UNIVERSIDAD NACIONAL DE SAN LUIS

FACULTAD DE PSICOLOGÍA

METODOLOGÍA DE LA INVESTIGACIÓN I

Licenciatura en Psicología

Licenciatura en Psicomotricidad

Prof. Responsable: Dr. HORACIO DANIEL GARCIA

Trabajos prácticos: Lic. DANIEL PITONI, Lic. MAXIMILIANO SAPINO y Mg. ELIANA ZÁRATE

Unidad 4: Medidas de posición, dispersión y forma
Parámetros y estadísticos. Medidas de tendencia central: media aritmética, mediana y moda. Medidas de posición no central: cuartiles, deciles y percentiles. Medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación. Medidas de forma: asimetría y curtosis.

Autor: HORACIO DANIEL GARCIA

Año 2023

Parámetros y estadísticos

Tal como habíamos visto con anterioridad, la posibilidad de recurrir a una muestra para indagar las características de una población, es una estrategia ampliamente utilizada por los investigadores. Igualmente, cuando es posible acceder a la totalidad de las personas que componen una población, siempre estaremos en mejores condiciones para obtener conclusiones sólidas; claro está que esta posibilidad requiere un esfuerzo claramente mayor.
Estas dos posibilidades se relacionan con los dos tipos de estrategias en investigación: la estadística descriptiva y la estadística inferencial.

- Estadística descriptiva: Como su nombre lo indica, busca detallar, representar, analizar los datos concretos que ha obtenido el investigador, sin pretender ir más allá y abordar algún supuesto que pueda derivarse de sus datos. Expresado de manera técnica, podemos decir que la estadística descriptiva se orienta a *analizar todo el conjunto de datos que ha recolectado el investigador, permitiéndole extraer conclusiones válidas, únicamente para ese conjunto en particular*.
  En otras palabras, si el investigador ha trabajado con los datos de una población (no con una muestra), podrá utilizar sus análisis para describir la población que ha investigado. En cambio, si el investigador ha obtenido la información de una muestra, no podrá generalizar los resultados con el objeto de describir a la población (en este caso estaría haciendo una inferencia), por lo que deberá limitarse a la descripción de los datos que posee, es decir: de la muestra. Desde esta estrategia, si necesitamos llegar a conclusiones que abarquen una población, solamente podremos hacerlo estudiando la totalidad de los participantes que se incluyen en ese espacio.
  Teniendo en cuenta este importante aspecto, podemos decir que los análisis más frecuentes que se hacen bajo esta modalidad, son diversas medidas de resumen que sirven para caracterizar las variables en estudio: Frecuencias, Porcentajes, Promedios, Medidas de dispersión, Medidas de forma, etc.

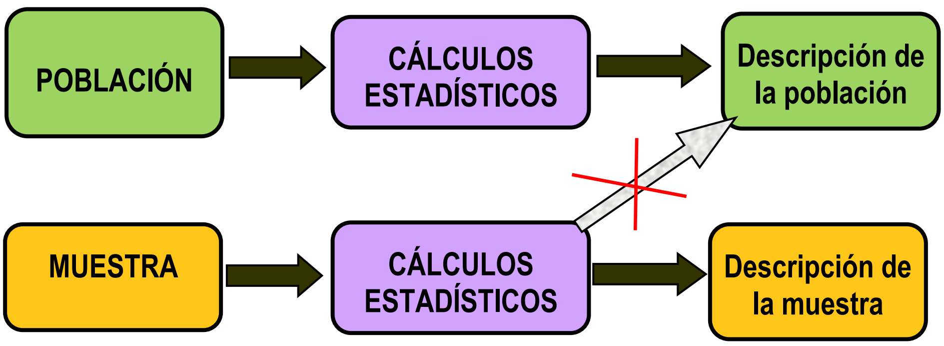


Figura 1: esquema procesual de la estadística descriptiva

- Estadística inferencial: La palabra inferencia alude a la suposición que hace el investigador en base a argumentos probabilísticos sólidos sobre un aspecto de la realidad que desconoce. Así, esta estrategia *pretende obtener conclusiones generales de una población, estudiando una muestra representativa sacada de ella*. O, dicho de otro modo, investiga a una población, valiéndose de los datos y resultados que se obtienen de una muestra representativa. Los análisis que suelen realizarse desde esta perspectiva son múltiples; en nuestra asignatura veremos: Estimaciones intervalares de parámetros, Diferencias entre grupos, Asociación entre variables, Predicción de resultados, etc.

