Criterios de precisión cartográfica

MIGUEL J. SEVILLA DE LERMA

INSTITUTO DE ASTRONOMÍA Y GEODESIA
FACULTAD DE CIENCIAS MATEMÁTICAS
UNIVERSIDAD COMPLUTENSE DE MADRID

Introducción

EN 1966, la Asociación Internacional de Cartografía dio la siguiente definición de Cartografía: "Conjunto de estudios y de operaciones científicas, artísticas y técnicas que, a partir de los resultados de observaciones directas o de la explotación de una documentación intervienen en la elaboración de cartas, planos y otros medios de expresión, así como en su utilización". Esta definición fue posteriormente adoptada por la UNESCO.

En el proceso cartográfico suelen distinguirse tres fases bien diferenciadas: *a*) *concepción*, constituida por el estudio teórico de las leyes, principios y sistemas de representación; también se llama *cartografía matemática*; *b*) *producción*, constituida por la selección de datos, escala y materialización de sistemas de proyección, incluye la cartografía automática numérica y la cartografía informatizada; está íntimamente relacionada con la geodesia que le proporciona la red de apoyo, fundamental para su precisión, y con la topografía y fotogrametría que le facilitan la obtención de datos sobre el terreno; y *c*) *utilización*, que no es otra cosa que la forma de facilitar la comunicación y uso de la información contenida en una carta o mapa y que conecta con todas aquellas actividades científicas o de otro tipo que en algún momento han de servirse de representaciones de la superficie terrestre.

El producto final de la cartografía es la *carta* o *mapa* para el cual podemos tomar la siguiente definición: "Representación reducida, generalizada y matemáticamente precisa de la superficie terrestre o de una parte de ella sobre un plano o un soporte informático, que muestra la situación, distribución y relaciones de elementos geométricos y de los diversos fenómenos naturales y sociales, escogidos y definidos en función del objeto de la carta o mapa. La carta permite igualmente mostrar las variaciones y los desarrollos de los fenómenos en el tiempo, así como sus factores de movimiento y de desplazamiento en el espacio".

A lo largo de la historia, las cartas se han utilizado, y en su principio casi exclusivamente, para materializar itinerarios terrestres o marítimos, o posiciones de estrellas en el firmamento; después también se han usado con fines geográficos y geodésicos. La aparición de necesidades de otro tipo: científicas, militares, demográficas, catastrales, etc., ha hecho que se hayan adaptado y construido cartas o mapas especiales adecuados a cada utilidad. Estos trabajos han permitido un desarrollo, sobre todo con el concurso de la informática, y una justificación crítica de los métodos utilizados y de la cartografía misma.

El contenido de un mapa y su modo de representación son función del objeto asignado al mismo y de las necesidades que debe cubrir. La representación plana de la superficie terrestre para fines en los que la geometría de la representación sea parte importante debe efectuarse en sistemas de proyección tales que con ellos se facilite el empleo de coordenadas rectangulares para la designación de puntos.

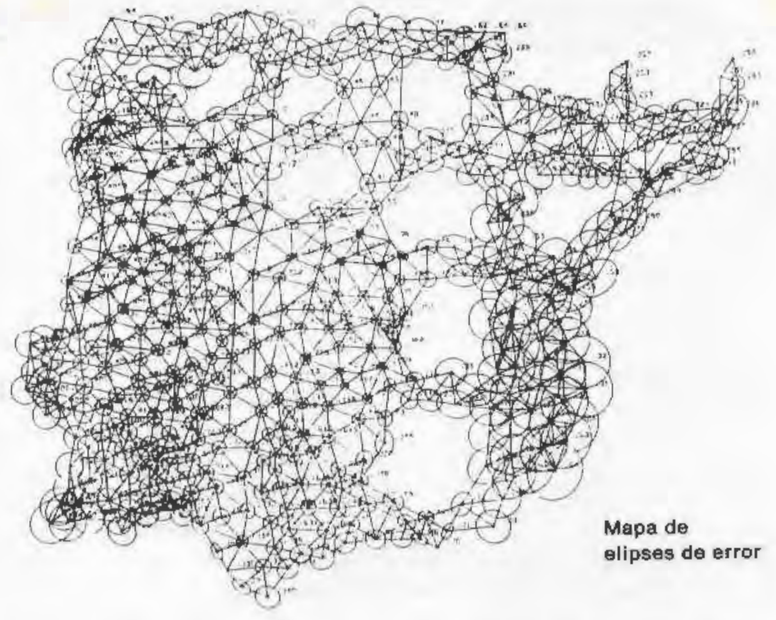
Una constante preocupación de diversas organizaciones nacionales e internacionales en el campo de la cartografía, topografía y fotogrametría es el establecimiento de unas normas claras que permitan definir y controlar la exactitud o precisión espacial de un mapa topográfico y en especial de un mapa de gran escala (1:10.000 o mayor). Asimismo, existe interés por establecer los procedimientos para evaluar, comprobar o controlar esta precisión. Este interés no es sólo de estos organismos; también lo es de los productores de mapas que deben cumplir unas normas de precisión y de los usuarios de los mapas que necesitan confiar en las precisiones que se dan.

El control de la calidad de un mapa en cuanto a su precisión o exactitud tiene por objeto garantizar que la precisión ofertada por el mapa se cumple en la realidad. La forma más simple de realizar este control es contrastar las medidas proporcionadas por el mapa en una escala conveniente con los correspondientes valores obtenidos por observación directa en el terreno.

