Estadísticos de Posición

Distribuciones y Promedios

Las tablas estadísticas nos presentan toda la información sobre las características de la población, o muestra, en estudio. Sin embargo, en la mayoría de las ocasiones, esta información es excesiva para que podamos captarla de forma conjunta y alcanzar una idea global que nos permita tomar decisiones. Además, presentada en forma de tablas, no somos capaces de realizar comparaciones directas entre distintos grupos de datos. Para obviar estos inconvenientes, se utilizan los promedios, que son medidas que sintetizan los datos, proporcionando un único valor numérico que nos da la posición global de la distribución. Una vez que dos, o más, distribuciones han sido resumidas en sendos números que representan a cada una de ellas, las comparaciones son fáciles e inmediatas.

Presentaremos únicamente los tres promedios más conocidos, que son la Moda, la Mediana y la Media.

Moda

Es el valor, o valores, de la variable estadística al que corresponde la mayor frecuencia.

En el caso de que exista un único valor de la variable que presente la máxima frecuencia de la distribución, se tiene una sola moda, y la distribución se dice unimodal. Si son dos los valores a los que corresponde la frecuencia máxima, la distribución es bimodal y así sucesivamente.

La moda puede calcularse para todos los tipos de variables; nominales, ordinales, cuantitativas discretas y cuantitativas continuas. En los tres primeros casos, la moda queda perfectamente definida, siendo simplemente el valor que se presenta con la mayor frecuencia. Si la variable estadística es continua, se define la clase o intervalo modal que corresponde a aquel intervalo que presenta mayor frecuencia rectificada.

Mediana

Es el valor de la variable estadística que divide al total de la población, supuesto que está ordenada, en dos efectivos iguales. Es decir, es el valor de la variable que tiene la propiedad de que la mitad de los individuos que componen la muestra o la población, toman valores inferiores o iguales a él, y la otra mitad de los individuos tienen valores superiores. Como se ve, el concepto de Mediana implica la existencia de un orden en los valores de la variable, por consiguiente, podrá ser obtenida para variables ordinales y cuantitativas, discretas o continuas.

Cálculo de la Mediana para variables ordinales y cuantitativas discretas

El procedimiento consiste en ordenar todas las observaciones y una vez hecho esto, el valor que ocupa la posición central, será la Mediana. Si el número de observaciones es impar, la Mediana existirá y será única. En el caso de que las observaciones sean un número par, si los valores centrales, ese será el valor de la Mediana, pero si ambos difieren no existirá Mediana, propiamente dicha, no obstante en aquellos casos en que los valores intermedios tengan sentido, se conviene en tomar como Mediana la semisuma de los valores centrales.

Ejemplo: A 34 participantes en un curso de técnicas de estudio, se les preguntó por el número de hermanos que tenían, obteniéndose los siguientes datos:

1, 4, 1, 0, 3, 0, 1, 2, 3, 1, 1, 2, 0, 0, 2, 1, 0, 3, 2, 2, 0, 0, 0, 2, 1, 1, 2, 1, 0, 1, 2, 1, 0, 0.

Ordenando estos valores tenemos:

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, **1**, **1**, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 4.

Como los dos valores centrales son 1, la Mediana es igual a 1.

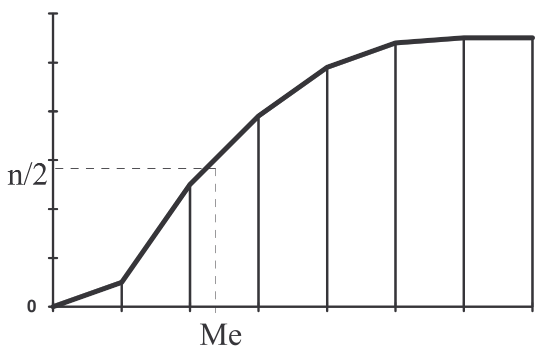
Este cálculo puede realizarse de forma equivalente cuando los datos aparecen agrupados en una tabla. En este caso la tabla sería:

Nº hermanos	Nº sujetos	Frec. acumulad.
0	11	11
1	11	**22** ←
2	8	30
3	3	33
4	1	34

La mitad de la muestra sería $34/2 = 17$. A continuación buscamos cual es la primera frecuencia acumulada que supera a 17, que es 22, y el valor correspondiente, 1, sería la Mediana.

