In vitro veritas 2013; 14:42-48 ISSN: 1697-5421

Document docent

Associació Catalana de Ciències de Laboratori Clínic
Secció d'Estadística i Metrologia[1]

Estadístics i paràmetres usats en les ciències de laboratori clínic per a variables contínues

Preparat per:
José Manuel González de Aledo Castillo, Ariadna Arbiol Roca
Laboratori Clínic, Hospital Universitari de Bellvitge, L'Hospitalet de Llobregat

Estadístics i paràmetres de tendència central

Els estadístics i els paràmetres de tendència central mostren quins són els valors numèrics al voltant dels quals s'agrupen les dades d'una mostra poblacional o d'una població i es classifiquen com s'indica a la Figura 1.

Com que la situació més freqüent és tractar amb mostres poblacionals, i no amb poblacions, en aquest text es parlarà només d'*estadístics*, donant per entès que si es tracta d'una població, s'està fent referència a un *paràmetre*.

[1] Membres de la Secció d'Estadística i Metrologia durant la preparació d'aquest document: A. Blanco Font, B. Candás Estébanez, X. Fuentes Arderiu (president), M. Martínez Casademont, M. Mosquera Parrado, J.M. Queraltó Compañó, L. Rami Brualla, R. Rigo Bonnin, H. Valbuena Parralejo.

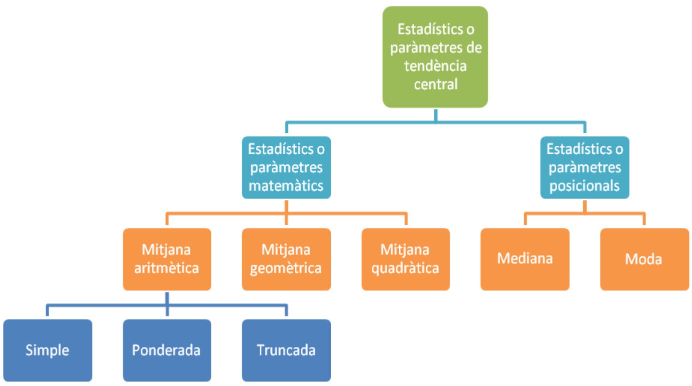

Figura 1. Esquema dels estadístics o paràmetres de tendència central més utilitzats al laboratori clínic per a les variables contínues.

Mitjana aritmètica simple

La mitjana aritmètica simple o, simplement, mitjana és la «suma de valors numèrics dels ítems d'una mostra poblacional, o d'una població, dividida pel nombre d'aquests valors» (1). Només és aplicable pel tractament de valors pertanyents a escales racionals, intervalars, fraccionals i absolutes, encara que la mitjana en aquestes últimes és fictícia, ja que els integrants de les escales absolutes han de ser nombres enters (3), i es calcula de la forma següent:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_i + \ldots + x_n}{n}$$

on $\bar{x}$ és la mitjana mostral, (si fos poblacional el símbol seria μ), x_i és qualsevol dels valors numèrics individuals i n és el nombre d'aquests valors.

La mitjana és l'estadístic de tendència central més utilitzat. És sensible a qualsevol canvi en les dades; és per això que és útil com a detector de les seves variacions. Degut al seu rigor matemàtic, és emprada sovint en altres càlculs estadístics, i representa el centre de gravetat d'un conjunt de dades. Pel contrari, és sensible als valors numèrics extrems i no és recomanable utilitzar-la quan la distribució de les dades és asimètrica.

Una de les aplicacions més destacades de la mitjana aritmètica al laboratori clínic està relacionada amb el control intern de la qualitat. A l'hora d'assignar valors numèrics a les magnituds que es mesuren en un material de referència (de control) no valorat, el laboratori obté una mitjana aritmètica [així com una desviació estàndard i un coeficient de variació dels que es tractarà més endavant]. A partir d'aquesta mitjana es pot establir l'interval de control amb el que s'haurà de comparar el valor mesurat en el material de control.

Exemple:

Es vol saber la mitjana de la concentració en substància d'ió sodi en el plasma d'un pacient que porta tres dies ingressat a l'hospital i durant aquest temps se li han fet deu extraccions de sang. Les concentracions en substància d'ió sodi en plasma són: 140, 140, 135, 147, 141, 140, 128, 138, 120 i 159 mmol/L. La mitjana aritmètica és 138,8 mmol/L.