La idea fundamental consiste en asignar a cada tipo de mapa un parámetro de precisión o calidad según el objeto para el que ha sido confeccionado. Los valores de este parámetro deben indicar los límites de las posibles diferencias existentes entre la geometría del mapa y la realidad del terreno que representa. Se dice que un mapa cumple los requisitos de precisión cuando las diferencias entre los valores de ciertas magnitudes geométricas, obtenidos a partir de la información del mapa y de los correspondientes valores reales obtenidos sobre el terreno, no sobrepasen un límite fijado por el parámetro de precisión. En caso contrario el mapa deberá calificarse de deficiente.

Criterios de precisión y definición de parámetros

Los criterios de precisión de un mapa deben establecerse en términos de conceptos de error entendibles y cuantificables por los usuarios y por los productores de mapas. Asimismo, deben ir acompañados de un claro procedimiento de verificación del producto final, y todo esto ha de estar basado en teorías matemáticas precisas. Un amplio consenso entre todas las parte implicadas es necesario para que los criterios o normas que se establezcan y los métodos de verificación sean asumidos y aplicados sin restricciones. En un proceso de establecimiento de normas es aconsejable crear grupos de trabajo compuestos por personas interesadas, tanto usuarios como productores, asesorados por personal científico cualificado; después, los documentos elaborados por estos grupos deben someterse a información y discusión todo lo amplia que sea posible y con las sugerencias recibidas se redactarán las normas finales de modo que obtengan el consenso antes mencionado.


Mapa de elipses de error

Normalmente, los criterios de precisión se establecen en el sistema de referencia y *datum* del terreno, es decir, en escala real. Un buen parámetro en función del cual pueden establecerse los criterios o normas de calidad es el error medio cuadrático de la magnitud considerada, como especificaremos más adelante, que tendrá en cuenta todos los errores cometidos en el proceso cartográfico.

Para mapas topográficos de gran escala, los criterios de calidad se referirán fundamentalmente a los errores en las coordenadas planimétricas X e Y y en la coordenada altimétrica Z, pues estos datos siempre son conocidos y muy utilizados en aplicaciones de ingeniería y catastrales. Además, estos datos pueden deducirse fácilmente a partir de la información dada por el mapa, y esto puede hacerse, también fácilmente, por cualquier usuario. Los criterios sobre errores en otras magnitudes tales como distancias, áreas y volúmenes pueden establecerse a partir de los errores en coordenadas, aunque esto no lo trataremos aquí.

Supongamos que se ha elegido un conjunto de n puntos en el mapa, que se han extraído sus coordenadas; para fijar ideas tomemos la coordenada X y que estas coordenadas se han transformado al sistema y *datum* del terreno, así tenemos un conjunto de coordenadas que representamos por el vector $\underline{X}^m$,

$$\underline{X} = (X_1^m, X_2^m, \dots, X_n^m)^T,$$

el superíndice ^{T} indica vector traspuesto, es decir, $\underline{X}^m$ es un vector columna; estas coordenadas las llamamos *coordenadas mapa* y por eso les ponemos el superíndice ^{m}.

Por otra parte, estos puntos seleccionados, que han sido perfectamente identificados y señalizados en el terreno, constituyen la *red de control*. Como resultado del tratamiento de esta red, vamos a disponer de otro conjunto de coordenadas de los mismos puntos en el sistema y *datum* del terreno; las llamamos *coordenadas terreno* y las representamos por el vector $\underline{X}^t$,

$$\underline{X} = (X_1^t, X_2^t, \dots, X_n^t)^T,$$

para indicar coordenadas terreno hemos puesto ahora el superíndice ^{t}.

Con las coordenadas de los mismos puntos, provenientes del mapa y de observación, definimos las *discrepancias* por medio del vector diferencia

$$\Delta\underline{X} = \underline{X}^m - \underline{X}^t,$$

que también podemos escribir una a una por

$$\Delta X_i = X_i^m - X_i^t, \quad i = 1, 2, \dots, n.$$

donde n es el número de puntos considerados.

El valor medio de las diferencias, o *media muestral*, viene dado por

$$\Delta\overline{X} = \frac{1}{n} \sum_{i=1}^{n} \Delta X_i$$

y la *desviación típica* (error medio cuadrático) por

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (\Delta X_i - \overline{X})^2}.$$

Estos valores de la media y del error medio cuadrático de las discrepancias son los que utilizaremos para establecer el control de calidad de un mapa.

Precisiones planimétricas y altimétricas

La *precisión planimétrica* de un mapa queda definida por un número que indica el error medio cuadrático de las coordenadas horizontales (X, Y) y que se interpreta como el error medio cuadrático de las diferencias entre las coordenadas horizontales de un punto (X^m, Y^m) en el sistema del terreno obtenidas a partir de datos del mapa y las coordenadas (X^t, Y^t) del mismo punto en el terreno obtenidas por levantamiento de precisión. Se supone que los errores en las coordenadas X e Y son estadísticamente independientes e iguales, por lo que un solo valor del error medio cuadrático sirve para ambas coordenadas. Este valor lo designamos por σ_0.

El control planimétrico consiste en comprobar que este límite de precisión es efectivamente alcanzado en la realidad. Esto quiere decir que si de un mapa aceptado con una precisión σ_0 obtenemos las coordenadas X o Y de un punto y las transformamos al sistema del terreno, las diferencias de las coordenadas verdaderas del punto están estadísticamente por debajo de σ_0 con una probabilidad dada de antemano (en general el 68,3 por 100 están por debajo de σ_0 y el 99,7 por 100 por debajo de $3\sigma_0$, aunque estos límites se especificarán en cada caso concreto).

El límite σ_0 debe estar relacionado, entre otras cosas, con la escala del mapa; por ejemplo, unos valores de σ_0 que suelen utilizarse como referencia son los dados en la Tabla 1.

Esta Tabla 1 corresponde aproximadamente a decir que el 90 por 100 de los puntos del mapa están a menos de 0,43 milímetros de su posición correcta (en el mapa).

