es válido para las muestras más grandes que superan un tamaño de muestra de 30, que en estadística de muestreo se toma como umbral aproximado que distingue las muestras pequeñas de las grandes; los valores medios estimados de las muestras más pequeñas siguen la distribución t de Student. Para muestras grandes, podemos utilizar la distribución normal para determinar los límites superior e inferior del intervalo en el que esperamos el valor verdadero con una probabilidad definida (por ejemplo, el 95 %).

Los gráficos a continuación representan una distribución normal y una distribución t de Student ligeramente diferente. Ambos permiten calcular un intervalo en el que esperamos el valor paramétrico verdadero con una probabilidad definida.



Intervalos de confianza

Como parte del proceso de estimación, nos gustaría evaluar el nivel de confianza que tenemos en nuestras estimaciones. Esto se refleja en lo cerca que estaría la estimación del parámetro verdadero, para cada muestra tomada. Si para todas las muestras posibles las estimaciones estuvieran muy próximas al parámetro poblacional verdadero, tendríamos una alta confianza en nuestras estimaciones. Para evaluarlo, solemos utilizar los intervalos de confianza.

Formalmente, podemos afirmar que la probabilidad P, de que el parámetro verdadero, μ, esté dentro de un límite inferior y un límite superior es x %. Cuanto mayor sea esa probabilidad, mayor será la confianza en nuestra estimación. Por ejemplo, en el caso concreto de la estimación de la media, nuestros intervalos de confianza estimados (expresados en las mismas unidades que la estimación de la media) definirán nuestros límites como:

$$\bar{y} - C.I. \leq \bar{y} \leq \bar{y} + C.I.$$

donde el intervalo de confianza I.C. se define por el valor de la distribución t de Student y el error estándar de la estimación:

$$C.I. = t\, S_{\bar{y}}$$

Por lo general, se indican intervalos de confianza del 95 %. El origen de estos intervalos de confianza del 95 % se remonta muy atrás en la historia de la estadística y no existe ningún argumento perfectamente convincente a favor de la probabilidad de error del 5 %. También podría utilizarse otro intervalo de confianza (como el 90 %), siempre que se indique claramente.

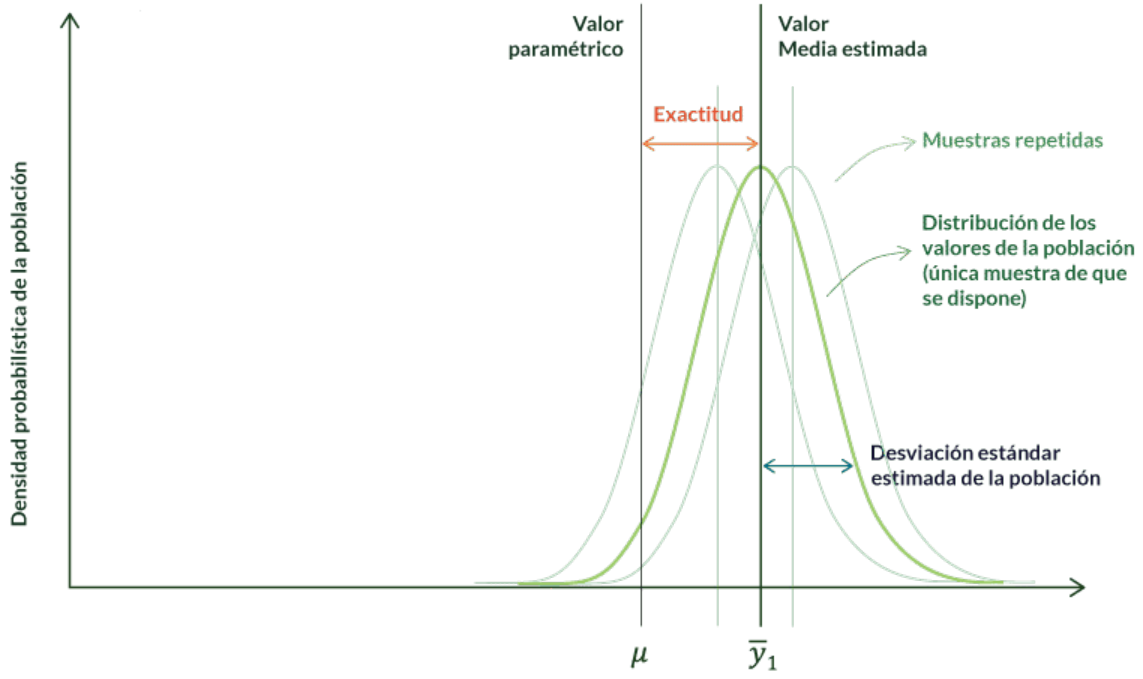
El error estándar de las estimaciones

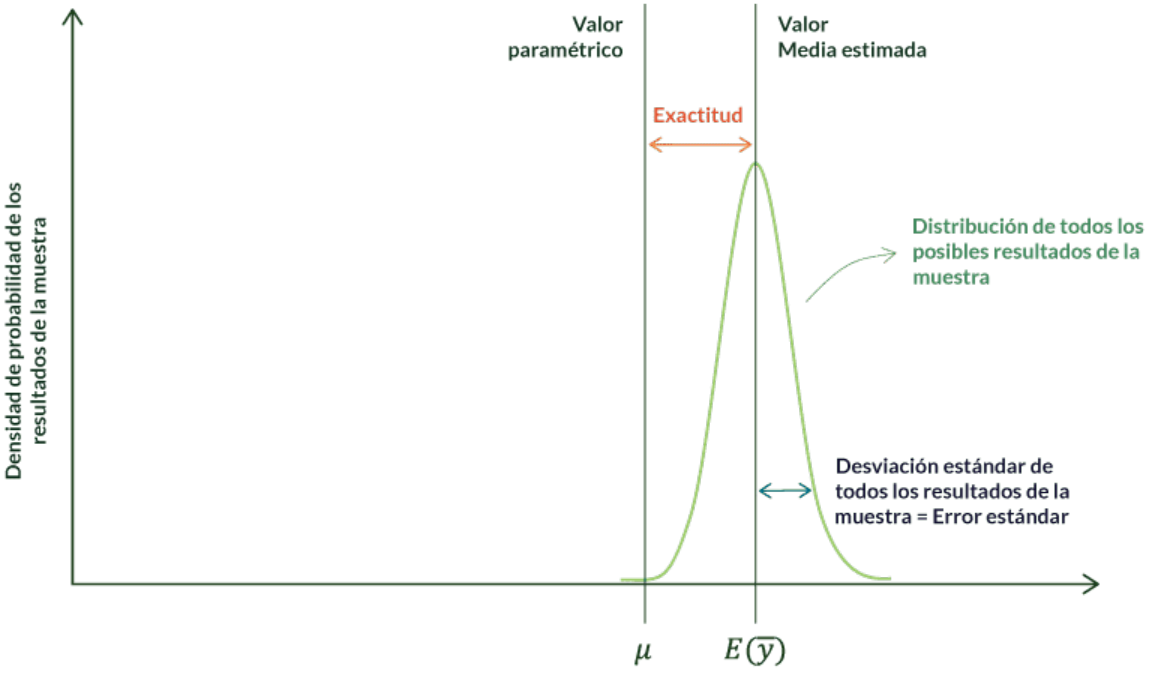
Si nos fijamos en la única muestra de que disponemos (el inventario que hemos realizado), ¿cómo podemos deducir una expectativa sobre la variación de todas las demás muestras posibles con el mismo diseño a partir de la misma población? En la práctica, por supuesto, no podemos repetir el IFN muchas veces.

Pues bien, hemos aprendido que podemos sacar conclusiones sobre la variabilidad de muestras repetidas (imaginarias) a partir de la muestra única que tenemos a disposición. La medida de esta variabilidad es la **varianza del error**.

El denominado **error estándar** es la raíz cuadrada de la varianza de este error. En otras palabras: es la desviación estándar estimada de todos los resultados posibles de la muestra. El error estándar es la medida de precisión de la estimación que se utiliza con mayor frecuencia. Al contrario de la varianza del error, viene con las mismas unidades que la estadística estimada. Por lo tanto, es más fácil de interpretar que la varianza del error.

La siguiente figura puede ayudar a desentrañar estas dos perspectivas diferentes. En el grafico superior podemos ver la distribución (variabilidad) de los elementos poblacionales (por ejemplo, los valores de las parcelas) a partir de una única muestra (línea en negrita). Sin embargo, esta única muestra que tenemos a disposición es sólo una de las muchas muestras posibles (verde claro) que podríamos extraer. En el grafico inferior se ve la distribución de todos los resultados potenciales de la muestra en torno al "valor esperado" y el error estándar es la desviación estándar de esta distribución.


Valor paramétrico
Valor Media estimada
Exactitud
Muestras repetidas
Distribución de los valores de la población (única muestra de que se dispone)
Desviación estándar estimada de la población
Densidad probabilística de la población
μ
$\overline{y}_1$


Valor paramétrico
Valor Media estimada
Exactitud
Distribución de todos los posibles resultados de la muestra
Desviación estándar de todos los resultados de la muestra = Error estándar
Densidad de probabilidad de los resultados de la muestra
μ
$E(\overline{y})$

Estimación mediante muestreo aleatorio simple (MAS)

Hemos llegado al último segmento de esta lección. En esta sección, veremos algunas explicaciones más detalladas que se tratarán en la próxima lección, y consideraremos algunos estimadores para el **muestreo aleatorio simple (MAS)**.

El muestreo aleatorio simple se refiere a una selección aleatoria independiente de cada elemento de muestreo. Significa que consideramos una selección aleatoria sin restricciones de los lugares de muestreo en una superficie forestal. El muestreo aleatorio sin restricciones significa que todos los elementos del muestreo tienen la misma probabilidad de selección. Este es el fundamento básico de la estadística de muestreo y muy apropiado para explicar los estimadores, porque es bastante sencillo determinar las probabilidades de selección, que aquí son iguales para todos los elementos.

Aunque rara vez se aplique en el inventario forestal, este diseño de muestreo (o procedimiento de selección) es fundamental para todas las estadísticas, porque los estimadores existentes son bastante sencillos y las características del muestreo estadístico pueden explicarse fácilmente, y es útil mencionarlo en aras de la exhaustividad.

En la siguiente tabla se muestra en el lado izquierdo la fórmula de cálculo del valor paramétrico (verdadero) de la población, que sigue siendo desconocido, y en el lado derecho el correspondiente valor estimado (basado en la muestra) de la población. Observe que el concepto que subyace a la varianza del error en la tabla ya se explicó anteriormente (Lección 1, Conceptos básicos del muestreo, Algunos conceptos y términos importantes, En qué se diferencia la varianza poblacional de la varianza del error).

Estadística	Cálculo paramétrico	Estimador basado en muestras
Media	$\mu = \frac{\sum_{i=1}^{N} y_i}{N}$	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$
Varianza	$\sigma^2 = \frac{\sum_{i=1}^{N} (y_i - \mu)^2}{N}$	$S_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$
Desviación estándar	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \mu)^2}{N}}$	$S_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}}$
Coeficiente de Variación (CV)	$CV = \frac{\sigma}{\mu}$	$CV = \frac{S}{\bar{y}}$
Error estándar	$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$	$S_{\bar{y}} = \frac{S_y}{\sqrt{n}}$
Varianza del error	Error estándar^2	

Los estimadores que se presentan aquí son los del MAS. En lecciones más adelante, aprenderá que se vuelven algo más complejos en cuanto consideramos otros diseños de muestreo.