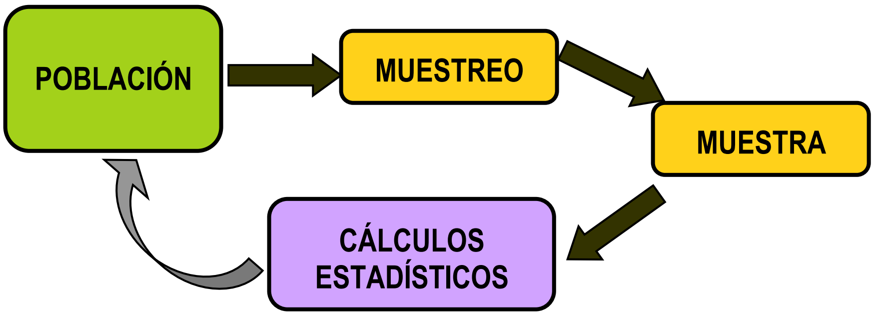


Figura 1: esquema procesual de la estadística inferencial

Ahora bien, ¿qué son los Parámetros y los Estadísticos? ¿Qué diferencia hay entre ellos?
Primero que nada, debes saber que cuando estamos hablando de Parámetros o Estadísticos nos estamos refiriendo a indicadores numéricos que se obtienen del análisis de los datos, y que la diferencia entre ambos es el universo de datos que refieren. Veamos a continuación…
Un *Parámetro* es un valor, una medida, un indicador representativo de una característica (de una variable) de la POBLACIÓN, mientras que el *Estadístico* es el valor, medida o indicador que describe una característica (una variable) de la MUESTRA y que, si se cumplen determinadas condiciones, puede llegar a servir para realizar una estimación de los parámetros de la población.

Como te puedes imaginar, con mucha frecuencia los datos se obtienen de una muestra que ha sido extraída de una población, por lo que de ella se obtienen los estadísticos que describirán la muestra. Algunas veces, si el investigador conoce los valores que caracterizan la población (parámetros), podrá compararlos con los obtenidos con la muestra (estadísticos). En el caso de que no conozca los parámetros, si se cumplen ciertas condiciones relacionadas con la representatividad de la muestra, podrá *inferirlos* (recuerda que de ahí proviene el nombre de estadística inferencial).

Figura 3: Diferencia entre Parámetro y Estadístico. Tomado de http://matematicas-tec31.blogspot.com/2014/10/estudio-estadistico-y-diseno-de-una.html

Entonces, como venimos viendo, Parámetro es cualquier medida descriptiva de una población. Para identificar esos indicadores se utilizan las letras griegas; por ejemplo, μ (se pronuncia mu) para la media poblacional y σ (se pronuncia sigma) para la desviación estándar poblacional. En tanto que cualquiera de las medidas descriptivas de una muestra

(Estadístico), se las representa mediante letras minúsculas de nuestro alfabeto; por ejemplo: media (x), desviación estándar (s).

Tanto los parámetros, como los estadísticos, se los puede agrupar en las siguientes categorías:

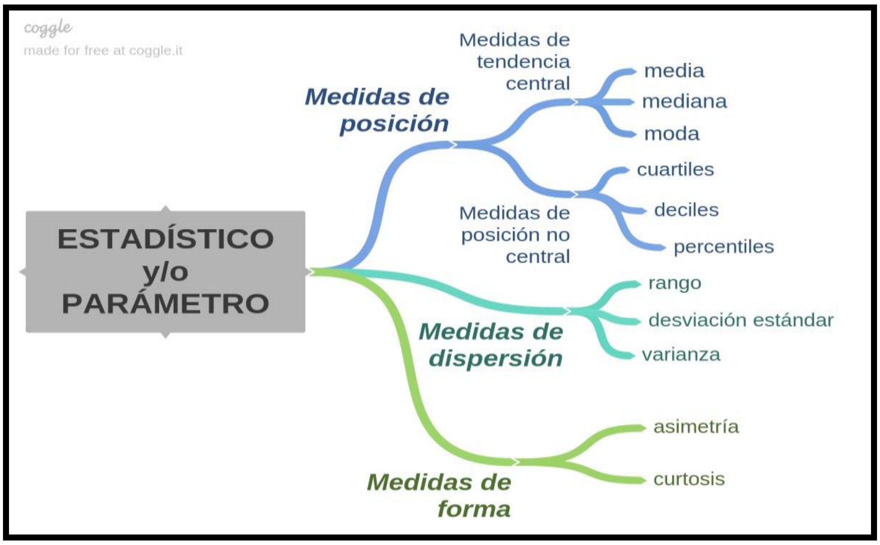


Figura 4: Clasificación de Parámetros y Estadísticos.