Mitjana aritmètica ponderada

La mitjana aritmètica ponderada o, simplement, mitjana ponderada, és la «suma dels productes de cada valor numèric pel nombre de valors corresponent, dividit per la suma de tots els nombres de valors numèrics» (1). La mitjana ponderada s'utilitza quan no hi ha el mateix nombre de valors numèrics de cadascuna de les dades considerades, i per calcular-la s'empra la fórmula següent:

$$\bar{x}_{\mathrm{w}} = \frac{\sum_{i=1}^{n} \bar{x}_i \cdot n_i}{\sum_{i=1}^{n} n_i} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \ldots + \bar{x}_i n_i + \ldots + \bar{x}_n n_n}{n_1 + n_2 + \ldots + n_i + \ldots + n_n}$$

on $\bar{x}_{\mathrm{w}}$ representa la mitjana ponderada, x_i són els valors numèrics individuals i n_i és el nombre de cadascun d'aquests valors.

Un exemple d'ús és l'estimació de la mitjana aritmètica anual a partir de les mitjanes mensuals dels valors mesurats d'una magnitud en un material de control, quan cada mes s'obté un nombre de valors mesurats diferent.

En el cas que hi hagi el mateix nombre de valors numèrics de cadascuna de les dades, aquesta mitjana pot ser una mitjana de mitjanes.

Pel que es refereix a la seva aplicació, aquesta mitjana té les mateixes limitacions que la mitjana aritmètica simple.

Mitjana aritmètica truncada

La mitjana truncada és la mitjana aritmètica obtinguda després d'eliminar una part dels valors numèrics més baixos i més alts. Així, una mitjana truncada al 10 %, és la mitjana aritmètica del 90 % dels valors numèrics centrals, descartant el 5 % dels valors numèrics més elevats i el 5 % dels més baixos. És útil per eliminar l'efecte de possibles valors numèrics aberrants.

Pel que es refereix a la seva aplicació, aquesta mitjana té les mateixes limitacions que la mitjana aritmètica simple.

Mitjana geomètrica

La mitjana geomètrica d'un conjunt d'*n* valors numèrics és l'arrel *n*-èssima del producte de cadascun d'ells en valor absolut. La mitjana geomètrica d'un conjunt de nombres positius és sempre menor o igual que la mitjana aritmètica. Considera tots els valors numèrics del conjunt i és menys sensible als valors numèrics extrems que la mitjana aritmètica. Presenta un significat estadístic menys intuïtiu que la mitjana aritmètica i el seu càlcul és més laboriós.

La seva fórmula és la que es presenta a continuació:

$$\bar{x}_{\mathrm{g}} = \left(\prod_{i=1}^{n} |x_i|\right)^{\frac{1}{n}} = \left(|x_1| \cdot |x_2| \cdot \ldots \cdot |x_i| \cdot \ldots \cdot |x_n|\right)^{\frac{1}{n}}$$

on $\bar{x}_{\mathrm{g}}$ representa la mitjana geomètrica, $|x_i|$ és qualsevol dels valors numèrics (prescindint del seu signe) i n és el nombre d'aquests valors.

Una de les poques aplicacions de la mitjana geomètrica al laboratori clínic, recomanada pel Comitè Internacional de Normalització en Hematologia i el Comitè Internacional de Trombosi i Hemostàsia, és l'assignació d'un temps mitjà de coagulació del plasma induït pel factor tissular («temps de protrombina») a una barreja de 20 o més mostres de plasma procedents de voluntaris presumptament sans. Aquesta barreja de plasmes serveix per calibrar els sistemes de mesura corresponents. .

Mitjana quadràtica

La mitjana quadràtica d'un conjunt d'*n* valors numèrics és l'arrel quadrada de la suma de valors numèrics elevats al quadrat dels ítems d'una mostra poblacional, o d'una població, dividida pel nombre d'aquests valors. La seva fórmula és:

$$\bar{x}_{\mathrm{q}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_i^2 + \ldots + x_n^2}{n}}$$

on $\bar{x}_{\mathrm{q}}$ és la mitjana quadràtica, x_i és qualsevol dels valors numèrics individuals i n és el nombre d'aquests valors.

