

55 RESPUESTAS A DUDAS TÍPICAS DE ESTADÍSTICA

2ª Edición



ROBERTO BEHAR GUTIÉRREZ
PERE GRIMA CINTAS

55

**Respuestas
a dudas típicas
de ESTADÍSTICA**

**Roberto Behar Gutiérrez
Pere Grima Cintas**

Segunda edición



ACERCA DE LOS AUTORES



Roberto Behar Gutiérrez es Licenciado en Educación en la especialidad de Matemáticas, por la Universidad Santiago de Cali, y es Estadístico por la Universidad del Valle (Cali, Colombia). Obtuvo el grado de doctor en la Universidad Politécnica de Cataluña. Es profesor de la carrera de estadística de la Universidad del Valle desde su fundación en 1978, donde también ha sido director del departamento de Producción e Investigación de Operaciones, director de la carrera de estadística, y director del Master en Ingeniería Industrial y de Sistemas.

Ha sido asesor estadístico para diversas instituciones colombianas en estudios sobre medio ambiente y desarrollo social. También ha asesorado a empresas sobre temas relacionados con el control estadístico de la calidad y estadística industrial. Entre sus publicaciones se encuentra el libro de texto que escribió junto a Mario Yepes: “Estadística: Un enfoque descriptivo”. Ed. Feriva, 1995. Ha escrito también numerosos artículos sobre técnicas estadísticas en revistas especializadas, algunos de los cuales tratan sobre su enseñanza, como el publicado con Mario Miguel Ojeda en el *Newsletter* (1997) del *International Statistical Institute*. (ISI): “*A Reformulation of the problem of Statistical Education: A learning Perspective*” o el que escribió con Pere Grima en la revista “Estadística Española” (2001): “Mil y una dimensiones del aprendizaje de la estadística”. También ha realizado numerosas conferencias sobre el uso de la estadística y sobre su enseñanza y aprendizaje, tema que es una de sus pasiones.



Pere Grima Cintas es doctor ingeniero industrial y profesor de la Universidad Politécnica de Cataluña, donde también es coordinador académico del Master en Gestión de la Calidad. Su especialidad son las técnicas estadísticas para el control y la mejora de la calidad, tema sobre el que ha asesorado a numerosas empresas e instituciones. También se ha ocupado de temas relacionados con la gestión de la calidad y en el año 2000 fue evaluador de la *European Foundation for Quality Management* (EFQM) para el premio europeo a la calidad que otorga esta institución. En el periodo en que ha estado trabajando en este libro, sus actividades de asesoramiento a empresas se han centrado mayoritariamente en la implantación de programas de mejora Seis Sigma.

Junto con sus compañeros Albert Prat, Xavier Tort-Martorell y Lourdes Pozueta escribió el libro “Métodos estadísticos. Control y mejora de la Calidad”, Ediciones UPC, del que ya se han realizado varias ediciones y que ha sido publicado en Iberoamérica por Editorial Alfaomega. Con Xavier Tort-Martorell escribió “Técnicas para la Gestión de la Calidad”, editado por Díaz de Santos en 1995. También le gusta dedicar parte de su tiempo a trabajar en temas relacionados con la divulgación de la estadística.

Presentación

Muchos de los que alguna vez hemos sido estudiantes de un curso de Estadística, recordamos momentos en los que intentábamos entender, no siempre con éxito, las razones por las cuales había que hacer las cosas de una determinada manera. ¿Por qué dividir por $n-1$ al calcular la desviación estándar? ¿Por qué no dividir por el número total de datos? El tiempo disponible en los cursos no permite explicarlo todo y en ocasiones el profesor, en su intento por dar una explicación al estudiante que pregunta, responde usando términos como “grados de libertad” o “estimador insesgado”, lo cual puede generar más dudas de las que aclara.

Por otra parte, a través de nuestra experiencia ayudando a profesionales de la Medicina, la Administración o la Ingeniería en el uso de métodos estadísticos, hemos comprobado que los conceptos o las técnicas con que se trabaja no siempre están del todo claras. ¿Cómo hay que interpretar el p-valor que da el listado del ordenador? ¿Es lo mismo diferencia significativa que diferencia importante? Esta falta de seguridad en su manejo hace que muchas veces se evite hacer uso de todas las posibilidades que brinda la estadística, con lo que se pierde la oportunidad de obtener una información que puede resultar muy útil para la toma de decisiones.