Aquí, y también en las lecciones siguientes, consideramos el **muestreo sin reemplazo** y el **muestreo a partir de una población infinita** - y por lo tanto ignoramos la llamada **corrección de población finita (fpc)**. Para más detalles sobre la fpc, consulte la ***wiki de la Universidad de Gottingen*** (en inglés) o cualquier libro de texto sobre muestreo estadístico.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- Los expertos deben hacer inferencias y sacar conclusiones sobre el estado actual y el cambio de las variables objetivo mediante observaciones en subconjuntos relativamente pequeños o "muestras" de la superficie forestal total, denominadas parcelas de muestreo.
- Cuando el muestreo sigue las reglas de la estadística, lo llamamos muestreo estadístico. La congruencia estadística es una de las principales características del muestreo estadístico aplicado al monitoreo forestal.
- Un "estimador" en el muestreo estadístico se refiere a la fórmula de cálculo que utilizamos para producir una estimación.

Lección 2: Elementos de diseño de un estudio de muestreo

Introducción de la lección

En esta lección se examinan los métodos y fórmulas necesarios para obtener estimaciones no sesgadas a partir de los datos recogidos siguiendo una determinada estrategia de muestreo. También le muestra cómo calcular el tamaño de la muestra asociada.

Recuerde que esta lección no le convertirá en un experto en ninguna de las técnicas aquí descritas, sino que mejorará su comprensión de los conceptos generales. Al igual que otras lecciones de este curso, ésta es sólo una "introducción" para los estudiantes que no tengan una base sólida en estadística, requisito indispensable para un entendimiento exhaustivo del muestreo estadístico.

Objetivos

Al final de esta lección, usted podrá:

- Describir los tres elementos del diseño técnico de un estudio de muestreo
- Describir el diseño de muestreo.
- Identificar los tipos de diseños de muestreo.
- Explicar los fundamentos y el enfoque de la estratificación.
- Describir el diseño de la parcela/de observación.
- Resumir el concepto de corrección por pendiente.

Tres elementos de diseño de un estudio de muestreo

La planificación de cualquier estudio de muestreo puede desglosarse en tres elementos básicos de diseño técnico que proporcionan un marco para los proyectos de muestreo. Recuerde que, para preparar un estudio de muestreo, es necesario tener en cuenta los tres elementos de diseño en toda su extensión. Veamos qué significa cada una de ellas.

Diseño de muestreo	El diseño de muestreo responde a la pregunta "¿Cómo se seleccionan los elementos de muestreo?". En el monitoreo forestal, los puntos muestrales seleccionados dentro del área de inventario son teóricamente infinitamente pequeños, por lo que se consideran sin dimensión. Estos puntos definen la posición de la(s) parcela(s) de muestreo.. 
Diseño de observación	El diseño de observación, también conocido como diseño de parcela o diseño de respuesta, responde a la pregunta "¿Cómo se recogen las observaciones en cada elemento de muestreo?" El diseño de la observación viene definido por las reglas que guían la forma de incluir los árboles de muestra en la parcela de muestreo, con referencia al punto muestral sin dimensión. 
Diseño de la estimación	El diseño de la estimación responde a la pregunta "¿Cómo se calculan las estimaciones y qué estimadores estadísticos deben utilizarse?". Se trata del conjunto de estimadores o fórmulas que se utilizarán para el muestreo y el diseño de parcela determinados. En el muestreo y el diseño de parcelas, usted es libre de elegir los diseños "óptimos" o que mejor se adapten a sus objetivos. Sin embargo, no puede elegir libremente los estimadores. Esto se debe a que tienen que coincidir con los diseños de muestreo y parcela seleccionados. Por lo general, sólo existen algunos de estos estimadores. 

Recuerde que en esta lección nos centraremos en los diseños clásicos de muestreo y parcela en los IFN. Los diseños de estimaciones se tratarán en la siguiente lección y en la última lección de este curso.

Determinar el tamaño de la muestra

Uno de los aspectos definidos en el diseño de muestreo es el número de elementos de muestreo (parcelas) que deben observarse. Esto también se denomina tamaño de la muestra. Desde un punto de vista puramente estadístico, hay dos criterios principales que determinan el tamaño de la muestra necesario para una precisión objetivo definido para una situación de inventario determinada:

① La variabilidad de la población, es decir, la varianza poblacional. Esto se puede estimar a partir de un estudio piloto, o tomarse de inventarios/inventarios anteriores en áreas comparables. Nos referimos aquí a la población de parcelas de muestreo y las varianzas poblacionales serán diferentes para distintos diseños de parcelas de la misma superficie forestal.

② La precisión objetivo deseada, que es una cuestión de definición. Generalmente, la precisión se define como la mitad de la amplitud del intervalo de confianza objetivo.

¿Qué ocurre cuando no hay conocimientos previos ni información sobre inventarios anteriores?

A falta de datos de inventarios anteriores o de estimaciones de la variabilidad de la variable objetivo, un estudio piloto puede ayudar a obtener la información pertinente. Puesto que la varianza estimada se refiere siempre al diseño específico de parcela utilizado, un número relativamente pequeño de parcelas podría distribuirse entre los distintos tipos de bosque típicos de un país. Un estudio piloto de este tipo puede proporcionar estimaciones sobre la varianza poblacional (aunque probablemente no sean muy precisas).

También es posible que se encuentren tipos de bosque similares en países vecinos, de los que pueden obtenerse estimaciones de la varianza poblacional que sirvan de base para nuestro diseño de muestreo.

Cuando ni siquiera se dispone de esta información, los estadísticos forestales pueden tener que recurrir a información alternativa, a menudo no basada en diseños probabilísticos, como la opinión de expertos o revisiones bibliográficas.

Para una muestra aleatoria, el tamaño de la muestra es el siguiente:

$$n = \frac{t^2 * S^2}{A^2} = \frac{t^2 * (CV\%)^2}{(e\%)^2}$$

donde A se refiere al intervalo de confianza, en valor absoluto, que pretendemos alcanzar en nuestras estimaciones (como porcentaje e% si se expresa como relativo a la media), t es el valor correspondiente de la distribución t de Student y S^2 (normalmente reestimado a partir de estudios piloto o información previa) es la varianza muestral de la variable de interés, como el volumen por ha. El porcentaje de CV es el coeficiente de variación de la información anterior, expresado en porcentaje respecto a la media. El siguiente ejercicio muestra un ejemplo práctico para calcular el tamaño de la muestra.

Ejercicio práctico

Queremos calcular cuántas parcelas serían necesarias para estimar las reservas forestales de carbono con una precisión del 10 % (referida al intervalo de confianza del 95 %). Diversos estudios han mostrado valores de biomasa aérea en torno a 100 t/ha con una desviación estándar de 70 t/ha (CV%=70). **¿Cuántas parcelas deberían medirse si consideramos un muestreo aleatorio simple**?

Para calcularlo, necesitamos el valor correspondiente de la distribución t para una probabilidad de error del 5 % (o 0,05, a dos extremos). Sin embargo, para determinar ese valor t, necesitamos conocer el tamaño de la muestra, que en realidad es lo que se busca. Podemos empezar con un valor t de 2 en la primera iteración - que corresponde a un tamaño de muestra grande de más de 30 y obtener: $2^2*70^2/10^2 = 196$.

Consultando la ***tabla T*** para este tamaño de muestra de n=196 (de la primera iteración), hemos llegado a un valor t de ~1,97 - y la estimación anterior puede calcularse de nuevo a $1{,}97^2*70^2/10^2 \sim 190$.

Tenga en cuenta que esta estimación del tamaño de muestra exigido sólo es válida para el MAS.

En realidad, sin embargo, nuestros recursos son limitados y sólo es factible un cierto número de parcelas. En este caso, intentamos conseguir el resultado más preciso con el presupuesto disponible. Como ya ha aprendido, el aumento del tamaño de la muestra aumentará la precisión, por lo que deberíamos intentar planificar el mayor número posible de parcelas en función del diseño de la parcela y de las restricciones prácticas.

¿Cuál es mi variable objetivo?

Un inventario forestal sólo puede optimizarse hacia una única variable objetivo (cuya precisión debe maximizarse para los recursos disponibles). Con frecuencia se utiliza como variable objetivo el área basal del rodal, que está altamente correlacionada con el volumen y la biomasa. No obstante, la consideración de múltiples propósitos requiere compromisos en el muestreo y en el diseño de las parcelas, y puede ocurrir que el tamaño de la muestra que optimiza la precisión de la estimación del área basal no sea óptimo para otras variables.

Diseño de muestreo

Hasta ahora, hemos visto los conceptos básicos del muestreo y considerado los tres elementos del diseño de muestreo. Veamos ahora algunas opciones del diseño de muestreo. El diseño de muestreo define el proceso de selección de los elementos de muestreo, es decir, cómo se seleccionan los elementos de muestreo y cuántos (tamaño de la muestra). El resultado de este proceso de selección es una lista de todas las coordenadas de los lugares de muestreo.

Aquí nos limitaremos a abordar exclusivamente algunos diseños de muestreo típicos utilizados en el contexto de los IFN. Recuerde que ya tratamos anteriormente el MAS (consulte la Lección 1, Estimación mediante muestreo aleatorio simple) como un diseño principalmente teórico que en la práctica rara vez se utiliza en los IFN, pero que resultó útil para establecer una referencia sencilla con la que comparar las siguientes opciones.

Muestreo sistemático: el diseño de muestreo más común en los IFN

La utilización de una cuadrícula sistemática de sitios de muestreo es el diseño de muestreo estándar en los IFN. Un muestreo sistemático de este tipo tiene la ventaja de que la superficie forestal queda cubierta uniformemente por los sitios de muestreo, y garantiza que todos los sitios mantengan una distancia mínima entre sí. Conduce a una "afijación proporcional" de los lugares de muestreo entre los

tipos de bosque existentes. Y, como cubre uniformemente toda el área de interés, cabe esperar que una cuadrícula de muestreo de este tipo genere una muestra "representativa" de la población.

Las consideraciones teóricas y los numerosos estudios de simulación han demostrado que el muestreo sistemático genera prácticamente siempre una mayor precisión que el MAS, dado el mismo número de puntos de observación.

Esto puede explicarse por el hecho de que el muestreo sistemático cubre uniformemente toda la población, de modo que todas las condiciones se cubren de manera casi uniforme; otra razón es que, en el muestreo sistemático, los puntos muestrales vecinos tienen siempre una distancia definida y no pueden estar muy próximos entre sí: en los bosques y muchas otras poblaciones naturales, las parcelas que están próximas entre sí suelen estar más auto correlacionadas que las distantes, lo que resulta ineficaz.