La mitjana quadràtica és molt útil quan es vol obtenir informació d'una variable que pot adoptar valors positius i negatius, i per tant donar lloc a una mitjana de valor nul o pròxim a zero. En les ciències de laboratori clínic, la mitjana quadràtica té interès aplicat en el camp del control intern de la qualitat i és la base del control estadístic de la qualitat aplicat oficialment al laboratori clínic a Alemanya (5).

Mediana

La mediana és el valor que divideix un conjunt de valors ordenats en dues parts, amb el mateix nombre de valors a cada part. La mediana coincideix amb el fractil 0,5 (1), com es comentarà més endavant. Només es pot calcular per a variables quantitatives.

Si el nombre de valors (*n*) és senar, la mediana és el valor que ocupa la posició (*n* + 1) / 2 una vegada que les dades han estat ordenades (en ordre creixent o decreixent).

Si *n* és parell, la mediana és la mitjana aritmètica de les dues dades que estan al centre de la mostra poblacional i ocupen les posicions *n* / 2 i (*n* / 2) + 1.

No és sensible als valors extrems, inclosos els eventuals valors aberrants. L'ús d'aquest estadístic és recomanable per a distribució de dades asimètriques, encara que no té el rigor matemàtic de la mitjana aritmètica.

Partint de les dades de l'exemple anterior de la mesura de la concentració en substància de l'ió sodi en el plasma, els valors mesurats ordenats de menor a major són: 120, 128, 135, 138, 140, 140, 140, 141, 147 i 159 mmol/L. Com que el nombre de valors numèrics és parell la mediana serà la mitjana aritmètica entre els valors numèrics de les posicions 5 i 6, és a dir, 140 mmol/L.

Moda

La moda indica el valor més freqüent d'un conjunt de valors numèrics. En el cas de que dos valors presentin la mateixa freqüència, es diu que la mostra poblacional, o la població, és bimodal; si hi hagués més de dues modes la mostra poblacional, o la població, seria multimodal.

L'*interval modal* és aquell interval que presenta la major freqüència absoluta. La moda, quan les dades estan agrupades, és un punt que divideix a l'interval modal en dues parts.

La moda es pot estimar per un càlcul senzill i d'interpretació molt clara. Al dependre només de les freqüències, pot estimar-se per a tota mena de variables, incloses les discretes. És per això que és l'estadístic més utilitzat quan al descriure una mostra poblacional no és possible calcular altres estadístics de tendència central. No sempre es situa cap al centre de les dades.

En l'exemple citat anteriorment la moda és 140 mmol/L, el valor més repetit.

Estadístics o paràmetres de dispersió

Tal com s'ha dit en el primer apartat d'aquest document docent, es parlarà només d'*estadístics* de dispersió, donant per entès que si es tracta d'una població, s'està fent referència a un *paràmetre* de dispersió.

Des del punt de vista estadístic, la dispersió és el grau de distanciament d'un conjunt de valors respecte la seva tendència central. Els estadístics de dispersió mostren la variabilitat de les dades dins d'una mostra poblacional, és a dir, permeten conèixer quantitativament si els diferents valors numèrics d'una variable estan molt allunyats entre si. A la Figura 2 es presenta una classificació d'aquests estadístics.

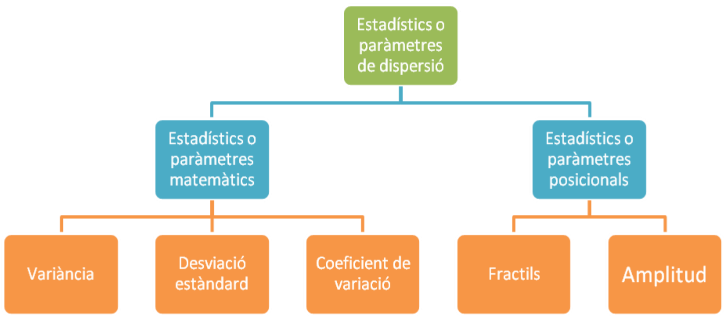

Figura 2. Esquema dels estadístics o paràmetres de dispersió més utilitzats al laboratori clínic per a les variables contínues.

Els estadístics de dispersió donen una perspectiva sobre un conjunt de valors numèrics complementària a la que donen els estadístics de tendència central. Per exemple, si en un conjunt de dades numèriques d'una població particular on la mitjana de la concentració de substància d'ió potassi en el plasma és 4,64 mmol/L, es pot arribar a pensar que hi ha persones dins de la mateixa en les que aquesta concentració pot ser 7,01 mmol/L i d'altres en les que pot ser 1,93 mmol/L. Si en aquest cas, a més a més de la mitjana aritmètica, es disposés d'un estadístic de dispersió, la informació seria molt més completa.