La precisión altimétrica se establece de forma análoga a la anterior, pero para las altitudes obtenidas a partir del mapa (coordenada Z). Ahora bien, esta precisión hay que referirla a las curvas de nivel entre las cuales se interpolan las altitudes de los puntos del mapa. Entonces, la precisión altimétrica se define como el error medio cuadrático de la coordenada Z (con las discrepancias $Z^m - Z^t$). Este límite estará en relación con la separación de las curvas de nivel; suele utilizarse un valor que es aproximadamente igual a un tercio de dicha separación. Un ejemplo de valores propuestos se da en la Tabla 2.

Si los puntos corresponden a señales de nivelación, esta precisión debe reducirse a la mitad.

Los límites de error establecidos aquí lo son para mapas de calidad superior, no obstante, en algunas aplicaciones basta con una calidad inferior; en estos casos, los límites dados deberán aumentarse multiplicándolos por unos factores de relajación de acuerdo con la precisión final deseada. También puede haber mapas de precisión mixta, por ejemplo, alta en planimetría y baja en altimetría o viceversa.

TABLA 1. Precisión planimétrica.

ESCALA DEL MAPA	DESVIACION TIPICA
1:10.000	2,500 metros
1:5.000	1,250 metros
1:2.000	0,500 metros
1:1.000	0,250 metros
1:500	0,125 metros
1:200	0,050 metros
1:100	0,025 metros

TABLA 2. Precisión altimétrica.

INTERVALO ENTRE CURVAS DE NIVEL	DESVIACION TIPICA
10 metros	3,04 metros
5 metros	1,52 metros
2 metros	0,61 metros
1 metros	0,30 metros
0,5 metros	0,15 metros

Red de control

La red de control está constituida por el conjunto de puntos del terreno seleccionados para realizar el control. La aplicación de técnicas y métodos geodésicos o topográficos permite la asignación de coordenadas a todos ellos y de sus correspondientes medidas de la precisión dadas generalmente por las matrices de varianza-covarianza o por las correspondientes elipses de error, en el caso plano.

En la adquisición de datos en el terreno deben aplicarse métodos de medida precisos y realizar las observaciones con la máxima garantía y fiabilidad. Deben prepararse cuidadosamente las observaciones, revisarse los equipos e instrumentos auxiliares, que suponemos están debidamente calibrados, y deben tomarse todas aquellos datos e informaciones que sean necesarios, no sólo para controlar las medidas sino también para su adecuada corrección e interpretación. Efectuadas las observaciones debe procederse al análisis de las observaciones brutas con el fin de detectar errores que pudieran haberse cometido, por ejemplo errores de centrado, de puntería, de lectura, errores en los datos meteorológicos, errores de alineamiento, variaciones de frecuencias, etc.; si se ha tenido en cuenta durante la observación muchos de estos errores se habrán minimizado.

La red de control (triangulación, trilateración o poligonales) debe establecerse en el mismo *datum* en el que está basado el mapa y ser debidamente compensada de acuerdo con el modelo matemático establecido (unidimensional, bidimensional o tridimensional). Los resultados compensados de los puntos (coordenadas compensadas) junto con sus correspondientes precisiones (elipses de error) son los que se utilizan como datos de terreno (vectores $\underline{X}^T$) para el control de calidad en precisión del mapa en cuestión.

Los errores resultantes del levantamiento y ajuste de la red de control deben ser menores (por ejemplo un tercio) que los permitidos en las normas de precisión del mapa correspondiente.

Contraste de hipótesis

El contraste de hipótesis estadísticas, o *test* de hipótesis, es un algoritmo establecido a partir de una muestra de valores aleatorios y de una hipótesis sobre su comportamiento, ordinariamente sobre la función de densidad de dichas variables aleatorias y sobre sus parámetros, que conduce a la decisión de aceptar o rechazar la hipótesis considerada. Toda hipótesis establecida, llamada *hipótesis nula*, tiene una (o infinitas) *hipótesis alternativa*. Si la hipótesis nula es aceptada, entonces se rechaza la hipótesis alternativa y viceversa. Por ejemplo, una hipótesis nula consistirá en que los parámetros poblacionales tienen unos valores particulares, cualquier otro valor distinto de los postulados constituirá una hipótesis alternativa.

Toda hipótesis tiene un estadístico asociado con ella. Un *estadístico* es una función de una o más variables aleatorias que no depende de ningún parámetro poblacional desconocido. Los valores que este estadístico tome nos conducirán a decidir sobre la aceptación o rechazo de la hipótesis con cuyos valores se ha calculado. Puede suceder que ninguna de las dos hipótesis, nula o alternativa, sea cierta; en este caso el test mostrará, por lo menos, cuál de las dos hipótesis es mejor.

Se llama *nivel de significación* de un test a la probabilidad α de rechazar la hipótesis nula cuando en realidad es cierta. Este valor debe seleccionarse antes de aplicar el test y naturalmente debe ser tan pequeña como sea posible. Valores ordinarios de α son 0,01 y 0,05. La probabilidad complementaria $1-\alpha$ es la medida de la confianza que se tiene en la decisión y se denomina *nivel de confianza*.

Seguidamente introduciremos los test que se aplicarán para el control estadístico de la precisión de un mapa, según todo lo expuesto en los apartados anteriores.

Test de la media de una población normal

En general, los test de la media se aplican para averiguar si la media muestral obtenida a partir de una serie de observaciones, que se sabe provienen de una población normal, es compatible con la media poblacional supuesta; de esta forma podrá estudiarse la presencia de posibles sistematismos. Consideraremos series de observaciones en las que todas las medidas sean de igual confianza y distinguiremos dos casos según que la varianza poblacional sea conocida o desconocida.

- *Primer caso:* Varianza poblacional σ^2 conocida.