También en un ámbito no estrictamente profesional existen muchas dudas “populares” en torno a la Estadística: ¿cómo es que con una muestra de 2.000 personas puede conocerse razonablemente bien la opinión de un país de 40 millones de habitantes?, o lo que es todavía más sorprendente, ¿cómo es que esas 2.000 personas también serían suficientes para una población de 100 millones? Y ligada con estas, si la Estadística es tan potente, ¿por qué cuesta tanto acertar en los sondeos electorales?

Este texto pretende dar respuesta a muchas de estas preguntas y nuestra intención es que sea útil tanto a los estudiantes de los cursos de estadística que se imparten en la universidad, como a los profesionales que están interesados en refrescar sus ideas o aclarar dudas concretas, y también a todas aquellas personas interesadas en esta disciplina que quieran resolver algunas de sus dudas.

Sin ser exhaustivos, pues siempre es posible aumentar la lista con nuevas preguntas, hemos tratado de cubrir un amplio espectro, tratando dudas en estadística descriptiva, distribuciones de probabilidad, estimación, contraste de hipótesis, comparación de poblaciones, correlación y regresión, diseño de experimentos, estudios de capacidad y control de procesos y un apartado para dudas varias, como las relacionadas con los grados de libertad y el teorema central del límite, entre otras.

Muchas preguntas tienen un carácter general e introductorio y son “aptas para todos los públicos”, pero otras tratan sobre temas específicos en el contexto de las ecuaciones de regresión, el diseño de experimentos o el control estadístico de procesos. En este último caso se requiere un cierto nivel de conocimientos sobre el tema, aunque si la pregunta despierta interés, seguramente ya se sabe lo suficiente para entender la respuesta. En todos los casos se ha intentado usar un lenguaje coloquial, recurriendo a la intuición y apoyándose en la metáfora, pero procurando que no haya pérdida en el rigor.

Se ha intentado también que cada respuesta sea lo más autocontenida posible, es decir, lo suficientemente completa, para que no requiera de otras para su adecuada comprensión. De todas maneras, en cada uno de los temas que se tratan, se han colocado las dudas y sus respuestas en el orden que consideramos más efectivo, de tal manera que un lector que desee leer todas las preguntas de un apartado en forma secuencial vaya ganando elementos para comprender mejor la siguiente.

Dejando claro que cualquier falta en la virtud de este trabajo es de exclusiva responsabilidad de los autores, deseamos poner de manifiesto nuestro agradecimiento a todos nuestros compañeros en las tareas docentes, seguramente la mejor fuente de información que hemos tenido. Lluís Marco, de la Universitat Politècnica de Catalunya y Guillermo de León, de la Universidad Veracruzana nos sugirieron algunas de las preguntas que se incorporan y también ideas y posibles enfoques para muchas respuestas, además de leer los originales y sugerir numerosas mejoras. Rafael Antonio Klinger y Eloina Mesa, de la Universidad del Valle, también leyeron los originales y realizaron muchas sugerencias que han mejorado notablemente la claridad de las respuestas.

Deseamos agradecer también a la Agencia Española de Cooperación Internacional (AECI) y a nuestras Universidades, la Universidad del Valle y la Universitat Politècnica de Catalunya, las ayudas y facilidades obtenidas para la realización de este trabajo.

Muy probablemente, no podremos evitar la frustración de algunos de nuestros lectores al buscar en vano alguna duda que no fue tratada aquí, o al no quedar del todo satisfechos con alguna respuesta. Nuestra aspiración es poder recoger todas las sugerencias y apreciaciones que nos permitan realizar un proceso mejora continua de nuestro trabajo, por lo que agradeceremos todos los comentarios y sugerencias que nos hagan llegar a través de la página web: www.55RespuestasEstadistica.com.

Prólogo a la segunda edición

En la primera edición corregimos las erratas detectadas pero manteniendo prácticamente el mismo contenido del texto.

En esta segunda edición hemos realizado una revisión más a fondo, con la idea de sustituir algunas preguntas por otras que a nuestro entender pudieran despertar mayor interés, aunque finalmente los cambios no han sido muchos. Además de retocar algunos detalles y añadir una nota aclaratoria al final de la pregunta 4, hemos cambiado la pregunta 21, que en ediciones anteriores se titulaba: "¿Qué significa la expresión: 'un intervalo de confianza del 95% es $27,5\% \pm 3,6\%$ '?" por una pregunta típica, seguramente de las más típicas, pero que habíamos pasado por alto en nuestra lista inicial: "Estoy realizando un estudio basado en la realización de una encuesta ¿Cuál debe ser el tamaño de la muestra para que las conclusiones sean fiables?". Respondiendo de forma directa a esta pregunta esperamos aclarar algunas dudas y mantenemos la explicación de los conceptos que se explicaban en la pregunta sustituida. Al cambiar esta pregunta también hemos cambiado la siguiente para que el contenido se complemente y no sea repetitivo.