Una vez que tenemos en claro a qué nos referimos cuando hablamos de parámetro y/o estadístico, podemos profundizar en algunos de los índices que existen, y que sirven tanto para describir una muestra o una población. Recuerda que la utilidad de estos (estadísticos o parámetros) se hace evidente cuando hemos recolectado información de diferentes observaciones y es necesario encontrar algún modo de resumirla para que cobre sentido y podamos interpretarla.

Medidas de Posición

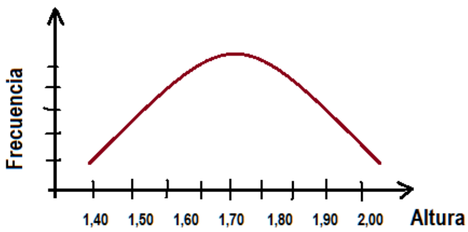
Normalmente, un valor puede ayudarnos a caracterizar suficientemente a una muestra, o bien a la población. En este sentido, las medidas de posición ofrecen un conjunto de indicadores de gran utilidad. Se las puede clasificar en dos grandes grupos: a) medidas de tendencia central, y b) medidas de posición no central.

Medidas de tendencia central: media aritmética, mediana y moda

Para entender mejor es conveniente comenzar imaginando que los (muchos) datos que tenemos de la muestra, o de la población, se distribuyen en una gráfica conforme los valores que se han obtenido. Imagina que estás investigando la altura de tus compañeros de estudio. Lo primero que vas a notar es que cada dato registrado es importante, pero, sin embargo, al instante te darás cuenta que necesitas alguna forma de resumir esa información porque a medida que vayas teniendo más datos te será más difícil poder explicarlos o describirlos.

Gráfico 1.

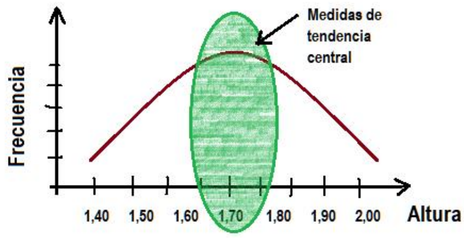
Distribución de la variable Edad



Como puedes observar en el gráfico de arriba, parece que los valores se distribuyen desde 1,40 mts. a poco más de 2 mts. Ahora bien, posiblemente te des cuenta, incluso con solo ver a tus compañeros, que muy pocos tienen la altura del valor mínimo, al igual que muy pocos son tan altos como el valor máximo que has encontrado; entonces posiblemente te interese saber cómo se caracterizan los valores intermedios. Justamente, las medidas de tendencia central tienen como función resumir en un solo valor, un conjunto de valores que representan un centro (un punto de equilibrio) de la distribución de los datos.

Gráfico 2.

Representación de la ubicación de las medidas de tendencia central



Ahora veamos cada una de ellas:

a) ***Media:*** es probablemente uno de los estadísticos más usado, posiblemente lo conozcas ya que también se le suele llamar promedio o media aritmética. Permite comparar poblaciones o muestras y se suele interpretar como un *punto de equilibrio de los valores que conforman el conjunto de datos*. Aunque es necesario acompañarla de una medida de dispersión, para que la información que brinda sea más completa, es una excelente medida de resumen que nos sirve para caracterizar o describir un fenómeno.

Definición: la media aritmética de una variable es la suma de todos los valores de dicha variable dividida por el número total de valores considerados. Se la representa con la letra $\bar{x}$

Fórmula: $$Media(X) = \overline{x} = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

Ejemplo: tenemos una muestra de niños con diferentes alturas….

Sumamos 120 + 117 + 128 + 111 + 138 + 141 + 132 = **887 cm,** ahora a este valor lo dividimos por la cantidad de datos que hemos considerado, que es 7.

887 / 7 = **126,71 cm**

Es decir, que la media (el promedio) de altura de los niños fue de 126,71 cm ($\bar{x}$ = 126,71).

b) ***Mediana:*** Este estadístico representa el *punto de equilibrio de la cantidad de datos*, ya que identifica el valor central que divide la distribución de datos justo por la mitad. El valor de la mediana va a determinar el lugar, el punto, donde se articula el 50% de los datos con valores menores a la mediana y el 50% de los datos con valores mayores a la mediana. La mediana se representa por (Me).