Consideremos una serie de n observaciones $l_1, l_2, \ldots, l_n$, estadísticamente independientes, tales que todas ellas sean normales

$$l_i \approx \mathbb{N}(\mu, \sigma^2),$$

es decir, suponemos que provienen de una población normal de media μ e igual varianza σ^2 conocida. La hipótesis nula de este test de la media puede enunciarse como sigue

$$\mathbb{H}_0: \quad \mu = \mu_0,$$

donde μ_0 es un valor fijo de la media. La hipótesis alternativa es

$$\mathbb{H}_1: \quad \mu \neq \mu_0.$$

Sea

$$\bar{l} = \frac{1}{n}\sum_{i=1}^{n} l_i$$

la media muestral que sabemos se distribuye según una ley normal de media μ y varianza σ^2/n, esto es

$$\bar{l} \approx \mathbb{N}(\mu, \sigma_2/n),$$

y definamos la media tipificada

$$\bar{l}^* = \frac{\bar{l}-\mu}{\sigma/\sqrt{n}}$$

que también sabemos se distribuye normalmente con media 0 y varianza 1,

$$\bar{l}^* \approx \mathbb{N}(0, 1).$$

En estas condiciones, para contrastar el valor μ_0, es decir para averiguar si estadísticamente $\bar{l} = \mu_0$, con los valores de la muestra $(\bar{l})$ y de la hipótesis nula $\mathbb{H}_0$ (μ_0), definimos el estadístico

$$y = \frac{\bar{l}-\mu_0}{\sigma/\sqrt{\mu}} = (\bar{l}-\mu_0)\frac{\sqrt{n}}{\sigma}.$$

Este estadístico debería distribuirse como $\mathbb{N}(0, 1)$. Entonces, fijado un nivel de significación α, se rechaza la hipótesis nula $\mathbb{H}_0$ si

$$|y| > \lambda_{\alpha/2},$$

donde $\lambda_{\alpha/2}$ viene dado por la ley normal, es decir,

$$\Phi_N = \int_{-\infty}^{\lambda_{\alpha/2}} \mathbb{N}(0, 1)\, d\lambda = 1 - \frac{\alpha}{2};$$

en la práctica este valor suele tomarse de las tablas habituales.

Intervalo de confianza. Con el valor de $\lambda_{\alpha/2}$ podemos definir el intervalo de confianza para μ escribiendo la probabilidad

$$\Pr\left(-\lambda_{\alpha/2} < (\bar{l}-\mu)\frac{\sqrt{n}}{\sigma} < \lambda_{\alpha/2}\right) = 1-\alpha = 2\Phi_N - 1,$$

por la simetría de la función densidad de probabilidad $\mathbb{N}(0, 1)$. Entonces,

$$\Pr\left(-\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{l}-\mu < \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha,$$

$$\Pr\left(-\lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{l} < -\mu < \right.$$

$$\left. < \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{l}\right) = 1 - \alpha,$$

en definitiva

$$\Pr\left(\bar{l} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \right.$$

$$\left. < \bar{l} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

y queda establecido el intervalo de confianza al $(1 - \alpha) \times 100\%$,

$$\left(\bar{l} - \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}},\quad \bar{l} + \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

La tabla de la distribución normal nos da los valores ordinariamente utilizados siguientes:

$\lambda_{\alpha/2}$	$1 - \alpha$	$1 - \alpha/2$
1,96	0,950	0,9750
2,57	0,990	0,9949
3,29	0,999	0,9995

• *Segundo caso:* Varianza poblacional σ^2 desconocida.

Supongamos ahora la misma serie de observaciones l_i, $i = 1, ..., n$, pero con la varianza σ^2 desconocida. $\bar{l}$ sigue siendo la media muestral y sea

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(l_i - \bar{l})^2$$

la varianza muestral. Como en el caso anterior se tiene

$$\bar{l} \approx \mathbb{N}(\mu, \sigma^2/n),$$

$$\bar{l}^* \approx \mathbb{N}(0, 1),$$

pero el estadístico $\mathcal{Y}$ no puede calcularse al desconocerse σ.

Definimos el nuevo estadístico

$$\mathcal{T} = \frac{\bar{l} - \mu}{s/\sqrt{n}};$$

este estadístico se distribuye como una t de Student con $n - 1$ grados de libertad, esto es

$$\mathcal{T} \approx t_{n-1}.$$

En efecto, sabemos que $\bar{l}^*$ se distribuye según la ley normal tipificada, pero no nos sirve por figurar σ, que es desconocida, en su expresión. La teoría de distribuciones estadísticas muestra que el estadístico definido por

$$\mathcal{Y} = (n-1)\, s^2/\sigma^2$$

se distribuye según una chi cuadrado con $n - 1$ grados de libertad, entonces el cociente

$$\frac{\bar{l}^*}{\sqrt{\mathcal{Y}/(n-1)}} = \frac{\bar{l} - \mu}{s}\sqrt{n} \approx t_{n-1},$$

que es el nuevo estadístico $\mathcal{T}$, se distribuye como se indica, por definición de la distribución t de Student.

La hipótesis nula en este caso sigue siendo $\mathbb{H}_0$: $\mu = \mu_0$, y la hipótesis alternativa $\mathbb{N}_1$: $\mu \neq \mu_0$.