En les ciències de laboratori clínic els estadístics de dispersió són imprescindibles per tal de comprendre teories i conceptes fonamentals com ara la teoria dels errors de mesura, la incertesa de mesura, el control de la qualitat, la variabilitat biològica, la teoria dels valors de referència biològics, entre d'altres.

Desviació

Per a un conjunt de valors numèrics, la desviació és la diferència entre un valor obtingut i la mitjana aritmètica. És un estadístic de dispersió de cada valor numèric que té en compte tots els valors numèrics del conjunt en qüestió. Matemàticament:

$$Desviació = x_i - \bar{x}$$

on x_i representa qualsevol valor numèric i $\bar{x}$ la mitjana mostral.

En metrologia el concepte d'error de mesura és molt proper al concepte estadístic de desviació.

Variància

Per a un conjunt de valors numèrics, la variància (símbol s^2 quan és un estadístic o σ^2 quan és un paràmetre) queda definida per la fórmula següent (1):

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

on s^2 representa la variància mostral, x_i cadascun dels valors numèrics, $\bar{x}$ la mitjana mostral i n el nombre de valors.

La variància és un estadístic (o un paràmetre) molt relacionat amb el concepte metrològic de precisió de mesura, propi dels valors mesurats i dels sistemes de mesura. Des d'aquest punt de vista, destaquen les propietats homoscedasticitat i heteroscedasticitat dels sistemes de mesura. Si la variància és constant sigui quin sigui el valor de la magnitud mesurada, la propietat d'un sistema de mesura s'anomena *homocedasticitat*; si pel contrari, la variància varia segons el valor de la magnitud mesurada, la propietat s'anomena *heteroscedasticitat*.

Desviació estàndard

Per a un conjunt de valors numèrics, la desviació estàndard (símbol s quan és un estadístic—que també s'anomena desviació estàndard experimental— o σ quan és un paràmetre) queda definida per la fórmula següent (1):

$$s = \sqrt{s^2}$$

on s representa la desviació estàndard i s^2 la variància.

La desviació estàndard és l'estadístic de dispersió més utilitzat en les ciències de laboratori i en la ciència i tecnologia en general. Des del punt de vista metrològic, cal destacar el seu ús per quantificar la precisió de mesura, és a dir, per estimar la imprecisió. Seguint amb l'exemple presentat pels estadístics de tendència central, podem calcular la desviació estàndard dels valors mesurats de la concentració en substància de l'ió sodi en el plasma (en mmol/L) com segueix:

s^2 =[(140-138,8)² + (140-138,8)² + (135-138,8)² + (147-138,8)² + (141-138,8)² + (140-138,8)² + (128-138,8)² + (138-138,8)² + (120-138,8)² + (159-138,8)²] / 10 = 96,96 (mmol/L)² i per tant, s = 9,85 mmol/L.

Quan es parteix de dades duplicades, una altra manera de calcular la desviació estàndard és mitjançant la fórmula de Dahlberg (6):

$$s = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{2n}}$$

on s és la desviació estàndard, d_i^2 és la diferència al quadrat entre un parell de valors mesurats duplicats i n és el nombre de parells de valors numèrics.

En el laboratori clínic, aquesta fórmula s'utilitza en l'estudi de la imprecisió d'un sistema de mesura quan es treballa a partir de la mesura de duplicats d'una mateixa mostra dins d'un conjunt de mostres.
En (7) es calcula la imprecisió intraserial de l'eritrosedimentació en l'analitzador TEST-1 mitjançant la fórmula de Dahlberg. Per fer-ho es mesuren diverses mostres per duplicat en un interval de temps determinat i es calcula la diferència entre els valors mesurats de la magnitud en estudi. El sumatori d'aquestes diferències es fa servir per calcular la desviació estàndard mitjançant la fórmula de Dahlberg.