Nota

El **tamaño de la muestra** se refiere al **número de elementos de muestreo seleccionados independientemente**, donde independientemente significa seleccionados mediante aleatorización. Dado que todos los sitios de muestreo de una cuadrícula sistemática son fijos una vez seleccionados el punto de partida y la orientación de la cuadrícula, una muestra sistemática basada en una aleatorización es sólo una (tamaño de la muestra = 1). Sin embargo, a partir de una única observación independiente, no podemos obtener una estimación de la varianza y, por tanto, ¡tampoco una estimación de la precisión!

Si nos fijamos en el estimador de la varianza para el MAS, si el tamaño de la muestra es n=1, entonces el denominador n-1 será cero y, por lo tanto, la varianza de la variable de interés S2y no está definida:

$$S_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

Con frecuencia, el estimador del MAS se utiliza para calcular la varianza del error de una muestra sistemática. Se sabe que esa varianza del error sobreestimará la verdadera varianza del error, por lo que decimos que el estimador del MAS es aquí un **estimador conservador**.

Esto significa que la precisión real es mayor que la que estimamos con el estimador del MAS, pero no podemos decir hasta qué punto es más precisa. Esta subestimación de la precisión también afectará a todas las demás estimaciones, como el tamaño de muestra necesario.

Estratificación

Ya hemos aprendido en lecciones anteriores que nuestro objetivo es **reducir** al máximo la distribución de las observaciones porque así aumentará la precisión de las estimaciones.

¿Qué más podemos hacer para que la población sea "más homogénea", a fin de aumentar la precisión*?*

La estratificación pretende subdividir la población en subpoblaciones más homogéneas. A estas subpoblaciones las llamamos estratos (en singular: estrato). En cada estrato se toma una muestra independiente. Cuando utilizamos el muestreo aleatorio simple en cada estrato, denominamos al diseño muestreo aleatorio estratificado. Es decir: no introducimos aquí un diseño de muestreo completamente nuevo, sino que aplicamos el MAS de forma independiente en cada estrato; lo nuevo es, en realidad, cómo combinamos al final las estimaciones por estratos para llegar a un total compuesto de todos los estratos. Para tener más precisión en este diseño, la estratificación debe ser "homogénea dentro de los estratos y heterogénea entre los estratos".

Observe el diagrama: la superficie forestal total se subdivide en dos estratos diferentes (claro y oscuro), y consideramos que cada uno de ellos es más homogéneo que la superficie total y que difieren claramente en sus valores medios. En el diseño de muestreo, se tratan como subpoblaciones independientes y se utilizan cuadrículas sistemáticas diferentes.

Hay muchas formas de subdividir una población en subpoblaciones: los criterios de estratificación son,

por ejemplo: tipos de bosque o regiones de crecimiento con condiciones homogéneas del sitio. A veces también se utiliza la demarcación administrativa, pero esto no conduce necesariamente a subpoblaciones más homogéneas ni a una mayor precisión. Sin embargo, puede utilizarse para facilitar la ejecución del inventario o para garantizar que se puedan ofrecer estimaciones más precisas por unidad administrativa.

Cálculo del tamaño de la muestra y afijación de las muestras a los estratos

Cuando se determina el tamaño de la muestra en el muestreo estratificado, es necesario responder a dos preguntas: cuántas muestras en total, y cómo distribuir/afijar las muestras a los estratos.

El tamaño de muestra requerido depende siempre del error permitido con una probabilidad de error dada y de la variabilidad dentro de la población; en la estratificación tratamos con una serie de subpoblaciones, y debemos considerar que las varianzas de las subpoblaciones difieren entre estratos.

Como los estratos suelen tener tamaños diferentes, estas diferentes varianzas de las subpoblaciones deben ponderarse al calcular el tamaño total de la muestra. Si hay un número de L estratos denotados por el subíndice h, y cada estrato tiene el tamaño (por ejemplo, en términos de área) N_h, el peso de cada estrato viene determinado por N_h/N.

Pero diseñar un inventario también puede implicar enfrentarse a limitaciones en cuanto a los costos del mismo. Así pues, aunque uno quiera afijar las parcelas de muestreo para minimizar la variabilidad, también puede pensar en el costo total incurrido en el inventario, donde C_h es el costo por unidad de muestreo en el estrato h.

Entonces, el tamaño total de la muestra puede calcularse como:

$$n = \frac{t^2 \sum \frac{N_h^2 S_h^2}{C_h}}{N^2 A^2}$$

Donde A es el error permitido, expresado como la mitad de la amplitud del intervalo de confianza objetivo. El error permitido es una cuestión de definición. De forma similar a la estimación del tamaño

de la muestra con MAS, proporcionada anteriormente, S y A pueden sustituirse por expresiones relativas: CV(%) y e(%).

Una vez calculado el tamaño total de la muestra, hay que afijar estas muestras a los distintos estratos. Para ello, se pueden considerar tres características de los estratos, individualmente o en conjunto:

1. **El tamaño del estrato:** cuanto mayor sea el estrato, más muestras se afijarán.
2. **La variabilidad del estrato**: cuanto más variable sea un estrato, más muestras se afijarán.
3. **El costo por unidad de muestreo**: cuanto mayor sea el costo, menos muestras se afijarán

Afijación proporcional	**Afijación de Neyman**	**Afijación óptima con minimización de costos**
Afijación de muestras en función únicamente del tamaño del estrato.	Considerando el tamaño de los estratos y la variabilidad dentro de los estratos para la afijación.	En esta opción, además del tamaño de los estratos y la variabilidad dentro de los estratos, también se incluyen las implicaciones de los costos (c).
$n_h = n \frac{N_h}{N}$	$n_h = n \frac{N_h S_h^2}{\sum_{h=1}^{L} N_h S_h^2}$	$n_h = n \frac{\frac{N_h S_h^2}{\sqrt{C_h}}}{\sum_{h=1}^{L} \frac{N_h S_h^2}{\sqrt{C_h}}}$

Nota

Recuerde que se puede **aplicar cualquier técnica de muestreo por estrato**. También puede haber diferentes técnicas de muestreo utilizadas en los distintos estratos. Es importante que para cada estrato las estimaciones puntuales y de intervalo de las variables objetivo puedan producirse, en el mejor de los casos, sin sesgo.

De hecho, l**a principal característica del muestreo estratificado es que consta de varios estudios de muestreo aplicados de forma independiente**. La única novedad es que hay que averiguar cómo combinar finalmente las estimaciones procedentes de los L diferentes estratos para poder generar estimaciones para toda la población.

Post- estratificación

También podemos estratificar el inventario después de haber aplicado una muestra no estratificada (por ejemplo, obteniendo estimaciones separadas para los distintos tipos de bosque). Esto se denomina post-estratificación y puede considerarse un tipo de agrupación de datos para los análisis.

Sin embargo, este análisis post- estratificado debe realizarse con cuidado, ya que la estimación deja de seguir estrictamente el diseño de muestreo. Por ejemplo: la agrupación de datos para el análisis no debe hacerse junto con la variable objetivo, por ejemplo, formando tres grupos igualmente amplios (post-estratos) de valores bajos, medios y altos; ¡este enfoque sería totalmente erróneo, aunque conducirá a valores de precisión altos (pero falsos)!

Antes de realizar un análisis post- estratificado con estimaciones de precisión de estimación, debe consultar a un experto en muestreo para evitar errores innecesarios e inferencias y conclusiones erróneas: todos los estimadores que se recomiendan en los libros de texto para los análisis post-estratificados vienen con algunos supuestos que se deben observar.

Muestreo en dos fases o muestreo doble

En el muestreo doble, se introduce una nueva característica: el uso de **variables auxiliares**, también

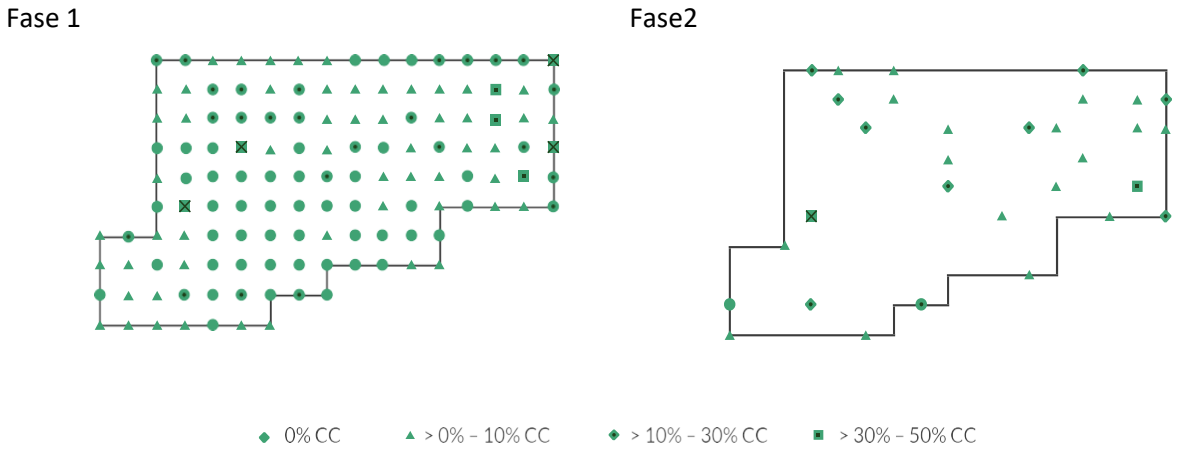
conocidas como **covariables**. Para aumentar la precisión de la estimación de la variable objetivo, es esencial comprender la correlación entre la variable objetivo y las variables auxiliares. La idea es recoger una muestra relativamente grande -pero de bajo costo- en la primera fase para obtener información de esas variables auxiliares, por ejemplo, mediante teledetección.

A continuación, en la segunda fase, se selecciona una muestra más pequeña en la que se observan tanto la variable objetivo como la auxiliar. Esto suele suponer un costo mucho mayor por parcela; entendamos esto mejor con un ejemplo.

Al estimar la biomasa forestal, es posible utilizar una primera fase de estimación de la variable auxiliar a partir de imágenes de teledetección y determinar un índice de vegetación en torno a muchos puntos muestrales. Esto es rápido y barato y también puede hacerse automáticamente.

A continuación, en la segunda fase, se instala en el campo una muestra mucho menor de parcelas de muestreo en las que se toman medidas y se estima la biomasa de la parcela. Esto es mucho más costoso que una muestra en la primera fase. Para todas las parcelas de campo, no sólo se determina la variable objetivo (biomasa, en este ejemplo), sino que también se observa la variable auxiliar (índice de vegetación, en este ejemplo) a partir de los datos de teledetección.

A continuación, se utilizan los pares de datos de la segunda fase de la variable objetivo y la variable auxiliar para establecer un modelo entre la variable objetivo y la auxiliar, a partir del cual se pueden producir las estimaciones. Los modelos más utilizados en este caso son la razón simple entre ambas variables o un modelo de regresión. Esto conduciría entonces a un doble muestreo con el estimador de razón y a un doble muestreo con el estimador de regresión, respectivamente.



Puede que ya haya quedado claro que cuanto mayor sea la correlación positiva entre las dos variables, **más eficiente será el estimador en términos de precisión**. Es decir, para una variable auxiliar eficiente, siempre buscamos una variable que esté altamente correlacionada positivamente con la variable objetivo. Al final, esto es, por supuesto, también una consideración del costo, porque la introducción de una primera fase también aumenta el costo del inventario. En la próxima lección sobre diseño de estimaciones, verá cómo esto mejora finalmente la precisión de la estimación de la variable objetivo.