Agradecemos a todos los lectores que nos han hecho llegar sus comentarios y sugerencias, gracias a las cuales hemos podido mejorar diversos aspectos del texto.

Los Autores

Índice

Acerca de los autores	VII
Presentación	IX
Prólogo a la segunda edición	XI
Estadística descriptiva	
1. ¿Para qué sirve la mediana, si ya tenemos la media aritmética?	3
2. ¿Tiene alguna aplicación práctica la media geométrica?	5
3. ¿Por qué en la expresión de la varianza se utiliza el cuadrado de las diferencias en vez de su valor absoluto?	7
4. ¿Por qué cuando se calcula la varianza de una muestra se divide por $n-1$ en vez de dividir por n ?	11
5. ¿Cuál es la forma “correcta” de calcular los cuartiles?	15
6. ¿En cuántos intervalos conviene dividir los datos para construir un histograma? ¿Qué otros aspectos hay que tener en cuenta?	17
7. ¿Cuándo conviene utilizar <i>boxplots</i> para analizar o describir datos?	21
8. En los <i>boxplots</i> las anomalías se marcan a partir de $\pm 1,5$ veces el rango intercuartílico (IQR) ¿De dónde sale el 1,5?	23
9. ¿Qué hay que hacer cuando nos encontramos con valores atípicos?	25
10. ¿Qué es la curtosis (o <i>kurtosis</i>) y para qué sirve?	29
Distribuciones de probabilidad	
11. ¿Cómo se sabe que una variable aleatoria concreta sigue una determinada distribución de probabilidad?	35
12. La media de una muestra es un número concreto. ¿Por qué se dice entonces que es una variable aleatoria?	39
13. ¿Por qué la función densidad de probabilidad de la distribución Normal es la que es?	41
14. ¿Por qué las probabilidades calculadas a través de la Normal estandarizada coinciden con las buscadas en la distribución de interés?	47
15. Yo mido 1,68. ¿Por qué la probabilidad de que una estatura sea 1,68 calculada con la distribución Normal es 0?	51
16. ¿Existen variables aleatorias que presenten un comportamiento “contrario” a la distribución Normal, siendo los valores más probables los de los extremos?	53
17. ¿De dónde sale la fórmula de la distribución de Poisson?	57
18. ¿Cómo se puede ver que la distribución de la varianza muestral está relacionada con la distribución chi-cuadrado?	59

19. ¿Por qué da un resultado distinto sumar k variables aleatorias de la misma distribución de probabilidad que tomar una y multiplicarla por k ? 61

Estimación

20. Sabemos que las características de una muestra (proporción, media, ...) varían de una muestra a otra. ¿Por qué entonces creer en los resultados de una muestra, sabiendo que si tomáramos otra esos resultados serían distintos? 65
21. Estoy realizando un estudio basado en la realización de una encuesta ¿Cuál debe ser el tamaño de la muestra para que las conclusiones sean fiables? 67
22. ¿Por qué el tamaño de la muestra no crece proporcionalmente con el tamaño de la población? 71
23. ¿Por qué cuesta acertar en los sondeos electorales? 75
24. ¿Qué es un estimador de máxima verosimilitud? 79

Contraste de hipótesis

25. ¿Qué es el p-valor y cuál es el significado de las otras palabras clave que aparecen en el contraste de hipótesis? 83
26. ¿A partir de qué p-valor es razonable rechazar la hipótesis nula? 87
27. ¿Qué tipos de error se pueden cometer en un contraste de hipótesis? 89
28. ¿Es correcto multiplicar por 2 el área de cola en los tests de igualdad de varianzas cuando H_1 es del tipo “distinto de”? 93
29. ¿Por qué respecto a la hipótesis nula se habla de “no rechazo” y no de “aceptación”? 95
30. ¿Es lo mismo diferencia significativa que diferencia importante? 97