Se calcula $\mathcal{T}$ con los valores muestrales $(\bar{l}, s^2)$ y con el parámetro de la hipótesis $\mathbb{H}_0$ (μ_0), y resulta

$$\mathcal{T} = \frac{\bar{l} - \mu_0}{s}\sqrt{n}$$

y fijando un nivel de significación α se rechaza la hipótesis nula $\mathbb{H}_0$ si

$$|\mathcal{T}| > t_{n-1,\alpha/2},$$

donde,

$$\Phi = \int_{-\infty}^{t_{n-1,\alpha/2}} t_{n-1}\, dt = 1 - \frac{\alpha}{2}.$$

Intervalo de confianza. Con el valor $t_{n-1,\alpha/2}$ podemos definir la probabilidad

$$Pr\left(-t_{n-1,\alpha/2} < \frac{\bar{l} - \mu}{s}\sqrt{n} < \right.$$

$$\left. < t_{n-1,\alpha/2}\right) = 1 - \alpha = 2\Phi - 1,$$

pues t_{n-1} es simétrica. Entonces,

$$Pr\left(\bar{l} - \frac{2}{\sqrt{n}}t_{n-1,\alpha/2} < \mu < \right.$$

$$\left. < \bar{l} + \frac{s}{\sqrt{n}}t_{n-1,\alpha/2}\right) = 1 - \alpha$$

y queda establecido el intervalo de confianza al $(1 - \alpha) \times 100\%$ siguiente

$$\left(\bar{l} - \frac{s}{\sqrt{n}}t_{n-1,\alpha/2}, \bar{l} + \frac{s}{\sqrt{n}}t_{n-1,\alpha/2}\right).$$

Los valores de $t_{n-1,\alpha/2}$ se obtienen de las tablas de la t de Student.

Si estos test sobre la media rechazan, en cada caso, la hipótesis nula puede pensarse en que la serie de observaciones no sigue la ley normal, en que $\bar{l}$ y μ_0 sean incompatibles, en que existan errores constantes o sistemáticos en las observaciones o en que dichas observaciones sean estadísticamente dependientes. Habrá que revisar cuidadosamente las observaciones y en último extremo repetir todo el proceso de medida.

Test de una cola. En algunos casos, en vez de utilizar test de dos colas, como los anteriores, interesan test de una cola, que sirven para contrastar no la igualdad del parámetro en cuestión (la media en este caso) sino una desigualdad. Entonces la hipótesis nula y la hipótesis alternativa quedan establecidas, en cada caso, de la forma siguiente:

1) Para σ^2 conocida

1.a) Hipótesis nula

$$\mathbb{H}_0:\ \mu \geqslant \mu_0$$

Hipótesis alternativa

$$\mathbb{H}_1:\ \mu < \mu_0$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$\mathcal{Y} = (\bar{l} - \mu_0)\frac{\sqrt{n}}{\sigma} < -\lambda_\alpha$$

1.b) Hipótesis nula

$$\mathbb{H}_0:\ \mu \leqslant \mu_0$$

Hipótesis alternativa

$$\mathbb{H}_1:\ \mu > \mu_0$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$y = (\bar{l} - \mu_0)\,\frac{\sqrt{n}}{\sigma} > \lambda_\alpha$$

2) Para σ^2 desconocida

2.a) Hipótesis nula

$$\mathbb{H}_0:\quad \mu \geqslant \mu_0$$

Hipótesis alternativa

$$\mathbb{H}_1:\quad \mu < \mu_0$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$y = (\bar{l} - \mu_0)\,\frac{\sqrt{n}}{\sigma} < t_{n-1,\,\alpha}$$

2.b) Hipótesis nula

$$\mathbb{H}_0:\quad \mu \leqslant \mu_0$$

Hipótesis alternativa

$$\mathbb{H}_1:\quad \mu > \mu_0$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$y = (\bar{l} - \mu_0)\,\frac{\sqrt{n}}{\sigma} > t_{n-1,\,\alpha}$$

Test de la varianza de una población normal

Los test de la varianza se aplican para averiguar si la varianza muestral s^2 obtenida a partir de una serie de observaciones, que se sabe provienen de una población normal, es compatible con la varianza poblacional supuesta σ_0^2; de esta manera podrá saberse si la serie de observaciones ha sido efectuada con la precisión global requerida para que los resultados finales tengan las garantías previstas. Estudiaremos series de igual confianza para todas las observaciones y, como antes, distinguiremos dos casos según que la media μ sea conocida o desconocida.

• *Primer caso:* media poblacional μ conocida.

Consideremos una serie de n observaciones $l_1, l_2, \ldots, l_n$, estadísticamente independientes, tales que todas ellas sean normales

$$l_i \approx \mathbb{N}\,(\mu, \sigma^2),$$

es decir, suponemos que provienen de una población normal de media μ conocida y varianza σ^2. La hipótesis nula de este test de la varianza puede enunciarse como sigue

$$\mathbb{H}_0:\quad \sigma^2 = \sigma_0^2,$$

donde σ_0^2 es un valor fijo de la varianza. La hipótesis alternativa es

$$\mathbb{H}_1:\quad \sigma^2 \neq \sigma_0^2.$$

Sea

$$s_\mu^2 = \frac{1}{n}\sum_{i=1}^{n}(l_i - \mu)^2$$

la varianza muestral teórica calculada con la media μ conocida. En estadística se demuestra que la relación

$$\sum_{i=1}^{n}\left(\frac{l_i - \mu}{\sigma}\right)^2 = \frac{ns_\mu^2}{\sigma^2} \approx \chi_n^2$$

se distribuye según una ley chi cuadrado con n grados de libertad.