El muestreo doble es una forma muy eficaz de aprovechar una variable auxiliar (barata de observar) para mejorar la precisión de la estimación de una variable objetivo (más cara de observar.

Muestreo doble para estratificación

El muestreo doble también es pertinente en el contexto de la estratificación. Hay casos de inventarios en los que se sabe o se asume que la estratificación puede aumentar la precisión, pero a veces no es posible delinear claramente los límites de los estratos (por ejemplo, en imágenes de teledetección), ya que se trata de transiciones "difusas" o continuas en lugar de líneas exactas. Además, esta delimitación lleva mucho tiempo y requiere conocimientos previos rigurosos.

Hasta ahora, para el muestreo estratificado, hemos considerado que los estratos se definirán antes del muestreo, por lo que también lo denominamos **pre- estratificación**. Al hacerlo, asumimos que dicha definición previa de los estratos está libre de errores; es decir: el tamaño de los estratos y sus ponderaciones en los estimadores no se consideran una fuente de error.

En el **muestreo doble para estratificación** (DSS) o **muestreo en dos fases para estratificación**, no es necesario definir los estratos antes del muestreo, sino que se definen durante el proceso de muestreo y se estiman los tamaños de los estratos.

Las dos fases del DSS son las siguientes: **en la primera fase se selecciona una muestra relativamente grande** (frecuentemente en imágenes de teledetección, ya que resulta bastante económica) y para cada punto muestral se determina a qué estrato pertenece. Es decir: la variable auxiliar que se observa en la primera fase es el **estrato**; en los IFN podría ser el **tipo de bosque**.

En la segunda fase, se selecciona una submuestra estratificada de las parcelas de la primera fase y se visitan estas parcelas para observar la variable objetivo -relativamente costosa-, lo que suele hacerse

sobre el terreno. La afijación del tamaño total de la muestra a los estratos se puede hacer siguiendo las mismas estrategias que con la pre- estratificación: uniforme, proporcional al tamaño, proporcional al tamaño y a la variabilidad, o proporcional al tamaño, a la variabilidad y al costo. La decisión sobre dicha afijación deberá tomarse a partir de la información disponible sobre la variabilidad esperada y el costo por parcela en los estratos que se hayan distinguido.

Dados los mismos tamaños de muestra en la muestra de la segunda fase y en la pre- estratificación normal, el DSS será menos preciso que la pre -estratificación. La razón es que en el DSS los tamaños de los estratos se estiman a partir de la muestra de la primera fase y dicha estimación del tamaño conlleva un error de muestreo que se propaga al error total. Esto también puede observarse con los estimadores para el DSS, que no se indican aquí, pero pueden encontrarse en los libros de texto sobre muestreo.

¿Sabía qué?

¿Podemos utilizar también una clasificación por teledetección para separar estratos?

Si, podemos, y a menudo se hace así. Imaginemos, por ejemplo, una clasificación basada en teledetección en distintos tipos de bosque, para la que esperamos diferencias en la biomasa forestal. Sin embargo, al igual que ocurre con la interpretación visual antes mencionada, toda clasificación tendrá errores. Dado que las estimaciones del área de los estratos contienen errores, ¡hay que tener en cuenta esa fuente adicional de incertidumbre en el estimador!!

Consejos prácticos

Nunca hay que "inventar" un nuevo diseño de muestreo o de parcela ignorando el problema de obtener un estimador estadístico no sesgado. La exactitud de un estimador depende de una cuidadosa reflexión sobre el proceso de selección e inclusión. Puede encontrarse fácilmente con trampas estadísticas sin solución con sólo hacer pequeños cambios en la parcela o en el diseño de

muestreo.

Por ejemplo, una simple regla como "ampliar la parcela de muestreo si se cumple una determinada condición" puede dar lugar a problemas estadísticos inesperados (¡la inclusión resultante de las probabilidades de los árboles no se puede calcular fácilmente!). Otras reglas, como **cambiar las parcelas que se superponen completamente en el límite del bosque hacia el interior del bosque**, infringen la definición de la población. Son simplemente erróneas y podrían dar lugar a estimaciones sesgadas.

Diseño de parcela o de observación

A continuación, examinaremos el diseño de la observación.

El diseño de muestreo describe cómo se seleccionan los puntos muestrales, mientras que el diseño de la parcela describe cómo se eligen los árboles que se van a muestrear alrededor del punto seleccionado. La pregunta es: ¿qué objetos (por ejemplo, árboles) deben incluirse en cada sitio de muestreo alrededor del punto muestral?

Como en prácticamente todos los pasos de diseño/planificación de un IFN, mientras se optimiza o adapta el diseño de la parcela a las condiciones específicas del bosque, hay que pensar cuidadosamente en cómo asignar los limitados recursos (tiempo, presupuesto y personal) de la manera más eficiente. La **eficiencia** puede considerarse como la **relación entre los costos y la precisión resultante de las estimaciones**. Si no se consideran suficientemente los recursos, esto puede comprometer la sostenibilidad de un IFN permanente.

Captar la variabilidad como objetivo principal en la planificación del diseño de las parcelas

Desde un punto de vista puramente estadístico en la optimización del diseño de las parcelas, nuestro objetivo es capturar **el máximo de variabilidad dentro de cada parcela**. La razón es que así conseguimos que la variabilidad entre la parcela sea pequeña. Y eso se traduce en una distribución ajustada de las observaciones de la parcela en torno a la media estimada (*compare la Lección 1 de este curso*), lo que significa: mayor precisión de la estimación.

Un concepto importante en este contexto es la **autocorrelación espacial**, que también observamos en

las poblaciones de bosques. Esto significa que los objetos (en este caso, las parcelas) que están más próximos tienden a tener observaciones más correlacionadas.

Una alta correlación significa que, conociendo el valor del primer objeto, se puede predecir bastante bien el valor del segundo objeto a una distancia espacial determinada. En ese caso, la medición de la segunda observación no es muy eficaz, ya que no aporta mucha información adicional; incluso puede ser una pérdida de dinero.

Considerar la importancia de la autocorrelación espacial en la planificación del diseño de los inventarios lleva a algunas conclusiones relativas al diseño de las parcelas, así como al diseño de muestreo:

- ✓ Es conveniente que haya cierta distancia entre las parcelas de muestreo. Las parcelas de muestreo espacialmente próximas no son eficaces.
- ✓ Es bueno tener un diseño de parcela que cubra una superficie más grande para que las observaciones dentro de la parcela muestren una menor autocorrelación:
    a) Por lo tanto, dada la misma superficie, las parcelas en franjas alargadas son estadísticamente más eficientes que las parcelas circulares o cuadradas; y
    b) Otra opción para aumentar la eficiencia de una determinada superficie de parcela es subdividir la parcela en subparcelas espacialmente separadas a cierta distancia unas de otras: es lo que llamamos "conglomerados".

Con estas dos opciones, recuerde que no sólo son importantes las consideraciones estadísticas, sino también las de costos. El costo por parcela será mayor en el caso de las parcelas alargadas en franjas o de los conglomerados que en el de una parcela compacta de la misma superficie: por lo tanto, en la práctica, estas consideraciones de optimización deben equilibrar siempre los criterios estadísticos y de costos. Por lo general, esta correlación espacial disminuye a partir de los 50-200 m (dependiendo del tipo de bosque y de la gestión).

Parcelas de superficie fija y parcelas de muestreo anidadas

El diseño de parcela más básico es la parcela de superficie fija. La forma y el tamaño de estas parcelas de superficie fija pueden variar en función del objetivo específico del inventario y de las condiciones del bosque. En general, las parcelas circulares son más comunes que las rectangulares en el monitoreo

forestal, mientras que en los estudios ecológicos la forma cuadrada es más común y, a veces, en los estudios ecológicos se utiliza el término "quadrat" en lugar de "parcela".

Desde un punto de vista teórico, cualquier forma de parcela es admisible; sin embargo, es crucial considerar cuidadosamente el objetivo del inventario y las condiciones del bosque a la hora de seleccionar el diseño de parcela adecuado, y equilibrar el costo y las consideraciones prácticas con la necesidad de una recopilación de datos exacta.

Nota

El factor de expansión de la parcela (o árbol)

Muchas variables de interés están relacionadas con la superficie, como el "número de árboles por hectárea". Esto significa que, por ejemplo, cuando se duplica la superficie de una parcela, se espera que el número de árboles encontrados también se duplique en promedio.

Para expandir o aumentar la escala de la observación a la unidad de notificación típica de una hectárea, estas observaciones por parcela relacionadas con la superficie deben multiplicarse por un factor de expansión resultante de la relación 1 ha/superficie de parcela.

Las variables relacionadas con la superficie suelen ser las asociadas a mediciones cuantitativas directas, como el volumen, la biomasa, el número de árboles o la densidad de regeneración.

Los árboles de distintos diámetros suelen aparecer con densidades diferentes (en los bosques naturales, por ejemplo, hay muchos más árboles pequeños que grandes). Los árboles grandes contienen gran parte de la biomasa forestal, pero son muy escasos. Si entonces utilizamos una parcela relativamente grande para estar seguros de que en promedio tenemos algunos árboles grandes dentro de la superficie de la parcela, tendríamos que medir una cantidad enorme de árboles pequeños. Es decir: una sola superficie de parcela no suele ser eficiente.

Una solución común en este caso es utilizar el **denominado diseño** de parcelas anidadas, en el que se anidan subparcelas de diferentes tamaños, de forma que se observan árboles de diferentes tipos de tamaño en diferentes superficies de subparcelas. Aquí es importante mantener una terminología

estricta para evitar confusiones: el conjunto (la combinación de todas las subparcelas) es la parcela, mientras que las diferentes formas anidadas de distinto tamaño son las subparcelas.

Las subparcelas anidadas no se evalúan una tras otra, sino que se comprueba (en un barrido, normalmente en el sentido de las agujas del reloj) para cada árbol si está incluido o no. Todas acabarán en la misma tabla de datos.

Nota

Contabilizar las probabilidades desiguales

La probabilidad de inclusión en un inventario forestal es la probabilidad de que un árbol esté incluido en una muestra. Dado que las unidades de muestreo son efectivamente parcelas, basadas en superficies, esta probabilidad es en efecto el inverso del factor de expansión.

Por lo tanto, un diseño con subparcelas dará lugar a probabilidades de inclusión desiguales, y tenemos que reflejar esto en el diseño de la estimación: el factor de expansión de la parcela será mayor para las parcelas más pequeñas, ¡donde se observan los árboles más pequeños!

Dado que las subparcelas tienen tamaños y superficies diferentes, los árboles tendrán factores de expansión distintos según la subparcela en la que se hayan contado. Por tanto, debemos calcular el factor de expansión correcto para cada árbol individualmente, basándonos en su dap en su pertenencia a una subparcela y su superficie correspondiente. Los factores de expansión estandarizan entonces todas las superficies a una única base por hectárea.