Comparación de tratamientos

31. ¿Cómo elegir la hipótesis alternativa que conviene plantear? 101
32. Si la hipótesis alternativa es del tipo “mayor que” o “menor que”, ¿cómo se sabe hacia qué lado hay que mirar el área de cola? 103
33. ¿Por qué el análisis de la varianza se llama así, cuando en realidad se trata de una técnica para comparar medias y no varianzas? 105
34. ¿Por qué para comparar k tratamientos se utiliza la técnica de análisis de la varianza, en vez del ya conocido test de la t de Student, aplicándolo a todas las parejas que se pueden formar con k tratamientos? 107

Correlación y Regresión

35. ¿Por qué cuando se ajusta una nube de puntos a una ecuación de regresión se utiliza siempre el criterio de minimizar la suma de los cuadrados de los residuos, y no otros como minimizar la suma de su valor absoluto? 111
36. Si los coeficientes de una ecuación de regresión son unos números concretos, ¿por qué se dice que son variables aleatorias? 113

37. ¿Por qué cuando se ajusta una recta que pasa por el origen no se utiliza el coeficiente de determinación R^2 como medida de calidad del ajuste? 117
38. ¿Por qué cuando se comparan ecuaciones de regresión con distinto número de variables regresoras no se utiliza R^2 sino el llamado R^2 ajustado? 121
39. ¿Cómo se pueden utilizar e interpretar variables cualitativas en una ecuación de regresión? 127
40. ¿Por qué del conjunto de variables candidatas a entrar en un modelo de regresión no necesariamente se seleccionan las que están más correlacionadas con la variable dependiente Y ? 133

Diseño de experimentos

41. ¿Por qué no es una buena estrategia ir moviendo las variables una a una cuando se trata de estudiar experimentalmente cómo estas afectan a una respuesta? 137
42. ¿Cómo es posible estudiar por separado el efecto de cada una de las variables que afectan a una respuesta si, tal y como se hace en los diseños factoriales, se mueven todas a la vez? 139
43. ¿Por qué funciona el algoritmo de Yates? 145
44. ¿Por qué cuando se representan valores en papel probabilístico normal (ppn), en la fórmula que da la ordenada se resta 0,5 del número de orden? 149
45. En los diseños factoriales, ¿cómo se puede escribir una ecuación para la respuesta a partir de los efectos? 153
46. ¿Qué es un diseño bloqueado? ¿Por qué en estos diseños no se tienen en cuenta las interacciones entre los factores de bloqueo y el resto de factores? ¿Qué ocurre si esas interacciones existen? 157
47. ¿Por qué es razonable suponer no significativas las interacciones de 3 o más factores? 161
48. ¿Qué hacer si al aleatorizar el orden de experimentación se obtiene el orden estándar de la matriz de diseño? 165

Estudios de capacidad y control estadístico de procesos

49. ¿Qué diferencia hay entre un estudio de capacidad a corto y largo plazo? ¿Cómo se estima la variabilidad en uno y otro caso? 169
50. ¿Por qué en los gráficos de control es más eficiente controlar medias que observaciones individuales? 173
51. En los gráficos de control, ¿la línea central debe ser el valor objetivo o el promedio obtenido al hacer el estudio de capacidad? 177

Varios

52. Cuando se habla de transformación logarítmica, ¿se refiere al logaritmo decimal o al neperiano? 181
53. ¿Qué significan los llamados “grados de libertad”? 183

54. ¿Debe decirse “Teorema central del límite” o “Teorema del límite central”? 187
55. ¿Cuál es la mejor estrategia para ganar la lotería (nacional, primitiva,...)? 189

Créditos y referencias

- ¿Cómo hemos resuelto nuestras dudas? 195
- Libros y páginas web que se citan 199

Estadística descriptiva

1

¿Para que sirve la mediana, si ya tenemos la media aritmética?

La media aritmética es una excelente medida de tendencia central. Sus buenas propiedades, junto con el hecho de ser fácil de entender y de calcular, la hacen muy usada y también muy apreciada (a veces demasiado, como cuando se pretende resumir solo en ella toda la información que contienen los datos), pero la mediana tiene unas propiedades de las que carece la media, por lo que es un buen complemento informativo e incluso en algunos casos puede ser una medida más útil. Estas propiedades son:

- Es más robusta que la media frente a la presencia de anomalías. Supongamos que nuestros datos son: 2, 5, 6, 7 y 9. La media es 5,6 y la mediana es 6. Si al introducir los datos al ordenador nos equivocamos y en último lugar en vez de 9 introducimos 99, la media pasa a ser de 23,8 mientras que la mediana sigue siendo 6.