Cálculo de la Mediana para variables cuantitativas continuas

Como vimos en el capítulo anterior, la curva de distribución nos da, para cada valor de la variable, el número de individuos que toman valores inferiores o iguales a él, por tanto el valor al que corresponda la ordenada $n/2$ será la Mediana de esa distribución, con lo cual la determinación gráfica es inmediata:



Si suponemos que dentro de los intervalos los individuos están uniformemente distribuidos, podremos determinar analíticamente la Mediana, por interpolación lineal.

De hecho el supuesto anterior, ya lo hemos usado de forma implícita, al construir la curva de distribución uniendo por segmentos rectilíneos los valores correspondientes a los extremos de los intervalos.

Pare el cálculo se procede en primer lugar a determinar el intervalo mediano, que será aquel cuya frecuencia acumulada sea la primera en superar el valor $n/2$.

Supongamos que el intervalo i-ésimo es el intervalo mediano, entonces tendremos que e_{i-1} y e_i serán sus extremos, entendiendo que son los límites reales, $a_i = e_i - e_{i-1}$ será la amplitud del intervalo y n_i , la frecuencia, nos da el número de observaciones comprendidas en el intervalo.

Por ser el intervalo mediano, sabemos que la Mediana, Me , será un valor comprendido entre e_{i-1} y e_i ; es decir $e_{i-1} \leq Me < e_i$ y también que la frecuencia acumulada de la clase anterior N_{i-1} será inferior a $n/2$. Por consiguiente la Mediana será igual a:

$$Me = e_{i-1} + x$$

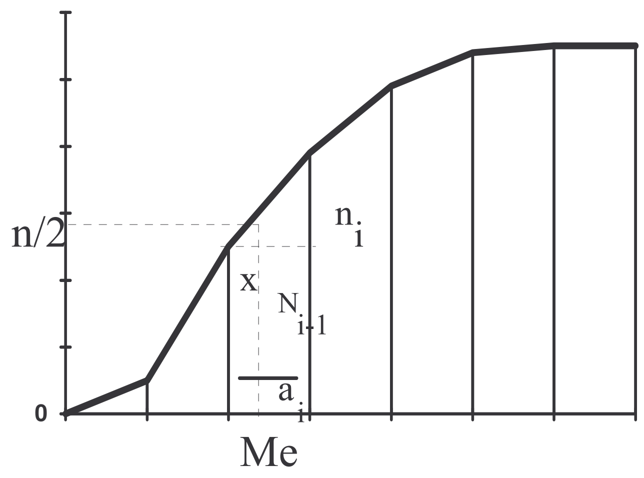
siendo x la parte proporcional de la amplitud del intervalo que corresponde al número de individuos:

$$\frac{n}{2} - N_{i-1}$$

Este valor x cumplirá la siguiente proporción:

$$\frac{x}{a_i} = \frac{\frac{n}{2} - N_{i-1}}{n_i}$$

En el gráfico siguiente se puede observar que esta proporción se deduce de una simple semejanza entre triángulos:



despejando:

$$x = \frac{\left(\frac{n}{2} - N_{i-1}\right) \cdot a_i}{n_i}$$

y por consiguiente:

$$Me = e_{i-1} + \frac{\left(\frac{n}{2} - N_{i-1}\right) \cdot a_i}{n_i}$$

Ejemplo: Si volvemos, una vez más, a las puntuaciones de la escala de Ansiedad Estado:

Puntuación	Nº de pacientes	Frec. acumulad.
0 - 10	10	10
10 - 20	40	50
20 - 30	28	**78** ←
30 - 40	20	98
40 - 50	10	108
50 - 60	2	110

observamos que n/2 vale 55 . Por consiguiente, la primera frecuencia acumulada que supera 55 es 78 , y la Mediana estará en el intervalo 20 - 30. Aplicando la formula, anteriormente deducida:

$$Me = 20 + \frac{(55 - 50) \cdot 10}{28} = 21{,}8$$

Resumiendo diremos que la Mediana es un promedio que depende de las observaciones solo por su orden y no por su valor. Es decir, si se disminuye el valor de una o varias observaciones inferiores a la mediana, o se aumenta el valor de una o varias observaciones superiores a ella, la Mediana permanece invariable. La Mediana presenta las ventajas de su facilidad de cálculo y de ser relativamente insensible a las fluctuaciones de los valores extremos de la distribución.

Media

La media, o media aritmética simple, de una variable estadística es una medida de posición que se define como la suma de todos los valores de la variable, dividida por el número de observaciones.

Las afirmaciones acerca de la media de una distribución solo tendrán significado cuando la variable haya sido medida, al menos, en una escala de intervalos, careciendo del mismo en el caso de variables nominales u ordinales.

La propia definición de media nos proporciona la formula para su cálculo, que será:

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

Si los datos vienen resumidos en una tabla estadística, esto nos indica que cada valor x_i se presenta tantas veces como indica su frecuencia n_i y por consiguiente la formula a aplicar en ese caso será:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n} = \sum_{i=1}^{k} x_i f_i$$

En el caso de variables continuas, cuyos valores estén agrupados en intervalos, utilizando una vez más el supuesto de que los valores están igualmente distribuidos dentro del intervalo, puede calcularse la media considerando como valores los centros de los intervalos, o marcas de clase, es decir:

$$\bar{x} = \frac{\sum_{i=1}^{k} c_i n_i}{n}$$

De entre las múltiples propiedades de la media, señalaremos solamente dos. La primera afirma que en cualquier distribución, la media de las diferencias a la media es cero.

La segunda nos indica que si sometemos una variable a una transformación lineal, su media experimenta igual transformación.

Ejemplo: Si las puntuaciones de 10 sujetos en una escala de autoconcepto son las siguientes:

52; 71; 52; 58; 55; 48; 53; 52; 57; 62

La media de estas puntuaciones es 56. Si ahora restamos, a todos los valores, 6 puntos y el resultado lo dividimos por dos, tendremos las siguientes puntuaciones:

23; 32,5; 23; 26; 24,5; 21; 23,5; 23; 25,5; 28

cuya media podemos saber que es 25, sin necesidad de calcularla, ya que:

$$\frac{56 - 6}{2} = 25$$

Cuantiles

El concepto de Mediana puede ser generalizado mediante los llamados cuantiles. Un cuantil de orden r, será el valor de la variable que deja por debajo de si una proporción r de las observaciones, supuestamente ordenadas. Es decir habrá r·n sujetos con valores inferiores o iguales al cuantil r. Así la Mediana resulta ser el cuantil de orden 1/2. Los cuantiles más utilizados son los cuartiles, deciles y percentiles.

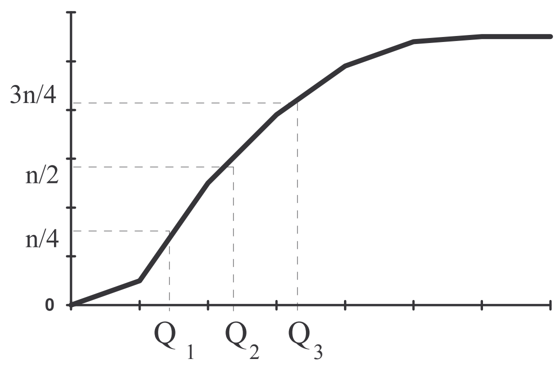
Cuartiles

Los cuartiles dividen a la población en cuatro efectivos iguales. Existen tres cuartiles, designados Q_1, Q_2, Q_3. El primer cuartil Q_1, se define como el valor de la variable que deja por debajo de si la cuarta parte de los individuos con puntuaciones inferiores.

El segundo cuartil Q_2, es el valor de la variable que deja por debajo de si las dos cuartas partes de la población, o de la muestra, con puntuaciones inferiores. Obviamente el segundo cuartil coincide con la Mediana.