El siguiente vídeo [en inglés] explica cómo establecer y evaluar parcelas de muestreo anidadas de superficie fija sobre el terreno: **https://www.youtube.com/watch?v=IA-PflXW9_k&t=2s**

Corrección por pendiente

La superficie a la que se refieren todas las observaciones y estimaciones es la superficie cartográfica, o proyección horizontal del terreno en el plano cartográfico.

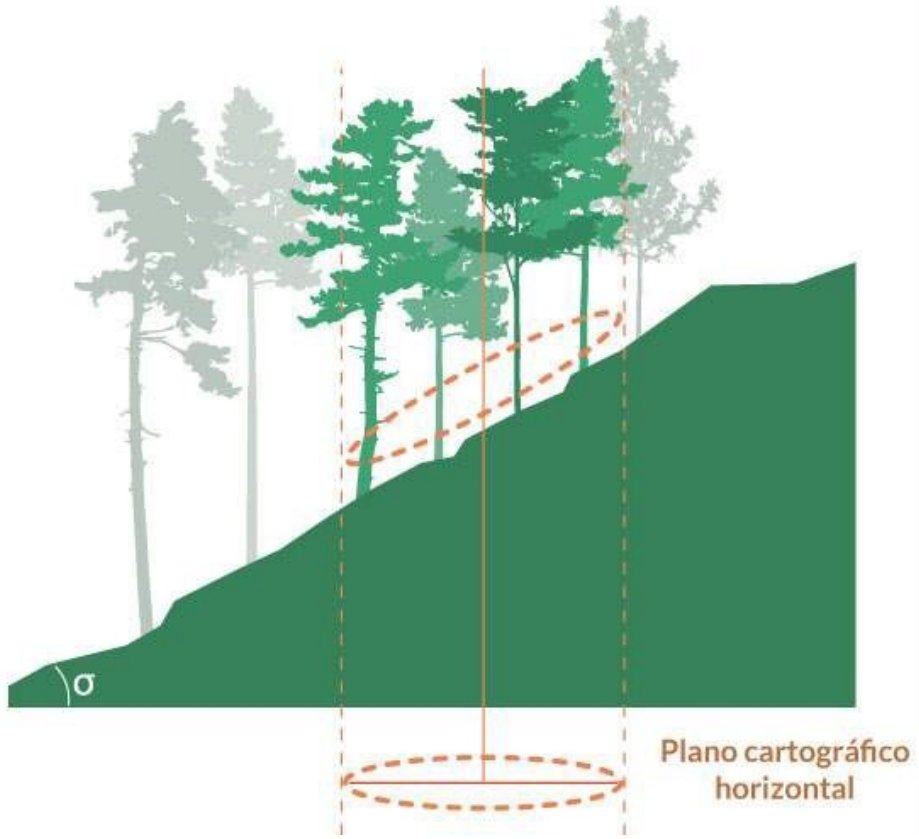
Cuando no es posible medir directamente las distancias horizontales al medir la parcela (utilizando instrumentos electrónicos modernos) y las distancias se miden a lo largo de la pendiente, la superficie de proyección horizontal de esta parcela es menor que la superficie prevista de la parcela y las distancias medidas desde el centro de la parcela hasta los árboles son mayores que las distancias horizontales proyectadas (salvo que midamos exactamente a lo largo de las líneas de contorno).

Para garantizar una superficie de parcela igual en el plano cartográfico horizontal, que es un requisito previo para obtener estimaciones no sesgadas relacionadas con la superficie, la superficie de parcela oblicua que constituye la parcela en el terreno debe ampliarse en función del ángulo de la pendiente.

Cuando se utilizan parcelas circulares de superficie fija, éstas se convierten en elipses cuando se proyectan en la pendiente. Para establecer estas parcelas en la pendiente hay esencialmente dos opciones:

1. bien se utiliza un distanciómetro electrónico que mide directamente la distancia horizontal: entonces, automáticamente se incluyen los árboles correctos dentro de la distancia horizontal definida (radio). Se establece una parcela elíptica sin necesidad de trazarla específicamente.
2. O -que es el enfoque tradicional- se calcula la superficie (más grande) de la elipse proyectada en pendiente y se establece en la pendiente un círculo con exactamente esa superficie. Para ello, hay que corregir la pendiente del radio nominal de la parcela en el plano horizontal con el factor que se indica a continuación para obtener el radio más grande del círculo que se trazará en la pendiente.

$$\sqrt{\cos \alpha}$$



Videos

How to correct for slope and how to deal with plots at the forest boundary [en ingles]

https://www.youtube.com/watch?v=InPERYNxQ0E&t=1s

En caso de que se haya omitido dicha corrección por pendiente durante el establecimiento de la parcela, las observaciones obtenidas de esta parcela podrán corregirse posteriormente (ya que las parcelas tienen tamaños desiguales en la proyección horizontal en función de la pendiente). Puesto que la superficie horizontal real es entonces menor que la prevista, habría que multiplicar el resultado por el factor de corrección 1/cos α. Sin embargo, es necesario haber medido el ángulo de la pendiente; de lo contrario, no es posible realizar una corrección.

En la mayoría de los IFN, la corrección por pendiente suele considerarse para ángulos de pendiente > 10

%, que es una de las condiciones con las que funcionan en la práctica los inventarios forestales. Además, en presencia de pendientes suaves de < 10 %, las mediciones de distancia a menudo se pueden realizar horizontalmente mediante nivelación manual.

Nota

La corrección por pendiente se aplica a cualquier diseño de parcela y siempre se debe tener en cuenta con antelación; las correcciones son bastante sencillas para las parcelas circulares de superficie fija. Por supuesto, los mismos principios de corrección por pendiente se aplican también a las parcelas cuadradas y rectangulares.

Y para estas dos formas de parcela, la corrección por pendiente es más laboriosa: para las parcelas cuadradas, hay que marcar en la pendiente las esquinas de una superficie de parcela efectiva más grande para que la superficie de parcela proyectada se corresponda con la superficie nominal. En el caso de parcelas rectangulares alargadas, solemos recorrer la línea central y medir los árboles a derecha e izquierda en una distancia definida: en este caso, es necesario corregir por pendiente en ambas direcciones de la parcela: la línea larga por la que deseamos recorrer y las mediciones a derecha e izquierda.

Muestreo con conglomerados

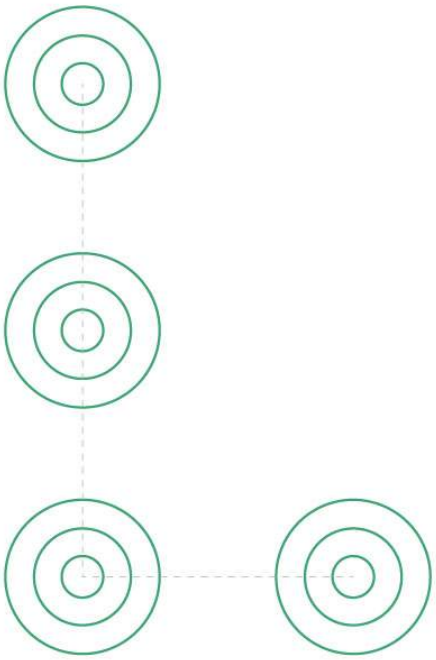
En los IFN, la ubicación y el desplazamiento a los lugares de muestreo es un factor de costo importante, sobre todo cuando la red de carreteras es deficiente. La cuadrícula de muestreo suele ser dispersa y las distancias entre las parcelas son grandes. Por eso queremos evaluar toda la información posible una vez que el equipo se encuentra en la parcela. Esto requiere parcelas grandes.

Sin embargo, hemos aprendido que, debido a la autocorrelación espacial, es bueno que las observaciones se encuentren a cierta distancia espacial unas de otras, por lo que, en lugar de establecer una parcela grande por punto muestral, los IFN suelen optar por establecer los **denominados conglomerados**: las parcelas grandes individuales se subdividen en subparcelas dispuestas cada una a cierta distancia espacial.

El resultado son subparcelas dispuestas según algún patrón geométrico (por ejemplo, las esquinas de un cuadrado o en forma de L). El conjunto de subparcelas conforma la parcela y conviene no confundir parcelas y subparcelas. Las parcelas son los elementos centrales del muestreo y el número de parcelas corresponde al sitio de muestreo; no al número de subparcelas. Veamos algunos ejemplos típicos de muestreo con conglomerados.

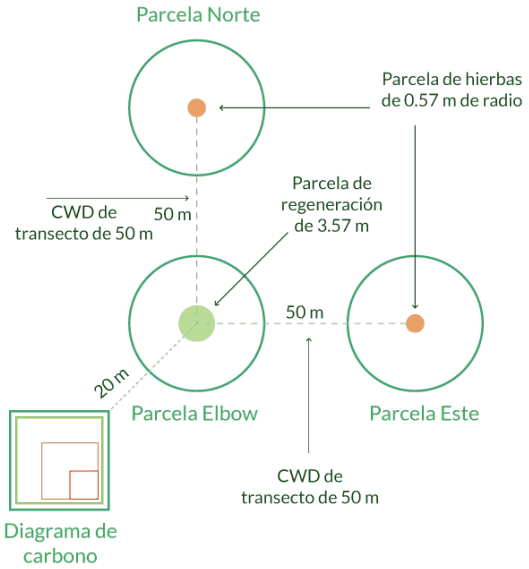
Ejemplo 1

Ejemplo de un conglomerado en forma de L con 4 subparcelas. Las subparcelas son parcelas anidadas con diferentes radios de parcelas circulares para diferentes tipos de diámetro de árboles.



Ejemplo 2

Conglomerado con 3 subparcelas anidadas en forma de L, que se utiliza en el IFN de Bután. CWD: Residuos leñosos gruesos



La distribución espacial y la distancia entre las subparcelas están dispuestas de tal forma que un conglomerado puede "capturar" mucha más variabilidad en comparación con una única parcela compacta del mismo tamaño de superficie. Para planificar el diseño de un conglomerado, debemos decidir ciertas características:

1. El número de subparcelas por conglomerado.
2. La distancia entre las subparcelas.
3. Tamaño y forma de las subparcelas.
4. La disposición espacial de las subparcelas.

Algunas consideraciones sobre la forma y el tamaño de las (sub)parcelas

Ya hemos llegado a la conclusión de que cada parcela de muestreo debe capturar la mayor variabilidad posible para aumentar la precisión global de la estimación. Si dispusiéramos de un número ilimitado de recursos, el mismo número de parcelas más grandes siempre sería mejor que el mismo número de parcelas más pequeñas.

En cambio, si disponemos de recursos limitados, tendremos que decidir si utilizamos un mayor número de parcelas más pequeñas o un menor número de parcelas más grandes. Aumentar el tamaño de las parcelas implica un aumento de los costos marginales de precisión por cada árbol adicional medido, mientras que aumentar el número de parcelas implica ganancias de precisión cada vez menores a cambio de un mayor tiempo de recorrido entre parcelas. Sin embargo, a partir de un determinado tamaño de parcela, el efecto del aumento del tamaño de la muestra sobre la precisión será más importante que el aumento del tamaño de la parcela.

¿Sabía qué?