Entonces, para contrastar el valor σ_0^2, es decir para averiguar si estadísticamente $\sigma_0^2 = s_\mu^2$, con los valores de la muestra (s_μ^2) y de la hipótesis nula $\mathbb{H}_0$ (σ_0^2), definimos el estadístico

$$c = \sum_{i=1}^{n}\left(\frac{l_i - \mu}{\sigma_0}\right)^2 = \frac{ns_\mu^2}{\sigma_0^2}$$

Este estadístico debería distribuirse como χ_n^0. Entonces, fijado un nivel de significación α, se rechaza la hipótesis nula $\mathbb{H}_0$ si

$$c < \chi^2_{n,\,1-\alpha/2} \qquad \text{o} \qquad c > \chi^2_{n,\,\alpha/2},$$

donde

$$\int_0^{\chi^2_{n,\,1-\alpha/2}} \chi_n^2(x)\,dx = \frac{\alpha}{2}$$

y

$$\int_0^{\chi^2_{n,\,\alpha/2}} \chi_n^2(x)\,dx = 1 - \frac{\alpha}{2};$$

en la práctica estos valores suelen tomarse de las tablas habituales. Obsérvese que esta función densidad no es simétrica.

Intervalo de confianza. Con los valores de $\chi^2_{n,\,1-\alpha/2}$ y $\chi^2_{n,\,\alpha/2}$, al no ser simétrica la función de densidad de χ^2, tenemos

$$\Pr\left(\chi^2_{n,\,1-\alpha/2} < \frac{ns_\mu^2}{\sigma^2} < \chi^2_{n,\,\sigma/2}\right) = 1 - \alpha,$$

entonces,

$$\Pr\left(\chi^2_{n,\,1-\alpha/2}\,\frac{1}{ns_\mu^2} < \frac{1}{\sigma^2} < \chi^2_{n,\,\alpha/2}\,\frac{1}{ns_\mu^2}\right) = 1 - \alpha,$$

$$\Pr\left(\frac{ns_\mu^2}{\mu^2_{n,\,1-\alpha/2}} > \sigma^2 > \frac{ns_\mu^2}{\mu^2_{n,\,\alpha/2}}\right) = 1 - \alpha,$$

$$\Pr\left(\frac{ns_\mu^2}{\chi^2_{n,\,\alpha/2}} > \sigma^2 > \frac{ns_\mu^2}{\chi^2_{n,\,1-\alpha/2}}\right) = 1 - \alpha,$$

y queda estaclecido el intervalo de confianza al $(1 - \alpha) \times 100\%$

$$\left(\frac{ns_\mu^2}{\chi^2_{n,\,\alpha/2}},\ \frac{ns_\mu^2}{\chi^2_{n,\,1-\alpha/2}}\right).$$

• *Segundo caso:* media poblacional μ desconocida.

Supongamos ahora la misma serie de observaciones l_i, $i = 1, \ldots, n$, pero con la media μ desconocida. Entonces, sustituimos la varianza s_μ^2 por la varianza muestral s^2 calculada con la media muestral $\bar{l}$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(l_i - \bar{l})^2.$$

En este caso interesa utilizar el estadístico y antes definido.

La hipótesis nula en este caso sigue siendo $\mathbb{H}_0$: $\sigma^2 = \sigma_0^2$, y la hipótesis alternativa $\mathbb{H}_1$: $\sigma^2 \neq \sigma_0^2$. Se calcula y con los valores muestrales ($\bar{l}$,

s^2) y con el parámetro de la hipótesis $\mathbb{H}_0$ (σ_0^2), y resulta

$$\wp = \frac{(n-1)\,s^2}{s_0^2} = \sum_{i=1}^{n} \left(\frac{l_i - \bar{l}}{\sigma_0} \right)^2,$$

que debería distribuirse como una chi cuadrado con $n-1$ grados de libertad (obsérvese que se ha perdido un grado de libertad al utilizar s^2). Entonces, fijado un nivel de significación α, se rechaza la hipótesis nula $\mathbb{H}_0$ si

$$\wp < \chi^2_{n-1,\, -\alpha/2} \quad \text{ó} \quad \wp > \chi^2_{n-,\, \alpha/2}$$

Estos valores se obtienen de las tablas de la función χ^2, es decir, son tales que

$$\int_0^{\chi^2_{n-1,\,1-\alpha/2}} \chi^2_{n-1}(x)\, dx = \frac{\alpha}{2}$$

y

$$\int_0^{\chi^2_{n-1,\,\alpha/2}} \chi^2_{n-1}(x)\, dx = 1 - \frac{\alpha}{2}.$$

Intervalo de confianza. De forma análoga al caso anterior, se obtiene el intervalo de confianza al $(1 - \alpha) \times 100\%$ siguiente

$$\left(\frac{(n-1)\,s^2}{\chi^2_{n-1,\,\alpha/2}},\ \frac{(n-1)\,s^2}{\chi^2_{n-1,\,1-\alpha/2}} \right).$$

Si estos test sobre la varianza rechazan, en cada caso, la hipótesis nula puede pensarse en que la serie de observaciones no sigue la ley normal, en que s^2 y σ^2 sean incompatibles, o en que dichas observaciones sean estadísticamente dependientes. Habrá que revisar cuidadosamente las observaciones y en último extremo repetir todo el proceso de medida.

Test de una cola. Cuando interese contrastar desigualdades deberán aplicarse test de una cola de la siguiente forma:

1) Para μ conocida
   1.a) Hipótesis nula

$$\mathbb{H}_0:\ \sigma^2 \geqslant \sigma_0^2$$

Hipótesis alternativa

$$\mathbb{H}_1:\ \sigma^2 < \sigma_0^2$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$\wp = \frac{ns^2}{\sigma_0^2} < \chi^2_{n,\,1-\alpha}$$

1.b) Hipótesis nula

$$\mathbb{H}_0:\ \sigma^2 \leqslant \sigma_0^2$$

Hipótesis alternativa

$$\mathbb{H}_1:\ \sigma^2 > \sigma_0^2$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$\wp = \frac{ns^2}{\sigma_0^2} < \chi^2_{n,\,\alpha}$$