¿Cómo se calcula? Veamos, tenemos los siguientes datos que representa la cantidad de integrantes que tiene las familias de nueve participantes que componen la muestra….

Tabla 1.
Matriz de datos

Participante	**Cantidad de integrantes que tiene en su familia**
01	1
02	9
03	6
04	4
05	4
06	5
07	6
08	3
09	2

- Primero, lo que tenemos que hacer es ordenar los datos de manera creciente

1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**

- Luego vemos el dato que deja igual cantidad de datos a la izquierda y a la derecha

1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**

- En este caso, como podrás observar es el **Cuatro (Me= 4)**

*Nota: observa que a la izquierda de la mediana quedan la misma cantidad de datos que a la derecha

Cuando la cantidad de datos es impar, como en el caso de arriba, es más simple que cuando la cantidad de datos es par…. Veamos que sucede cuando se da esta situación, para ello le agregaremos un número (9) a la derecha del ejemplo anterior

1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**	**9**

- En este caso el punto medio no cae en un valor específico sino que cae entre el 4 y el 5, por lo que tomaremos ambos números y calcularemos el valor de media. Esto es: (4 + 5) = 9 / 2 = **4,5**
- En este caso, el valor de la mediana sería 4,5 **(Me= 4,5)**

c) **Moda:** es el dato que más se repite o, dicho de otro modo, aquel que tiene mayor frecuencia absoluta. Se representa por Mo. Es importante que sepas que si hay dos valores que tienen la mayor frecuencia absoluta, ambos representarán la moda (en este caso decimos que la distribución de datos es bi-modal, o si son más de dos, diremos que es multimodal) y, en el caso que no se repita ningún valor, decimos que no hay Moda.

Veamos el ejemplo que tomamos para la mediana para ver qué Moda nos da

1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**

- En este caso observamos que tanto el 4 como el 6 se repiten dos veces, por lo tanto, decimos que la distribución de los datos es bi-modal y que sus valores son **4 y 6 (Mo= 4 y 6).**

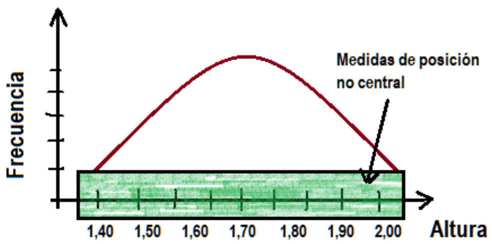
Medidas de posición no central: cuartiles, deciles y percentiles

Las medidas de posición no central identifican puntos característicos de una serie de valores, que no necesariamente tienen que ser centrales. Lo que buscan estas medidas *es "dividir el conjunto de datos en grupos con la misma cantidad de valores o de datos".*

Si bien la utilidad de estas medidas es variada, se los suele utilizar para dividir las muestras o poblaciones en grupos. Por ejemplo, si tomamos la variable salario quizás desee saber entre qué sueldos se encuentra el 25% de la población más humilde.

Gráfico 3.

Representación de la ubicación de las medidas de posición no central



a) ***Cuartiles:*** estos estadísticos dividen la distribución de los valores de la variable en 4 partes, cada una de las cuales engloba el 25% de los mismos. Los símbolos de estas medidas son: Q_1 (primer cuartil que deja a su izquierda el 25% de los datos); Q_2 (segundo cuartil que deja a su izquierda el 50% de los datos y coincide con la mediana), y Q_3 (tercer cuartil que deja a su izquierda el 75% de los datos). Abajo te mostramos cómo se calcula. Retomaremos el ejemplo del cálculo de la mediana:

- Primero, ordenamos los datos de manera creciente

1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**

- Luego vemos el dato que deja igual cantidad de datos a la izquierda y a la derecha

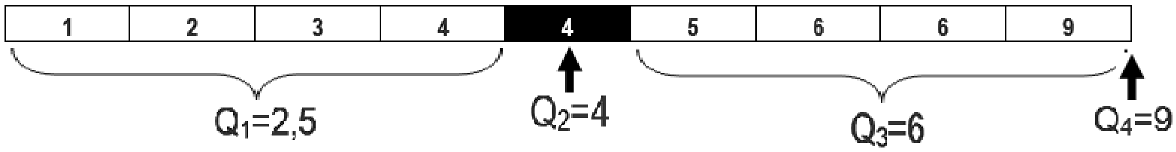
1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**

- Con ello, hemos identificado la mediana **(Me= 4),** que siempre coincidirá con el límite superior del cuartil 2. Entonces diremos que $Q_2=4$