En algunos casos, como cuando se trabaja con datos todavía no depurados, fijarse en la mediana puede ser más recomendable porque la información que da está menos afectada por las posibles anomalías que puedan existir.

- Por su propia definición, la mediana deja un 50% de las observaciones por encima y otro 50% por debajo y esto le da unas ventajas que la media no tiene. Por ejemplo, si queremos saber si en nuestra empresa estamos entre los que cobran más o entre los que cobran menos, debemos comparar nuestro salario con la mediana, y no con la media. Si sólo hay 10 trabajadores y los salarios son (pongamos que en miles de euros): 0,8; 0,8; 0,9; 0,9; 1,0; 1,0; 1,1; 1,1; 1,2 y 10, todos menos 1 (en este caso el 90%) están por debajo de la media, que es 1,88. Esto no pasa nunca con la mediana, si estamos por encima de la mediana, estamos con el 50% de los que más cobran

Otro ejemplo. Si un examen se aprueba sacando una nota igual o superior a 5 y la nota media que han sacado los estudiantes es de 5, no sabemos cuantos han aprobado. Si se han examinado 50 estudiantes, puede ser que 41 hayan suspendido con un 4; 8 estudiantes hayan sacado un 10 y uno haya obtenido un 6. Esto da media 5, aunque es verdad que son unas notas muy raras. Si la mediana es 5, seguro que la mitad han aprobado.

Además, puestos a criticar la media, podemos decir que su uso conduce a algunas situaciones paradójicas, como aquella que dice que la mayoría de los hombres tiene un número de piernas superior a la media.

Si la distribución de los datos es simétrica, la media y la mediana coinciden, y entonces todo son ventajas. Por ejemplo, en una distribución Normal la media y la mediana son iguales [$P(X > \mu) = 0,5$] y por tanto, si los valores que tenemos provienen de una Normal, y no hay anomalías, la media y la mediana no andarán muy lejos la una de la otra. En cualquier caso, si disponemos de un medio de cálculo fácil, podemos calcular las dos y aprovechar lo mejor de cada una.

2

¿Tiene alguna aplicación práctica la media geométrica?

Sí la tiene, pero su aplicación es menos frecuente que la media aritmética. Un caso muy típico en que la media geométrica resulta útil es para calcular promedios de tasas de crecimiento. Por ejemplo: una población que tenía 10.000 habitantes en el año cero, creció el primer año a una tasa del 5%, el segundo creció a una tasa del 20% y el tercer año al 50% ¿A qué tasa promedio ha crecido la población en estos 3 años?

Año	Población inicial	Tasa crecimiento	Factor de expansión	Población al final del año
1	10.000	0,05	1,05	10.500
2	10.500	0,20	1,20	12.600
3	12.600	0,50	1,50	18.900

Si calculamos la media aritmética de la tasa de crecimiento tenemos: $(0,05 + 0,20 + 0,50) / 3 = 0,25$ y el factor medio de expansión sería 1,25. Pero si la población hubiera crecido de esta forma los 3 años, no se llegaría al mismo resultado final:

Año	Población inicial	Tasa crecimiento	Factor de expansión	Población al final del año
1	10.000	0,25	1,25	12.500
2	12.500	0,25	1,25	15.625
3	15.625	0,25	1,25	19.531

Por tanto, la media aritmética no es un buen indicador de la tasa media de crecimiento.

Si la población crece a una tasa constante i , para que al final del tercer año tenga el mismo efecto que las tasas del ejemplo, se debe verificar que:

$$10.000(1+i)(1+i)(1+i) = 10.000(1+0,05)(1+0,20)(1+0,50)$$

De donde:

$$(1+i) = \sqrt[3]{1,05 \cdot 1,20 \cdot 1,50} = 1,2364$$

Si se hubiera tenido este factor de expansión cada año (nótese que es la media geométrica), hubiera conducido a una población final exactamente igual a la que tenemos. Es decir, que la tasa media de crecimiento ha sido del 23,64%.

Curiosidades sobre la media geométrica son:

- A diferencia de la media aritmética, la media geométrica solo se define para números positivos.
- La media geométrica nunca es mayor que la media aritmética. La demostración para el caso de 2 valores es fácil por reducción al absurdo. Supongamos que $\sqrt{ab} > (a+b)/2$, entonces $ab > (a^2 + 2ab + b^2)/4$ de donde $0 > a^2 - 2ab + b^2$. Como $a^2 - 2ab + b^2 = (a-b)^2$ es imposible que este valor sea negativo, luego es imposible que $\sqrt{ab} > (a+b)/2$.