Tamaño de la parcela y eficacia estadística

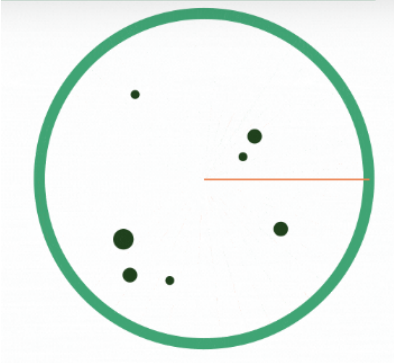
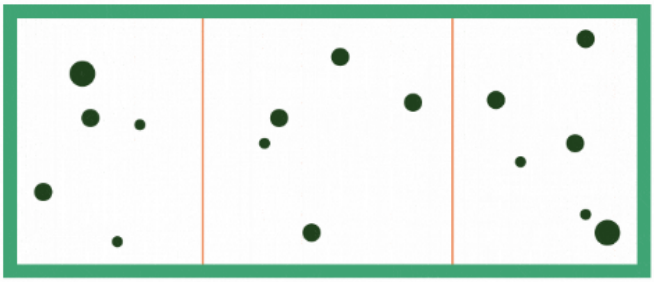
La ganancia marginal de información que podríamos esperar de la medición de un árbol adicional por parcela disminuye con cada nuevo árbol. Imagine que ya hemos medido 99 árboles en una parcela de

muestreo, ¿esperaría que aprendiéramos algo nuevo midiendo el árbol número 100? Probablemente no, porque sólo añadiría información redundante.

Por otra parte, los costos de evaluación en la parcela aumentarán linealmente con cada unidad adicional de superficie observada (o número de árboles). Pero ¿**dónde debemos trazar el límite**?

Tanto la experiencia como los estudios empíricos sugieren que evaluar más de 15-20 árboles por (sub)parcela ya no es eficiente. En ese caso, es mejor invertir los recursos en aumentar el tamaño de la muestra (¡es mejor tener más parcelas pequeñas que menos parcelas grandes!).

Existen diversos argumentos prácticos y estadísticos que determinan la forma de las parcelas (o subparcelas), y también hay que tener en cuenta las tradiciones y normas comunes en las distintas partes del mundo. Los siguientes lineamientos generales se aplican a las mismas superficies de parcela/subparcela con diferentes formas:

Las parcelas de muestreo circulares son:	**Parcelas rectangulares largas:**
	
• fáciles de aplicar: prácticamente todo puede medirse desde el centro; • relativamente fácil en cuanto a la corrección por pendiente; pero • muy compactas y probablemente capturen menos variabilidad.	• necesitan más trabajo para marcar las parcelas (por ejemplo, marcando con una cinta el transecto central y recorriéndolo a pie); • tienen, en promedio, más árboles limítrofes que verificar; • intersecan, en promedio, más a menudo con los

límites entre tipos de bosque y necesitan una mayor consideración de las correcciones de los límites;

- requieren más tiempo para la corrección por pendiente; probablemente capturará más variabilidad; y
- son buenas cuando la visibilidad es baja (el sotobosque es demasiado denso), ya que sólo pueden observarse distancias cortas a derecha e izquierda de la línea central.

Sobre el número de subparcelas por conglomerado

La agrupación de subparcelas en una observación conjunta siempre será menos eficiente que seleccionar el mismo número de subparcelas como parcelas seleccionadas independientemente en toda la región de inventario. El muestreo con conglomerados es un compromiso utilizado para reducir los costos de desplazamiento y observar mayores superficies de parcelas en cada lugar de muestreo, reduciendo al mismo tiempo la redundancia causada por la autocorrelación espacial mediante la distribución espacial de las subparcelas.

Por lo tanto, son válidos los mismos argumentos que para la planificación del diseño de parcelas únicas: aumentar la superficie observada significa aumentar los costos, mientras que el error estándar se reducirá hasta un cierto límite, más allá del cual apenas habrá una mayor reducción.

Así pues, invertir cada vez más tiempo y esfuerzo en una sola parcela no tiene, a partir de cierto punto, ningún efecto significativo sobre la precisión. Por lo general, no hay un gran efecto en la precisión después de un número de 3-5 subparcelas (dependiendo de la variabilidad espacial), y la medición de más parcelas por agrupación se vuelve ineficaz.

Comprobación de la realdad

La viabilidad como argumento orientador

Podemos derivar muchas consideraciones estadísticas a partir del tamaño de la muestra y la parcela, pero al final, el argumento más importante es la viabilidad. En la mayoría de los casos, nos vemos obligados a examinar los recursos disponibles y sacarles el máximo partido.

A efectos de planificación, sería beneficios que, en promedio, un solo equipo de campo pudiera medir todo un conglomerado en un solo día. Esto afectará al número de subparcelas que sean factibles y a su tamaño, teniendo en cuenta las subparcelas relativamente pequeñas (en muchos inventarios forestales, las consideraciones estadísticas llevan a tamaños de parcela que incluyen unos 15-20 árboles en promedio).

Es una situación habitual en los IFN que se necesite mucho tiempo para llegar al punto muestral y para caminar de una subparcela a otra. Podemos considerar estos tiempos de recorrido como ineficientes con respecto a la medición de nuestras variables objetivo: con frecuencia, la mayor parte del tiempo en terreno se utiliza para este tipo de recorridos ineficientes. Entonces, uno puede imaginar fácilmente que más de 4-5 subparcelas en muchos casos ya se convertirán en un desafío en términos de consumo de tiempo.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección:

- La planificación de cualquier estudio de muestreo puede desglosarse en tres elementos básicos de diseño técnico: diseño de muestreo, diseño de observación/parcela y diseño de la estimación.
- Uno de los aspectos definidos en el diseño de muestreo es el número de elementos de muestreo (parcelas) que deben observarse.
- Por lo general, un inventario forestal debería optimizarse hacia una única variable objetivo (cuya precisión debe maximizarse para los recursos disponibles). Con frecuencia se utiliza como variable objetivo el área basal del rodal, que está altamente correlacionada con el volumen y la biomasa.
- En los inventarios forestales, el diseño del muestreo define cómo se seleccionan las muestras de la población y cuál es el tamaño de la muestra.
- La estratificación consiste en "subdividir" la población total (superficie forestal) en subpoblaciones más homogéneas que denominamos "estratos" (en singular "estrato").
- El diseño de la parcela define lo que se va a hacer en cada punto muestral; también define las reglas para incluir los árboles de muestra que se van a observar.

Lección 3: Diseño de estimaciones

Introducción de la lección

En esta lección estudiaremos el diseño de la estimación, que consiste en los métodos y fórmulas que aplicamos para obtener estimaciones no sesgadas **a partir de los datos recogidos en un diseño de muestreo y en un diseño de parcela.**

Objetivos

Al final de esta lección, usted podrá:

1. Describir los estimadores básicos de los métodos de muestreo comunes.
2. Explicar la importancia de aplicar el estimador correcto.

Diseño de la estimación

Empecemos esta lección examinando algunos diseños típicos de estimaciones. Algunas de estas alternativas son específicas para el diseño de muestreo utilizado, y otras pueden aplicarse en función de diferentes diseños de muestreo

En algunos casos, también tenemos la libertad de aplicar diferentes estimadores a los datos recogidos con un diseño de muestreo determinado. Por ejemplo, podemos incluir datos auxiliares con un **estimador de razón** (analizado en secciones más adelante en esta lección). También podríamos obtener una estimación sin tener en cuenta la variable auxiliar si ésta no ayuda a producir una estimación más precisa.

Corresponde entonces al analista de datos decidir qué estimador utilizar. Si se pueden producir múltiples estimaciones alternativas, la elección suele recaer en el diseño de estimación que conduzca a una mayor precisión (lo que equivale a "menor error estándar de las estimaciones").

Inferencia basada en el diseño, asistida por el modelo y basada en el modelo

En la Lección 1 analizamos el término "**inferencia**". A veces, los términos **inferencia** y **estimación** se utilizan indistintamente, porque cada estimación significa hacer inferencias sobre los valores verdaderos

de la población. Por ello, algunos expertos en inventarios prefieren hablar en general de inferencia cuando se refieren a la estimación, porque la inferencia implica algo más que sólo la estimación: también se refiere al propósito de la estimación. Veamos ahora los tres paradigmas inferenciales: inferencia basada en el diseño, **inferencia basada en el modelo** e **inferencia asistida** por el modelo.

Inferencia basada en el diseño	Inferencia basada en el modelo	Inferencia asistida por el modelo	
No hacemos supuestos sobre la estructura (espacial) de la población. Suponemos esta estructura como desconocida y nuestro objetivo es estimar las características de esta población fija. La ausencia de sesgo está garantizada exclusivamente por el diseño del muestreo y de la parcela, es decir, por la aleatorización.	La población se considera una realización de un proceso estocástico, y durante la estimación se pueden considerar supuestos sobre el proceso o modelo subyacente. El supuesto es que estamos considerando sólo una de las muchas poblaciones posibles (que constituyen una superpoblación). Dado que ningún modelo puede describir perfectamente esta población, la incertidumbre persistirá incluso después de un censo completo, y procede de la "calidad" del modelo utilizado, no del diseño de muestreo	Se utiliza un modelo como apoyo a la estimación basada en el diseño, situándose a medio camino entre la inferencia basada en el diseño y la inferencia basada en el modelo. Esto significa que, aunque el modelo no esté bien especificado, no introducirá sesgos, pero afectará a la precisión de la estimación. Algunos ejemplos son el estimador de razón y el de regresión, que utilizan modelos simples durante la estimación estableciendo una relación entre una variable auxiliar y la variable objetivo.	**Supuestos de la población**

La validez de las estimaciones (ausencia de sesgo) depende exclusivamente del diseño de muestreo (selección de parcelas de muestreo, aleatorización). La teledetección o los datos auxiliares no se integran en la fase de estimación, pero quizá sí en la de planificación, por ejemplo, para la estratificación. Las estimaciones se producen únicamente a partir de observaciones en parcelas de las variables objetivo.	Las observaciones de campo se utilizan para establecer una relación (modelo) con variables auxiliares que suelen ser índices teledetectados. A continuación, el modelo se utiliza para predecir la variable objetivo a partir de una cobertura completa de estos índices. La validez de la estimación depende totalmente de la validez del modelo. **Ejemplo**: Para cada píxel de una imagen satelital un modelo predice biomasa/ha, las estadísticas se derivan posteriormente como agregado de los valores de los píxeles.	Se consideran las observaciones de campo de la variable objetivo más las variables auxiliares de las parcelas. La validez de las estimaciones depende del diseño de muestreo, pero la precisión de la estimación se puede aumentar integrando la información adicional procedente de la variable auxiliar. **Ejemplo**: En lugar de estimar la biomasa directamente, se estima una relación entre la biomasa y, por ejemplo, el NDVI, donde los valores del NDVI sirven como variable auxiliar y están disponibles para toda el área de bosque (población).	**Validez de las estimaciones**

Estimación con conglomerados

Contrariamente a muchos libros de texto sobre muestreo, no nos referimos al muestreo por conglomerados como un diseño de muestreo en sí mismo, sino al muestreo con conglomerados, es decir, lo consideramos como un diseño de parcelas, ya que es más coherente con la terminología que

utilizamos para el diseño de parcelas.