La desviació estàndard també es pot estimar a patir dels fractils 0,25 ($x_{0,25}$) i 0,75 ($x_{0,75}$), dels quals es tractarà més endavant), segons la fórmula (8):

$$s = (x_{0,75} - x_{0,25}) / 1,349$$

Coeficient de variació

El coeficient de variació (*CV*) és la desviació estàndard d'un conjunt de valors numèrics dividida per la mitjana aritmètica d'aquests valors, que habitualment es multiplica per cent per expressar-lo com a percentatge (1):

$$CV = \frac{s}{\bar{x}} \cdot 100$$

on CV representa el coeficient de variació, s la desviació estàndard i $\bar{x}$ la mitjana mostral.
A major valor del coeficient de variació, major heterogeneïtat dels valors del conjunt que es tracti; i a menor coeficient de variació, major homogeneïtat.
El coeficient de variació és útil per comparar la variabilitat d'un o més conjunts de dades entre ells, encara que tinguin valors numèrics molt diferents. Per exemple, permet comparar la imprecisió de diferents sistemes de mesura.
Seguint amb l'exemple anterior, el

$$CV = \frac{9,85}{138,8} \cdot 100 = 7,1\%$$

Amplitud

L'amplitud és la diferència entre el valor màxim i el valor mínim d'un conjunt de valors numèrics (1). És l'estadístic de dispersió menys robust i és per això que s'utilitza poc.
En l'exemple anterior citat, l'amplitud és: 159 mmol/L – 120 mmol/L = 39 mmol/L

Fractils

Tal com s'ha dit en el primer apartat d'aquest En un conjunt de valors numèrics ordenats de menor a major, un fractil (o quantil) és el valor, existent o calculat, d'aquest conjunt que deixa a la seva esquerra una fracció determinada de valors més petits que ell. El fractil d'ordre 0,50 ($x_{0,5}$) correspon a la mediana.

En general, els fractils són uns estadístics de dispersió molt importants a l'hora de decidir valors numèrics decisoris, així els fractils $x_{0,025}$ i $x_{0,975}$ tenen una gran transcendència en la teoria dels valors de referència biològics.

Quartils

El quartil és el fractil d'ordre 0,25 o fractil d'ordre 0,75 (2).

La distància interquartílica, és a dir, la diferència entre $x_{0,75}$ i $x_{0,25}$, també s'utilitza com a estadístic de dispersió.

Percentils

En un conjunt de valors numèrics ordenats de menor a major, un percentil és el valor, existent o calculat, d'aquest conjunt que deixa a la seva esquerra un percentatge determinat de valors més petits que ell. En són exemples el percentil 1 ($x_{1\%}$) que deixa a la seva esquerra l'1 % de les dades, el percentil 50 ($x_{50\%}$) que deixa a la seva esquerra el 50 % de les dades (i coincideix amb la mediana) i el percentil 99 ($x_{99\%}$) que deixa a la seva esquerra el 99 % de les dades.

Bibliografia

1. International Organization for Standardization. Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability. ISO 3534-1:2006. Geneva: ISO; 2006.
2. International Organization for Standardization. Statistics — Vocabulary and symbols — Part 2: Applied statistics. ISO 3534-2:2006. Geneva: ISO; 2006.
3. Fuentes-Arderiu X, Miró Balagué J. Naturalesa de les propietats biològiques examinades al laboratori clínic. In vitro veritas 2011;12:150-9. <http://www.acclc.cat/continguts/ivv135.pdf> (accés: 2013-02-26).
4. World Health Organization. WHO Expert Committee on Biological Standardization. Guidelines for thromboplastins and plasma used to control oral anticoagulant therapy. WHO Technical Report Series. No. 880. Geneva, Switzerland: World Health Organization; 1999.
5. Richtlinie der Bundesärztekammer zur Qualitätssicherung laboratoriumsmedizinischer Untersuchungen. <http://www.bundesaerztekammer.de/downloads/RiliBAEKLabor201205.pdf> (accés: 2013-02-26).
6. Dahlberg G. Statistical methods for medical and biological students. New York: Interscience Publications; 1940.
7. A. Padró-Miquel, X. Fuentes-Arderiu. Quality control strategy to verify day-to-day imprecision of length of blood sedimentation reaction measurements with the TEST-1 analyzer. Clin Chem Lab Med 2007;45(7):930-1.
8. Centre Suisse de Contrôle de Qualité. Parmètres statistiques utilisés dans les rapports de CQE. 2009. <http://www.cscq.ch/SiteCSCQ/FichierPDF_FR/z-score-f.pdf>(accés: 2013-02-26).