1	**2**	**3**	**4**	**4**	**5**	**6**	**6**	**9**

$Q_2=4$

- A continuación, para encontrar el cuartil 1 y 3 buscaremos el valor deja igual cantidad de datos a la izquierda y a la derecha en los dos corchetes. En ambos casos veremos que tenemos cuatro datos, el valor que dividiría con exactitud estos casos se encuentra entre 2 y 3 para el segmento de la izquierda y entre 6 y 6 para el de la derecha. Como el punto de equilibrio no cae en un número determinado, sino entre dos, lo que debe hacerse es promediar cada par. Por ejemplo: el promedio de 2 y 3 es 2,5, en tanto que el promedio de 6 y 6 es 6.



b) ***Deciles:*** se obtienen de manera similar que los cuartiles. Son 9 valores que segmentarán la serie de datos, ordenada de forma creciente o decreciente, en diez grupos iguales de datos. Cada Decil contiene el 10% de los datos.

c) ***Percentiles:*** en este caso son 99 valores que dividen los datos, ordenados de forma creciente o decreciente, en cien partes iguales, concentrando cada uno de ellas el 1% de los datos.

Medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación

Las medidas de dispersión, como sugiere la palabra, miden el grado de dispersión de los valores de la variable. Nos dicen qué tanto difieren los datos entre sí, es decir, cuánto se distancian unos de otros. Por este motivo, las medidas de dispersión son un excelente complemento a las medidas de tendencia central, ya que nos informan si los datos que tenemos tienden a presentarse de manera concentrada (datos homogéneos), o si por el contrario están muy dispersos entre sí (datos heterogéneos).
Si las puntuaciones de una variable están muy alejadas de la media, mayor será la variabilidad de los datos obtenidos; en cambio, si las puntuaciones están bien cerca de la media los datos estarán menos dispersos y por lo tanto serán más homogéneos respecto de la media. De este modo, se sabe cuán parecidos o cercanos son los valores de los datos, o si varían mucho entre ellos.

Para explicártelo de una forma más concreta vamos a recordar el ejemplo que te ofrecimos para la media de altura:
120 + 117 + 128 + 111 + 138 + 141 + 132 = **887 cm**
887 / 7 = **126,71 cm**

Ahora, observá qué sucede si cambiamos algunos valores….
<u>80</u> + 117 + 128 + 111 + 138 + 141 + <u>172</u> = **887 cm**
887 / 7 = **126,71 cm**

Como podrás observar, cambiamos el primer y último valor (los que están subrayados). Sin embargo, habrás notado que conservamos el mismo valor de media (es decir que, a pesar de tener datos más separados o heterogéneos obtuvimos la misma media). De esta manera, las medidas de dispersión vienen a aportar información que las medidas de tendencia central no pueden ofrecer.

a) ***Rango o recorrido***: es la medida de dispersión más sencilla de calcular. Es la diferencia que existe entre el valor mayor y el valor menor que toma la variable. Con este indicador tenemos información de la amplitud de valores que toma la variable, dándonos una idea de la variabilidad de los datos

Tomaremos como ejemplo los valores del cálculo de media del primer caso:
120; 117; 128; 111; 138; 141; 132

Identificamos el valor mayor (141) y luego el menor (111); posteriormente realizamos la resta de los dos valores: 141 – 111 = **30**

Podemos decir entonces que el **rango de la variable Altura es 30 cm. (caso 1).**

Ahora haremos lo mismo con los valores que modificados:

80; 117; 128; 111; 138; 141; 172

Identificamos el valor mayor (172) y luego el menor (80); restamos 172 – 80 = **92**

En este caso tenemos que el **rango de la variable Altura es 92 cm. (caso 2)**. Como puedes observar, en el caso 2 tenemos una mayor dispersión o variabilidad de los datos.

b) ***Varianza:*** su propósito es establecer una medida de la variabilidad de los valores de la variable. Mide la dispersión de los datos en una variable (cuánto se separan) respecto a la media, calculando la media de los cuadrados de las distancias de todos los datos.