Sin embargo, el significado de ambos es el mismo: un único elemento de muestreo consta de varias subelementos, que se seleccionan conjuntamente en un único paso de aleatorización. Dado que las subparcelas en un conglomerado no se seleccionan independientemente unas de otras, **el tamaño de la muestra se refiere al número de conglomerados seleccionados y no al número de subparcelas**. El conglomerado puede considerarse como una única parcela de "forma inusual", cuya forma inusual se debe a la disposición espacial inconexa de la parcela.

Para el muestreo aleatorio simple de conglomerados: cuando las observaciones de subparcelas se agregan a nivel de conglomerado = nivel de parcela (sólo un valor, ya sea una media o un total por conglomerado), la estimación posterior puede seguir los mismos estimadores que los introducidos en la lección 1 (muestreo aleatorio simple: MAS). Sin embargo, a menudo ocurre que tenemos que considerar conglomerados de diferentes tamaños (= diferentes números de subparcelas), ya que no siempre todas las subparcelas están dentro de la población objetivo.

Entonces, el estimador de razón sería una opción, utilizando el tamaño del conglomerado (el número de subparcelas) como variable auxiliar.

Sin embargo, también puede ser interesante hacer un análisis por subparcela dentro de los conglomerados; esto no cambiará nada en cuanto a los resultados de la estimación puntual y por intervalos, pero permite realizar análisis adicionales de la estructura espacial de los bosques y de la eficacia del diseño de conglomerados.

Enséñame las matemáticas

Estimación con conglomerados

Los conglomerados pueden tener un número igual o desigual de subparcelas (m) para todos los conglomerados. En esta sección, presentamos el estimador sólo para la situación de conglomerados con igual número sometidos a muestreo aleatorio. Para los conglomerados con tamaños desiguales, es necesario volver a ponderarlos. La media estimada por subparcela se puede calcular a partir de la media estimada por conglomerado y el número medio de subparcelas por conglomerado de la

siguiente manera:

$$y = \frac{\bar{y}_{cl}}{\bar{m}}$$

y la varianza del error estimada de la media por subparcela es::

$$\hat{var}_{cl}(\bar{y}) = \frac{1}{\bar{m}^2}\frac{S^2_{y_i}}{n}$$

Donde y_i son observaciones por subparcela. La varianza estimada por conglomerado se puede obtener calculando la varianza sobre las observaciones por conglomerado con el estimador conocido para el MAS:

$$S^2_{y_i} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_{cl})^2}{n-1}$$

Los resultados totales estimados, como es usual, proceden de multiplicar la media por el total. En el muestreo por conglomerados, se puede tomar la media por conglomerado y el número de conglomerados (***N***), o la media por subparcela y el número de subparcelas (***M***):

$$\tau = N * \bar{y}_{cl} = M * \bar{y}$$

Y la varianza del error respectiva para el total se puede calcular como:

$$var(\hat{\tau}) = N^2 \hat{var}(\bar{y}_{cl}) = M^2 \hat{var}(\bar{y})$$

Eficacia del muestreo con conglomerados-Correlación dentro del conglomerado

La similitud de las observaciones dentro de un conglomerado se puede cuantificar mediante el **Coeficiente de Correlación dentro del Conglomerado (ICC**), a veces denominado Coeficiente de Correlación Intraclase. Cuanto mayor sea esta correlación, más redundantes serán las observaciones de las distintas subparcelas y menor será la información obtenida.

Este análisis es muy instructivo a la hora de comprender y analizar el desempeño del muestreo por conglomerados para poblaciones con diferente estructura de autocorrelación espacial, ya que ésta se

refleja directamente en el coeficiente de correlación dentro del conglomerado. Cuando el CCI es alto, se puede considerar la posibilidad de aumentar las distancias entre las subparcelas (lo que aumenta el tiempo de recorrido y los costos) o reducir el número de subparcelas.

Sin embargo, hay que tener en cuenta que los ICC pueden ser diferentes para distintas variables y que un diseño de conglomerados óptimo para una variable no es necesariamente igual de óptimo para otra. Es práctica común utilizar el área basal como **variable orientadora** en estas optimizaciones, ya que se correlaciona bien con otras variables del árbol (por ejemplo, la biomasa).

Para los conglomerados compuestos por varias subparcelas, resulta que si:

- ICC = 0 (observaciones no correlacionadas), no hay diferencia en el desempeño del muestreo por conglomerados de n conglomerados y el MAS con n*m subparcelas. Nuestro objetivo es mantener el CCI en un nivel bajo. Pero acercar el ICC a cero es imposible en la práctica, ya que la distancia entre las subparcelas resultaría demasiado grande.
- ICC < 0 (correlación negativa entre subparcelas), el muestreo con n conglomerados sería más eficaz que con n*m subparcelas seleccionadas de forma independiente. Esta situación es muy poco probable en los inventarios forestales (debido a la autocorrelación espacial).
- ICC > 0 (cierta redundancia dentro de los conglomerados) es el caso más típico. El muestreo con conglomerados es menos eficaz que la selección independiente de parcelas individuales.

Sin embargo, si se incluyen los costos de inventario, es probable que la eflciencia total sea mayor debido a la reducción de los desplazamientos

Muestreo estratificado

Ha aprendido que la estratificación tiene por objeto subdividir la población total en subpoblaciones más homogéneas, en las que se aplican estudios de muestreo independientes. Al combinar las estimaciones únicas de los distintos estratos, debemos recordar que estos estratos tienen tamaños diferentes. Por lo tanto, tenemos que ponderar todas las estimaciones de los estratos con los respectivos tamaños relativos de los estratos. En el monitoreo forestal, el tamaño de los estratos suele indicarse en términos de superficie, y la suma de las ponderaciones de todos los estratos sería 1 (es decir, igual al área total).

El muestreo estratificado no introduce un nuevo diseño de muestreo, pero lo que sí es nuevo es el marco utilizado para integrar las estimaciones de los distintos estratos en una estimación del área total. Por lo tanto, el muestreo estratificado introduce en realidad una variación del diseño de la estimación: combinar estimaciones independientes de L estratos en una única estimación de la población total.

Enséñame las matemáticas

Estimación con muestreo estratificado

En lo sucesivo, utilizaremos la notación ***h*** como índice para un estrato y L para el número total de estratos. A continuación, se puede estimar una media no sesgada sobre múltiples estratos como una suma ponderada. Las ponderaciones aquí son las proporciones de superficie de los distintos estratos N_h/N:

$$\bar{y} = \sum_{h=1}^{L} \frac{N_h}{N} \bar{y}_h = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{y}_h$$

El estimador para la varianza del error es:

$$\hat{var}(\bar{y}) = \sum_{h=1}^{L} \left\{ \left(\frac{N_h}{N} \right)^2 \hat{var}(\bar{y}_h) \right\} = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 \frac{S_h^2}{n_h}$$

La raíz cuadrada de esta varianza del error es el error estándar. El total se calcula como:

$$\hat{\tau} = N\bar{y} = \sum_{h=1}^{L} \frac{N_h}{N} \hat{\tau}_h = \sum_{h=1}^{L} N_h \bar{y}_h$$

Y la varianza del error del total estimado es:

$$\hat{var}(\hat{\tau}) = \hat{var}(N\bar{y}) = N^2 \hat{var}(\bar{y})$$

Eficacia del muestreo estratificado

Las consideraciones estadísticas revelan que la estratificación es más eficaz para aumentar la precisión de la estimación de la media cuanto más diferentes son las medias de los estratos. Estos efectos positivos (es decir, una mayor precisión global) tienden a reducirse a medida que aumenta el número de estratos.

Desde un punto de vista estadístico, la conformación de más de seis estratos no suele tener un efecto significativo para mejorar la precisión de la estimación. Sin embargo, puede haber algo más que argumentos estadísticos para conformar los estratos. También se plantea la interrogante de si una post-estratificación sería más indicada en ese caso.

Muestreo doble para estratificación

El muestreo doble para estratificación ya se mencionó en la Lección 2. Se trata de un diseño de muestreo en dos fases para estimar el tamaño de los estratos (que no pueden delimitarse o predefinirse fácilmente). Dado que las áreas de los estratos (y las ponderaciones) se estiman a partir de la muestra de la primera fase, el error de muestreo de la estimación de estas áreas debe tenerse en cuenta al estimar la media y la varianza de toda la población..

Enséñame las matemáticas

Estimación en el muestreo doble para estratificación

Suponiendo que las ponderaciones de los estratos se estimen a partir de la muestra de la primera fase (indicada con el apóstrofe) como:

$$w'_h = \frac{n'_h}{n'}$$

Y una media sin sesgo se puede estimar como:

$$\bar{y} = \sum_{h=1}^{L} w'_h \bar{y}_h$$

Ignorando la corrección por población finita y suponiendo que n' es grande, la varianza del error respectivo se estimaría como:

$$\hat{var}(\bar{y}) = \sum_{h=1}^{L} \left({w'_h}^2 * \frac{{s_h}^2}{n_h} + w'_h * \frac{(\bar{y}_h - \bar{y}')^2}{n'} \right)$$

Este estimador de la varianza es muy similar al estimador del muestreo aleatorio estratificado, excepto por el último término entre paréntesis: en este caso, se añade un componente de error debido a que los tamaños de los estratos sólo se estiman y no se conocen.

La herramienta ***Collect Earth*** (en inglés), que forma parte del sistema Open Foris de la FAO, es útil en este contexto y se ha aplicado en numerosas ocasiones. Se diseñó para hacer uso de las imágenes satelitales y aéreas disponibles y georreferenciadas de Google Earth, Bing y otros, para una interpretación visual de los sitios de muestreo o parcelas.

Con ayuda de esta herramienta, se puede visitar una gran cantidad de puntos y clasificarlos visualmente en diferentes estratos. Posteriormente, el tamaño del área de los estratos puede estimarse como proporción de puntos muestrales por estrato. Se puede extraer una varianza de esta estimación e incorporarla a la estimación como se muestra más arriba.

El estimador de razón-utilizando información cuantitativa auxiliar

Existen situaciones en el muestreo de inventarios forestales en las que se sabe (o se sospecha) que el valor de la variable objetivo está bien correlacionado con otra variable (denominada covariable o variable auxiliar).

Si esa variable auxiliar se puede observar en la parcela sin demasiado esfuerzo ni costos (por ejemplo, mediante un análisis por teledetección), será eficaz observarla también y utilizar la correlación con la variable objetivo para, con el tiempo, mejorar la precisión de la estimación de la variable objetivo. Aquí es donde se aplica el estimador de razón.

Nota

Imagine que se ha realizado una clasificación de imágenes satelitales para producir una **predicción continua del porcentaje** de cubierta de copas para un área de bosque con densidad de copas variable. Suponiendo una alta correlación con el volumen de la parcela o la biomasa, este sería un caso en el que se aplicaría el estimador de razón. En superficies forestales cerradas con una cubierta forestal completa en todos los lugares, esto no tendría ningún sentido, porque allí el porcentaje de cubierta de copas no variaría, sino que sería constantemente del 100 %, de modo que la correlación con la biomasa entre el **porcentaje de cubierta de copas** de la variable auxiliar y la **biomasa** de la variable objetivo sería cercana a cero. En lugar de estimar directamente la biomasa en pie por unidad de superficie a partir de las parcelas de campo, el estimador de razón utiliza un desvío: estimamos una razón, r, de las dos medias, que nos da el **porcentaje de biomasa/cubierta de copas**, y a continuación, utilizamos la cubierta de copas conocida para obtener una estimación de la biomasa. La biomasa media podría entonces estimarse como **r*Porcentaje medio de cubierta de copas**.