El tercer cuartil Q_3, será el valor de la variable que deja por debajo de si las tres cuartas partes de los sujetos con menores puntuaciones. Esto es lo mismo que decir que quedarán por encima del tercer cuartil, la cuarta parte de los sujetos con mayores puntuaciones.

Los cuartiles pueden determinarse gráficamente en la curva de distribución, sin más que buscar los puntos de ordenada, n/4, 2n/4 y 3n/4, como vemos en la gráfica siguiente:



El cálculo de los cuartiles es totalmente semejante al cálculo de la mediana, veámoslo en el siguiente ejemplo:

Supongamos que en una muestra de 40 familias hemos registrado el número de hijos, una vez ordenados los datos, tenemos los valores siguientes:

0, 0, 0, 0, 1, 1, 1, 1, 1, **1**, **1**, 1, 1, 1, 1, 1, 1, 1, 1, **1**, **1**, 2, 2, 2, 2, 2, 2, 2, 2, **2**, **2**, 2, 3, 3, 3, 3, 3, 4, 4, 5

No hay más que contar la cuarta parte de los sujetos (10), las dos cuartas partes (20) y las tres cuartas partes (30), para determinar que el primer cuartil es 1, el segundo cuartil 1 y el tercer cuartil 2.

Si los datos anteriores los agrupamos en una tabla tendremos:

Nº hijos	Nº familias	Frec. acumul.
0	4	4
1	17	21 ←
2	11	32 ←
3	5	37
4	2	39
5	1	40

Si observamos la tabla, vemos que 21 es la primera frecuencia acumulada que supera a 10 (La cuarta parte de la muestra), por consiguiente el primer cuartil es el valor de la variable correspondiente a esa frecuencia acumulada 1. También es 21 la primera frecuencia acumulada que supera a 20 (Las dos cuartas partes de la muestra) y por ello el segundo cuartil también vale 1. Por último, 32 es la primera frecuencia acumulada que supera a 30 (Las tres cuartas partes de la muestra) y el valor correspondiente 2 es el tercer cuartil.

En el caso de variables, cuyos valores aparezcan agrupados en intervalos, se procede a realizar una interpolación lineal dentro del intervalo que contenga al correspondiente cuartil. Esto da origen a expresiones, para el cálculo de los cuartiles, análogas a la obtenida para la Mediana. Damos a continuación estas expresiones para el primer y tercer cuartil:

$$Q_1 = e_{i-1} + \frac{\left(\frac{n}{4} - N_{i-1}\right) \cdot a_i}{n_i}$$

$$Q_3 = e_{i-1} + \frac{\left(\frac{3n}{4} - N_{i-1}\right) \cdot a_i}{n_i}$$

Ejemplo: La siguiente tabla nos da las puntuaciones de 78 personas, participantes en un curso, en el factor verbal del T.E.A.

Puntuaciones	Nº personas	Frec. acumulad.
7 - 13	5	5
14 - 20	21	26
21 - 27	31	57
28 - 34	17	74
35 - 41	4	78

Para obtener el primer cuartil, comenzamos por calcular la cuarta parte de la población que es 78/4 = 19,5. Como 26 es la primera frecuencia acumulada que supera a 19,5, sabemos que el primer cuartil se encuentra en el intervalo 14 - 20. Aplicando la formula anterior tenemos:

$$Q_1 \quad = \quad 13{,}5 + \frac{(19{,}5 - 5)\cdot 7}{21} \quad = \quad 18{,}3$$

obsérvese que en los cálculos se han utilizado los límites reales del intervalo 13,5 - 20,5 y por ello la amplitud es 7. Análogamente el tercer cuartil sería:

$$Q_3 \quad = \quad 27{,}5 + \frac{(58{,}5 - 57)\cdot 7}{17} \quad = \quad 28{,}1$$

<u>Deciles y Percentiles</u>

Los deciles son medidas de posición que dividen a la población, o la muestra en diez efectivos iguales. Habrá nueve deciles que se denotan D_1, D_2, ... D_9. De forma general, el decil h, D_h será el valor de la variable que deja por debajo de si las h décimas partes de los sujetos con menores puntuaciones.