2) Para μ conocida
   2.a) Hipótesis nula

$$\mathbb{H}_0:\ \sigma^2 \geqslant \sigma_0^2$$

Hipótesis alternativa

$$\mathbb{H}_1:\ \sigma^2 < \sigma_0^2$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$\wp = \frac{(n-1)\,s_\mu^2}{\sigma_0^2} < \chi^2_{n-1,\,1-\alpha}$$

2.b) Hipótesis nula

$$\mathbb{H}_0:\ \sigma^2 \leqslant \sigma_0^2$$

Hipótesis alternativa

$$\mathbb{H}_1:\ \sigma^2 > \sigma_0^2$$

Se rechaza $\mathbb{H}_0$ y se acepta $\mathbb{H}_1$ si

$$\wp = \frac{(n-1)\,s_\mu^2}{\sigma_0^2} > \chi^2_{n-1,\,\alpha}$$

Los test estadísticos que se acaban de definir en general, serán aplicados en el apartado siguiente al control de la calidad del mapa.

Control de calidad de un mapa

Decimos que un mapa pasa el control de calidad cuando los errores planimétricos y/o altimétricos no exceden los límites de error permitidos (σ_0) establecidos por las correspondientes normas dadas en el apartado "Precisiones planimétricas y altimétricas". Si esto no ocurre el mapa debe ser rechazado.

Todo método directo de contrastar la calidad de un mapa por comparación de magnitudes obtenidas del mapa y del terreno estará basado en un muestreo de la zona representada, es decir, se deberá elegir un conjunto depuntos del terreno debidamente representados en el mapa. Estos puntos se denominarán *puntos de control*, y su definición y establecimiento es de extrema importancia por la trascendencia que tendrán después. Evidentemente, el contraste debe realizarse por métodos estadísticos, unidos que permiten obtener conclusiones con un nivel de confianza previamente determinado.

El control se hace de forma independiente para las coordenadas X, Y y Z, para X e Y se suponen los mismos límites de error (Tabla 1), por eso sólo hablaremos de la coordenada X y el proceso de repite para las coordenadas Y y Z (Tabla 2). Para efectuar este control se elige en el mapa un conjunto de 20 o 25 puntos bien definidos, identificados en el terreno y uniformemente distribuidos por toda la hoja. Se calculan o determinan las coordenadas mapa $\underline{X}^m$ y las coordenadas terreno $\underline{X}^t$ de esta muestra de puntos y se calculan las discrepancias $\Delta\underline{X}$, y los parámetros media $\Delta\overline{X}$ y desviación típica de las discrepancias s.

Los valores de las discrepancias ΔX_1 constituyen los valores de la serie de observaciones (l_i) y los valores de la media $\Delta\overline{X}$ y varianza s^2 son los parámetros que utilizaremos para realizar el contraste por aplicación de los test expuestos en la Sección anterior.

En todos los casos hemos supuesto que la serie de observaciones está normalmente distribuida.

En primer lugar hemos de estu-

Observación de la Red Geodésica (Láser).

diar la presencia de sistematismos en la serie de discrepancias. Esto se hace averiguando si la media de la disstribución de las observaciones es cero, que es lo que debe ocurrir si una población normal y contrastar la hipótesis nula

$$\mathbb{H}_0: \quad \mu = \mu_0,$$

frente a la hipótesis alternativa

$$\mathbb{H}_1: \quad \mu \neq \mu_0.$$

Si suponemos que la varianza poblacional es conocida (por ejemplo tomando σ_0), el estadístico que hay que usar es

$$y = \Delta \overline{X} \frac{\sqrt{n}}{\sigma^0}.$$

y fijado un nivel de significación $\alpha = 0{,}05$, se rechaza la hipótesis $\mathbb{H}_0$ si

$$|y| > 1{,}96.$$

Entonces, se admite la hipótesis alternativa y concluimos que las observaciones presentan sistematismo o sesgo; en tal caso hay que revisar el proceso de obtención de las discrepancias para averiguar a qué puede ser debido y eliminarlo. Si la hipótesis nula $\mathbb{H}_0$ es admitida concluimos que no existe sesgo o sistematismo.

Si suponemos que la varianza poblacional es desconocida, el estadístico que hay que usar es

$$y = \bar{l} \frac{\sqrt{n}}{s}$$

y con el mismo nivel de confianza, se rechaza la hipótesis nula si

$$|y| > t_{n-1,\, \alpha/2},$$

con las mismas consecuencias que en el caso anterior.

Para averiguar si la precisión del mapa (precisión de la serie de discrepancias) es aceptable debemos contrastar la varianza por medio de los test de la varianza de una población normal. Nos interesa saber si la varianza del mapa es menor o igual que el valor σ_0^2 fijado por las normas de precisión del mapa, en cuyo caso el mapa es aceptable, es decir, se trata de contrastar la hipótesis nula

$$\mathbb{H}_0: \quad \sigma^2 \leqslant \sigma_0^2,$$

frente a la hipótesis alternativa

$$\mathbb{H}_1: \quad \sigma^2 > \sigma_0^2.$$

Si suponemos que la media poblacional es conocida (en buenas condiciones debe ser $\mu = 0$), se usa el test de una cola y el estadístico

$$y = \frac{n s_\mu^2}{\sigma_0^2},$$

donde,

$$s_\mu^2 = \frac{1}{n} \sum_{i=1}^{n} \Delta X_i^2,$$

siendo n el número de puntos; fijado un nivel de significación de $\alpha =$ $= 0{,}05$ (nivel de confianza del 95 por 100), se rechaza la hipótesis $\mathbb{H}_0$ si

$$y > \chi^2_{n,\, \alpha}$$

en cuyo caso se admite la hipótesis alternativa $\mathbb{H}_1$. Esto quiere decir que, si todos los supuestos hechos (distribución normal y media cero) son ciertos, el mapa no cumple las normas de precisión establecidas (σ_0). Si el test admite la hipótesis $\mathbb{H}_0$ concluimos que el mapa es de la calidad establecida.