Valores elevados de varianza indican que los datos están distantes de la media, más separados, más dispersos. Supongamos que la varianza de la variable altura en la muestra A es de 12^2, mientras que en la muestra B es de 10^2; con esos datos podríamos decir que la muestra A tiene mayor dispersión de datos (más lejos de la media). La varianza se representa con el símbolo σ^2 (sigma al cuadrado) para el universo o población, y con el símbolo s^2 (s al cuadrado), cuando se trata de la muestra.

c) ***Desviación estándar:*** es la raíz cuadrada de la varianza. Se representa por σ (sigma) cuando pertenece al universo o población, y por "s", cuando pertenece a la muestra. Como verás tiene mucha relación con la varianza ya que se obtiene de ella calculándole la raíz cuadrada. La interpretación es similar a la varianza; solo que ésta se expresa en unidades de variable al cuadrado y la desviación estándar simplemente en unidades de variable. La varianza es menos usada porque no expresa las mismas unidades que los datos, ya que las desviaciones están elevadas al cuadrado. Si los datos fueran en metros, la varianza denotaría metros cuadrados y eso induciría a confusiones con una medida de superficie. En cambio, en la desviación estándar esto se corrige, ya que expresa las mismas unidades que los datos.

d) ***Coeficiente de variación:*** se utiliza para comparar distintos conjuntos de datos ya que nos ofrece un valor que representa la dispersión relativa. Como la correcta interpretación de la desviación estándar depende del valor de media, cuando comparamos dos conjuntos de datos tenemos que observar el valor de media y la desviación estándar de un conjunto y luego compararla con la media y desviación estándar del otro conjunto.

Como ese proceso es complejo, y muchas veces impreciso, se ha propuesto sintetizar los valores en un coeficiente. Es decir que en el caso de comparar dos conjuntos de datos ya no tendremos dos valores de media y dos valores de desviación estándar, sino que tendremos un coeficiente de variación por cada conjunto de datos.

El coeficiente de variación elimina las posibles distorsiones de las medias de dos o más poblaciones. Se obtiene de dividir la desviación típica por el valor absoluto de la media y por lo general se expresa en porcentaje para su mejor comprensión. Se calcula del siguiente modo: $C_V = \frac{s}{|\bar{x}|} \cdot 100$

Medidas de forma: asimetría y curtosis

Estos estadísticos permiten comprobar la forma que se distribuyen los datos en una variable determinada. Esto es muy importante, tanto como para determinar el comportamiento de los datos, como para adoptar las herramientas más adecuadas para su posterior procesamiento estadístico.

Es importante que recuerdes que las variables aleatorias continuas suelen caracterizarse por el hecho de que los datos tienden a acumularse en torno a un valor central, coincidente con la media, y que van decreciendo su frecuencia de forma aproximadamente simétrica a medida que se alejan por ambos lados de dicho valor. Los gráficos (histogramas) de estas variables continuas se asemejan a lo que se conoce como campana de Gauss, que es el modelo matemático de la distribución normal, siendo la distribución que con mayor frecuencia aparece en multitud fenómenos reales.

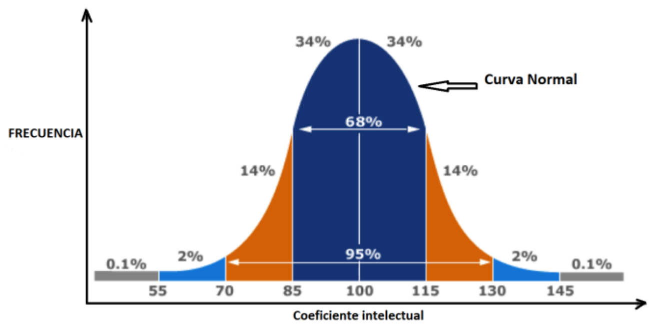


Figura 3: Ejemplo de una Distribución Normal (Campana de Gauss) para la variable Coeficiente intelectual.

Como puedes ver en la gráfica; el 0,1% de la población tiene un Coeficiente intelectual de hasta 55 puntos; un 2% tiene entre 55 y 70; un 14% entre 70 y 85 puntos; un 68% entre 85 y 115; un 14% entre 115 y 130; un 2% entre 130 y 145; y sólo un 0,1% presenta un coeficiente intelectual por arriba de 145.

Ahora bien, no siempre la distribución de las variables asume la forma de una *curva normal*. En ocasiones, los datos suelen estar concentrados no en el centro, sino en los valores más bajos, otras veces en los valores más altos, y/o puede verse la curva con forma puntiaguda, elevada o bien aplanada. De aquí surgen las nociones de asimetría y de curtosis:

a) ***Asimetría:*** indica dónde se ubica la mayor concentración de datos. En una distribución simétrica, la media, mediana y moda coinciden en su valor (x = Me = Mo), y la mayoría de los datos se encuentran en la mitad del recorrido de la variable. En una distribución asimétrica positiva, la moda es menor a la mediana, y ésta a su vez menor que la media (Mo < Me <x), por lo que los datos se concentran en los valores más bajos. En una distribución asimétrica negativa, la moda es mayor a la mediana, y ésta a su vez mayor que la media (Mo > Me > x), por lo que los datos se concentran en torno de los valores más altos de la variable.