Otro caso típico para el estimador de razón es si una cierta proporción de parcelas grandes (o conglomerados) se inclinan más allá de los límites de la región del inventario y sólo están parcialmente dentro de la población objetivo. En ese caso, la superficie de la parcela dentro del bosque no es idéntica para todas las parcelas, y se puede aplicar el estimador de razón, con la superficie de la parcela como

variable auxiliar para tener esto en cuenta. De hecho, podemos suponer entonces que la superficie de la parcela estará altamente correlacionada con las variables de existencias (incluida el área basal, el volumen, la biomasa, el carbono y el número de árboles) registradas en la superficie de la parcela. Para estimar la precisión, necesitamos conocer el valor paramétrico (media) de la variable auxiliar (en este caso, el porcentaje medio de cubierta de copas sobre la superficie forestal total, o la superficie forestal total que se va a inventariar en el ejemplo del tamaño de la parcela).

Enséñame las matemáticas

Estimación con el estimador de razón

La razón paramétrica entre la variable objetivo y la variable auxiliar **x**

$$R = \frac{\mu_y}{\mu_x}$$

se estima sobre la base de la muestra de

$$r = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

La varianza estimada de esta razón estimada es:

$$var(r) = \frac{1}{n}\frac{1}{{\mu_x}^2}\frac{\sum_{i=1}^{n}(y_i - rx_i)^2}{n-1}$$

El total estimado se calcula a partir de:

$$\tau_y = r\tau_x$$

con una varianza del error asociada de:

$$var(\hat{\tau}_y) = {\tau_x}^2 \hat{var}(r)$$

Dada la razón estimada, r, la media de la variable objetivo podría estimarse como:

$$y_r = r\mu_x$$

Esta media estimada conlleva una varianza estimada de :

$$var(\bar{y}_r) = {\mu_x}^2 v\hat{a}r(r) = \frac{1}{n}\left\{{s_y}^2 + r^2{s_x}^2 - 2r\hat{\rho}s_x s_y\right\}$$

Enséñame las matemáticas

Diseño de la estimación con el estimador de regresión

Mientras que el estimador de razón modela la relación entre la variable objetivo y una variable auxiliar, el estimador de regresión utiliza un modelo de regresión con coeficiente de intercepción y de pendiente. Recuerde: en el estimador de razón, se supone que el intercepto es cero. La media se estima a partir del estimador de regresión de la siguiente manera:

$$y_L = \bar{y} + b\left(\mu_x - \bar{x}\right)$$

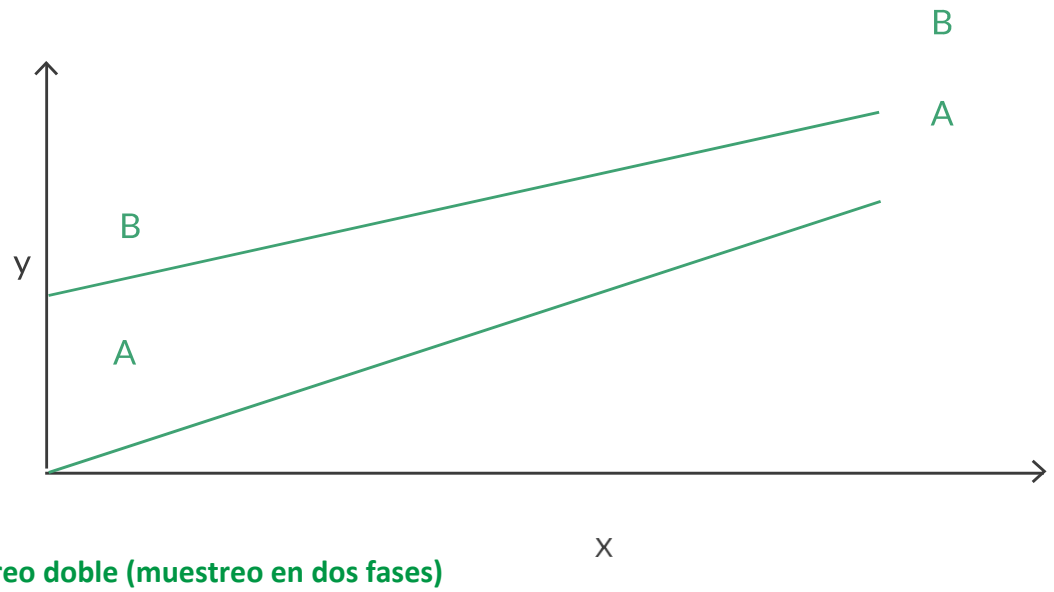
La varianza estimada de esta media estimada está dada con:

$$var(\bar{y}_L) = \frac{1}{n}\frac{1}{n-2}\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2 - b^2\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}$$

Estimador de razón vs. estimador de regresión

El estimador de razón utiliza una razón fija simple, lo que significa que la variable objetivo, y, será cero si la variable auxiliar, x, es cero. Sin embargo, hay situaciones en las que esto no es correcto. Imaginemos que podemos encontrar árboles pequeños en aquellas parcelas cuya cubierta de copas no se detectó en las imágenes de teledetección (por ejemplo, debido a la baja resolución espacial). En este caso, una línea de regresión con un coeficiente de intercepción que no esté forzado a ser cero (como con el estimador de razón) sería más apropiada; si, por ejemplo, el porcentaje de la cubierta de copas es cero, todavía podría haber una biomasa considerable en el suelo. En este caso, el estimador de regresión utiliza un modelo lineal simple.

En ambos casos, la estimación por razón o por regresión, la eficacia total depende de la correlación entre las variables objetivo y las auxiliares, que debe ser altamente positiva. A veces resulta que esta correlación es relativamente baja y que las expectativas eran demasiado altas, después de que, por ejemplo, se compraran imágenes de teledetección muy costosas



Muestreo doble (muestreo en dos fases)

Para el estimador de razón y regresión, es necesario conocer la media paramétrica o el total paramétrico de la variable auxiliar. Si esta información no está disponible, se pueden estimar estos valores a partir de una muestra.

En esto consiste exactamente el muestreo doble, también denominado muestreo en dos fases: en la

primera se estima la variable auxiliar, por lo general con una muestra grande de una variable que puede observarse con relativa facilidad y a bajo costo, y que se sabe que está altamente correlacionada positivamente con la variable objetivo.

A continuación, en la segunda fase de muestreo, se toma una muestra más pequeña de la variable objetivo, que suele ser una variable mucho más cara o mucho más difícil de observar. A continuación, se puede establecer una relación entre una variable objetivo y una variable auxiliar (ya sea una razón simple o una regresión, que sería un muestreo doble con el estimador de razón, o el estimador de regresión, respectivamente).

En este caso, cuanto más alta sea la correlación positiva con la variable auxiliar, menor será el tamaño de muestra necesario en la segunda fase, cuando se observe la variable objetivo más compleja/ costosa/difícil.

A continuación, abordamos las fases dependientes, en las que la muestra de la segunda fase es un subconjunto de la primera (y no una muestra seleccionada de forma independiente). Los estimadores presentados son exclusivamente para el MAS.

Enséñame las matemáticas

Estimación en el muestreo doble

Para el muestreo doble con el estimador de razón, la media de y puede estimarse como:

$$\bar{y}_{md.r} = \frac{\bar{y}}{\bar{x}}\bar{x}' = r\bar{x}'$$

Con una varianza estimada de la media estimada de:

$$\hat{var}(\bar{y}_{md.r}) = \frac{S_y^2 + r^2 S_x'^2 - 2rS_{xy}}{n} + \frac{2rS_{xy} - r^2 S_x'^2}{n'} - \frac{S_{\cdot}^{\cdot}}{N}$$

Y para el estimador de regresión, la media se estima como:

$$\bar{y}_{md.reg} = \bar{y} + b(\bar{x}' - \bar{x})$$

Con una varianza estimada de la media de:

$$\hat{var}(\bar{y}_{md.reg}) = \frac{S_y^2}{n}\left\{ 1 - \frac{n' - n}{n'}\hat{\rho}^2 \right\}$$

Donde ρ es el coeficiente de correlación estimado entre *x* e *y*

En ambos casos, la varianza del error del total se calcula, como es usual, como:

$$\hat{var}(\hat{\tau}) = N^2 \hat{var}(\bar{y}$$

La eficacia total del muestreo doble depende de la relación de costos entre la observación de las muestras de las fases 1 y 2 y de la correlación entre las dos variables. De hecho, nos esforzamos por explotar al máximo la variable auxiliar, para poder reducir el número de muestras (costosas) de la segunda fase. Cuanto mayor sea la correlación y más caras sean las observaciones en la segunda fase, menor será la muestra de la segunda fase.

¿Sabía qué?

Elegir entre estimadores alternativos

Dependiendo del diseño de muestreo aplicado, podrían aplicarse estimadores alternativos. Por ejemplo, una estimación se podría producir sólo a partir de muestras de campo, o considerar variables auxiliares adicionales. O se puede aplicar o no una post- estratificación a los datos. En estas situaciones, la producción de diferentes estimaciones válidas, con estimadores alternativos, debería dar como resultado la misma media, pero diferentes estimaciones de la precisión. En caso de que existan varios estimadores no sesgados, preferiríamos el que produzca el menor error estándar de las estimaciones.

Resumen

Antes de finalizar, aquí están los puntos clave de aprendizaje de esta lección.

- Para la **inferencia basada en el diseño** no hacemos supuestos sobre la estructura (espacial) de la población. Suponemos esta estructura como desconocida y nuestro objetivo es estimar las características de esta población fija.
- En la **inferencia basada en el modelo**, las observaciones de campo se utilizan para establecer una relación (modelo) con variables auxiliares que suelen ser índices teledetectados. A continuación, el modelo se utiliza para predecir la variable objetivo a partir de una cobertura completa de estos índices.
- En la **inferencia asistida por el modelo**, se utiliza un modelo como apoyo a la estimación basada en el diseño, situándose a medio camino entre la inferencia basada en el diseño y la inferencia

basada en el modelo.

- Cuando se utilizan conglomerados, un único elemento de muestreo consta de varios subelementos, que se seleccionan conjuntamente. Dado que las subparcelas en un conglomerado no se seleccionan independientemente unas de otras, el tamaño de la muestra se refiere al número de conglomerados seleccionados y no al número de subparcelas.
- La similitud de las observaciones dentro de un conglomerado se puede cuantificar mediante el Coeficiente de Correlación dentro del Conglomerado (ICC), a veces denominado Coeficiente de Correlación Intraclase.
- El muestreo estratificado no es un nuevo diseño de muestreo, sino un marco para integrar estimaciones basadas en muestras generadas independientemente a partir de distintos estratos en una estimación de la población total, es decir: es más bien una variación del diseño de estimación.
- En el muestreo doble para estratificación, los tamaños de los estratos no se determinan antes del muestreo, sino que se estiman en la primera fase de muestreo.