El test de la varianza también puede aplicarse en el segundo caso de suponer la media desconocida (lo que suele ser más corriente). Entonces, para contrastar la hipótesis $\mathbb{H}_0$: $\sigma^2 \leqslant \sigma_0^2$, utilizamos el test de una cola y el estadístico

$$y = \frac{(n-1)\, s^2}{\sigma_0^2}$$

y al mismo nivel de confianza, se rechaza $\mathbb{H}_0$ si

$$y > \chi^2_{n-1,\, \alpha}$$

Las conclusiones son análogas a las del caso anterior. Si se desea en cualquiera de los casos, pueden establecerse los intervalos de confianza correspondientes.

Ejemplo práctico

Supongamos una hoja de un mapa de escala 1:2.000 al que las normas de precisión le exigen una desviación típica de 0,500 metros. Para proceder al control de calidad en precisión del mapa se han seleccionado 20 puntos uniformemente distribuidos y, por los procedimientos indicados anteriormente, se han calculado para la coordenada X las discrepancias que figuran en la segunda columna de la Tabla 3.

Con estos datos se han calculado los siguientes valores, algunos de los cuales se encuentran en las columnas de la Tabla 3.

Suma de las discrepancias:

$$\sum \Delta X_1 = 3,33$$

Media aritmética de las discrepancias:

$$\Delta \overline{X} = 0,1665$$

Cuadrados de las discrepancias: columna 3.

Suma de los cuadrados de las discrepancias:

$$\sum \Delta X_1^2 = 7,4819$$

Media de los cuadrados de las discrpancias:

$$(\overline{\Delta X^2}) = 0,374095$$

Resíduos de las discrepancias respecto a su media: columna 4.

Cuadrados de los resíduos: columna 5.

Suma de los cuadrados de los resíduos:

$$\sum (\Delta X_i - \Delta \overline{X})^2 = 6,927455$$

Varianza respecto de una media cero:

$$s_\mu^2 = 0,374095$$

Varianza muestral:

$$s^2 = 0,3646029$$

Aplicación de los test de hipótesis

Test de la media con varianza conocida:

$$(\sigma_0^2 = 0,50)$$

Valor del estadístico:

$$\mathcal{Y} = 0,1665 \times \sqrt{20/0,50} = 1,489$$

Valor crítico:

$$\lambda_{0,025} = 1,96$$

Como $\mathcal{Y} < \lambda_{0,025}$ la hipótesis nula es *aceptada*; no existen sistematismos.

Test de la media con varianza desconocida.

Valor del estadístico:

$$\mathcal{Y} = 0,1665 \times \sqrt{20/0,3646029} = \\ = 2,041$$

Valor crítico:

$$t_{19,\ 0,025} = 2,093$$

Como $\mathcal{Y} < t_{19,\ 0,025}$ la hipótesis nula es *aceptada*.

Test de la varianza con media conocida ($\mu_0 = 0$)

Valor del estadístico:

$$\mathcal{Y} = 200 \times 0,3740950/0,25 = 29,9276$$

Valor crítico:

$$10\chi^2_{20,\ 0,05} = 31,410$$

Como $\mathcal{Y} < \chi_{20,\ 0,05}$ la hipótesis nula es *aceptada*. Precisión correcta.

Test de la varianza con media desconocida.

Valor del estadístico:

$$\mathcal{Y} = 19 \times 0,3646029/0,25 = 27,717$$

Valor crítico:

$$\chi^2_{19,\ 0,05} = 30,144$$

Como $\mathcal{Y} < \chi^2_{19,\ 0,05}$ la hipótesis nula es *aceptada*. Precisión correcta.

Los valores críticos de las distribuciones están tomados de la siguiente Tabla 4 donde se dan estos valores para los números de elementos más frecuentemente usados para estos test.

TABLA 3.

N	ΔX	ΔX^2	$\Delta X - \Delta\overline{X}$	$(\Delta X - \Delta\overline{X})^2$
1	0,52	0,2704	0,3535	0,12496225
2	−0,71	0,5041	−0,8765	0,76825225
3	0,88	0,7744	0,7135	0,50908225
4	−0,37	0,1369	−0,5365	0,28783225
5	0,46	0,2116	0,2935	0,08914225
6	0,39	0,1521	0,2235	0,04995225
7	−0,50	0,2500	−0,6665	0,44422225
8	0,65	0,4225	0,4835	0,23377225
9	−0,90	0,8100	−1,0665	1,13742225
10	−0,30	0,0900	−0,4665	0,21762225
11	0,72	0,5184	0,5535	0,30636225
12	0,61	0,3721	0,4435	0,19669225
13	−0,48	0,2304	−0,6465	0,41796225
14	−1,00	1.0000	0,8335	0,69472225
15	0,68	0,4624	0,5135	0,26368225
16	−0,39	0,1521	−0,5565	0,30969225
17	0,70	0,4900	0,5335	0,28462225
18	0,49	0,2401	0,3235	0,10465225
19	0,38	0,1444	0,2135	0,04558225
20	−0,50	0,2500	−0,6665	0,44422225

TABLA 4.

n	$t_{n,\ 0,05}$	$t_{n,\ 0,025}$	$\chi^2_{n,\ 0,05}$	$\chi^2_{n,\ 0,025}$
18	1,734	2,101	28,869	32,346
19	1,729	2,093	30,144	33,687
20	1,725	2,086	31,410	35,020
21	1,721	2,080	32,671	36,343
22	1,717	2,074	33,924	37,659
23	1,714	2,069	35,172	38,698
24	1,711	2,064	36,415	40,270
25	1,708	2,060	37,652	41,566