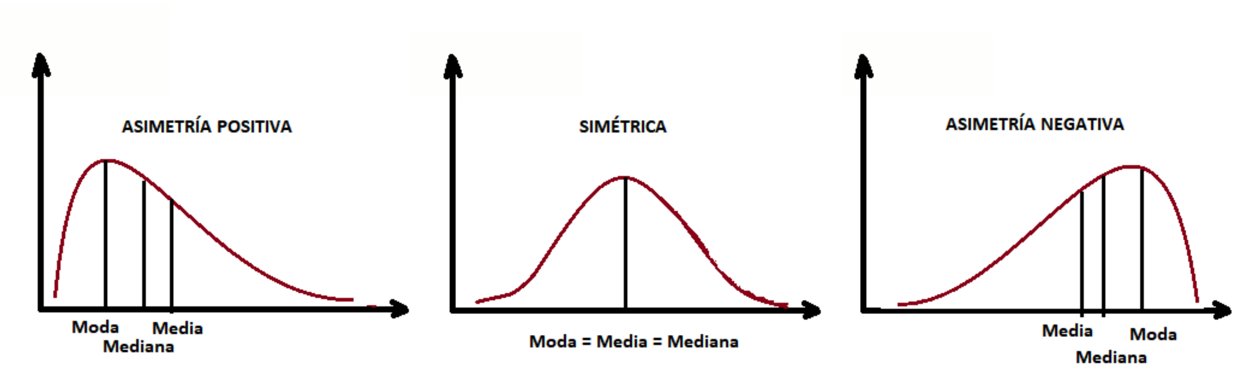


Figura 4: Tipos de Asimetría

El coeficiente de asimetría es un estadístico que permite determinar el grado de asimetría de la distribución con respecto a la media. Los valores positivos de este estadístico reflejan que la variable presenta una Asimetría Positiva. Por el contrario, un valor negativo nos dice que la distribución presenta una Asimetría Negativa. En general, se suele tomar que el coeficiente, en un rango de ±0,5, se considera indicador de una distribución simétrica. Valores menores a -0,5 describen una distribución asimétrica negativa, en tanto que mayores a +0,5 describen una distribución asimétrica positiva.

b) ***Curtosis:*** este estadístico nos dice qué tan puntiaguda o achatada se encuentra una distribución en comparación con la distribución normal. Supongamos que los datos se han concentrado, y mucho, en torno a la media, entonces tenemos una distribución leptocúrtica. En este caso la gráfica suele mostrar una curva con mucha pendiente, bien alta y puntiaguda.

   Si por el contrario los datos están muy dispersos, la distribución es platicúrtica (porque tiene forma de plato aplanado, chato). En este caso veremos que los datos tienden a concentrarse menos en torno a los valores medios de la variable.

   La curva normal, la que no es ni puntiaguda ni aplanada, se la denomina mesocúrtica. Del mismo modo que sucede con la asimetría, es poco frecuente encontrar un coeficiente de Curtosis de cero (0), por lo que se suelen aceptar que los valores comprendidos entre ±0,5 describen distribuciones mesocúrticas. Valores menores a -0,5 describen una distribución platicurtica, en tanto que mayores a +0,5 describen una distribución leptocurtica.

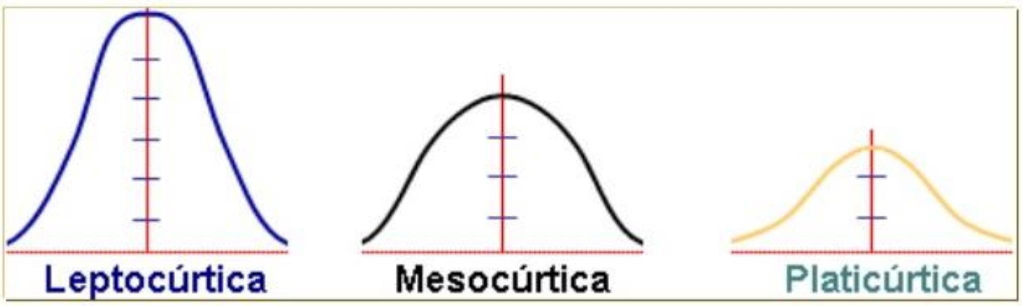


Figura 5: Tipos de Curtosis

Muy bien, ahora veremos un análisis realizado con Jamovi

Tomaremos de la base de datos la variable Afecto Positivo y le solicitaremos a Jamovi calcular todos los estadísticos descriptivos posibles. El procedimiento sería: Análisis>Exploración> Descriptivas