Análogamente, los Percentiles serán aquellos valores de la variable que dividen a los sujetos en 100 efectivos iguales. Habrá noventa y nueve Percentiles, que se denotarán por P_1, P_2, ... P_{99}. Siendo cada uno de ellos el valor de la variable que deja por debajo de si el correspondiente tanto por ciento de los sujetos con menores puntuaciones. Por ejemplo, el percentil treinta y dos, P_{32} será un valor tal que el 32% de los individuos toman valores inferiores o iguales a él.

El cálculo de deciles y percentiles será idéntico a los que hemos realizado para Mediana y cuartiles. En concreto para el caso de una variable cuyos datos vienen agrupados en intervalos las fórmulas de cálculo serán:

$$D_h \quad = \quad e_{i-1} + \frac{\left(\frac{h\cdot n}{10} - N_{i-1}\right)\cdot a_i}{n_i}$$

$$P_g \quad = \quad e_{i-1} + \frac{\left(\frac{g\cdot n}{100} - N_{i-1}\right)\cdot a_i}{n_i}$$

Ejemplo: Calculemos el percentil treinta y dos de la distribución de las puntuaciones en el factor verbal del T.E.A. que vimos antes.

Puntuaciones	Límites reales	Nº personas	Frec. acumulad.
7 - 13	6,5 - 13,5	5	5
14 - 20	13,5 - 20,5	21	26
21 - 27	20,5 - 27,5	31	57
28 - 34	27,5 - 34,5	17	74
35 - 41	34,5 - 41,5	4	78

Calculamos el 32% de los sujetos, que es 24,96. Buscamos la primera frecuencia acumulada que supere 24,96, que es 26, correspondiente al intervalo de límites reales 13,5 - 20,5, por consiguiente el percentil 32 será:

$$P_{32} \quad = \quad 13{,}5 + \frac{(24{,}96 - 5) \cdot 7}{21} \quad = \quad 20{,}15$$

Escalas de Cuantiles

Hemos visto que si los datos se colocan ordenados la Mediana los divide en dos efectivos de igual cantidad de sujetos, que podríamos definir como la mitad con puntuaciones bajas y la mitad con puntuaciones altas. De igual manera, los cuartiles clasifican a los sujetos en cuatro grupos que podemos denominar, la cuarta parte con puntuaciones más bajas, la cuarta parte con puntuaciones bajas, la cuarta parte con puntuaciones altas y la cuarta parte de los sujetos con mayores puntuaciones. Análogamente, los deciles y percentiles dividirán a los sujetos en 10 y 100 grupos. Este hecho es utilizado habitualmente en Psicología para construir escalas fácilmente interpretables.

Por ejemplo, la escala de Ansiedad Estado del Cuestionario STAI, que hemos utilizado como ejemplo, proporciona una puntuación en una escala de 0 a 60 puntos. Sin embargo, esta puntuación no es interpretable directamente. Si decimos que un adolescente varón ha alcanzado una puntuación de 30, no podemos afirmar que su ansiedad sea media, o que otro con una puntuación de 18 tenga una ansiedad baja, puesto que los sujetos no tienen porqué estar distribuidos uniformemente a lo largo de la escala, pudiendo agruparse en un extremo u otro, o estar muy concentrados en el centro.

Por el contrario, una escala basada en los deciles o percentiles es interpretable en términos de un grupo normativo. Si decimos que en una muestra de 146 adolescentes varones normales los deciles fueron los siguientes:

D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
10	12	14	17	20	24	28	34	38

observaremos que la puntuación de 30, a pesar de estar en la mitad de la escala de posibles puntuaciones, representa una ansiedad notable, pues más del 70% de los sujetos normales presentan un nivel de ansiedad inferior. De igual manera, una puntuación de 18 no es tan baja como aparentaba, el 40% de los sujetos del grupo tienen un nivel de ansiedad inferior a 18.