Y ya puestos a hablar de otras medias, podemos hacer un comentario sobre la media armónica, mucho menos conocida pero también útil en algunos casos.

Se define la media armónica de x_1, x_2, \dots, x_N como:

$$Mh = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_N}}$$

Parece que esto sea un retorcimiento sin ningún interés, pero no. Por ejemplo, si un coche recorre una cierta distancia a una velocidad media de 100 km/h y vuelve por el mismo camino a 120 km/h, la velocidad media a que ha realizado el viaje es:

$$\frac{2}{\frac{1}{100} + \frac{1}{120}} = 109,1 \text{ km/h}$$

y no 110 km/h, como en principio se podría pensar.

Observe que la velocidad es distancia recorrida partido por el tiempo tardado en recorrerla, es decir $v = d/t$ y por tanto $t = d/v$. En nuestro caso, si la distancia a recorrer es d , el tiempo tardado en la ida es $t_1 = d/100$ y el tiempo tardado en el regreso es $t_2 = d/120$. De esta manera el tiempo total invertido en todo el recorrido ($2d$) será $t = t_1 + t_2$ y la velocidad media se calcula de la forma:

$$\text{Velocidad media} = \frac{\text{Distancia total recorrida}}{\text{Tiempo total gastado}} = \frac{2d}{\frac{d}{100} + \frac{d}{120}}$$

Cancelando d en la expresión anterior se obtiene la fórmula de la media armónica.

Otro ejemplo: Un avión recorre 3.000 km. Los 1000 primeros a 700 km/h, los 1.000 siguientes a 800 km/h, y los 1.000 restantes a 900 km/h ¿Cuál ha sido su velocidad media? No ha sido 800 km/h, sino 791,6 km/h.

3

¿Por qué en la expresión de la varianza se utiliza el cuadrado de las diferencias en vez de su valor absoluto?

El problema de la varianza es que sus unidades son el cuadrado de las unidades de los datos, y esto dificulta su interpretación. Por eso hacemos su raíz cuadrada, la desviación estándar¹, que es la medida que más usamos para referirnos a la variabilidad.

Claro que podríamos evitar el tener dos medidas –varianza y desviación estándar– utilizando solo una calculada con el módulo (valor absoluto) de la distancia de cada valor con respecto a la media (en vez del cuadrado) y así ya tendría las mismas unidades que los datos. Pero no lo hacemos porque la varianza tiene unas propiedades extraordinarias que ni de lejos presenta esa nueva medida. Vamos a desarrollar unas ideas que nos permitirán justificarlo.

Utilizaremos los datos representados en la Figura 3.1, en la que también hemos representado un valor a , en principio arbitrario, con el propósito de descubrir dónde conviene colocarlo para que sea un “buen representante” del conjunto de estos datos. Empezaremos diciendo que a puede ser cualquier número real y después le vamos a exigir algunos requisitos asociados con nuestra idea de lo que significa “buen representante” lo cual restringirá el conjunto de valores que pueda asumir. Veamos 2 criterios para seleccionar el valor de a .

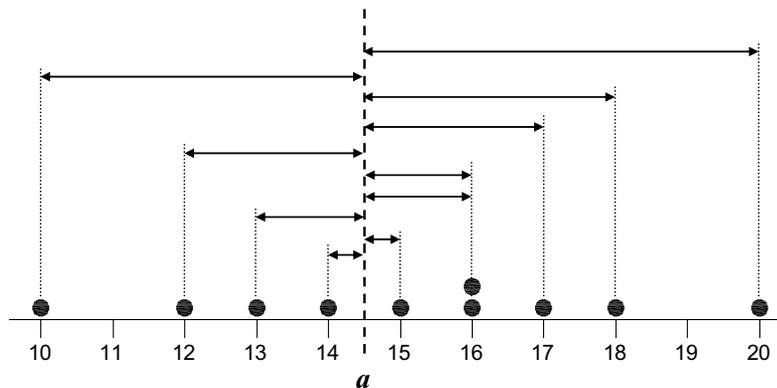


Figura 3.1. Muestra aleatoria de 10 valores, con sus distancias a un presunto valor central

Criterio 1

De todos los posibles valores posibles de a , escogemos aquel que minimice la función $f(a)$, definida de la forma:

$$f(a) = \frac{\sum_{i=1}^N |x_i - a|}{N}$$

¹ También desviación tipo, o desviación típica.