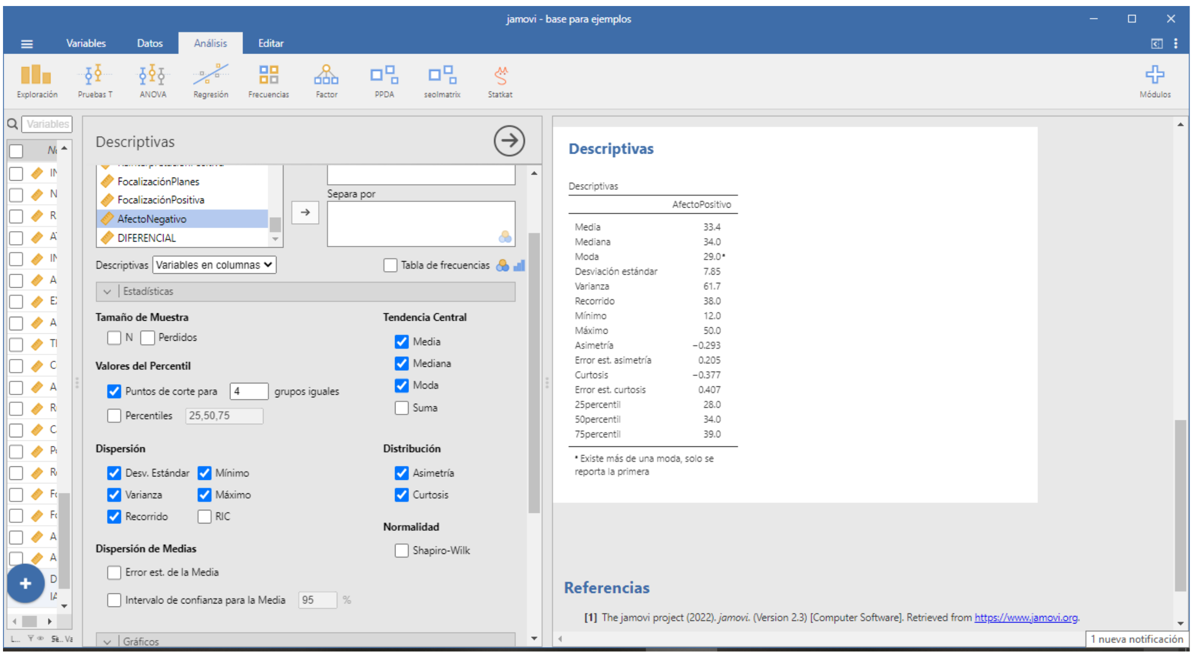


Figura 6: *Captura de pantalla del análisis descriptivo realizado mediante el programa estadístico Jamovi*

Tabla 1.

Estadísticos descriptivos para la variable Afecto positivo

	Afecto Positivo	
Media	33.4	**Medidas de tendencia central**
Mediana	34.0	
Moda	29.0 ?	
Desviación estándar	7.85	**Medidas de dispersión**
Varianza	61.7	
Mínimo	12.0	
Máximo	50.0	
Asimetría	-0.293	**Medidas de forma**
Curtosis	-0.377	
25percentil	28.0	**Medidas de posición no central**
50percentil	34.0	
75percentil	39.0	

? Existe más de una moda, solo se reporta la primera

¿Te animas a interpretar algunos de los valores que ofrece la tabla 1?... hazlo y consúltanos en clase… te ayudará mucho a fortalecer tus conocimientos!!

BIBLIOGRAFÍA

Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, P. (2010). *Metodología de la Investigación* (5ta edición). México D.F.: McGraw Hill.

Caballero-Romero, A. E. (2009). *Metodología de la Investigación Científica, Diseños con Hipótesis Explicativas*. Lima – Perú: Editorial UDEGRAF.

The Jamovi project (2022). *Jamovi (Version 2.3.13) [Computer Software].* https://www.jamovi.org.

SIMULADOR DE DISTRIBUCIONES online

http://www.shodor.org/interactivate/activities/SkewDistribution/